MEDICAL SIGNAL PROCESSING

Medical signal and image processing entails the acquisition and processing of information-bearing signals recorded from the living body. The extraction, enhancement, and interpretation of the clinically important information buried in these physiological signals and images have significant diagnostic value for clinicians and researchers. This is due to the fact that they help probe the state of the underlying physiological structures and dynamics. Because the signals are in general acquired noninvasively (i.e., from the surface of the skin, using sensors such as electrodes, microphones, ultrasound transducers, optical imagers, mechanical transducers, and chemical sensors etc.), medical signal and image processing offers attractive clinical possibilities.

Particularly in noninvasive (and therefore indirect) measurements, the information is not readily accessible in the raw recorded signal. To yield useful results, the signal must be processed to compensate for distortions and eliminate interference. The medical environment is commonly noisy, and the recorded signals are noise corrupted. Signals and systems engineering knowledge, particularly signal processing expertise, is therefore critical in all phases of signal collection and analysis.

Engineers are called upon to conceive and implement processing schemes suitable for medical signals. They also play a key role in the design and development of medical monitoring devices and systems that match advances in signal processing and instrumentation technologies with medical needs and requirements.

This article provides an overview of contemporary methods in medical signal processing that have significantly enhanced our ability to extract information from vital signals.

MEDICAL SIGNALS

The need for noninvasive measurements of signals generated by the body presents unique challenges that demand a clear understanding of medical signal characteristics. Therefore, the success of signal processing applications strongly depends on a solid understanding of the origin and the nature of the signal.

Medical signals are generally classified according to their origin. Bioelectrical, bioimpedance, biomagnetic, bioacoustical, biooptical, biomechanical, and biochemical signals refer to the physiological sources that they originate from or to the induction process imposed by the measurement. The dynamic range and the frequency content of these signals have been tabulated (1).

The source of bioelectric signals detected from single excitable nerve and muscle cells, for example, is the membrane potential. When bioelectric signals are sensed noninvasively by skin electrodes as in the case of surface electrocardiogram (ECG), which emanates from the beating heart (see also ELECTROCARDIOGRAPHY), a multitude of excitable cardiac cells collectively create an electric field that propagates through the biological medium before giving rise to the measured signal. Other clinically used spontaneous bioelectric signals recorded from surface electrodes include the electroencephalogram (EEG), which records the electrical activity of the brain (see also ELECTROENCEPHALOGRAPHY), surface electromyogram (EMG) detected from the muscles (see also ELECTROMYOGRAPHY), electrooculogram (EOG) from the eye, and electrogastrogram (EGG) from the gastrointestinal track.

Evoked potentials (EP) which measure brain response to specific visual, auditory, and somatosensory stimulus are also bioelectric signals. Bioelectric signals measured invasively (i.e., transcutaneously) with needle electrodes include the electroneurogram (ENG) that represents the electrical activity of a nerve bundle, electroretinogram (ERG) of the retina, electrocorticogram of the cortex, and single fiber EMG.

In bioimpedance measurements, by contrast, high-frequency (>50 kHz), low-amplitude (10 μA to 10 mA range) currents are applied to the tissue, organ, or whole body, and the impedance is measured. The impedance characteristics provide information including tissue composition, blood volume, and blood distribution.

The brain, the heart, and the lungs also emit extremely weak magnetic fields that often complement and augment information obtained from bioelectric signals. The main impediment to wide clinical use of biomagnetic signals is the very low signal-to-noise ratio, which presents engineering challenges in detection.

Many physiological phenomena create sound signals. The bioacoustic signals originate from blood flow in vessels, air flow in the lungs, the joints, and the gastrointestinal tract and are noninvasively recorded by microphones. The development of fiber optics and photonics has given rise to a multitude of clinical applications based on biooptical signals. Blood oxygenation, for instance, can be monitored by detecting the backscattered light from tissue at a variety of wavelengths.

The mechanical functions of physiological systems give rise to motion, displacement, flow, pressure, and tension signals. Biomechanical signals are often detected by local transducers and hence require invasive measurements because mechanical signals are rapidly attenuated. Monitoring ion concentration or partial pressure of gases in living tissue or in extracted samples yield biochemical signals. Recent developments in biochemical signal detection underlie the design of new-generation biosensors intended for medical as well as environmental and industrial applications.

Processing of Medical Signals

The rapid pace of development in discrete-time signal processing (2) (see also DISCRETE TIME FILTERS) coupled with the advent of digital computing and, in particular, the rise of vir-

tual instrumentation have given rise to efficient, flexible, and cost-effective methods to acquire and analyze biomedical data in digital form (3). Nonelectrical medical signals are generally transduced to electrical form to readily digitize and process them by digital means. Continuous-time signals are preprocessed by an analog antialiasing filter prior to analog–digital conversion (see also ANALOG-TO-DIGITAL CONVERSION). This presampling amplifier stage is introduced in order to satisfy the sampling theorem, which demands that the sampled signal be band-limited and that the sampling frequency exceed twice the highest frequency contained in the signal. Standards for a minimum sampling rate in clinical recordings have been introduced (4). In medical signal acquisition, it is critically important to preserve the time morphology of the waveforms; thus, phase distortions must be avoided by selecting a presampling amplifier with linear phase within the passband (5).

Although generic signal acquisition, time-domain analysis (see also TIME-DOMAIN NETWORK ANALYSIS), digital filtering [see also DIGITAL FILTERS; FIR FILTERS, DESIGN; IIR FILTERS], and frequency-domain and spectral analysis (see also SPECTRAL ANALYSIS) methods are often used, the low amplitudes and strong interference from other bioelectric signals and noise contamination typical in medical signals require special treatment (6–14). Fourier decomposition implemented using the fast Fourier transform (FFT) is widely used to study the frequency or harmonic content of medical signals. When the frequency content of the interfering signal (e.g., high-frequency muscle contraction or EMG noise) does not overlap with that of the signal under study (e.g., ECG), digital filters can be used to separate the two signals (e.g., ECG can be low-pass filtered to reject the EMG). Similarly, slow baseline drift resulting from low-frequency motion artifacts and effects of breathing in bioelectric recordings can be eliminated by high-pass filtering. The 50 Hz or 60 Hz power-line noise can be drastically reduced by notch filtering. As mentioned earlier, the need to prevent phase distortion in medical signal processing suggests the use of linear-phase finite impulse response (FIR) filters with center symmetric impulse response or, equivalently, filter parameters. Several powerful design methods have been developed for linear-phase FIR filters and are available in commercially available digital signal processing libraries.

Digital filters can not be used, however, when the frequency content of the signal and the interference are not distinguishable. A case in point is the extraction of evoked potentials from the background EEG. An understanding of the deterministic or stochastic nature of the underlying physiological process is necessary for the classification of a medical signal as such, hence the selection of the appropriate processing methods and algorithms. Signals described by exact mathematical formulation, graphical form, rule, or look-up table are recognized as deterministic signals. Nondeterministic or stochastic signals defy mathematical formulation and evolve in an unpredictable random manner (see also STOCHASTIC SYSTEMS). In the detection of EP from the EEG, it is natural to assume that EP is a deterministic signal embedded in EEG, which is assumed to be a zero-mean stochastic signal. Signal averaging with proper time alignment of the EPs is a commonly used technique whenever the signal (EP) and the noise (background EEG) are uncorrelated and the noise is characterized by a white (i.e., constant) spectral distribution.

The nonparametric spectral analysis of medical signals, which can be assumed to be stationary stochastic processes, can be accomplished using algorithms based on direct Fourier transform of the time series or the inverse Fourier transform of the autocorrelation sequence. To taper the edges of data records to reduce leakage, time-windows are applied to time sequences and lag-windows to autocorrelation sequences. Parametric spectral analysis casts the estimation problem into a framework in which the medical signal is modeled as the response of a rational system function to a white Gaussian noise signal. Powerful methods to identify the model parameters have been introduced, and test criteria to select the appropriate order and type of the model have been developed. Depending on the problem at hand, these models can have all-pole or auto-regressive (AR), all-zero or moving average (MA), or pole-zero or ARMA form (15).

Conventional linear signals and systems theories have been founded on assumptions of stationarity (or, equivalently, time-invariance in deterministic systems) and Gaussianity (or, equivalently, minimum-phaseness). Also, the assumption of a characteristic scale in time constitutes an implicit assumption. Investigators engaged in the study of medical signals have long known that these constraints do not hold under the test of practice. There is evidence that new methods are better suited to capture the characteristics of medical signals.

The investigation of higher-order statistics and polyspectra in signal analysis and system identification fields has provided processing techniques able to deal with signal nonlinearities and non-Gaussian statistics as well as allowing for phase reconstruction. In cases where the power spectrum is of little help, higher-order spectra (HOS) calculated from higher-order cumulants of the data can be used. The aim behind using HOS rather than any other system identification method is threefold:

1. Suppress Gaussian noise of unknown mean and variance in detection and parameter estimation problems;
2. Reconstruct phase as well as the magnitude response of signals or systems;
3. Detect and characterize the nonlinearities in the signal (16,17).

