VEHICLE NAVIGATION AND INFORMATION SYSTEMS

Travel in a vehicle has traditionally involved following the directions given by a person, posted along the road, or found from a paper map. With the rapid advance of computer, control, communication, and information technologies, these travel methods will be gradually augmented by much more powerful and convenient ones. Systems built using these advanced technologies are called vehicle navigation and information systems. The goal of these systems is to guide vehicle occupants to their destinations safely and efficiently, with less congestion, pollution, and environmental impact. This is a goal shared with intelligent transportation systems (ITS), which are transportation systems that apply advanced technologies to help their operations. In working toward this goal, ITS can take different forms. As an important portion of ITS, vehicle navigation and information systems provide the foundation necessary to understand, design, and implement advanced ITS.

A variety of navigation and information systems have been developed to assist vehicle operators, such as traffic display boards, variable message signs, rollover advisory systems, parking guidance, and ramp metering (breaking up the platoons on the ramp). In this article more emphasis is placed on the technologies used in the vehicle, in particular those relevant to guide the surface vehicle (navigation) and to advise the user about it (information) (1).

The earliest vehicle navigation and information system can be traced according to legend, back to around 2600 B.C. in ancient China. It was called "south-pointing carriage." This carriage had a two-wheeled cart on which was mounted a wooden human figure. No matter which way the cart was moving, the figure was kept continually pointed toward the south (2,3). Another interesting invention was the "li-recording (distance-measuring) drum carriage." Two wooden human figures sat on the carriage along each side of the drum. As the carriage moved, the wheels turned to drive the arms of these figures via a set of gears. An arm of one figure would strike one side of the drum once every li (about half of a kilometer); the other figure would strike the other side once every 10 li (2,3). The working mechanisms of these carriages are similar to the modern positioning technologies.

Basic positioning and navigation technologies have been incorporated gradually into modern automobiles (1,4). In the early twentieth century, mechanical route-guidance devices were introduced into modern automobiles. During World War II, an electronic vehicle navigation system was developed in the United States for military vehicles. In the late 1960s, an electronic route guidance system (ERGS) was proposed in the United States to control and distribute the traffic flow with wireless route guidance capability (centralized dynamic navigation). Similar projects were later developed and tested in Japan and Germany in the 1970s. During the same period, autonomous navigation systems were developed in the United States and United Kingdom. In the 1980s, improved autonomous navigation and information systems began to appear in Japanese, European, and American markets. These systems have been further improved in the 1990s to include many modern technologies discussed below.

There is a range of vehicle navigation and information systems from the very low end to the very high end. Most of them consist of some or all of the basic modules depicted in Fig. 1. To facilitate the discussion, we first present subsystems of vehicle navigation and information systems as separate modules and then discuss how they work together.

The modules shown in Fig. 1 can be implemented by different hardware and software components. A *positioning* module automatically determines the position of the vehicle. It can employ either integrated sensor data or radio signals to identify the coordinates of the vehicle (position) or the placement of the vehicle relative to landmarks or other terrain features (location). A variety of sensor fusion methods and radio-signal-based methods have been developed for positioning. A *digital map database* contains digitized map information. It can be processed by a computer for map-related functions, which are very similar to those provided by conventional paper maps and travel guides. Many expanded features may also be facilitated by the computerized map. *Map matching* is to use a position (or route) on a map to match the position (or trajectory) determined by a positioning module. This method can improve the accuracy of the positioning module, provided

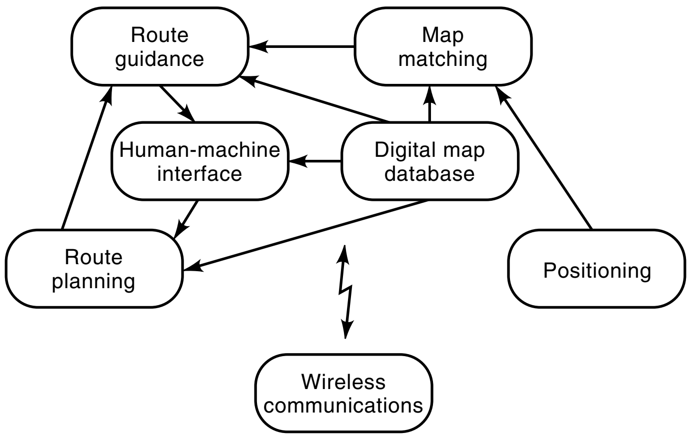

Figure 1. Basic modules for a vehicle navigation and information system. The wireless communications module interacts with the system over the air while the rest of the modules interact with each other on board. This modular approach makes system study, design, and implementation easy to accomplish. The same approach can be applied to the vehicle side and the service (traffic management/dispatch) center side of the system which is connected by a wireless communications network.

the map database has sufficient precision, normally within 15 m of "ground truth."

Route planning helps vehicle drivers plan a route prior to or during their journey. Different criteria can be used to find a minimum-travel-cost route. These include time, distance, and complexity. *Route guidance* directs the driver along the route calculated by the route-planning module. It often generates instructions either via audio or video devices to guide drivers to their destinations. A *human–machine interface* permits users to interact with the navigation and information computer and devices. Many key navigation activities must rely on this interface to control the system and to inform the user. A *wireless communications* subsystem connects the vehicle and its users with a service or a traffic management center that can further enhance the performance and increase the functionalities of the system. Certain technologies of this subsystem can also be used for vehicle-to-vehicle communications. As society continues its rapid advance to an information age, more and more people and their vehicles will depend on wireless technologies to keep them connected with others and to facilitate safe and efficient travel. Recently, Europeans coined a new term for this exciting field: *telematics,* that is, the use of computers to receive, store, and distribute information over a telecommunication system.

Based on whether remote hosts (centralized computing facilities) and wireless communications networks are involved, one can broadly divide the navigation system into either the autonomous system or centralized system. Note that different criteria may result in other classifications of these systems. In *autonomous navigation systems,* all the navigation capabilities are located solely on the vehicle. The system is responsible for a single-vehicle navigation. In *centralized navigation systems,* communications networks, host facilities, and other infrastructures work together to navigate. The system is, in general, responsible for multivehicle navigation operations.

For a large and complex system, implementation typically goes through certain well-defined phases. A top-down approach is often used, which typically includes identifying system requirements, determining functions and system architecture, specifying appropriate modules, selecting hardware components and software tools, and designing, implementing, integrating, and testing the system. Each of these individual phases needs to be iterated a few times to come up with solutions agreed upon by all the project teams and customers involved.

SUBSYSTEM TECHNOLOGIES

In this section subsystem technologies that can be used to construct a vehicle navigation and information system will be discussed. As in the last section, we first present these as separate modules, and then discuss how they are integrated into a working system.

Positioning

The positioning module is a vital component of any location and navigation system. In order to provide accurate navigation information to the vehicle occupants, the system must determine the coordinates of the vehicle on the surface of the earth (position) or the placement of the vehicle relative to landmarks or other terrain features (location).

There are three most commonly used technologies: stand-alone, satellite-based, and terrestrial radio-based. A typical stand-alone technology is dead reckoning. A typical satellite-based technology is global positioning system (GPS). A typical terrestrial radio-based technology is the "C" configuration of the LOng RAnge Navigation (LORAN-C) system. The principles behind these technologies are discussed below.

Stand-alone technology is differentiated from the others in that it does not require a communications receiver to determine vehicle position or location. A very primitive stand-alone technology is *dead reckoning,* which determines the vehicle location (or coordinates) relative to a reference point. Dead reckoning depends on the system deriving an initial vehicle position. It uses distance traveled and directional (heading) information provided by vehicular sensors to calculate a relative coordinate in two-dimensional planar space (Fig. 2). Relative distance measurements are commonly derived from the vehicle odometer, and directional information is usually provided by a gyroscope or magnetic compass.

