Design of TravTek Auditory Interface

LINDA G. MEANS, REBECCA N. FLEISCHMAN, JANETH T. CARPENTER, FRANCIS E. SZCZUBLEWSKI, THOMAS A. DINGUS, AND MARK K. KRAGE

In an advanced driver information system, computer voice messages and nonverbal auditory signals provide a means of imparting information to the driver without creating a visual distraction. In the TravTek system, synthesized voice has been used extensively as a supplement to the visual display, providing route guidance instructions, navigation assistance, and traffic information. Special consideration has been given to strategies for maximizing the acceptability of synthesized voice to drivers. TravTek voice messages are designed as concise utterances that aid the driver in reaching the destination easily, quickly, and safely. Data collected from onboard computers, controlled experiments, and driver questionnaires will be analyzed to evaluate user acceptance and feature preferences.

Computer-generated voice has been widely identified as a useful way to impart information to drivers in an advanced driver information system (ADIS) (1). A typical ADIS may use voice to present navigation and traffic information without creating a visual distraction. There is, however, a common perception in the North American automotive community that people do not like talking cars. In the TravTek ADIS, the authors have attempted to understand the objections that drivers may have to the use of voice in cars and to develop strategies that enhance the user acceptance of synthesized voice.

TRAVTEK DEMONSTRATION

TravTek is an ADIS operational field test conducted jointly by General Motors, FHWA, the American Automobile Association, the Florida Department of Transportation, and the city of Orlando, Florida. TravTek has been in operation in Orlando with 100 vehicles available as rental cars and for use by local drivers. ADIS features of the TravTek system include route planning and guidance, real-time traffic information, navigation assistance, and an onboard services and attractions database. Fleischman et al. (2) and Carpenter et al. (3) describe the functionality and human factors design of the TravTek driver interface. Rillings and Lewis (4) and Krage (5) provide details of the TravTek program and the architecture of the vehicle subsystem.

VOICE FUNCTIONS IN TRAVTEK

Synthesized voice is used extensively in the TravTek system, providing route guidance instructions, navigation assistance, and traffic information. Nonverbal auditory signals are also used in a limited way as feedback for button presses, as well as prompting for screen glances when voice functions are deactivated. Voice functions are controlled by the driver through the use of four buttons on the steering wheel. These buttons are labeled “Where Am I?”, “Repeat Voice,” “Traffic Report,” and “Voice Guide.” The voice controls allow a driver to select the types of information desired through voice, allowing all voice messages to be disabled if necessary.

Where Am I?

The Where Am I? function provides information on the vehicle’s current street location and the name of the next cross street ahead of the vehicle. Each press of the Where Am I? button elicits a single message with location information. Special messages have been formulated to accommodate situations in which the vehicle is not situated on a known street or no known cross streets are ahead of the vehicle.

Repeat Voice

The Repeat Voice function enables the driver to replay the most recently spoken voice message. The message, when repeated, is prefaced by the statement “The last message was . . . .” The replay is a literal repeat as opposed to a functional one. Even if the information has changed since the message was originally spoken, the original text is spoken in the replay. The repeatability of a message times out after a short period, to avoid repeating messages in which the information content is extremely outdated. After the timeout, a button press of Repeat Voice produces the message “No recent message to repeat.”

Traffic Report

The Traffic Report function provides a verbal traffic advisory that reports known traffic problems. Traffic data are broadcast to the vehicles once a minute from the TravTek traffic management center (TMC). Traffic data contain information on congestion problems within the TravTek map area, as well as details on incidents and construction when such information is known.

Onboard the vehicle, the traffic data are displayed on a color map using incident and congestion icons. Traffic problems are filtered for relevance to the vehicle's location and route. With each new broadcast, a voice traffic message is formulated to describe each geographically relevant traffic problem. Voice traffic messages contain information on the location, cause, and severity of traffic problems. The set of relevant traffic messages constitutes the voice traffic report. When the Traffic Report function is activated by a button press, the voice begins to speak the current traffic report, with messages ordered by urgency and proximity to the vehicle. A subsequent button press will terminate the report. While the Traffic Report function remains activated, as subsequent TMC broadcasts are received, all new, relevant traffic problems are spoken.

At the start of a trip, the Traffic Report function defaults to an off state. Voice traffic reports are spoken only if requested by the driver. When Traffic Report is switched off and then back on, the full, relevant, ordered set of traffic messages is respoken, thus enabling the driver to hear all traffic problems.