The power of neural-network based methods in medical signal processing is widely acknowledged. In essence, neural networks mimic many of the functions of the nervous system. Simple networks can filter, recall, switch, amplify, and recognize patterns and hence serve well many signal processing purposes. Neural networks have the ability to be trained to identify nonlinear patterns between input and output values of a system. The advantage of neural networks is the ability to solve algorithmically unsolvable problems. Neural networks are trained by examples instead of rules. The most widely used architecture of neural networks is the multilayer perceptron trained by an algorithm called backpropogation (18,19). (See also NEURAL NET ARCHITECTURE.)

Time-frequency representation (TFR) of a one-dimensional signal maps the time-varying (nonstationary) signal into a two-dimensional set where the axes are time and frequency. There are certain properties that a TFR should satisfy: covariance, signal analysis, localization, and inner products. Many innovative TFRs have been proposed to overcome the problems inherent in others, including the spectogram, Wigner

distribution, Cohen's class of TFRs, affine class of TFRs, and scalograms (20,21).

Many physiological dynamics are broadband and hence evolve on a multiplicity of scales. This explains the recent success and popularity of medical signal processing applications based on time-scale analysis and wavelet decomposition (22,23). The wavelet transform responds to the need for a tool well suited for the multiscale analysis of nonstationary signals by using short windows for high frequencies and long windows for low frequencies (24). Continuous wavelet transform (CWT) decomposes a random signal onto a series of wide bandpass filters that are time-scaled versions of a parent wavelet. The parent wavelet is in fact a bandpass filter with a given center frequency and parametrized by a scale factor. As this factor is changed, the time domain representation of the parent wavelet is either compressed or expanded. As the scale factor is increased in value, the bandwidth of the signal in the frequency domain is compressed and vice versa. This duality between the time-frequency domains is the advantageous side of the wavelet transform that results in an increased time-frequency localization. However, the signal is expressed in the wavelet-domain as a function of time and scale rather than frequency components as is the case for the Fourier transform. This property is of utmost advantage when the signal in interest has multiscale characteristics.

The discrete wavelet transform (DWT) uses orthonormal parent wavelets for its basis in handling multiscale characteristics of signals. When a dyadic orthonormal basis is chosen, a statistically self-similar, wide sense stationary random signal can be decomposed by DWT to yield wavelet coefficients parametrized by the scaling (dilation) factor and the translation factor. It has been demonstrated that orthonormal bases are well suited for the analysis of $1/f^{a}$ processes. $1/f^{a}$ or power-law processes are characterized by a power spectrum that obeys the power-law attenuation (25).

A large class of naturally occurring phenomena exhibit power-law or $1/f^{a}$ type spectral attenuation over many decades of frequency and, hence, "scale" or reveal new details upon time magnification. Medical examples include biological time series such as voltages across nerve and synthetic membranes and physiological signals such as the EEG and the heart rate variability (HRV) (26,27). HRV is analyzed to diagnose various disorders of the cardiovascular and the autonomous systems. HRV is derived from the heart beat signal by detecting the peaks of the QRS waveforms in the ECG. If the duration between consecutive beats is presented as a function of the beat-number, the resulting HRV is referred to as a tachogram.

The application of scaling concepts and tools such as wavelets to medical signals has uncovered a remarkable scaling order or scale-invariance that persists over a significant number of temporal scales. Methods to capture scaling information in the form of simple rules that relate features on different scales and to relate them to physiological parameters are actively investigated. As in the case of spatial fractals that lack a characteristic length scale and consist of a hierarchy of spatial structures in cascade, $1/f^{a}$ processes cannot be described adequately within the confines of a characteristic time scale. This property is beginning to be recognized as the dynamical signature of the underlying "complex" physiological phenomena (28).

Compression of Biomedical Signals

Medical signal processing applications generate vast amounts of data, which strain transmission and storage resources. The need for long-term monitoring as well the desire to create and use multipatient reference signal bases places severe demands on storage resources. Future ambulatory recording systems and remote diagnosis will largely depend on efficient and preferably lossless biomedical data compression. Data compression methods tailored to biomedical signals aim to eliminate redundancies while retaining clinically significant information (29).

Signal compression methods can be classified into lossless and lossy coding. In lossless data compression, the signal samples are considered to be realizations of a random variable, and they are stored by assigning shorter or longer codewords to sample values with higher or lower probabilities, respectively. In this way, savings in the memory volume are obtained. Entropy of the source signal determines the lowest compression ratio that can be achieved. In lossless coding, the source signal can be perfectly reconstructed by reversing the codebook assignment. For typical biomedical signals, lossless (reversible) compression methods can achieve compression ratios on the order of 2 : 1, whereas lossy (irreversible) techniques may produce ratios of 10 : 1 without introducing any clinically significant degradation. [See also DATA COMPRESSION, LOSSY.]

The promise of medical signal and image processing will be fulfilled when its techniques and tools are seamlessly merged with the rising technologies in multimedia computing, telecommunications, and eventually with virtual reality. Remote monitoring, diagnosis, and intervention will no doubt usher the new era of telemedicine. The impact of this development on the delivery of health care and the quality of life will be profound.

MEDICAL IMAGING

Medical images are of many forms: X ray, ultrasound, and microscopic, to name a few. Increasingly, the formation, storage, and display of these images is based on extensive and intensive utilization of signal and image processing. Some *imaging modalities* such as X-ray imaging onto film are mature and stable, whereas other modalities, such as magnetic resonance imaging, are recent and undergoing rapid evolution. There is a vast range of procedures: projection X-ray images of the body, dynamic studies of the beating heart with contrast medium, and radioisotope images that depict glucose utilization, a biochemical activity.

This section consists of the following subsections:

1. X-ray imaging,
2. Computer tomography,
3. Radionuclide imaging,
4. Magnetic resonance imaging,
5. Ultrasound, and
6. The ecology of medical imaging.

The first five sections deal with the medical imaging modalities. For each modality, we will describe the physical and imaging principles, discuss the role of image processing, and

describe recent developments and future trends. The last section will present the clinical and technical setting of medical imaging and image processing and indicate the relative importance of the various image modalities.

An article of this length cannot fully describe all medical image processing areas. We have selected the more important methods that are currently used in modern hospitals. We do not review microscope imaging, endoscopy, or picture archival and communications systems (PACS). Microscopy is the essential tool of pathology, and much image pattern recognition work has been done in this area. Endoscopic imaging is optical visualization through a body orifice (e.g. brachioscopy, viewing lung passages; sigmoscopy, viewing the intestine) or during minimally invasive surgery. PACS, comprising electronic storage, retrieval, transmission and display of medical images, is a very active field, which includes teleradiology (the remote interpretation of radiological images) is described in a separate article on TELEMEDICINE.

X-Ray Imaging

X-ray imaging is the most widely used medical imaging modality. A radiograph is a projection image, a shadow formed by radiation to which the body is partially transparent. In Fig. 1 we show radiographs of a hand. The quantitative relation between the body and the radiograph is quite complex. Some radiographs have extremely high resolution and contrast range. Both static and dynamic studies are performed, and contrast agents are used to enhance the visibility of internal structures. Recent developments in instrumentation are computed radiology and solid-state image enhancement systems. The most commonly used image processing technique is subtraction. Even though extensive research has been performed on image enhancement and recognition of X-ray images, there has been very little practical acceptance of these techniques. [See the article on X-RAY APPARATUS.]

Physical Principles. X rays are electromagnetic waves whose frequencies are higher than those of ultraviolet light. The frequency of X rays is normally not denoted in hertz but rather in photon energy in (kilo) electron volts, which is related to frequency by Planck's formula $E = h\nu$, where E is the photon energy, h is Planck's constant, and ν is the frequency. Photon energies for normal diagnostic use range from 10 keV to 100 keV. X rays are a form of *ionizing radiation;* low exposures to X rays may cause cancer whereas high exposures can be fatal. X-ray imaging systems can be analyzed by using ray optics because diffraction effects are negligible.

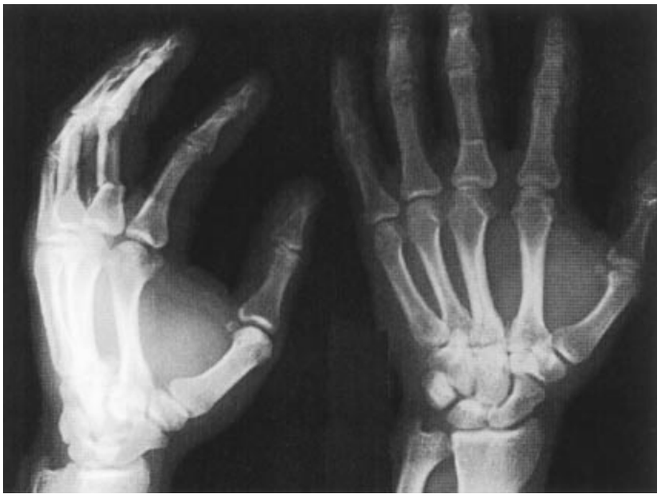

Figure 1. Two radiographic views of a hand. A fracture in the middle finger can be seen more easily in the lateral view on the left. Courtesy of Dr. Harold Kundel, University of Pennsylvania.