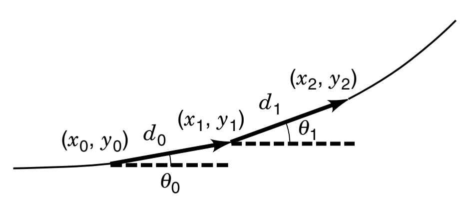

Figure 2. Dead-reckoned positioning. It is a method to continuously integrate successive displacement vectors. This method must know the starting position and all previous displacements to calculate the current position. For a vehicle, sensors are used to measure the distance traveled (d_n) and direction of travel (θ_n) or angular velocity at time t_n to a known position (x_n, y_n). Next, vehicle position (x_{n+1}, y_{n+1}) is then calculated from these measured data.

For dead reckoning, the vehicle position (x_n, y_n) and orientation (θ_n) at time t_n can be calculated from the equation

$$x_n = x_0 + \sum_{i=0}^{n-1} d_i \cos\theta_i$$

$$y_n = y_0 + \sum_{i=0}^{n-1} d_i \sin\theta_i$$

$$\theta_n = \sum_{i=0}^{n-1} \omega_i$$

where (x_0, y_0) is the initial vehicle position at time t_0, d_i is the distance traveled or the magnitude of the displacement between time t_{n-1} and time t_n, θ_i is the direction (heading) of the displacement vector, and ω_i is the angular velocity for the same time period. In practice, a constant is assumed for the sampling period (positioning data processing cycle of the embedded microcontroller).

Due to sensor inaccuracy and the assumption that the heading remains constant over the sampling period, dead reckoning generally accumulates errors as the vehicle travels. This will make the derived position of the vehicle less and less accurate. Many techniques are available to eliminate the accumulative errors, such as map-matching algorithms, short-range beacon networks (signpost), and complementary or redundant sensor compensation. Some of them are discussed in later sections.

There are a variety of sensors available to detect the distance traveled and direction of the vehicle (1,5–7). In general, surface vehicles are much less expensive than space vehicles or marine vehicles. Therefore, designers of the dead-reckoning-based system for the surface vehicle usually select low-cost sensors. There are relative sensors and absolute sensors. Relative sensors measure the distance or directional change based on predetermined or previous measurement. Absolute sensors provide the distance and directional data of the vehicle with respect to a coordinate system affixed to the earth.

For automotive location and navigation, transmission pickup sensors and wheel sensors are used for relative distance measurement. Low-cost gyroscopes and electronic magnetic compasses are utilized for direction measurement. As their names indicate, transmission pickups and wheel sensors obtain their measurements from different places in the vehicle, one from the transmission shaft and the other from the wheel shaft. Their operational principles are very similar; they convert mechanical motion into electronic signals by measuring the angular position of the shaft. Variable reluctance, Hall effect, magnetoresistive, and optical technologies are among many distance sensors that can be used for the measurement. People refer to such distance-measuring sensors as odometers. Obtaining the number of pulse counts per revolution and a proper conversion scale factor, the output of the sensor can be converted into distance traveled. Commonly used sensors are variable reluctance and Hall effect sensors. They are both electromagnetic pulse pickups, which consist of toothed wheels (exciter rings) mounted directly on the rotating component. As the toothed wheels rotate, they produce voltage waves in a sensing circuit as the teeth pass a magnet. The variable reluctance sensor uses a permanent magnet with a wire coil wound around it, which typically produces a sine wave. The Hall effect sensor uses a probe with a biasing magnet and a circuit board, which typically produces a square wave. These waveforms can be easily converted to the number of pulse counts per revolution. Due to its low cost and relative reliability, the variable reluctant sensor is still the most popular wheel sensor for the antilock braking system (ABS).

Combining the outputs from two distance sensors (odometers), one each for a pair of front or rear wheels, a differential odometer is formed. It can obtain both relative distance traveled and heading change information. Knowing the initial vehicle position, the traveled distance in the current sampling period is determined by averaging the left and right wheel rotation counts during the period and multiplying by a proper scale factor. The travel direction change in the current sampling period is determined by the difference between the counts for the left and right wheels multiplied by the same scale factor and divided by the axle length.

Gyroscopes are relative direction sensors. A popular low-cost gyroscope is the vibration gyroscope. A vibration gyroscope measures the Coriolis acceleration that is generated by the angular rotation of a vibrating bar or fork. The acceleration can be represented by the Coriolis force. This force is detected and converted to a voltage by the detector element attached on the vibrating bar and a simple circuit. Since the mass of the bar and the vibrational velocity of the bar are known quantities, the angular velocity can be easily calculated and provided to the user. Integrating this angular velocity over time, the change in the vehicle direction is obtained.

Compasses are absolute direction sensors that measure the earth's magnetic field. For on-board navigation applications, an electronic compass is better than a conventional one because of its quick response, portability, and high-vibration durability. The fluxgate compass is commonly used. Its measurement is obtained from the gating action imposed by an alternating current (ac)-driven excitation coil, which induces a time-varying permeability in the sensor core. Permeability is the property of a magnetizable substance (often made by a core) to modify the magnetic flux of the surrounding magnetic field. These varying flux lines induce positive and negative electrical current spikes in a sense coil. Two orthogonal sensing coils can be configured in a symmetrical fashion around a common core. The integrated direct current (dc) output voltages (obtained from the spikes) of these two orthogonal coils are then converted to an angle by taking the arctangent of the quotient. Since the compass itself cannot distinguish between the earth's magnetic field and other magnetic fields present, algorithms must be applied to extract the earth's magnetic field from the measurement. These other magnetic fields include those of the vehicle and of nearby objects, such as power lines, big trucks, steel structures, and reinforced concrete buildings and bridges. Nearby operational devices, such as the rear-window defroster or automatic car wash brushes, can also change the compass measurement temporarily.

Satellite-based technology uses satellites emitting radio signals to the receiver to determine the position of the receiver, often on the surface of the earth. A satellite system typically consists of a space segment (satellites), a user segment (receivers), and a control segment (monitor and control stations) as shown in Fig. 3.

Since the early 1990s, the satellite-based GPS receiver has become a dominant device for vehicle navigation and information systems. It provides an affordable means to determine

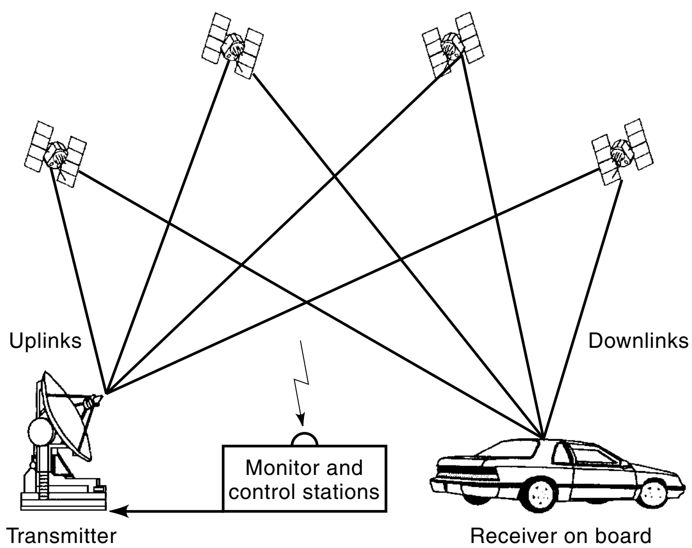

Figure 3. Satellite-based positioning. This is a typical GPS-based positioning system. The four satellites shown here are a part of a space segment, which emit radio signals from space. A GPS receiver on the vehicle is a part of a user segment, which receives the radio signals to calculate its position. The monitor and control stations are a part of a control segment, which control and monitor all the satellites in the system. These three segments make up a satellite system.

position, velocity, and time around the globe (8,9). Its distance and direction measurements are absolute (with respect to the earth). GPS was developed and is maintained by the US Department of Defense. Its constellation consists of 24 satellites orbiting at an altitude of 20,183.61 km above the earth's surface (equatorial radius of 6378.137 km). The positioning measurements of the GPS receiver are based on the time of arrival (TOA) principle. When four or more satellites are in the line of sight of the receiver, the latitude, longitude, and altitude of the receiver are determined. Two GPS service levels are provided. *Standard positioning service* (SPS) is for civilian users and *precise positioning service* (PPS) is for military users. The SPS is deliberately degraded by *selective availability* (SA). As documented, SPS provides horizontal position accuracy within a circle of 100 m radius 95% of the time (10). On the other hand, PPS provides horizontal position accuracy within a circle of 21 m radius 95% of the time. In 1996, the US government decided conditionally to phase out SA placed on SPS starting in the year 2000, subject to annual review by the US president. Even with SA, a much better accuracy can be obtained by using differential correction techniques. Differential GPS (DGPS) can reduce the position error to under 15 m, while SA is in effect. It uses a master receiver at known (surveyed) coordinates to send correcting information to a mobile receiver over a communications link for deriving a more accurate position. Typically, the receiver separation of DGPS is up to 50 km.