**Voice Guide**

The Voice Guide button enables and disables voice route guidance instructions. Voice guidance messages describe upcoming maneuvers on the planned route and indicate an off-route condition and destination arrival.

Up to three voice messages may accompany the upcoming maneuver. If the distance to the maneuver is so great that the driver need not attend to it yet, the voice guidance message specifies only the distance to the maneuver, corresponding to a straight-ahead arrow on the visual guidance display: for example, “Ahead, next turn in three and four-tenths miles.” At a closer distance to the maneuver, when the driver must get into the appropriate lane in anticipation of a turn, a “near-turn” message is spoken. The near-turn message contains the distance to the maneuver and the name of the turn street and the type of maneuver (e.g., “make a hard left” or “bear right”). A typical near-turn message is “In eight-tenths miles, turn right onto the ramp to I-4 east.” This corresponds with a change in the visual guidance display, which now deploys the geometry of the maneuver intersection and displays the name of the turn street. Just before the maneuver intersection, at a point at which the driver can be expected to identify visually the turn street, voice guidance speaks an “at-turn” message, which contains the same information as the near-turn message but eliminates the distance information. This informs the driver that the maneuver is imminent.

When an upcoming maneuver is followed closely by another maneuver, the message for the first maneuver alerts the driver to prepare for the second one. An example is “Bear left to follow the correct branch of Sand Lake Road, then prepare to turn right.” This helps the driver position correctly after the first maneuver in anticipation of the second one.

When Voice Guide is switched off and then back on again, a distance-appropriate guidance message is spoken. In this way the driver has the ability to force the system to speak an instruction for the next maneuver at any time. The Voice Guide function is automatically enabled at the start of a trip with a planned route, as voice guidance is intended to reduce glances at the visual displays.

**TRAVTEK PHILOSOPHY ON VOICE ACCEPTABILITY**

The auditory mode, if implemented effectively, has great potential as a means of imparting complex information to the driver. Voice messages elaborate on the information depicted on the visual display, providing an eyes-on-the-road mode of informing the driver. Voice messages also draw the driver's attention to the fact that new information is now available, so the driver need not glance at the visual display frequently to check for an update. In TravTek, voice functions are considered to be a supplement to the visual display, which can be used as a stand-alone system.

The TravTek driver interface design team has strived to make the application of voice to an ADIS palatable to the driver. The precepts that the authors have applied to the design of the auditory interface include

- Minimizing voice “chattering” and “nagging”;
- Maximizing voice intelligibility;
- Providing timely, useful information through voice; and
- Allowing significant driver control of voice functions.

**Talking Cars**

“People don't like talking cars.” This bit of folklore has presumably arisen from negative reactions to vehicles that use voice to warn driver of open doors and unbuckled seat belts. An examination of the use of voice for such purposes reveals violations of some of the aforementioned TravTek precepts.

Drivers may not be receptive to the use of voice for system warnings unless the condition is urgent, such as a collision warning. In the case of an open door, a nonverbal auditory signal or a telltale on the instrument panel is probably enough to alert the driver to the condition. The driver may perceive the voice in this instance as nagging, because the voice only speaks to say that the driver has done something wrong. Drivers may have various reasons to suppress a voice system at times, and these wishes must be accommodated by giving the driver control over volume as well as activation of voice functions.

Past negative reactions to misapplied voice technology do not necessarily bode poorly for future well-considered use of voice on passenger cars. The initial experiences with the TravTek auditory system indicate that the computer-generated voice that provides useful information at the appropriate time may be welcome to drivers, especially for route guidance.

The TravTek evaluation plan provides a mechanism for assessing user acceptance and actual usage of the voice functions (6). Driver interaction data that are logged onboard the vehicle serve as one source of feedback on how drivers use the system. Drivers also express subjective reactions to their TravTek driving experience through debriefings and questionnaires. Controlled experiments were designed that may analyze benefits provided by voice guidance. By analyzing TravTek data from all these sources, the authors expect to learn a great deal about driver behavior and preferences for
ADIS features. In particular, analysis of the use of voice controls will provide a measure of the usefulness and the affective impact of the auditory system. Preliminary evaluation of TravTek guidance displays indicates that drivers use voice guidance and perceive it as very helpful and that navigation with visual guidance displays is improved when combined with voice (7,8).