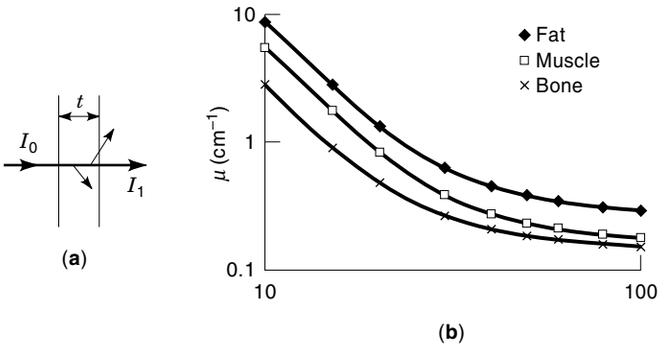

Figure 2. (a) Interaction between X-rays and a homogeneous object. An incident X-ray photon can be absorbed or scattered. The X-rays that are neither absorbed nor scattered constitute the attenuated beam, useful for imaging. Scattered radiation impairs image quality. (b) Linear attenuation coefficient for soft tissue (muscle, fat, blood) and for bone as a function of photon energy. Linear attenuation coefficients of different soft tissues are proportional to their densities, which range from about 0.95 g/cm^3 for fat to about 1.04 g/cm^3 for muscle.

Diagnostic X radiation is generated with X-ray tubes, which produce polyenergetic radiation (radiation over a broad range of frequencies). The primary energy (center frequency) is determined by the voltage applied to the tube and the amount of X-ray power by the tube current. The amount of X-ray radiation per unit area is called *exposure,* measured in roentgens, a unit proportional to the energy flux per unit area, where the proportionality factor depends on photon energy. The amount of X-ray energy per unit mass absorbed by a body is called the *dose,* measured in Grays.

If a monoenergetic (single-frequency) X-ray beam falls on a homogeneous object as shown in Fig. 2(a), the relation between the incident and the transmitted flux is

$$I_1 = I_0 e^{-\mu t} \tag{1}$$

where I_0 is the incident flux, I_1 is the transmitted flux, t is the thickness of the object, and μ is the *linear attenuation coefficient,* a property of the object that depends on its atomic composition, density, and the photon energy. Linear attenuation coefficients for body tissues decrease over the diagnostic energy range: plots of attenuation coefficients for soft tissue and bone are given in Fig. 2(b). X rays at different energies combine in an additive way. If a homogeneous body of constant thickness is illuminated with an X-ray beam with spectral intensity $I_0(V)$ in units of energy per unit area and photon energy, the transmitted energy per unit area is given by

$$E = \int I_0(V) e^{-\mu(V)t}\, dV \tag{2}$$

where V is the photon energy, $\mu(V)$ is the attenuation coefficient as a function of photon energy, and t is the thickness of the object.

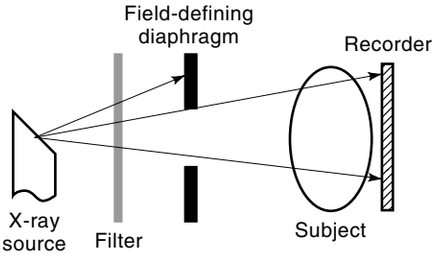

Figure 3. A conventional X-ray imaging system.

X-ray interaction with an object in the diagnostic energy range can occur in two ways: absorption and scattering. In absorption the photon is stopped and gives up all its energy in the vicinity of the interaction point, whereas in scattering the photon gives up part of its energy and continues in a different direction. Equation (1) includes both absorbed and scattered photons. However, scattered photons can leave the body and impinge on the detector. Scattered radiation is undesirable: it contributes to noise in the image and reduces contrast.

X-Ray Imaging. A typical X-ray imaging system is shown in Fig. 3. An X-ray beam from a generator is restricted to the appropriate region with a field-defining diaphragm, passes through the body being imaged, and is detected with a recording system. In *radiography* one or several still images are recorded. In *cineradiography,* a moving image sequence is recorded on film or videotape. In *fluoroscopy,* the image is viewed by a radiologist or a surgeon while the patient is in the X-ray apparatus. Images may be formed from the intrinsic contrast between tissues or by injecting or ingesting a *contrast agent,* a substance that enhances the radiographic difference between tissues (such as between blood and surrounding soft tissues). Contrast studies are *invasive;* they typically cause greater discomfort or danger than *noninvasive* procedures. We will discuss the factors that affect the image contrast and resolution and describe some X-ray image recording systems.

Image Contrast and Resolution. Image contrast is best at low photon energies, and the energy available for image formation increases with photon energy. These considerations lead to the use of lower voltages for imaging thin body sections, such as limbs and the female breast, and high voltages for thick body parts, such as the abdomen. To see how radiological physics affects this choice, consider the configuration shown in Fig. 4, which shows a schematic view of a uniform soft tissue region of thickness t containing a blood vessel of diameter d. The attenuation coefficients of tissue and blood are μ_1 and μ_2, respectively. Applying the law given in Eq. (1), the intensity of the beam passing through the tissue is $I_1 = I_0 \exp(-\mu_1 t)$, whereas the rays passing through the center of the blood vessel have intensity $I_2 = I_0 \exp[-\mu_1 t - (\mu_2 - \mu_1)d]$. For the blood vessel to be visible the *contrast* C

$$C = \frac{|I_1 - I_2|}{I_1} = |1 - \exp[-(\mu_2 - \mu_1)d] \approx |(\mu_2 - \mu_1)d| \qquad (3)$$

should be high. To obtain high contrast there should be a large difference between the attenuation coefficients of the blood vessel and surrounding tissue. Also, contrast increases with d, the size of the object being imaged. From Fig. 2(b), we see that it is desirable to use a low photon energy because attenuation coefficient values decrease as energy increases. On the other hand, I_0 is limited by the need to limit the dose to the patient, while the ability to form the image requires a sufficiently large value of transmitted energy, which is proportional to $\exp(-\mu_1 t)$. Because the attenuation coefficient decreases with photon energy, we see that for large values of t (thick body sections) we need to use higher-voltage X rays.

Contrast can be improved by introducing a material that changes the attenuation coefficient. *Contrast angiography* is the visualization of blood vessels by using externally injected agents. Elements with high atomic number, such as iodine, have a much higher attenuation coefficient than normal body tissues. By injecting such a material into the circulatory system, we can increase μ_2 in Eq. (2), making blood vessels visible in thick body sections that require high-voltage X rays. Contrast agents, however, disturb the body so that these procedures may be used only when the risk of not treating the disease outweighs the danger of the diagnostic procedure.

One of the factors affecting resolution (the ability to see small objects) is the focal spot, the size of the area on the generator that emits X rays. The issues in this phenomenon can be seen in Fig. 5(a), which shows a shadow cast by a small object when irradiated with a finite source. The shadow exhibits a penumbra, a region around the edges of the projected image of the objects in which the contrast is reduced. If the object is too small the penumbra may be greater than the image. The contrast of the object is reduced and may be too low to see the object. The extent of degradation depends on the size of the X-ray generator source (focal spot), the size of the structure being imaged, the distance from the source to the object, and the distance from the object to the recording system. For best resolution the focal spot should be small, and the object-recorder distance should be smaller than the source-object distance: if the object is thin and in contact with the recording surface, no blurring is caused by focal spot size. Focal spot size is limited by energy dissipation in the X-ray generator, and X-ray intensity on the detector varies inversely as the square of the source-detector distance, so that an optimal resolution is obtained by an appropriate balance of these factors.

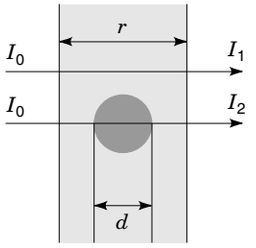

Figure 4. A schematic view of a blood vessel surrounded by soft tissue. The contrast depends on the size of the blood vessel and on the difference between attenuation coefficients (see text).