In addition to GPS, there are other satellite-based systems. The most notable is the global navigation satellite system (GLONASS). GLONASS was developed by the former Soviet Union and is now maintained by Russian Military Space Forces (11,12). The system does not have any measures to intentionally degrade system accuracy. Therefore, it provides a better position fix to civilian users than does GPS (SPS with SA on). Combining GLONASS with GPS as a dual receiver should provide position fixes most of the time, even in urban canyons where tall buildings are crowded in dense areas. The dual receiver can take advantage of 48 satellites instead of only 24 for a receiver based on GPS satellites alone. Other existing and planned systems include many low-earth-orbit (LEO), medium-earth-orbit (MEO, or ICO in Europe), geosynchronous (GEO) satellite systems, and DGPS-based systems, such as INMARSAT, OmniTRACS, Orbcomm, Iridium, Globalstar, ICO, wide area augmentation system (WAAS, US), European geostationary navigation overlay system (EGNOS, Europe), and multifunction transport satellite (MTSAT, Japan).

Terrestrial radio-based technology typically uses base stations or devices emitting radio signals to the mobile receiver to determine the position of its user. Signals can also be emitted from the mobile device to the base stations. Commonly used techniques are short-range beacon positioning, polar positioning, angle of arrival (AOA) positioning, time of arrival (TOA) positioning, and time difference of arrival (TDOA) positioning. The latter four are shown in Fig. 4. All these methods

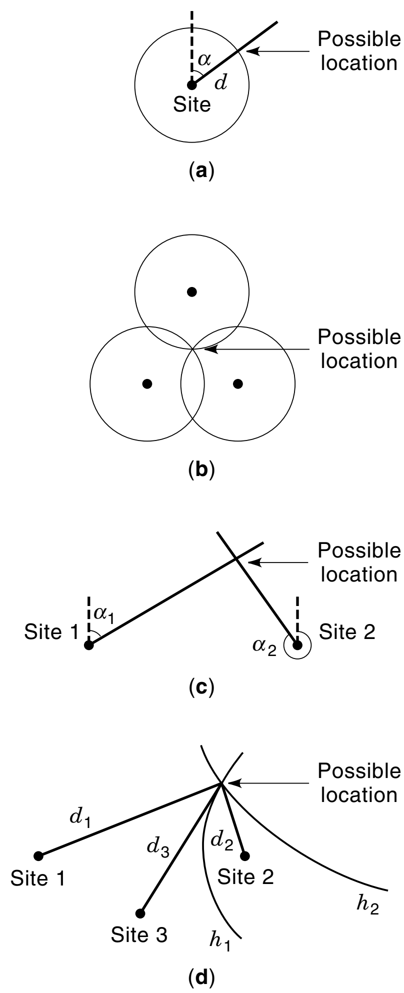

Figure 4. Terrestrial radio-based positioning: (a) Polar positioning, (b) time of arrival (TOA) positioning, (c) angle of arrival (AOA) positioning, and (d) time difference of arrival (TDOA) positioning. Multiple radio transmitters and receivers are used to determine the location of a mobile device, which can be attached to a human, a vehicle, or other objects. One or more of these four positioning technologies may be utilized.

require communications transmitters, receivers, or transceivers. In other words, they depend on emitting and receiving radio signals to determine the location of an object to which a simple reflective element, a receiver, or a transceiver is attached. To make the position determination, applications generally require that one end of the positioning system is fixed and the other end is moveable such as a mobile device attached on the vehicle. For performance improvement, hybrid methods (various combinations of the techniques discussed or with additional techniques) have also been used.

The *short-range beacon system* determines the position of the mobile device by sensing the device close to a reference point. Hence, the name of *proximity beacon* system is used in some of the literature. Since the beacon head of this system is sometimes installed on a signpost along the road, some people refer to it as the *signpost* system. Beacon systems do not always restrict themselves to signposts. One example is the loop detector, which is embedded in the road surface. Popular short-range beacon systems tend to use microwave or infrared frequency as their communication medium. Despite its robustness and low-cost mobile device, the beacon system has limited communications zones, discontinuous communication, as well as high installation and maintenance costs.

The *polar system* determines the position from a single base station that measures the distance to and the direction of the mobile device. The distance can be derived from the round-trip time to the device, which defines a circle around the station. The direction measurement forms a radiation line. If these two measurements are error free, their intersection defines the position. Despite its simplicity, for a long radius situation, even a relatively small error in a direction measurement can produce a very large error in geometric accuracy.

The *time of arrival (TOA) system* determines the position based on the intersection of the distance (or range) circles. The range is calculated from the signal transmission time, that is, multiplying the time by the speed of the signal. Two range measurements provide an ambiguous fix and three measurements determine a unique position. Within the triangle formed by the centers of the three circles, geometric accuracy is the highest. Away from the triangle, the accuracy gradually decreases. The same principle is used by GPS, where the circle becomes the sphere in space and the fourth measurement is required to solve the unsynchronized-receiver-clock bias. Because of this unsynchronized satellite and receiver clock issue, the signal transmission time determined by the GPS receiver is not accurate so the actual measurement is a pseudorange measurement.

The *angle of arrival (AOA) system* determines the position based on triangulation. The intersection of two direction measurements defines a unique position. When the two directional lines cross at right angles, geometric accuracy is the best. When they cross at 180° angle as a connected straight line, the accuracy is the worst. This technique requires only two stations to determine a position, but is susceptible to signal blockage and multipath reflection. Furthermore, a phased array of antennas is needed, which adds additional cost to the system without the phased array.

The *time difference of arrival (TDOA) system* determines the position based on trilateration. This system is commonly referred to as the hyperbolic system, where a time difference is often converted to a constant difference in distance to two base stations (as foci) to define a hyperbolic curve. The intersection of two such curves defines the position. Therefore, two pairs of base stations [at least three, as in Fig. 4(d)] are needed for positioning. Geometric accuracy is a function of the relative base station locations. Use of the TDOA technique for real-time location calculations requires a synchronized time among all base stations. On the other hand, it requires fewer antennas and is less susceptible to signal blockage and multipath reflection than using AOA.

As discussed above, many sensors can be used for positioning. However, no single sensor is adequate to provide position data to the accuracy often required by a navigation and information system. The common solution is to fuse data from a number of different sensors. These sensors usually have different capabilities and independent failure modes. People often refer to this technique as sensor fusion, which can provide the system with complementary, sometimes redundant data for its navigation and information task. Typical methods include various Kalman filters and other filters, such as low-pass, high-pass, and complementary filters (1). Fuzzy-logic, neural network, statistical decision, probability reasoning, and many other inference methods are also candidates for sensor fusion, as long as they are efficient enough for actual implementation in an embedded real-time system.

Digital Map Database

A digital map database is a module that provides map-related functions. In addition to presenting rich spatial information, as does a paper map, a digital map can be manipulated to support many navigation activities, such as locating an address or destination, calculating a travel route, guiding along a precalculated route, matching a vehicle trajectory with a known road, and providing travel information.

Two types of digital maps are used for computers: raster-encoded maps and vector-encoded maps. These maps are digitized from various paper maps, aerial photographs, census bureau data, field data, and many other sources. A raster encoded map stores in each pixel the value of each parameter of interest in a matrix over space and is digitized using a scanner. The map is displayed on the video screen systematically; the electron beam repeatedly paints a series of thin and horizontal stripes while moving from the top to the bottom of the screen. A vector-encoded map stores in a data structure the value of road network features using Cartesian geometry and it is digitized using a digitizer. The map is displayed on the screen unsystematically; the electron beam traces the outlines of the map directly, one line segment at a time.