Chattering and Nagging

As anthropomorphism is inevitable in a talking car, the authors have chosen a conservative approach to the sort of “personality” that may be attributed to this ADIS. Anthropomorphism may be lessened by designing voice systems that are somewhat machine-like in their expression. ADIS with excessively long voice messages, or messages that exceed strict bounds of usefulness, may be accused of chattering or nagging. Drivers are more likely to take kindly to chattering and nagging cars than they do to passengers who exhibit the same characteristics.

“Auditory clutter” is the term that Stokes et al. (9) use to describe the overuse of the auditory channel, resulting in potential distraction from the driving task and in annoyance. To minimize auditory clutter, the authors avoid voice feedback for correct maneuvers, driving speed, and system status—uses that Davis advocates (10). Initial road tests of the TravTek driver interface convinced the authors to reduce further the length and number of voice messages, which had been thought to be quite minimal to begin with.

VOICE INTELLIGIBILITY

Computer-generated speech for ADISs may be either synthesized or digitized. Digitized speech has the significant advantage of intelligibility and naturalness, along with the disadvantage of prohibitive limitations in recording and storing large amounts of text. In TravTek, synthesized voice has enabled the use of a large variety of messages in an implementation that achieves an acceptable level of intelligibility.

Female Versus Male Voice

Although the TravTek voice sounds indisputably male, this should not be interpreted as a deliberate design decision. The selection of a speech synthesis product was based on hardware requirements for durability in an automotive environment. The authors had little choice regarding voice characteristics and had to settle for the voice available in a product that satisfied our constraints. Although the synthesizer does allow for programmer control over speech attributes such as rate of speech, pitch, and voice gain, it does not provide a choice between a male and female voice. It does, however, offer the choice of a large person, a medium-sized person, and a small person. The male-sounding TravTek voice is the medium person (the other two voices were largely unintelligible).

A female digitized voice was preferred for use in military aircraft in the 1960s (11) because it contrasted with the predominantly male voices in flight crews and control tower radio communications of that era. This motivation for using female voice clearly does not apply to automobiles, where voices coming from passengers, radio, or car phones are equally likely to be male or female. Subsequent research suggests that the human female voice is less intelligible than the human male voice (12). Additional research is needed to assess the relative intelligibility of male and female synthesized speech, as well as machine speech that is not characteristic of either sex.

Intelligibility of Synthesized Speech

Synthesized speech is decidedly less intelligible than digitized human voice. The reduced intelligibility has been demonstrated to stem from the absence of many prosodic features that are found in natural human speech (13). The prosodic element of speech is that which gives natural speech its rhythm. A state-of-the-art voice synthesizer applies some reasoning to the insertion of pauses and variations in pitch and stress; however, the inability of the system to perform a syntactic parse of input sentences severely limits the prosodic results. Many commercially available voice synthesizers provide a way to insert prosodic markers in text, giving the programmer some control over the intonational pattern of a synthesized utterance.

Similarly, a speech synthesizer typically contains a large dictionary of stored pronunciations for known words, as well as a program that generates a pronunciation for an unknown word on the basis of its spelling. In English, spelling is not a good predictive measure of pronunciation, and no known algorithms will produce consistently accurate results in pronouncing unknown English words. For the TravTek voice messages, prosody has been specified carefully to enhance intelligibility. The authors have also listened to the voice synthesizer pronunciation of all words that may be spoken by the system, including more than 12,000 Orlando street names, storing corrected pronunciations as needed. This effort has resulted in large improvements in intelligibility, and this preprocessing is considered to be essential for public acceptance of synthesized voice.

Other strategies are also effective in increasing the intelligibility of synthesized speech. Simpson and Hart discuss the importance of providing sufficient linguistic context for synthesized warnings and commands to enhance comprehension (14). The authors have found that in TravTek route guidance and traffic messages, the street names are the least intelligible part of the utterance. To aid in the comprehension of street names, it is useful to speak the street name suffix (e.g., “Colonial Drive” as opposed to “Colonial”). Alerting prefixes are thought to be effective in attracting the listener’s attention to an impending voice message. Various studies have indicated a reduced response time for prefixed messages, despite the increased length of the messages because of the prefix (15,16).

The mechanical-sounding characteristics of synthesized voice may have some advantages over digitized human voice in automotive applications. Because the voice does not sound human, it is easily and immediately distinguishable from other voices in the automotive environment. In this way, the perceptual contrast makes the voice somewhat self-alerting; it is obviously the car speaking. The machine-like voice also may
reduce the tendency toward anthropomorphism, as mentioned before.