A second limitation on resolution is imposed by the X-ray recording system: currently available technologies require a trade-off between resolution and sensitivity. As an example, consider a screen-film system for radiography, which consists of an *intensifying screen* placed in contact with a photographic film [see Fig. 5(b)]. The intensifying screen is made from a material with a high attenuation coefficient that fluoresces in

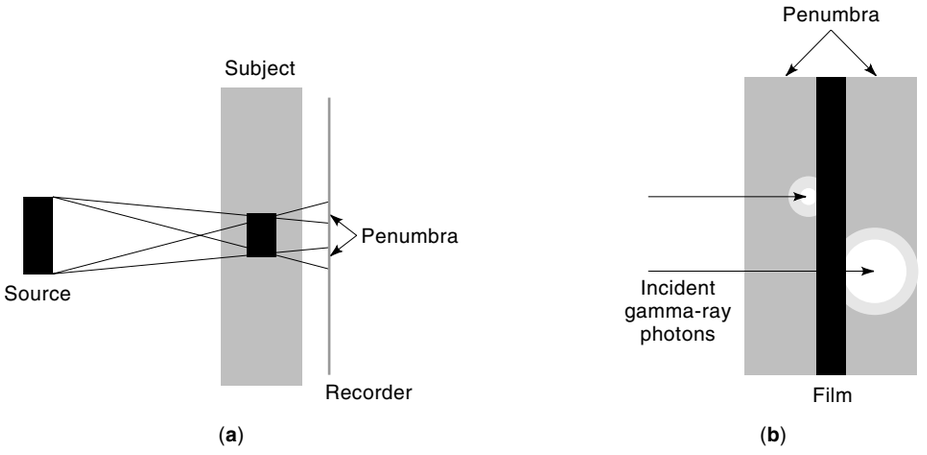

Figure 5. Source and recorder effect in X-ray resolution. (a) The blurring of an image depends on focal spot size and distances between the object, X-ray source, and the recording medium. (b) A screen-film system for radiography. X-rays are captured by screen materials that emits bursts of light photons which, in turn, expose photographic film. The cassette shown uses two screens to provide higher X-ray detection efficiency.

response to the absorption of X-ray photons: one X-ray photon can produce as many as 1000 light photons. Resolution is limited by the distance between the fluorescing spot and the film: the size of the area illuminated by one X-ray photon is proportional to the distance between absorption point and the film emulsion. An ideal intensifying screen should absorb most incident X-ray photons, should have adequate conversion gain (light photons per X ray), and should be thin enough to produce negligible blurring. Currently available materials and manufacturing techniques require compromises between these factors.

Image Recording Systems. The primary categories we consider are radiographic systems (still images) and fluoroscopic systems (moving images). In radiography, one may use direct film recording, screen-film systems, digital radiography systems, or screens coupled to recording cameras. Direct film recording is used for dental radiography and for imaging thin structures, such as the hand. Even though the sensitivity of direct film recording is not as high as for screen-film systems, these procedures produce a relatively low dose because the structure being imaged is relatively thin. The screen-film recorder was described in the previous section.

A computed radiography system captures (records) X rays on a photostimulable phosphor screen. This forms a latent image that is read by scanning the plate with a laser beam that induces fluorescence proportional to X-ray exposure. The scanner produces a digital image, which may be viewed on a monitor. The plate is cleared by exposing it to visible light and reused. This system has a wider latitude (dynamic range) than film, although the resolution is not as high as the best screen-film combinations. Even though the plates and scanner involve considerable capital investment, computed radiography can lead to savings in operating expense, especially when images are stored and distributed electronically. Some radiology departments are planning for a future when no more film will be used.

In screen-film and computed radiography systems a latent image is formed that requires further processing. In digital radiography the image could be recorded with a sufficiently large electronic camera. Another option is to optically reduce the image produced by an intensifying screen, and to capture the reduced image with a conventional charged-coupled-device (CCD) camera. However, with present technology, this leads to a reduction in image quality.

In fluoroscopy, where a moving image is viewed in real time, best performance is achieved by using image intensifiers. Fluoroscopic exams require prolonged exposure and viewing so that dose reduction is critical. Much of this dose reduction is achieved by using systems that trade resolution for sensitivity. Traditional radiological image intensifiers are vacuum tubes that contain, on the input side, intensifying phosphors coated with photoemissive targets. X rays produce an optical image that falls on the photoemissive material, which in turn produces an electron image emitted into the vacuum on the interior of the intensifying tube, where they are imaged with electron optics onto a small phosphor target (similar to that in a cathode-ray tube display). This small optical image is easily coupled to a television camera with no subsequent loss in image quality. The television signal can be displayed on a monitor, recorded onto videotape, or digitized and stored on computer media. Another option in fluoroscopy is to use digital scan converters: rather than use continuous exposure of the patient, images can be captured occasionally, as required, or at a slow frame rate. These images are digitized and stored in random-access memory (frame buffer) which subsequently generates a conventional television signal that drives a television monitor.

X-ray Image Processing. The most common image processing procedure in X-ray imaging is subtraction angiography, where an image without contrast material is subtracted from a radiogram produced with an injected contrast agent, thereby enhancing the visibility of the structures containing the contrast agent. In Fig. 6 we see a radiogram of the head before and after contrast injection, and the difference image. The arteries are clearly seen. This concept was introduced in the 1930s, when photographic processing was used to produce the difference image, but contemporary digital image processing techniques vastly improve the speed and convenience of these procedures.

Image processing for radiological image enhancement was first proposed in the 1960s. It seems plausible that such processing can improve contrast and correct for image blurring caused by finite focal spot size. Although appropriate contrast manipulation for image enhancement is effective in fluoroscopy systems, where an electronic image is viewed on a monitor, resolution enhancement has met no success, and there are as yet no accepted techniques for enhancing film radiograms. This is probably due to the fact that film radiography

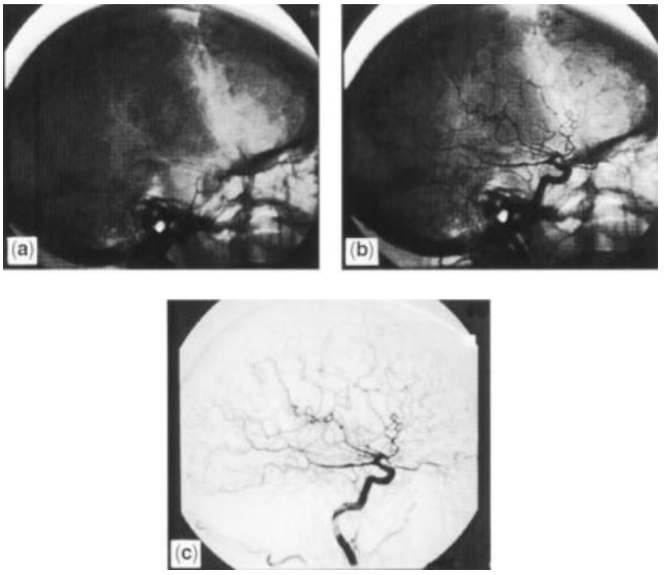

Figure 6. Subtraction angiography. Radiograph of a head before (a) and after (b) contrast agent injection. The difference image (c) shows only the arteries. Courtesy Dr. Barry Goldberg, Thomas Jefferson University.

systems are optimized with the proper choice of X-ray focal spot size and recording system sensitivity and resolution, so that detail enhancement increases noise as well as resolution producing no noticeable improvement in image quality.

One of the hurdles in successful implementation of X-ray image enhancement is the high resolution and contrast range in film images. For example, mammography (imaging of the female breast for cancer screening) produces images 25 cm $\times$ 20 cm, which may have significant detail as small as 50 μm in extent. This may require digitization to 6000 $\times$ 6000 pixels with 16-bit density resolution. Merely scanning such images in a way that preserves all the visual detail is a substantial task with present technology. Obtaining images that can be enhanced to show detail not apparent on the original film image may require even higher-quality digital image acquisition.

Extensive research has been done on systems for automatic interpretation of radiograms. Some success has been achieved in the measurement of the cardiothoracic ratio (the ratio of the width of the heart to the width of the chest), measurement of arterial stenosis (narrowing) in contrast arteriograms, and detection of degenerative change in lungs. These methods have not yet achieved practical acceptance. Some reasons for the lack of progress is the extreme complexity of X-ray images, which are projections (shadows) of three-dimensional bodies, and many structures may overlap in one area. Radiologists, who undergo extensive training, visualize the three-dimensional structures in the body, compare a specific image with others that they have seen before, and base their diagnosis on an extensive knowledge of anatomy and pathology. It is, at present, not feasible to construct an automatic system that can operate at this level. One recent line of research is in promising computer-assisted diagnosis (CAD) which is being explored for screening mammography interpretation. A processing algorithm examines a scanned image and indicates suspicious areas to a radiologist, who makes the final decision. This type of interactive diagnosis system has the promise of combining a radiologist's extensive knowledge and ability to interpret complex images with a machine's capability to perform a thorough, systematic examination of a large amount of data.

Computer Tomography

A tomogram is an image of a section through the body. Tomographic X-ray images were first formed in the 1930s by moving the source and film in such a way that one plane through the body remained in focus while others were blurred, but they were used for only very specialized investigations. X-ray computer tomography, which is abbreviated as Cat or X-Cat, was introduced in 1973 and gained almost instantaneous acceptance.

In Cat, transmitted X-ray intensity measurements are made for a large number of rays passing through a single plane in the body. These measurements are subsequently processed to compute (reconstruct) the values of attenuation coefficient at each point in the plane. This, in effect, produces a map of the density through the tomographic section plane. The geometric arrangement for performing these measurements is shown in Fig. 7. A ring of X-ray detectors surrounds the body being scanned, and an X-ray generator rotates in an orbit between the body and the detector ring. For each position of the source, intensities of the radiation passing through the body and impinging on the detectors are digitized and captured by a computer. A sectional image is computed from the data collected while the source rotates through a full circle.