Because it requires less storage space and is easier to manipulate, the vector-encoded map is very popular in navigation. The road network features are typically represented by one or more primitives: points, lines, and polygons. These primitives are encoded in a computer data structure as node (point), segment (line), and area (polygon) records, together with their respective attributes. Points, lines, and areas are graphic information. They are stored as coordinates, symbols, and rules. Attributes are nongraphic information. They are stored as alphanumeric characters to characterize, qualify, and link the graphic map features with their appropriate spatial locations. Examples of attributes are speed limit, street name, address range, road type, driveability, area name, city name, city range, state name, and zip code. The digital map used by vehicular navigation and information systems can also be viewed as a special subset of a geographic information system (GIS) (13), but with additional navigation attributes. GIS is a computer system used to support the capture, management, manipulation, analysis, and display of spatially referenced data for solving complex planning and management problems. The map used in turn-by-turn navigation applications, also referred to as a *navigable map* database, has great accuracy and includes attributes typically not included (or not as good) in GIS, such as nearly flawless road connectivity, one-way streets, and turn restrictions.

A map represents the geometry of the surface of the earth. The earth is an oblate ellipsoid. To represent this complex spatial information on a two-dimensional computer screen as a digital map, one must know the datum, projection, and production techniques used to make it happen. Otherwise, one may encounter some unexplainable problems later on in the design and development of the navigation and information system.

A *datum* is a set of parameters to define the location and orientation of the reference ellipsoid used to model the earth. There are global and local (regional) datums. Local datums are developed to match, as closely as possible, within the regions under consideration. Gobal datums are triggered by the earth-orbiting satellite technology. Unlike most local datums, the origin for global datum is at the center of the earth. In practice, sources for digitized maps are typically based on local datums. Popular satellite-based positioning systems rely on global datums. For instance, many North American maps use local datum North American Datum (NAD) 83 or NAD 27. GPS uses global datum World Geodetic System (WGS) 84. Blindly mixing the coordinates derived from these different datums will result in a displacement of up to 1,500 m or even more for a point on the earth's surface. More specifically, latitudes and longitudes referenced to different ellipsoids define different coordinate systems and cannot be mixed. Transformation equations between different ellipsoidal coordinate systems or between ellipsoidal and Cartesian coordinates systems are available (1,14). They are also included in some application software.

A *projection* is a technique to transform spatial data into a planar representation. For centuries, people have had difficulty representing all the features of the earth in their true relationship to each other. Projection of three-dimensional geographical information onto a two-dimensional flat plane inevitably introduces distortions to the map. Various projection techniques have been developed over the years but none of them can exactly model the reality. One can only preserve the angles (true directions) or the areas (a constant scale for distances). A valuable projection is the Mercator projection that preserves the angles. Its popular modification is the universal transverse Mercator (UTM) projection. UTM is often used for map projections involving large countries or continents. For instance, many paper maps used in the United States for local surveying and other mapping operations are based on this projection.

There are many production techniques used to generate the original source materials used to digitize the map. These techniques are beyond the scope of this article and will not be discussed here. For producing a vector-encoded map, after the map source collection, these data need to be digitized, validated, and updated. The whole process is a tedious and labor-intensive effort. A complete digital map database can be stored on a computer hard disk, a large Flash ROM, a CD-ROM disk, or a PCMCIA (PC) card. When employing map data sources, one must convert them into one unified coordinate system based on one preselected datum or reference ellipsoid. From the discussion in the beginning of this section, it is already known how to produce a digital map based on these sources. There are no widely accepted standards available to define the format of the results of either the scanning or digitizing process. For the vector-encoded maps used in navigation, quite a few standards have been proposed or are being proposed. They include the geographic data files (GDF), Japan Digital Road Map Association (JDRMA; also as an organization), spatial data transfer standard (SDTS), physical storage format (PSF) and application program interface (API). It is expected that a universal standard will be available in the near future for navigation software to interact with map databases stored on different media. For the moment, a common practice is for a map database vendor and its customers to agree on a (proprietary) software interface.

Map Matching

Map matching is a computer algorithm that utilizes a digital map to make the position determination more reliable and accurate. Unlike navigation in air and sea transportation, land transportation vehicles are basically constrained to a finite network of roads with only occasionally excursions into parking lots, driveways, or other off-road conditions. This makes it possible to correlate the trajectory of the vehicle with the road on which the vehicle travels. For fast processing and easy analysis, a vector-encoded map is commonly used. Furthermore, since the map-matching-assisted positioning systems depend heavily on the map, the accuracy of the vehicle position derived should be similar to that of the digital map database. A typical map precision requirement for urban areas is within 15 m of "ground truth."

All three positioning techniques discussed above can be assisted by map-matching algorithms. As an example, dead reckoning tends to accumulate errors with additional distance traversed. As time proceeds, the actual vehicle position will not agree with the dead-reckoned vehicle position that may already have drifted away from the road the vehicle is on. By comparing the vehicle trajectory with road segments stored in the digital map database, the error can be corrected by snapping the dead-reckoned position back to the matched position on the road of the map.

The basic principle of the map-matching algorithm is to identify the road segment on which the vehicle is traveling and determine the position of the vehicle on this segment to update the position derived by a positioning subsystem. In particular, a possibility is assigned to each candidate road segment. Low-possibility candidates are removed and the highest-possibility candidate is retained. This candidate is presented to the system to determine the vehicles position. Therefore, any inference method can be used to evaluate the possibility, as long as it is fast enough for real-time execution. Two methods have been used and proposed: probabilistic and fuzzy-logic-based.

The probabilistic method evolved from the semideterministic method (15). This conventional algorithm requires that the positioning errors be statistically propagated into the positioning determination. These errors and error models are used to define confidence regions. It is assumed that these regions may contain the actual vehicle locations. If only one road is contained in the region, a matching segment is found. If more than one road is within the region, road segment heading, connectivity, and closeness are compared against similar data received from the positioning subsystem and previous processing cycles. Finally, one most probable road segment is identified and presented to the system, along with the most likely vehicle position on that segment. Since none of the unconventional algorithms, such as the fuzzy-logic-based algorithm, have been deployed in commercial products, they will not be discussed further here. For more information about these algorithms, readers may refer to Refs. 1 and 16.

Besides matching, the digital map database can also be used to calibrate positioning sensors. For instance, the road segment length can be compared against the distance detected by the distance sensor. The road segment direction can be compared against the direction detected by the direction sensor. Once the error is above a certain threshold, corresponding correction and calibration can be done. Because many sensors could be affected by their working environments, recalibration becomes necessary. With a map, this can be done dynamically during the journey. The map and dead-reckoning sensors can also be used to detect GPS blunders caused by erratic satellite signals and fill the voids caused by signal blockage. There are many other techniques to utilize maps or other means to improve the positioning subsystem. In summary, a digital map is a good complement to other positioning technologies and has been demonstrated to produce much more robust position performance.

Route Planning

Route planning encompasses planning a route prior to or during a journey. It can be classified as single-vehicle route planning or multivehicle route planning (1). The former plans one route for a single vehicle based on the current location and a single destination (or multiple destinations). The latter plans multidestination routes for all vehicles on a particular road network. This classification is analogous to the single-source shortest path and all-pairs shortest path problems discussed in the computer science literature (17).

Travelers often prefer different route optimization criteria. Some may want to follow a shortest-distance route. Others may opt to minimize the travel time. Some may prefer to avoid expressways, minimize toll charges, or to impose a limit on number of turns and traffic lights during the trip, and so on. All these factors are referred to as the travel cost. The selection of optimization criteria can be done either by the system itself during execution or by the user prior to the planning. Actual planning needs to utilize a digital map. The travel cost must be obtained from the attributes in the digital map. For instance, distance can be derived from the road segment length. Travel time can be derived from the division of the segment length by the speed limit. As defined in the map, a street intersection or a dead-end is defined by a node, and a piece of roadway between two nodes is defined by a segment.