USEFULNESS OF INFORMATION

Useful information in an ADIS is that which enables the driver to optimize the performance of the driving and navigation tasks. Route guidance instructions are useful if they help the driver to follow a route safely and without error. Traffic information is useful if it enables a driver to avoid congestion, minimize travel time, and drive safely in unavoidable traffic congestion. There are many open questions as to the appropriate information content of guidance and traffic messages in an ADIS.

Route Guidance

People typically include the names of turn streets in route guidance instructions. Are street names a necessary or useful piece of information? The authors believe that they are, given the difficulty of timing turn messages accurately enough to prevent erroneous turns in areas with closely spaced cross streets. Street names also help drivers to orient themselves in unfamiliar territory.

At an intersection where two streets cross at right angles, the instruction describing the maneuver is easily formulated: "turn right" or "turn left." Complex intersections entail maneuvers that are more difficult to describe unambiguously. Although this is another justification for the use of street names, it is important to note that street signs are not always easily visible. For this reason, voice guidance, graphical representation of the maneuver intersection, and error recovery are all important elements of robust route guidance. Combined in a coherent system, they work effectively to keep the driver on the planned route.

It is useful for drivers to know how far they are from their next maneuver. Individual drivers, however, differ in their abilities to reason about distance. Some football fans might find meaning in distances measured in yards, but not in fractions of a mile. Davis (10) and Streeter et al. (17) have discussed the potential ambiguity of other measures of distance. For instance, "in two blocks" may be ambiguous when the next cross street does not intersect the driven street on both sides; does it delimit a block? Upon hearing an instruction like "Turn left at the third light," will the driver count the light at the intersection he or she is passing through when the message is spoken? The authors conclude that distance expressed unambiguously (e.g., in fractions of a mile) is more likely to help drivers who can gauge it than confuse drivers who cannot.

Navigation Assistance

The Where Am I? function was designed primarily to provide assistance when the driver is navigating along a self-planned route. It helps drivers to identify the intended turn street and to orient themselves when uncertain of the current location and heading. Where Am I? can be a useful supplement to route guidance as well, identifying turn streets when street signs are missing at intersections or visibility is obscured. It also indicates navigation system accuracy without reference to the map display.

An example of the Where Am I? message is "Approaching Lee Road. Headed north on Orlando Avenue." At first glance, it may seem counterintuitive to answer the question "where am I?" by first stating what you are near, then where you are. The authors chose to word the message in this way to impart the presumably more urgent piece of information first. Cross street information is considered to be more urgent, because it may identify the location of the intended maneuver and it changes more frequently than driven street data.

The Where Am I? function is self-interrupting. It aborts its current message and restarts itself when the button is pressed during Where Am I? output. Repeated, quick button presses will result in a series of cross street notifications, with the location information clipped. This functionality enables the driver to use Where Am I? to listen to cross streets in quick succession while proceeding down the road. The tenseness of the Where Am I? message further serves this purpose.

Research Issues in Traffic Advisories

There is not yet a good understanding of which pieces of information about a traffic problem are necessary and useful to drivers. Specifically, it is not clear whether knowing about lane closures, or the cause of a congestion problem, may cause drivers to modify their behavior. Perhaps it is beneficial to suppress clarification of a spectacular incident such as a fire, because it could encourage gawkers to travel to the scene. On the other hand, it also alerts drivers to the possibility of emergency vehicles in the area.

There are a variety of ways to express the severity of a congestion problem. Terms such as "heavy traffic," "sluggish traffic," "1-minute delay," "bumpier to bumper," "stop and go," "slow and go," and "merging delays" are used in traffic reports broadcast by radio stations. Radio traffic reports also occasionally provide an estimate of the length of a congestion queue. More research is needed, however, to determine how drivers use an estimate of a backup queue, whether it can be reliably estimated, and how best to describe it to drivers: in miles, or number of traffic signals, or from Street X to Street Y. The word "congestion" itself may be ambiguous: do drivers interpret it consistently as slow traffic, or can it also refer to a heavy volume of traffic moving at the posted speed?

The location of a traffic problem can also be expressed in various ways. Can the relevance of the problem be assessed more easily by the driver if its location is described relative to the vehicle, or in absolute terms?