Medical CT images typically have a resolution of 500 $\times$ 500 pixels, which is about the same as that of television, and much lower than the resolution of radiograms. Radiograms are projection images and superpose many anatomical structures onto one plane, but all structures in a tomogram are shown in their correct geometrical relationship. Figure 8 shows a CT scan through the upper abdomen. The spleen (lower right) contains a hematoma (arrow), a pool of blood caused by an injury. The hematoma density is only about 5% higher than that of surrounding tissues, and it would not be visible in a conventional radiograph.

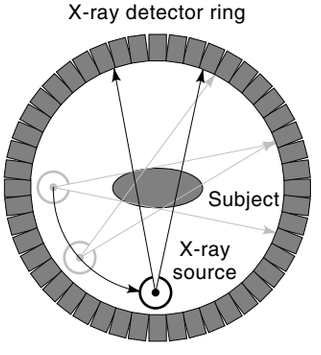

Figure 7. Schematic diagram of an X-ray tomography (CT) scanner. Signals are collected from 500 to 1000 detectors located on a ring. The X-ray source rotates on an orbit between the detectors and the subject.

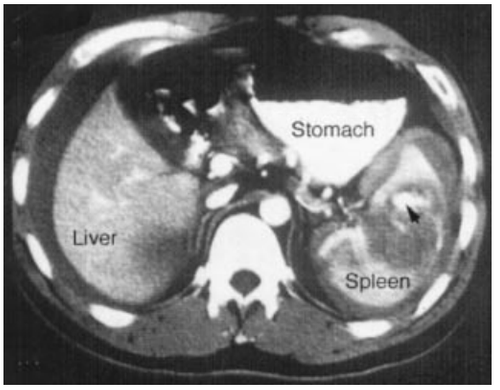

Figure 8. CT scan through the upper abdomen. The arrow points to a hematoma (a pool of blood) in the spleen. The density difference between the blood and surrounding tissues is about 5%. Courtesy of Dr. Richard Wechsler, Thomas Jefferson University.

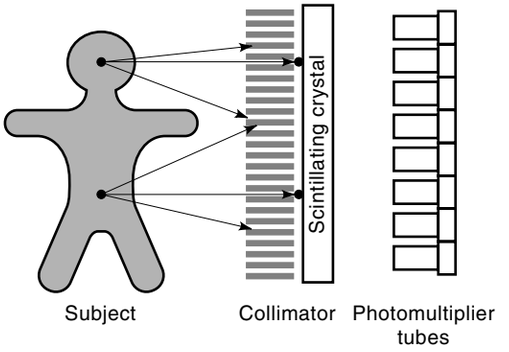

Figure 9. Schematic diagram of a gamma camera. Only gamma rays normal to the camera are admitted by the collimator. An electronic system collects signals from the photomultiplier tubes and computes the coordinates of gamma-ray photons detected by the scintillating crystal.

CT reconstruction algorithms are perhaps the most successful use of image processing in medicine. Unlike X-ray projection images, which are inherently analog, CT images are inherently digital and are impossible to obtain without sophisticated signal and image processing technology.

Radionuclide Imaging

Radionuclides are atoms that are unstable and decay spontaneously. Certain decay reactions produce high-energy photons, which are called *gamma rays.* There is no difference between gamma and X rays; the difference in terminology reflects the process that produces the radiation, and not its physical nature. In nuclear medicine a *radiopharmaceutical agent* (a compound containing a gamma-ray-producing radionuclide) is injected into a body. Images formed from the emitted radiation can be used to visualize the location and distribution of the radioactivity. Radiopharmaceutical agents are designed to migrate to specific structures and to reflect the functioning of various organs or physiological processes. Consequently, radionuclide imaging produces *functional images,* in contrast to the *anatomical images* formed by conventional X rays. Some nuclear medicine procedures require only bulk measurements of radioactivity; for example, the time-course of radioactively labeled hippuric acid in the kidney is an indication of kidney function. In this article, we will concentrate on the procedures that require image formation. We will describe the operation of a *scintillation camera,* the most commonly used device for radionuclide imaging. We will also discuss single-photon emission computer tomography (ELT), an imaging method that produces three-dimensional data. We will not discuss positron emission tomography (PET) scanners, which have unique technical and biological advantages over ECT but are more expensive and are used primarily for research.

Projection images of radionuclide distributions are most commonly formed with a *scintillation camera,* which consists of two major components: a *collimator* and a position-encoding radiation detector, as shown schematically in Fig. 9. The collimator is made form a highly absorbing substance, such as lead, and contains a set of parallel holes that admit only radiation normal to the collimator. These gamma rays fall on a plate of scintillating material, such as sodium iodide. Each gamma ray photon produces a pulse of light, which is detected by a set of photomultiplier tubes. The current pulses from the tubes differ; the output of the detector closest to the point of scintillation is largest, and those farther from the source are smaller. The pattern of these current pulse intensities is used to compute the location of scintillation at a resolution much finer than that of the tube spacing. The energy of the gamma ray is estimated by computing the total amount of light falling on the protomultiplier tubes. Since scattered gamma rays, which degrade the image, have lower energies and are excluded by the processing system. These gamma rays form a severely defocused image of the object, thus reducing the contrast and increasing the noise level. The image of the radiation distribution is developed by recording individual detected gamma rays and their positions.

Tomographic images of radionuclide distribution can be formed by collecting images while a gamma camera moves along an orbit around the body. A typical system is shown, schematically, in Fig. 10, where three gamma cameras collect projection data while they rotate around a patient. Because each camera collects data for many planes through the body, the data from such an examination facilitates reconstruction of a three-dimensional or volumetric image. The algorithms used for emission computer tomogaphy (ECT) are similar to those for CT, but the reconstruction problem is more complex: emissions from deeper structures are attenuated more than

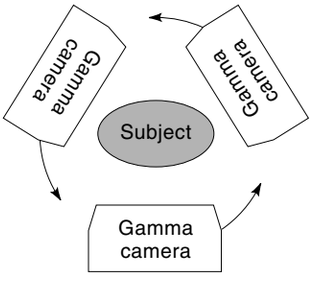

Figure 10. Schematic diagram of an Emission Computed Tomography (ECT) system. Three gamma cameras collect projection data from the subject while rotating about a common center.

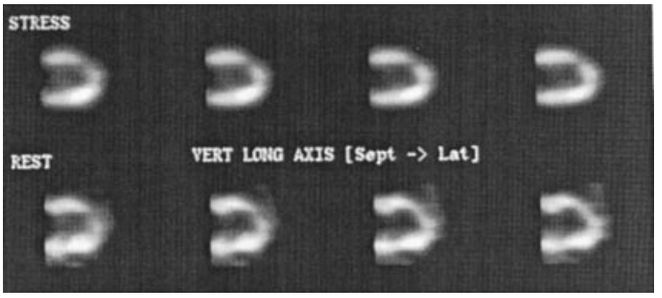

Figure 11. ECAT images of a heart. Shown are a set of slices through the left ventricle obtained after exercise (top) and at rest (bottom). The darker region in the upper right of the stress images indicates an ischemic (reduced blood supply) region of the heart muscle. Courtesy of Dr. David Friedman, Thomas Jefferson University.

those from the surface, and reconstruction algorithms must correct for this effect. There is no elegant solution to the general attenuation correction problem for ECT, although empirically developed methods produce satisfactory results.

Figure 11 shows several sections through the left ventricle of a heart obtained with an ECT scanner. The radiopharmaceutical is absorbed by metabolically active muscle. Two scans are obtained: one after exercise and one after a period of rest. The image has lower intensities in regions that are not receiving blood circulation, such as parts of the heart that may have been injured by a heart attack—a blockage of circulation due to an occlusion in an artery that supplies oxygen and nutrients to the heart muscle. The image is formed over a period of about 10 minutes. This is necessary because the amount of radioactivity is very low, so that photons must be collected for a long period to produce sufficiently high signal-to-noise ratio in the imaging. Blurring caused by heart motion is reduced by *cardiac gating;* the electrocardiogram identifies times at which the heart is in the same position, and data from these time intervals are pooled. The resolution in Fig. 11 is very low, but the technique provides unique information about muscle functioning that is not available from anatomical images.

Signal formation and image processing are an integral part of radionuclide imaging. The formation of the image from a gamma camera requires extensive real-time signal processing. A gamma camera produces a sequence of x–y coordinates of detected gamma rays that are processed to produce a pixel array of gray values. The image so obtained typically contains subtle geometric distortions and nonuniformities, which are corrected with appropriate algorithms. Further processing is required if cardiac gating is used or if tomographic images are produced. Certain studies measure parameters based on dynamic (time-course) data and use sophisticated statistical parameter estimation techniques. The viewing of three-dimensional data from ECT requires rendering methods that produce sectional or projection images.