For *single-vehicle route planning,* the most popular algorithm for a single-destination solution is the heuristic search algorithm. For multiple destinations (*vehicle routing*), Dijkstra's algorithm, traveling-salesman algorithms, genetic algorithms, tabu searches, robust algorithms, and other algorithms can be used. Heuristic search has information concerning the number of steps or the cost from the initial state and current state to the destination state. One very popular heuristic search is the A* algorithm. Unlike Dijkstra's algorithm, which finds the optimal route from the origin node to every other node in the road network, the A* algorithm finds the optimal route from the origin node to the destination node while saving computation time and memory space. It determines which is the "most promising" node, which successors to generate, and which irrelevant search branches to prune. The decisions are made based on the heuristic information, which provides an estimate of how far a node is from the destination node. From this information, the algorithm decides the likelihood of a particular node on the best solution route and which nodes to search. It has been demonstrated that if the heuristic evaluation never overestimates the actual cost of reaching the destination node, the A* algorithm will find the optimal solution route, provided that one exists.

Generalization of the search algorithms leads to the bidirectional search algorithm. The A* algorithm usually provides a solution route from a given origin node to a destination node, which is a forward search. Reversing the search origin and destination, it becomes a backward search. If both forward and backward searches are conducted at the same time, the algorithm becomes bidirectional. When implemented on a sequential machine (single processor), the bidirectional search must switch repeatedly between the forward and backward searches. Therefore, it needs to determine the alternating period and stopping criterion to force the two searches to meet somewhere in the middle of the prepared route. If these two criteria are chosen properly, a bidirectional search may require less search time and space than a unidirectional search.

Because of the hierarchical nature of roads, a popular technique is to switch among different layers of a map that is constructed on the basis of the road hierarchy to make the search more efficient. The algorithm is named as the hierarchical search algorithm. It has been demonstrated that this algorithm can reduce the exponential complexity of the search to linear complexity. A requirement for this algorithm is that the digital map itself must be constructed hierarchically. For instance, one could make a four-layer map as follows: Layer 1 includes all the roads; Layer 2 includes collector roads, arterials, and highways; Layer 3 includes arterials and highways; Layer 4 includes highways only. In this way, one can search major roads first, starting on Layer 4, and then fills in the details for any necessary portion of the route. Note that before reaching Layer 4, the search algorithm may need to determine the nearest starting and ending nodes on that layer from given origin and destination pair in Layer 1. For route planning in a very large map database, the bidirectional heuristic-based hierarchical search is the most popular algorithm.

For *multivehicle route planning,* there are many good candidates among the algorithms proposed in all-pairs shortest path problems, dynamic programming, and operations research. This requires a central service (control) center to provide such service to the vehicle on the road via a communications network. The algorithm often computes optimal or near-optimal routes for every possible (or selected) origin-and-destination (O-D) pair or designated-zone pair in the road network. For on-demand calculation, the route computation is done for each driver on request based on the O-D pair submitted. For periodic calculation, the route computation is done periodically for all (or selected) O-D pairs or for designated zone pairs.

Route Guidance

Route guidance directs the driver along the route generated by the route-planning subsystem. There are pretrip guidance and en route guidance. En route guidance can provide turn-by-turn driving instructions in real time and is much more useful than pretrip guidance (1). To complete an en route guidance, the system requires a navigable map database, an accurate positioning, and the planned route.

For real-time route guidance, the system needs to monitor the current vehicle position and heading, and compare them with the best route generated by the route-planning subsystem. As a turn or maneuver approaches, the guidance system informs the driver with visual signals, audible signals, or driving instructions. Visual signals shown on a screen-based system could be in one of the two display formats: turn arrow or route map. The turn-arrow format displays a limited amount of information such as a turn arrow, the shape of the next intersection, and the distance to the next intersection along with a countdown bar. The route-map format displays a detailed map overlaid with the planned route. Audible signals are used to alert the driver that the system is ready to announce an instruction. These instructions are often generated from a speech synthesizer or a prerecorded set of voice messages. A typical approach is to announce an "early" instruction such as "Drive 5 kilometers to Main Street" after finishing a previous maneuver. When moving close to the next maneuver, announce a "preparing" instruction such as "Right turn half a kilometer ahead, bear right," to tell the driver to move to the appropriate lane and to begin the maneuver preparation. Once the vehicle is very close to the maneuver, inform the driver with an "approaching" instruction to tell the driver to perform a maneuver such as "Turn right at the traffic light onto Main Street." The last instruction is the most critical one. Per common sense, the guidance system must avoid announcing upcoming maneuvers either too early or too late. Otherwise, it would be difficult to convince drivers that this system could increase safety and improve performance.

There are static route guidance (or navigation) systems and dynamic route guidance (or navigation) systems. The *static navigation system* calculates the route based on historical traffic information and precollected map data. The historical traffic information may include average daily travel time for each road, turn delay data for each intersection, and other collected data. All these data can be stored in the digital map as attributes. The static system cannot respond to unpredictable road conditions. These conditions may include traffic incidents, road congestion, and road closure. Conversely, a *dynamic navigation system* calculates the route based on the real-time traffic data. It can respond properly to real-time traffic conditions. To receive real-time traffic data, there must be a wireless communications network to connect the vehicle with a service center (or traffic management center). The real-time data processing could be either in the vehicle or in the center. The network can take different forms, as will be discussed later. Real-time traffic data collection can also take different forms, such as from loop detectors, video cameras, short-range beacons, traffic reports, weather data, and equipped vehicles (as probes).

Human–Machine Interface

The human–machine interface enables the user to interact with the navigation and information computer and devices. It is a special challenge to design and implement a human–machine interface for vehicular applications (1,18,19,20). Surface vehicles, especially automobiles, have a very diverse user community. Crowded instrument panels (or dashboards) create a highly distracting environment. Any on-board interface equipment must operate under environmental constraints such as varying light conditions, varying temperature, vibration, traffic noise, and contamination by dirt or grease. In addition, the interface presents the vital first impression to the potential vehicle owner. Safety and ease of use are the most important design considerations for the human–machine interface.

Two important technologies for the in-vehicle human–machine interface are visual display and voice (input/output). For a visual display, there are many control and display technologies available. Control devices include buttons, joysticks, switches, knobs, touch screens, and voice. Voice-based interfaces are discussed shortly. One interesting control device is the touch screen. It works as a multifunction transparent switch to reduce the number of input and output devices. The touch screen is operated by the touch of a finger or stylus. It can be overlaid on any flat panel device discussed below. Display technologies include the cathode-ray tube (CRT), electroluminescent display (ELD), heads-up display (HUD), light-emitting diode (LED), plasma display panel (PDP), and vacuum fluorescent display (VFD). Recently, the field-emitter display (FED) has appeared on the market. It looks sharp from any angle, has better or equivalent brightness, contrast, and resolution, and is thinner and lighter compared with popular LCDs.

In addition to the selection of control and visual display technologies, there are many other design considerations. One of them, discussed earlier, is the route map versus the turn arrow. Studies have shown that the route map makes the driver glance more often at the display screen than the turn arrow, suggesting that it is too much of a distraction while driving. The turn-arrow display, augmented with voice guidance, could address this problem. Other design considerations are the use of an interlocking mechanism, a dimmer mechanism, and a rotational mechanism. The interlocking mechanism automatically freezes most of the control functions while the vehicle is moving. The dimmer mechanism allows the driver to control the screen brightness. The rotational mechanism is for adjusting the limited-viewing-angle display device. Additional design considerations include, but are not limited to, reducing the glance number and duration on the display screen, when, where, and how the user should be informed of an upcoming turn, what minimum text size should be used, and how the intersection should be portrayed. In the future, reconfigurable steering wheels and reconfigurable displays may be used to reduce the number of control and display devices required. Techniques to monitor the driver's state will be deployed. These may include monitoring the driver's eye and foot movements or the driving behaviors in order to prevent accidents. A warning can then be fed back to the driver, other vehicles, or traffic personnel in the surrounding area.