Onboard computer-generated traffic advisories can provide information on demand that is filtered for relevance to a given vehicle location and planned route. Some issues that arise in relevance filtering include the criteria that are applied to determine relevance, the upper limit on the amount of information that should constitute an on-demand traffic report, and the possibility of enabling drivers to tailor traffic reports to their own needs and interests.
TravTek Approach to Traffic Reporting

TravTek traffic advisories report lane closures and the cause of an incident when known. Congestion is characterized as heavy or moderate, depending on the degree to which travel time on the affected road varies from free-flow travel time (18). The location description of a traffic problem varies according to whether the vehicle is on a planned route and whether the traffic problem is related to an incident or not. The location of an incident (e.g., accident, disabled vehicle, malfunctioning signal, construction) can be pinpointed, whereas congestion is so volatile that it cannot be delimited precisely and reliably. When the vehicle is on a planned route, only problems ahead of the vehicle on the route are reported. If an on-route problem is an incident, its location is described as "[street name] ahead on [street name]," where distance is expressed in miles. A nonincident congestion problem on the route is specified as "[street name] ahead on [street name]," which avoids an estimation of the distance to the tail of the congestion queue. When the vehicle is off the planned route, or no route has been planned, traffic problems are located as "On [Street X] between [Street Y] and [Street Z]." The visual display indicates incidents and congestion with icons on the local area map, thus clarifying the location of problems reported by the voice traffic report.

The collection, dissemination, and in-vehicle use of traffic data is a process whose design is subject to many interdependencies. The TravTek solution to onboard presentation of traffic data was largely driven by outside constraints imposed by the organization of the TravTek TMC. The design of the TMC was itself constrained by local availability of information sources. In future ADIs, it would be preferable to base the information content of traffic advisories on solid research in the usefulness of the information, constrained by the general feasibility of data collection in most large urban areas.

**DRIVER CONTROLS**

It is essential that drivers be allowed to select functions for which they receive voice output, control the volume of the voice, and suppress all voice messages. In the TravTek system, when a voice message is about to be spoken, the radio is muted for the duration of the voice output. Activation of the radio volume control during voice output adjusts the volume of voice messages; to adjust radio volume, the driver uses the volume control during radio output. This capability enables differing volume levels for radio broadcasts and for TravTek voice functions.

In initial testing, the authors have found that voice messages for TravTek functions are generally welcome when the driver is not listening to the radio or conversing with a passenger. Because the voice synthesizer suppresses radio output and tends to interrupt conversations, drivers may occasionally want to turn off all or some voice functions. Separate controls for voice guidance, traffic reports, and navigation assistance allow the driver to reduce the amount of voice selectively, instead of having only a single voice on/off control.

While a driver's attention is occupied by the driving task and competing thought processes, the driver may not immediately attend to a voice message that is issued automatically. Because the auditory display is inherently ephemeral, the Repeat Voice function provides a necessary mechanism to recapture information that may not be initially apprehended.

**NONVERBAL AUDITORY SIGNALS**

Considerable research has been done in the use of nonverbal auditory warnings in aircraft cockpits [see work by Patterson (19) for a comprehensive discussion]. Some of the knowledge that has accrued from aircraft research may pertain to passenger vehicles (e.g., appropriate volumes and temporal characteristics for auditory tones). However, principles guiding the use of auditory systems in aircraft must not be applied indiscriminately to passenger vehicles. It is important to remember the essential differences between highly trained cockpit personnel and automobile drivers who range widely in age, driving abilities, physical condition, and so on.

When ADIs become so commonplace to be available to untrained drivers, the meaning of auditory signals must be easily learned and retained, with minimal potential for confusion. In the TravTek system, three nonverbal auditory signals are used. Two are tied directly to driver actions: a feedback signal for touchscreen key presses, and an error tone for inappropriate steering wheel button presses (for instance, pressing Voice Guide when no destination has been entered). The third signal must be taught. When the Voice Guide function is turned off, a glance-at-the-screen tone is sounded to inform the user that new information is presented: when the next maneuver is first depicted, or when the car has left its planned route. The glance-at-the-screen tone is soft and unobtrusive; it should not startle drivers or disrupt conversations.

**CONCLUSIONS**

Computer-generated voice is a desirable component of ADIs; however, the North American automotive community has little experience in marrying voice to cars. The TravTek auditory interface strives to maximize the acceptability of the application of synthesized speech to an ADI. The voice messages were designed as concise utterances that help the driver reach the destination quickly, easily, and safely. Data logged on onboard computers and extracted from driver questionnaires will provide feedback on user acceptance and actual usage of the various voice functions. Further research is needed to determine the optimal information content of route guidance and traffic messages.

**REFERENCES**