Magnetic Resonance Imaging

X-ray and radionuclide imaging use electromagnetic radiation of a frequency so high that the human body is partially transparent. Magnetic resonance imaging (MRI) uses electromagnetic frequencies in the range of 40 MHz to 100 MHz, which also penetrate living tissues. At these frequencies, the wavelength is too long to allow focusing as an imaging principle. MRI uses the nuclear magnetic resonance effect to induce signals from the body. Signals from different locations in the body are frequency-encoded by applying spatially varying magnetic fields, and the emitted signals are reconstructed to produce images. The principles used for MRI are unlike those used for any other imaging technique. We will first describe the physical principles of nuclear magnetic resonance, which govern the production of the magnetic resonance signal. We will subsequently describe the methods used to generate the signals that form the images and also describe some of the processes and properties of the body that are viewed in medical MRI.

Nuclear Magnetic Resonance. Nuclear magnetic resonance is a process associated with the magnetic moments possessed by the nuclei of many atoms. There magnetic moments are present, typically, in nuclei whose atomic number (sum of the number of protons and neutrons) is odd. Any of these nuclei can be imaged through MRI. Medical MRI is almost exclusively based on the imaging of hydrogen nuclei or protons, so that in the subsequent discussion we will talk about the magnetic resonance and imaging of hydrogen nuclei. Each nucleus also has an angular momentum, which is aligned with the magnetic moment. Magnetic resonance is based on the joint action of these magnetic and rotational moments.

Thermal motion causes nuclear magnets to take random orientation, but an external magnetic field (the *dc field*) pulls at least some of these nuclei into alignment with the field direction. This alignment process is not instantaneous but follows an approximately exponential time course whose time constant, the so-called the *longitutional relaxation time* (T_1), depends on the physical and chemical state of the material, and ranges from about 0.25 s to 1 s for protons in biological tissues. Should such a nuclear magnet be deflected from the direction of external field, a torque that tends to move it toward alignment is generated. The nuclear angular momentum resists this so that the nucleus behaves like a spinning top: its axis maintains a constant angle in a relation to the dc field and spins (precesses) around it. This phenomenon is called *nuclear magnetic resonance.* The rate of this precession (the Larmour frequency) is proportional to the magnitude of the magnetic field and the proportionality constant, for protons, is 42.6 MHz/T. The magnitude of this oscillating magnetization decays exponentially at the so-called *transverse relaxation time* (T_2), which has a range from 20 ms to 100 ms for protons in biological tissues, depending on tissue type. The resonance of this oscillation is very sharp: the quality factor Q, which is a product of the resonance frequency and T_2, ranges from about 5×10^6 to 25×10^6 for protons in biological tissues.

The precessing magnetic moments set up an external oscillating field that can be detected by placing appropriate conductor loops called *pickup coils* near the body. These detect an induced voltage (due to the Faraday effect), which is proportional to the product of the magnetic moment and the frequency. Both the magnetic moment and the frequency are proportional to the external magnetic field; consequently, the voltage is proportional to the square of the field. It therefore is desirable to have as high a dc field as possible, and magnets with 1.5 T are now fairly common. To deflect (excite) nuclear magnetic moments from their equilibrium direction, an oscillating magnetic field (the *RF field*) is applied at right angles

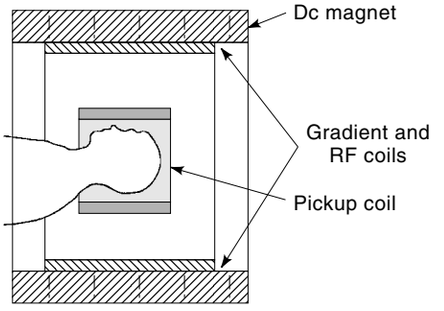

Figure 12. Schematic diagram of a MRI scanner.

to the dc field. If this field is at the Larmour frequency, nuclear magnets tip away from the direction of the dc field. The deflection angle from equilibrium (*flip angle*) is proportional to the product of the RF field strength and duration.

Magnetic Resonance Imaging Principles. A schematic diagram of a magnetic resonance imager is shown in Fig. 12. The bulkiest and most expensive component is the dc magnet. The RF coils excite the MR signal, and the pickup coils detect it. To form an image *gradient coils* impose linearly varying magnetic fields that are used to select the region to be viewed and to modulate the signal during readout. In a typical data collection step, the RF field is applied in the presence of a gradient field (slice selection gradient), so that only one slice through the body is at the Larmour frequency. After the magnetic moments are excited, another gradient field is applied (readout gradient) so that the moments in different portion of the slice radiate at different frequencies. The signal from the pickup coil is amplified and digitized. The Fourier transform of this signal gives a map of the amount of magnetic materials at regions of different gradient field value. A number of such data collection steps, each with a different readout gradient, are required to collect enough data to produce a sectional image. This sequence of data collection is programmed by applying pulses of current to slice selection coils, RF coil, and readout gradient coils. A magnetic resonance imager is a flexible instrument: image characteristics, such as slice thickness and resolution are determined by this *pulse sequence.* The slice orientation, thickness, and resolution can be varied.

Image intensity is basically proportional to proton concentration, or the amount of water in different tissues. Figure 13 shows a magnetic resonance image of the human head. This sectional plane, located along the plane of symmetry of the head, cannot be obtained with a single CT scan. Because bone contains less water than the soft tissues, the skull shows a lower image brightness. Further contrast is provided by differences in water content and chemical characteristics of various brain tissues, such as gray matter, white matter, and cerebrospinal fluid. Appropriate pulse sequences can provide images weighted by T_1 and T_2, which show characteristic tissue-specific variations. Other pulse sequences are designed to enhance signals due to differences in chemical composition (chemical shift imaging) which produce small changes in resonance frequency. Pulse sequences can also differentiate between moving and stationary tissues. MRI can produce angiograms (images produced by flowing blood) without using contrast agents. A relatively recent use of MRI is to form *functional images,* which show areas of brain activity that are related to sensory, motor, or cognitive brain activity. Functional imaging can be used to identify areas of the brain affected by stroke before the tissue has undergone a physical change caused by loss of blood circulation. This can guide treatment to restore circulation and avoid brain injury.

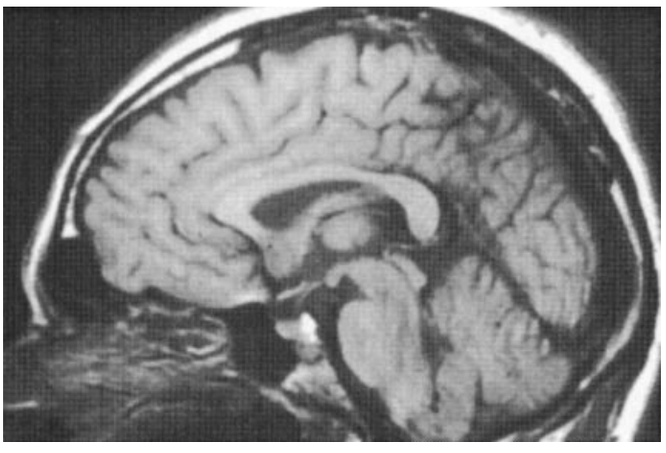

Figure 13. Sectional MRI image of a head.

From this description, we see that an MRI system functions through electronic and electromagnetic devices such as coils, amplifiers, and digital/analog converters that produce pulse sequences and detect signals. The signals from an MRI scanner are digitally processed to form images and to extract information such as proton density, relaxation times, motion, and chemical composition. Further processing is required to correct for magnetic field nonuniformity and motion-induced artifacts. Postprocessing of magnetic resonance images and image sets is an important area of signal processing and pattern recognition applications. Sets of images obtained with different pulse sequences are combined to enhance signals caused by tumors, neurological plaques, to provide three-dimensional reconstructions of complex structures and to plan surgical procedures.

In contrast to X-ray and radioisotope imaging, MRI uses low-frequency electromagnetic radiation, which is nonionizing. Therefore there is no concern about producing tumors or cancer; the only known source of harm from MRI is heating produced by the RF fields, which can easily be monitored and does not reach harmful levels in clinical equipment. On the other hand, high magnetic and RF fields in MRI may interfere with implanted electronic devices such as cardiac pacemakers, and MRI may be unusable or dangerous in the presence of implanted metal prostheses. MRI exams are long, uncomfortable, and expensive. Because MRI scanners are large and require extensive shielding they are available only in hospitals and other special facilities. Magnetic resonance image resolution is lower than that obtainable from X-rays. Because of these factors, in spite of its versatility and safety, MRI is not likely to replace other modalities in the near future.