For voice-based interfaces, the basic technologies used are speech synthesis and speech recognition. Speech synthesis and speech recognition are suitable for output and input, respectively. Speech synthesis is a computer-aided technology that translates text into understandable human language. It is a stable technology, with error rates as low as 3.25% for individual word perception. However, the available synthesizers still do not sound natural, although progress in synthesizers appears to have reached an asymptote. Some navigation systems have chosen to prerecord human-voice driving instructions and replay them during the guidance. However, this method is not flexible enough to adapt to varying environments and may take more memory space to store voice instructions. Speech recognition is a computer-aided technique that understands human language. It is a useful technique for hands-free control while driving. The available technologies are speaker-dependent and speaker-independent with isolated-word, connected-word, and continuous-speech capabilities. Speech recognition is still not suitable for critical inputs. In general, speech recognition products are not as robust as speech synthesizer products of comparable complexity and cost. In particular, speech recognition needs further improvement in order to recognize voice commands in noisy environments. Once the word error rate has been reduced below 5% in its intended environment, a speech-recognition product is ready to be used.

Clearly, future drivers will be much better informed. Safety and ease of use will be even more important for the future human–machine interface.

Wireless Communications

Reliable communication has gradually become an integral part of vehicle navigation and information systems. Wireless communications permit many quality services to be provided to drivers (1,21,22). Vehicle users can contact each other, traffic management centers, or service centers, and receive services not available to them in the past. The vehicle has been evolving from a transportation tool to a multipurpose tool that combines transportation, information, and mobile computing. Future vehicles may be connected with the Internet, electronic-mail (e-mail), fax, video conferencing, and remote diagnosis and provide games and movies for rear-seat passengers.

Vehicle communications require an infrastructure for voice and data that can reliably and efficiently deliver real-time traffic reports and other information. These information exchanges could be between an infrastructure (base stations) and mobile devices (portable radios) or between mobile devices themselves. The mobile device can be attached to the vehicle or to the person as a wearable item. For an ideal deployment, certain key communications network attributes are necessary. These include excellent coverage, high capacity, low cost, full connectivity, and secured access.

There are many different communications technologies available and a variety of applications to apply them to, such as traffic management, emergency management, intermodal travel, public transportation operations, commercial vehicle operations, electronic payment, and advanced (or intelligent) vehicle control. Each of these applications has specific needs that may likely be satisfied by a single communications technology. On the other hand, a regional or statewide communications network is likely to require a hybrid implementation of communications technologies. As an initial guide, a high-level overview of existing wireless technologies is presented in Table 1 (1).

Some of communications techniques are now briefly discussed. *Paging* was originally developed as a one-way commu-

Table 1. Communications Technology Matrix

Technology	Reliability	Coverage	Avg. Data Transfer Rate	Equipment and Airtime Costs	Security	Simplex or Duplex	Real Time or Store-and-Forward	Access
Paging	E	Metro, some rural	2.4–3.6 kbps	E	E	Simplex/duplex	Store-and-forward/ Real time	Public
Cellular	F-E	Metro, some rural	9.6–64 kbps	F-E	P-E	Duplex	Real time	Public
PCS	E	Metro	≥8 kbps	E	E	Duplex	Real time	Public
Private land mobile radio systems	VG-E	Metro, rural	1.2–64 kbps	F-E	P-E	Simplex/duplex	Real time	Private/public
Radio data networks	VG-E	Metro	2.4–256 kbps	F-E	E	Duplex	Real time	Public
Broadcast subcarriers	E	Metro	1.2–19 kbps	E	E	Simplex	Real time	Private/public
Short-range beacons	E	1–100 m	64–1,024 kbps	F	VG	Simplex/duplex	Real time	Private
Satellites	E	Worldwide	2.4–64 kbps	P-E	E	Duplex	Real time	Private
Cordless telephony	E	Near the base	≥32 kbps	VG	E	Duplex	Real time	Public
Radio LAN networks	VG-E	Indoors/outdoors; 40–11,263 m	64–5,700 kbps	VG	E	CSMA/CD	Near real time	Private
Infrared LAN networks	E	Indoors/outdoors; 9–6,436 m	1,000–10,000 kbps	F-E	E	CSMA/CD	Near real time	Private
Meteor burst	F	Worldwide	2–32 kbps	E	E	Duplex	Store-and-forward	Private
Microwave relays	VG-E	Metro, rural; 40,000 m hops	8,448–250 × 1,544 kbps	VG-E	VG-E	Duplex	Real time	Private

P = Poor; F = Fair; VG = Very Good; E = Excellent; Metro = Metropolitan Areas; CSMA/CD = Carrier Sense Multiple Access with Collisions Detection.

nications service. Recently, two-way paging has been introduced to the market, which is better for vehicular communications. Paging is mainly an urban service and traditionally a one-way service of short data messages. One interesting application of paging is to superimpose paging signals on AM/FM radio or TV signals (broadcast subcarrier) to transmit short traffic messages. Because of possible delay of queued messages, paging networks may not be suitable for vital services requiring immediate response.

Cellular technology has been rapidly moving to the digital domain with the channel access methods of GSM, CDMA, and TDMA leading the way. For analog networks, modems are used to modulate circuit data over cellular links. Cellular was originally designed for one-to-one voice communications. Recent developments have enabled cellular networks to carry more data. This will enhance their role in vehicular applications. A network overlaying the existing analog cellular network (CDPD; see below) has been developed to address packet data applications.

Personal communications services (PCS) are newcomers. They are proposed as a multienvironment, multioperator, and multiservice infrastructure. The basic concept is to provide highly competitive wireless communications services without any restrictions on providers' capabilities. Like cellular and other handset-based technologies discussed below, PCS handsets can be used by drivers to obtain traffic information and other services.

Private land-mobile radio systems include conventional, trunked, and specialized mobile radio (SMR). Distinction between public and private systems has become increasingly blurred. Some recent systems intend to serve both the private and public sectors, such as the *integrated Dispatch Enhanced Network* (iDEN). iDEN is a digital communications system that integrates voice dispatch, wireless phone, paging, and data-transmission capabilities into one handset with cellular-like coverage. Traditionally, private land-mobile radio systems have been applied to many transportation, public safety, and industrial applications.

Popular *radio data networks* (RDN) include ARDIS, Mobitex (or RAM Mobile Data in the United States), cellular digital packet data (CDPD), and Ricochet. Both ARDIS and Mobitex were developed for short wireless message communications. Ricochet has less coverage than both ARDIS and Mobitex. It uses an unlicensed band, which may not be suitable for mission-critical applications because it must yield to licensed users when an interference occurs. As mentioned above, CDPD is a system that overlays existing analog cellular networks. The National ITS Architecture Team of the US Department of Transportation selected CDPD as the preferred network for providing traffic and route guidance information to motorists, out of those available now or near-term in the US. This technology is suitable for applications requiring wireless short messages, burst messages, or lengthy sessions, which can afford sporadic communications.

Broadcast subcarriers provide communications from a traditional broadcast station (AM, FM, or TV) without a specially allocated frequency spectrum. Unlike receiving the traffic data directly from AM/FM radio such as the highway advisory radio (HAR), subcarriers multiplex traffic data over regularly broadcast signals on a sideband frequency for textual display on specially designed radios. The most popular system is the *radio data system* (RDS, or RBDS in the United States). A special channel, the *traffic message channel* (RDS-TMC), has been assigned to handle traffic data. Transmitted traffic messages can be converted into a selected language for display in textual format on the radio. RDS can also handle location and navigation messages such as DGPS correction information.

Short-range beacons provide a communications link between vehicles and road infrastructures. Applications include proximity positioning, travel information, dynamic navigation, electronic toll collection (ETC), electronic road pricing (automatic fee collection), parking management, fleet management, and many others. Beacons can also be used as complementary devices to correct dead-reckoning positioning errors. Beacon systems typically depend on dedicated short-range communication (DSRC) protocols. Like the digital map, there is no universal agreed upon standard available for DSRC although international standard organizations are actively working in that area. Beacon communications provide high transmission rates, effective position calibration, location-oriented traffic information, and the ability to detect the vehicle parameters such as types. However, beacon communication coverage zones are very limited and communication is not continuous over time. System installation and maintenance are very costly.

SYSTEM TECHNOLOGIES

This section covers how the modern technologies discussed earlier can be used to construct a vehicle navigation and information system. There are very primitive systems that only show the map of the road network, or a moving dot representing the vehicle or the mobile device on the map. There are advanced systems that can automatically generate an optimal route to guide the vehicle or mobile device user. To facilitate this discussion, navigation systems are divided into two categories: autonomous and centralized. An autonomous system provides all the navigation activities on board the vehicle. A centralized system needs assistance from a centralized facility to conduct its navigation activities.

To design and construct a sophisticated navigation and information system, a good system architecture is very important. The system architecture determines the allocation of functions to specific subsystems and provides foundation for interface standards. It is made up based on its intended working environments. Different subsystem hardware and software components must be harmonized to work in these environments. Besides meeting the current system requirements and specifications, a good architecture can provide a stable basis for the future evolution of the system. Knowing the principles of each individual subsystem (or module), one can integrate them together into a working system under the overall system architecture.

Autonomous Navigation Systems

An autonomous navigation system provides guidance and information to the user, based solely on its on-board processing capabilities and data resources. To provide navigation, accurate vehicle location is the essential component.

As discussed earlier, GPS is the dominant technology for vehicle location determination. However, in addition to the effect of SA and receiver noise, GPS signals are subject to multipath and blockage problems. To compensate for these problems, dead-reckoning sensors are commonly used. Integration of GPS and dead-reckoning sensors requires sensor fusion technologies such as simple low-pass and high-pass filters, complex Kalman filters, and many other fusion techniques mentioned in subsystem technologies. In addition, map matching can be used to provide additional location accuracy if a digital map is available. Table 2 (1) presents a comparison of various location technologies. Some of them have also been used in centralized systems, discussed in the next section.

To provide visual aids, a human–machine interface can be integrated into the vehicle location system. It can be as simple as a black-and-white text-based display or as complex as a color pictorial display that interacts with a digital map database. The interfaces can use various input and output devices. Popular ones are switches, buttons, knobs, voice, and touch screens.

Vehicle navigation systems consist of some or all the subsystems (or modules) depicted in Fig. 1. A complete autonomous system diagram is shown in Fig. 5. Besides algorithm design, software implementation must deal with the real-time and embedded system issues. These include time, memory, and device-driver management, system synchronization, interprocess communications, mutual exclusion, error handling, multitasking execution, and so on. In addition, hardware designs post a great challenge for the in-vehicle components. They must meet the rigid specifications of the automotive environments such as vibration tests, shock limits, humidity limits, and temperature limits, normally −40° to +85°C (−40° to 185°F).

As the progress of vehicle navigation systems continues, trends indicate that many navigation and entertainment components will be shared. All the components in the shaded blocks of Fig. 5 could be used both in navigation and entertainment systems. For instance, one radio could have the functions of a AM/FM radio, a RDS radio, a digital audio broadcasting (DAB) radio, a mobile transceiver, a GPS receiver, and receivers for other guidance and information media. A CD-ROM player could be used to contain the navigation software and map on one CD-ROM, or to listen to the music stored on another CD-ROM. A memory card reader could be used for map storage or for other mobile office devices. A display monitor could be used to display a map or to watch TV programs. A computer or microcontroller could be shared with other electronic devices. Many input and output devices could also be shared by both navigation and entertainment systems.

Table 2. Comparison of Different Location Technologies

Technology	Performance
Dead reckoning (DR)	Poor (poor longer term, but good short term)
Terrestrial radio	Fair (150–2,000 m, will improve)
GPS	Fair (100–300 m with SA and without blockage)
DR + map matching (MM)	Fair (20–50 m without loss)
DGPS	Good (10–20 m without blockage)
GPS + DR + MM	Better (15–50 m continuous)
DGPS + DR + MM	Best (10—15 m continuous)

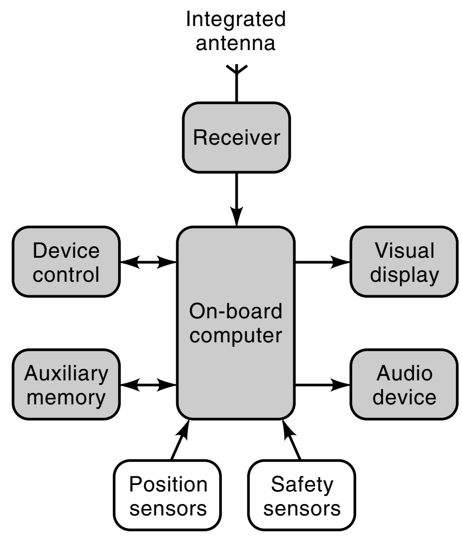

Figure 5. Simplified block diagram of an autonomous navigation system. An autonomous system can guide the user based on an onboard map to a pre-selected destination without any intervention. The guidance instructions can be displayed as a visual route, announced as turn-by-turn directions, and so on. The on-board computer is the "brain" of the system. Important actions are finalized there. The visual display, audio device, and device control are a part of a human–machine interface. The auxiliary memory can be used to store digital map data, software, or as an information and entertainment medium. The receiver may be integrated to receive both GPS positioning data and entertainment radio signals. Position sensors are used to measure vehicle distance traveled and direction of travel. Safety sensors are those that provide safety features to the vehicle user, such as the air bag deployment sensor. All the components in the shaded blocks could be used in both navigation and entertainment.

Research and development are under way to further improve driving conditions. Collision avoidance can steer the vehicle out of potential collision situations or inform the driver to take proper actions. Lane holding can keep the vehicle in its current lane. Blind-spot elimination can eliminate areas that cannot be observed by the driver. Vision enhancement can improve the driver's vision of traffic. Automated highway system (AHS) can automatically control vehicles on the road without intervention from drivers or AHS operators (23). Other existing devices, such as cruise control and air bags, will have intelligence built in to operate more adaptively to their intended environments.

Centralized Navigation Systems

A centralized navigation system provides guidance and information to the user with the assistance of a remote host or a centralized computing facility. It relies on the communications network to present additional services that autonomous systems cannot provide.

In addition to information transmission, communications systems can also provide location capabilities. The terrestrial radio-based technique is an example. It often involves the participation of one or more communications infrastructures in the position determination (see Fig. 4). In 1996, The US Federal Communications Commission (FCC) mandated that, by the year 2001, wireless communications systems must be able to locate the caller within a radius of 125 m in 67% of all cases. Many telecommunications companies are actively looking for technical solutions that can satisfy this regulation. This will certainly boost the positioning performance of terrestrial radio-based technologies.

For centralized systems, location and navigation intelligence can be placed either on the mobile unit or on the host. For instance, the mobile unit can play an active role in location determination with an on-board GPS receiver. It can also play a passive role by simply receiving a location determined by the remote infrastructure.

Automatic vehicle location (AVL) *systems* have been widely used in fleet-management systems. An AVL system tracks the position of each vehicle in a fleet and reports the information to a host via a wireless communications network. It can be used for private fleets, secure-delivery fleets, emergency vehicles, and public transportation. One popular approach is to integrate a simple positioning subsystem, a human–machine interface subsystem, and a voice and data communications subsystem in the vehicle. On the host end, a map database subsystem, a human–machine interface subsystem, and a voice and data communications subsystem are integrated into a dispatch or management center. Because of the availability of the communications network, the positioning device used on board is usually a GPS receiver that periodically receives broadcast differential corrections (DGPS) from the center. The driver has a mobile display terminal, while the dispatcher in the center has a graphical display showing the vehicle location on the map. In this way, the whole fleet can be easily controlled by the dispatch center either by voice, data (such as text commands), or both. An AVL system can be as simple as a stolen-vehicle-tracking system or as complex as a dispatching system. For a public-transit system, additional functions could be added such as a route-by-route transit schedule, enroute information (on board and at the bus station), transfer management, fare-collection registration, passenger counting, vehicle diagnosis, emergency alert, paratransit management, and on-board video surveillance. For other fleets, route planning and guidance, vehicle diagnosis, and emergency alert capabilities can be provided by the center.