ULTRASOUND

Acoustical imaging uses ultrasound (mechanical vibrations at frequencies above the range of human hearing) to form echo images of the interior of the human body. The operation of an ultrasound scanner is illustrated in Fig. 14. A transducer emits acoustic pulses that propagate along narrow beams

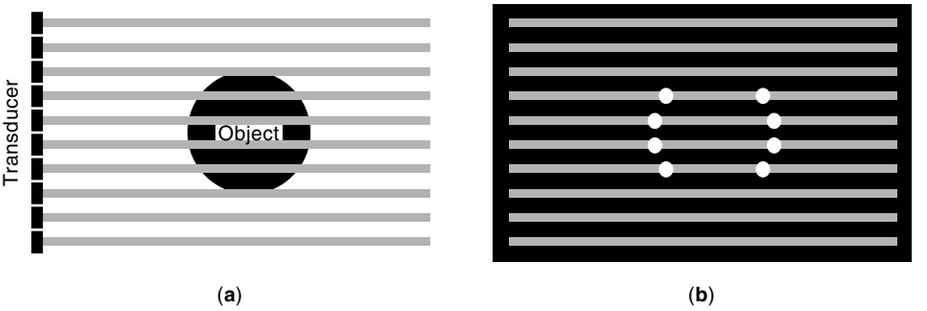

Figure 14. Schematic diagram of an ultrasound scanner. In (a), a transducer emits a sequence of acoustic pulses that travel along paths shown in gray. Echoes are produced when these pulses impinge on the front or back surface of the object. In (b) an image is formed by displaying echo signals along monitor scan lines.

through the body. Reflections (echoes) from tissues radiate toward the transducer, which converts them to electrical energy. An image is formed by collecting a set of echoes from beams sent out along parallel lines. We will describe the main physical properties that govern propagation of sound in tissues and relate them to factors that affect the quality of images obtained with ultrasound. [See also the article on ULTRASONIC MEDICAL IMAGING.]

Propagation of ultrasound through tissue is governed by the acoustical wave equation, and the wavelength of acoustical waves places critical limits on the resolution of ultrasound scanners. Wavelength, frequency, and acoustic velocity are related by $\lambda f = c$; in soft tissues $c \approx 1500$ m/s, so that, for example, the wavelength at 5 MHz is 0.3 mm. The resolution of the scanner in the depth direction is limited by the duration of the acoustic pulse and in the lateral direction by the width of the sound beam. In Fig. 15 we show the schematic pattern of acoustical energy produced by a concave (focused) transducer excited with a constant frequency. The transducer has diameter D and the focus (center of curvature of the surface) is at a distance F from the transducer. The field pattern is quite complex: approximately, the acoustic beam converges toward the focus where its diameter $w \approx 0.6\lambda F/D$. The width of the beam is relatively constant over a depth of field $d \approx 2wF/D$. To both sides of the region of focus the beam width increases linearly with distance. To obtain high resolution (small w), one should use a short wavelength (high frequency) and or a large transducer of diameter D, which leads to a small depth of field. This physical limitation can be overcome by the principle of dynamic focusing with an array transducer. A flat transducer is decomposed into elements. When an echo is received, signals from various transducer elements are delayed by appropriate amounts, producing the same sort of focusing obtained with a curved transducer surface. These delays are varied with time, allowing focus to be maintained for a large range of depth.

In practice, only soft tissues (muscle, fat, and fluid-filled cavities) can be imaged with acoustical imaging. Bone and air-filled spaces such as lungs do not propagate ultrasound well enough and are, in effect, opaque. Because resolution is limited by the wavelength, it is desirable to use as high a frequency as possible. Usable frequency is limited by attenuation. Sound power in a plane wave propagating through tissue is given by the formula $P(z) = P_0 \exp(-2\beta z)$, where P_0 is the power at the surface of the body, $P(z)$ is the power at depth z in the body, and β is the attentuation coefficient, which depends on frequency. The attenuation of ultrasound in soft tissue is approximately proportional to frequency, so that higher frequencies are attenuated more strongly and deeper structures can be viewed only with lower frequencies. The frequency used for abdominal ultrasound is about 5 MHz, and a resolution of about 0.3 mm in the depth direction and about 0.4 mm laterally can be obtained to a depth of about 20 cm. Smaller organs, such as the female breast, are scanned with higher frequencies (7.5 MHz), and still higher frequencies are used to examine the eye.

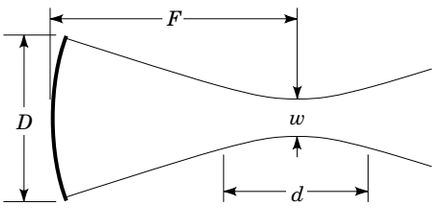

Figure 15. Depth dependent beam width in a focused ultrasound transducer.

The time to receive echoes from structures at a depth of 20 cm is about 270 microseconds, so that it is possible to emit a sequences of acoustic scanning beams and construct images in real time, at approximately video rates. One of the important uses of ultrasound is to view the heart in real time. Pathologies such as abnormal heart valve motion can be diagnosed from the temporal appearance of the moving image.

The relation between the acoustic echo and tissue characteristics is complex and not well understood. There are two main sources of acoustic echoes. Reflections are produced at tissue boundaries, where there is a change in tissue density or acoustic velocity. These echoes appear as lines at organ boundaries, or at boundaries between tissue and tumor. Such reflections are not seen if the surface is parallel to the beam direction. So-called soft-tissue echoes arise from microscopic inhomogeneities such as blood cells and muscle fibers. An acoustical image typically contains both types of signals. Figure 16 shows an ultrasound image of a pregnant uterus. The head and chest of the fetus are easily seen. The body is outlined with surface echoes, while soft-tissue echoes are produced by the lungs. Much of the image is composed of speckle, which is largely due to soft-tissue echoes, but may be caused by focusing defects or multiple reflections. The interpretation of echosonograms requires extensive training and experience, perhaps more than for other image modalities.

In addition to conventional scanners, there are many other techniques for producing acoustic images. In *Doppler imaging,* frequency differences between incident and reflected pulses are used to measure blood velocity and produce images that allow diagnosis of circulatory problems. *Harmonic imaging* produces signals from nonlinearities of acoustic propa-

gation. *Acoustic contrast agents* allow increased visibility of tissue differences. Very small imaging transducers can be introduced into body cavities or blood vessels. They operate at very high frequencies and produce high resolution images of anatomy of blood flow.

The formation of acoustical images is impossible without signal processing. Originally, acoustical scanners were constructed with analog signal processing circuitry, but modern instruments make extensive use of digital signal processing. Dynamic focusing of the transducer, Doppler signal detection, and harmonic signal analysis must be performed at high rates in real time. There is growth in the use of image processing to enhance images produced by acoustical scanners: for example, promising techniques have been proposed for reducing acoustical speckle.

Medical ultrasound images have a relatively low resolution and noisy appearance. Nonetheless, echosonography offers a number of important advantages. Ultrasound radiation is nonionizing. Even though high levels of ultrasonic energy can produce heating or mechanical injury, there are no known harmful effects from acoustic radiation at the levels used in diagnostic scanners. Consequently, ultrasonography is the preferred—almost exclusive—imaging modality for examining the pregnant uterus, where there is great concern about possible birth defects or tumors in the fetus that may be caused by X radiation. Ultrasound images are formed in real time so that the body can be explored to locate the region to be imaged and moving structures such as the beating heart can be viewed.

ECOLOGY OF MEDICAL IMAGING

What is an ideal medical imaging instrument? Cost, convenience, and safety are important considerations. However, the most important issue is whether the imager can provide definitive information about the medical condition of the patient. To have a better appreciation of these issues, it is necessary to review the conditions under which medical imaging is performed.

Medical imaging may be used as part of therapeutic procedures, to detect disease where no symptoms are present (screening), or to monitor the process of healing. The most common use of medical imaging is for differential diagnosis, where a patient exhibits symptoms that can be caused by one or more diseases, and there is a need to ascertain the specific condition so that proper treatment can be applied. The physician orders a specific *imaging procedure,* which may require patient preparation (for example, ingestion of a contrast material), the use of specific imaging instruments and settings (X rays of a certain voltage, intensifying screens, and films), and proper positioning of the patient. Most procedures are performed by properly trained technicians, but some very dangerous or invasive procedures, such as injecting contrast agents directly into the heart, are performed with direct participation of a medical doctor. The majority of imaging procedures are administered by a radiology department of a hospital, and images are interpreted (read) by radiologists, medical doctors who undergo post-MD training (residency) in this specialty. The choice of procedure for a given medical problem is based on medical knowledge and on guidelines promulgated by medical and/or health insurance organizations. We see that most medical imaging tasks are highly specialized and are performed in a very structured setting.

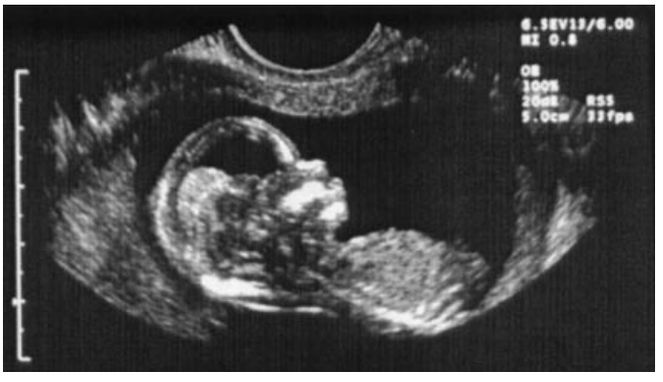

Figure 16. Ultrasound image of a pregnant uterus. The head and chest of a festus are shown. Courtesy of Dr. Barry Goldberg, Thomas Jefferson University.