Dynamic route guidance (navigation) is a typical centralized application. As was earlier discussed, a dynamic navigation system can guide the vehicle or its users based on real-time traffic data. When the navigation intelligence is placed on the host and infrastructure, the vehicle passively receives maneuvering instructions and guidance icons. The guidance data can be transmitted to the vehicle via various wireless media, such as short-range-beacons and mobile radios. When the navigation intelligence is placed on the vehicle or mobile device, all the navigation related activities are conducted on board. The vehicle then receives real-time traffic data to assist the on-board guidance. One example of such a system (Chicago ADVANCE system) with on-board components is shown in Fig. 6. In either case, vehicles equipped with a dynamic navigation unit can act as a traffic probe to collect real-time data for the traffic or service center.

Integration of all the system components is a very significant challenge. If the project involves multiple organizations, the challenge is even bigger. Meanwhile, additional systemwide impacts must be addressed. For example, if the dynamic navigation intelligence is placed in the vehicle and all the vehicles in the area are equipped with the same such system, the *Braess paradox* could occur. This phenomenon will cause all of these vehicles to be dispatched to a previously uncongested road, which may make the traffic situation in the area even worse. Despite the complicated integration tasks, the advantages of the dynamic navigation system thoroughly justify the pain of implementation.

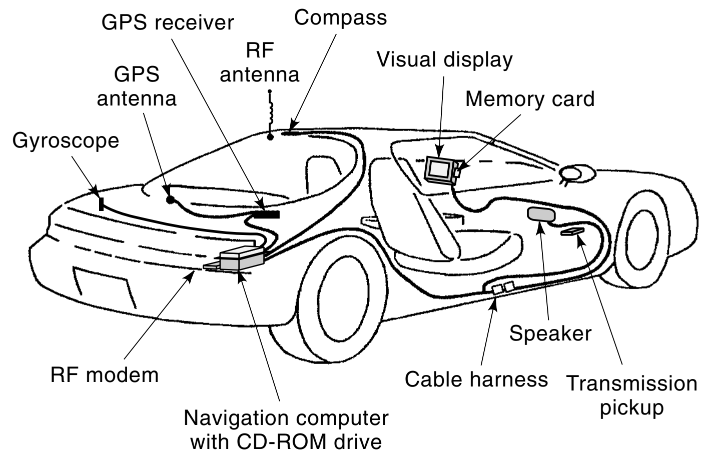

Figure 6. On-board components of a dynamic navigation system. This particular system communicates with a service (or traffic management) center via its RF antenna and modem. Real-time traffic information is used to provide for a safe and efficient journey. A GPS receiver, a transmission pickup, a gyroscope, and a compass are used for vehicle positioning. A speaker is used to announce maneuver instructions. Visual display and its touch-screen-based control are used to display information such as a map and to control the navigation computer and devices. A CD-ROM drive and a memory card are used to store and retrieve the digital map data and traffic data, respectively.

One recent development is the *mayday system,* whose on-board system integrates a GPS receiver with a cellular transceiver. The system is activated either by the user or by an emergency event to seek assistance from a service center. Therefore, a mayday system does not need to keep the communications channel open on the regular basis as most AVL systems do. Only when it is activated is the communications channel established to transfer GPS-detected location data for the verification on the center's map. It then keeps the user in voice contact with a human operator for emergency roadside assistance and other help. Some people refer to this system as the *emergency call* or *distress call* system. The mayday system can be further improved by including other services such as remote door unlocking, theft detection, stolen-vehicle tracking, vehicle diagnosis, route guidance, and travel information. As the installed base keeps increasing and the hardware unit price keeps shrinking, even more services can be added.

CONCLUSION

Great mobility has long been a dream of civilization. Modern vehicle navigation and information systems not only help to achieve this dream, but also make it more safe and efficient. More and more people have embraced this new technology. Modern autonomous vehicle navigation systems began to appear in the Japanese consumer market in the late 1980s and in the European and American markets in the mid-1990s. Dynamic navigation systems have been marketed in Japan and mayday systems have been sold in the United States since 1996. All of these systems have shown great user acceptance. In the future, vehicles will be equipped with additional devices to enhance performance, which may include collision avoidance, lane holding, blind-spot elimination, vision enhancement, vehicle-to-vehicle communications, and so on. As the explosive advance of communication, computer, control, and information technologies continues, the day when people in the vehicle can be connected with the Internet, e-mail, fax, video conferencing, remote diagnosis, and other office and entertainment devices will not be far away. Increasingly, people will benefit from intermodal travel, mobile offices, and other conveniences that modern technologies will bring. Vehicle navigation and information systems will be in every corner of the world in the twenty-first century.

BIBLIOGRAPHY

1. Y. Zhao, *Vehicle Location and Navigation Systems,* Norwood, MA: Artech House, 1997.
2. T. Tuo and T. Alu (eds.), *Song Shi* (*History of The Song Dynasty,* in Chinese), Yuan Dynasty, China, 1345.
3. J. Needham, *Science and Civilization in China,* Vol. 4, Part II, Cambridge, U.K.: Cambridge Univ. Press, 1965.
4. R. L. French, From Chinese chariots to smart cars: 2,000 years of vehicular navigation, *Navigation,* **42** (1): 235–257, 1995.
5. R. Jurgen (ed.), *Automotive Electronics Handbook,* New York: McGraw-Hill, 1995.
6. C. O. Nwagboso (ed.), *Automotive Sensory Systems,* London: Chapman & Hall, 1993.
7. W. B. Ribbens, *Understanding Automotive Electronics,* 4th ed., Carmel, IN: SAMS, 1992.
8. B. W. Parkinson et al. (eds.), *Global Positioning System: Theory and Applications,* Washington, DC: American Institute of Aeronautics and Astronautics, 1996.
9. E. Kaplan (ed.), *Understanding GPS: Principles and Application,* Norwood, MA: Artech House, 1996.
10. DoD/DoT, *1994 Federal Radionavigation Plan,* Springfield, VA: National Technical Information Service, Report No. DOT-VNTSC-RSPA-95-1/DOD-4650.5, May 1995.
11. R. B. Langley, GLONASS: Review and update, *GPS World,* 46–51, July 1997.
12. Russian Military Space Forces, *GLONASS* [Online]. Available at http://www.rssi.ru/SFCSIC/glonass. html, 1998.
13. M. N. DeMers, *Fundamentals of Geographic Information Systems,* New York: Wiley, 1997.
14. C. Drane and C. Rizons, *Positioning Systems in Intelligent Transportation Systems,* Norwood, MA: Artech House, 1998.
15. R. L. French, Map Matching Origins, Approaches and Applications, *Proc. 2nd Int. Symp. on Land Vehicle Navigation,* 1989, pp. 91–116.
16. C. Nwagboso (ed.), *Advanced Vehicle and Infrastructure Systems: Computer Applications, Control and Automation,* Chichester, UK: Wiley, 1997.
17. T. H. Cormen, C. E. Leiserson, and R. L. Rivest, *Introduction to Algorithms,* Cambridge, MA: MIT Press, New York: McGraw-Hill, 1990.
18. B. Peacock and W. Karwowski, *Automotive Ergonomics,* London: Taylor & Francis, 1993.
19. Y. I. Noy (ed.), *Ergonomics and Safety of Intelligent Driver Interfaces,* Mahwah, N.J.: Lawrence Erlbaum Associates, 1997.
20. W. Barfield and T. A. Dingus (eds.), *Human Factors in Intelligent Transportation Systems,* Mahwah, N.J.: Lawrence Erlbaum Associates, 1998.
21. S. D. Elliott and D. J. Dailey, *Wireless Communications for Intelligent Transportation Systems,* Norwood, MA: Artech House, 1995.
22. J. D. Gibson (ed.), *The Mobile Communications Handbook,* Boca Raton, FL: CRC Press, 1996.
23. P. A. Ioannou (ed.), *Automated Highway Systems,* New York: Plenum Press, 1997.

YILIN ZHAO
Motorola, Inc.