Table 1. Modalities, Subdivisions, and Annual Case Load in a Large Teaching Hospital

Modalities	Subspecialities	Studies
X ray	Chest, bone, gastrointestinal, genitourinary, neurological, neurosurgical, angiography, mammography	157,958
Computer tomography	Body, neurological (head)	14,343
Nuclear medicine		10,772
Magnetic resonance imaging		12,380
Ultrasound		15,844

Table 1 shows the divisions of a radiology department of a large teaching hospital. The divisions have been grouped by modality. It also shows the number of studies performed in one year in each modality. It is clear that X-ray imaging is still the predominant technique. Other modalities are growing more rapidly, but none are likely to displace X rays in the near future.

Medical imaging procedures and modalities exist in an extremely competitive environment. For example, heart disease may be diagnosed either with angiography or nuclear medicine. Each procedure, to survive, must fit an ecological niche: it must satisfy a need where it has advantages over other methods. The same issue arises when new image processing techniques are introduced: images produced with these procedures must have an advantage over other imaging procedures. The advantage will probably be application dependent. An image processing technique that is effective for mammography may not be of much value for cardiology, and vice versa.

IMAGE PROCESSING AND INTERPRETATION[1]

Signal and image processing play an essential and growing role in medical imaging. To date, most of the applications of

[1] The work reported in this section is supported by the National Cancer Institute and the National Institutes of Health under grant numbers CA52823 and P41 RR01638.

signal processing have been in design and construction of imaging devices: indeed, novel modalities such as CT, radioisotope imaging, MRI, and acoustic scanning are impossible without signal processing. There is a modest but growing application of *post-processing,* the use of image processing techniques to produce novel images from the output of conventional imaging devices. The most common medical images, radiograms, are largely formed with analog methods. However, there is great promise for digital storage and transmission of these images. This is likely to be a large area for development and application of image compression techniques. There is also the promise, to date largely unrealized, of using pattern recognition techniques to improve on human interpretation of medical images.

As in other biomedical areas, medical image processing is driven by the needs and constraints of the health care system. Innovations must compete with existing methods and require extensive clinical testing before they are accepted.

In conclusion, the fusion of several indices and local decisions leads to a more reliable global decision mechanism to improve on diagnostic imaging.

In this part of the article, we gave a few examples of medical applications that make direct use of various aspects of machine vision technology. We also showed how to formalize many medical applications within the paradigm of image understanding involving low-level as well as high-level vision. The low-level vision paradigm was illustrated by considering the problems of extracting a soft tissue organ's structure from an ultrasound image of the organ, and that of registering a set of histological 2-D images of a rat brain sectional material with a 3-D brain. For the high-level vision paradigm, we considered the problem of combining local decisions based on different aspects or features extracted from an ultrasound image of a liver to arrive at a more accurate global decision.

BIBLIOGRAPHY

1. A. Cohen, Biomedical signals: Origin and dynamic characteristics; Frequency–domain analysis, in J. Bronzino (ed.), *Biomedical Engineering Handbook,* Boca Raton, FL: CRC Press, 1995.
2. A. V. Oppenheim and R. W. Schafer, *Discrete-Time Signal Processing,* Englewood Cliffs, NJ: Prentice-Hall, 1989.
3. W. J. Tompkins (ed.), *Biomedical Digital Signal Processing,* Englewood Cliffs, NJ: Prentice-Hall, 1993.
4. M. D. Menz, Minimum sampling rate in electrocardiography, *J. Clin. Eng.,* **19** (5): 386–394, 1994.
5. L. T. Mainardi, A. M. Bianchi, and S. Cerutti, Digital biomedical signal acquisition and processing, in J. Bronzino (ed.), *Biomedical Engineering Handbook,* Boca Raton, FL: CRC Press, 1995.
6. R. E. Challis and R. I. Kitney, The design of digital filters for biomedical signal processing. I. Basic concepts, *J. Biomed. Eng.,* **4** (4): 267–278, 1982.
7. R. E. Challis and R. I. Kitney, The design of digital filters for biomedical signal processing. II. Design techniques using the z-plane, *J. Biomed. Eng.,* **5** (1): 19–30, 1983.
8. R. E. Challis and R. I. Kitney, The design of digital filters for biomedical signal processing. III. The design of Butterworth and Chebychev filters, *J. Biomed. Eng.,* **5** (2): 91–102, 1983.
9. N. V. Thakor and D. Moreau, Design and analysis of quantised coefficient digital filters: Application to biomedical signal processing with microprocessors, *Med. Biol. Eng. Comput.,* **25** (1): 18–25, 1987.
10. R. E. Challis and R. I. Kitney, Biomedical signal procesing. I. Time-domain methods, *Med. Biol. Eng. Comput.,* **28** (6): 509–524, 1990.
11. R. E. Challis and R. I. Kitney, Biomedical signal processing. II. The frequency transforms and their inter-relationships, *Med. Biol. Eng. Comput.,* **29** (1): 1–17, 1991.
12. R. I. Kitney, Biomedical signal processing. III. The power spectrum and coherence function. *Med. Biol. Eng. Comput.,* **29** (3): 225, 1991.
13. C. L. Levkov, Fast integer coefficient FIR filters to remove the AC interference and the high-frequency noise components in biological signals, *Med. Biol. Eng. Comput.,* **27** (3): 330–332, 1989.
14. M. Akay, *Biomedical Signal Processing,* New York: Academic Press, 1994.
15. M. Akay, *Detection and Estimation of Biomedical Signals,* New York: Academic Press, 1996.
16. C. L. Nikias and A. P. Petropulu, *Higher-order Spectra Analysis: A Nonlinear Signal Processing Framework,* Englewood Cliffs, NJ: Prentice-Hall, 1993.
17. A. P. Petropulu, Higher-order spectra in biomedical signal processing, in J. Bronzino (ed.), *Biomedical Engineering Handbook,* Boca Raton, FL: CRC Press, 1995.
18. A. S. Miller, B. H. Blott, and T. K. Hames, Review of neural network applications in medical imaging and signal processing, *Med. Biol. Eng. Comput.,* **30** (5): 499, 1992.
19. E. M. Tzanakou, Neural networks in biomedical signal processing, in J. Bronzino (ed.), *Biomedical Engineering Handbook,* Boca Raton, FL: CRC Press, 1995.
20. X. Wang, H. H. Sun, and J. M. Van De Water, Time-frequency distribution technique in biological signal processing, *Biomed. Instrum. Technol.,* **29** (3): 203, 1995.
21. G. F. Boudreaux-Bartels and R. Murray, Time-frequency signal representations for biomedical signals, in J. Bronzino (ed.), *Biomedical Engineering Handbook,* Boca Raton, FL: CRC Press, 1995.
22. M. Akay, *Time-Frequency and Wavelets in Biomedical, Signal Processing,* Piscataway, NJ: IEEE Press, 1997.
23. N. V. Thakor and D. Sherman, Wavelet (time-scale) analysis in biomedical signal processing, in J. Bronzino (ed.), *Biomedical Engineering Handbook,* Boca Raton, FL: CRC Press, 1995.
24. M. Vetterli and J. Kovacevic, *Wavelets and Sibband Coding,* Upper Saddle River, NJ: Prentice-Hall, 1995.
25. G. Wornell, *Signal Processing with Fractals,* Upper Saddle River, NJ: Prentice-Hall, 1996.
26. R. W. DeBoer et al., Comparing spectra of a series of point events particularly for heart rate variability data, *IEEE Trans. Biomed. Eng.,* **31**: 384–387, 1984.
27. H. G. Steenis et al., Heart rate variability spectra based on nonequidistant sampling: Spectrum of counts and the instantaneous heart rate spectrum, *Med. Eng. Phys.,* **16**: 355–362, 1994.
28. B. Onaral and J. P. Cammarota, Complexity, scaling, and fractals in biomedical signals, in J. Bronzino (ed.), *Biomedical Engineering Handbook,* Boca Raton, FL: CRC Press, 1995.
29. E. Getin and H. Köymen, Compression of Digital Biomedical Signals, in J. Bronzino (ed), *Biomedical Engineering Handbook,* Boca Raton, FL: CRC Press, 1995.

BANU ONARAL
OLEH TRETIAK
FERNAND COHEN
Drexel University

MEDICAL TELEMETRY. See BIOMEDICAL TELEMETRY.

MEDICAL ULTRASOUND. See BIOLOGICAL EFFECTS OF ULTRASOUND.

MEISSNER EFFECT AND VORTICES IN SUPERCONDUCTORS. See SUPERCONDUCTORS, TYPE I AND II.

MEMORIES, SEMICONDUCTOR. See SRAM CHIPS.