System Safety

Emerging Multimodal Technology

Role in enhancing combat and civilian system safety

By Ellen C. Haas

OVER THE PAST FEW YEARS, the equipment being developed and fielded is making military systems increasingly dynamic and more cognitively demanding, requiring soldiers to simultaneously monitor multiple displays, operate multiple controls and process large amounts of information. U.S. Army, Navy and Air Force researchers have been exploring advanced control and display technologies to make military systems safer and more effective to use under such challenging conditions. These technologies include spatial auditory displays, skin-based haptic and tactile displays, and automatic speech recognition (ASR) voice input controls. When used by themselves or collectively, displays involving more than one sensory modality (also known as multimodal displays) can enhance soldier safety in a wide variety of applications. Multimodal controls and displays are also found in civilian applications.

Spatial Audio Displays

Until the advent of electronic digital technology made it easy to design and synthesize auditory cues, most system displays were visual. However, when a user's visual system is overburdened (e.g., when a pilot must simultaneously monitor an aircraft's visual displays and look out of the cockpit window when flying in challenging meteorological conditions), s/he may not detect a critical event in the environment. System designers discovered that audio cues were useful—by themselves or as a supplement to visual feedback—because they can increase awareness of surroundings, cue visual attention and convey a variety of complex information to the user, especially when the user experiences a high visual load (Shilling & Shinn-Cunningham, 2002).

Technological advances have made it possible for auditory cues to be presented spatially. With spatial audio displays, also known as 3-D audio displays, a listener perceives spatialized sounds that appear to originate at different locations and distances from outside the head (just like audio cues that naturally occur in the environment). These 3-D audio displays permit sounds to be presented in different horizontal, vertical and distance locations that are meaningful to the listener.

Often, earphones are used to present spatial audio cues. Although loudspeakers may be used, their use may be problematic (Shilling & Shinn-Cunningham, 2002). Before the audio cues reach the earphones, they are filtered through computerized sound filter functions known as head-related transfer functions (HRTFs). These HRTFs provide the sound with specific time, intensity, phase and reverberation cues. The result is sound that upon output is heard at different locations in space. A headtracker is often used to provide a stable reference point for the audio cues.

Because each sound is presented in a different spatial location, listeners can use that spatial cue to selectively attend to more than one sound at a time as well as to sounds at designated locations. Relevant 3-D audio safety applications include mitigating aircraft pilot spatial disorientation and providing meaningful direction-related system warnings.

Spatial Disorientation

For military and civilian pilots, spatial disorientation (SD) is defined as the failure to correctly sense the attitude, motion and/or position of the aircraft with respect to the surface of the earth. SD is a temporary condition caused by the pilot being deprived of an external visual horizon critical to maintaining a correct sense of up and down while flying (Benson, 2003). It has been found to be caused by flight into weather conditions with low or no visibility, nighttime flight, instrument failure and conditions with high pilot visual load, such as combat conditions (Collins & Harrison, 1995). If not appropriately resolved, SD can cause aircraft crashes and fatalities.

As a result, SD is an expensive safety concern. Perceived or actual errors in aircraft control are estimated to cost the U.S. Air Force $150 million to $200 million per annum in aircraft accidents (Collins & Harrison, 1995). Thirty percent of U.S. Army military helicopter accidents involved SD as a significant factor (Braithwaite, Durnford, Crowley, et al., 1998). Results

Ellen C. Haas, Ph.D., *works at the U.S. Army Research Laboratory at Aberdeen (MD) Proving Ground where she heads the Multimodal Controls and Displays Laboratory. She holds a B.A. from the University of Colorado, an M.S. from California State University, Long Beach, and a Ph.D. in Industrial and Systems Engineering from Virginia Polytechnic Institute and State University. She has published widely on the topic of auditory and tactile displays and speech recognition, and coedited the book* The Ergonomics of Sound. *Haas is the founder and chair of the Auditory Ergonomics Technical Committee of the International Ergonomics Association, a member of the Human Factors and Ergonomics Society and an associate member of Division 21 of the American Psychological Association.*

Abstract: *This article describes how spatial audio and skin-based control and display technologies can be used to enhance system safety in military and civilian applications. The article describes how automatic speech recognition controls are relevant to military and civilian system safety, and shares practical considerations associated with the efficient use of these technologies, including technology demands and limitations.*

of a helicopter pilot survey by Braithwaite, et al. indicated that 78% of aircrews reported having been spatially disoriented—8% to the extent that flight safety was threatened, especially during combat operations, when their eyes were busy and they could not maintain instrument scans (Braithwaite, et al.).

SD is also a problem in civil aviation. A search of National Transportation Safety Board SD-related civil aviation mishaps and "SD and aircraft control not maintained" categories indicate a total of approximately 60 mishaps per year from 1990 through 1997 (Veronneau & Evans, 2004). These data indicate that 94% of these cases occurred on general aviation flights (all aviation other than scheduled airline flights and military aviation), while the remaining 6% occurred on commercial air carriers, including four mishaps on domestic carriers, two on foreign carriers and 27 on air taxi carriers (for-hire air carriers that fly to any destination on demand). This indicates that commercial aviation flights, especially air taxis, are not free of risk from SD.

Auditory alerting devices may be a good solution for mitigating the effects of SD for pilots who experience high visual load, because audio signals can call attention to changing conditions without demanding visual attention (Braithwaite, et al., 1998). Endsley and Rosiles (1995) investigated the use of spatial audio cues presented through pilot headphones to provide information to pilots on aircraft spatial orientation. They suggested that spatial auditory elevation cues could provide a direct indicator of aircraft pitch and roll. In a study using U.S. Air Force fighters and transport pilots, they varied several spatial audio signal tone types and frequencies to provide signals containing aircraft elevation information and distance from the horizon. Endsley and Rosiles found that the localizable auditory tones enhanced the spatial orientation of aircraft pilots who were visually loaded and failed to maintain instrument scans. They concluded that spatial auditory elevation cues might be very useful in maintaining pilot spatial orientation and, thus, enhancing system safety in both civilian and military aircraft.

Auditory Warnings

Spatial auditory displays also may be used as meaningful location-based warnings of system malfunctions. For example, in a helicopter cockpit, a signal for "fire in left engine" may be presented to the pilot's left, while the signal for "shaft-driven compressor failure" may be presented from the direction of the visual warning indicator for that function. Because each warning is presented in a different and meaningful spatial location, listeners may selectively attend to one message at a time, which is advantageous when multiple warnings sound simultaneously.

Haas (1998) investigated the use of spatial auditory warnings in an AH-64A (Apache) helicopter cockpit simulator. She conducted a study to determine how quickly helicopter pilots can process different warning signals when they were presented as visual signals only, as visual signals accompanied by 3-D audio speech, and as visual signals accompanied by 3-D auditory icons (i.e., meaningful but nonverbal sounds used to convey information about events, such as the use of a fire engine sound to convey a warning of system fire).

Haas tested four different helicopter warnings (fire in left engine, fire in right engine, chips in transmission, shaft-driven compressor failure), presented with three different types of display conditions (visual display only, visual display plus 3-D audio speech signals, and visual display plus 3-D auditory icons). The results indicate that pilot response time to visual cockpit malfunction signals was reduced when the visual signals were accompanied by 3-D audio speech or auditory icons.

Posttest interviews indicated that pilots felt that spatial auditory icon and speech warnings helped them maintain flight better during emergency conditions, while not diverting their attention from other events inside and outside the cockpit. One pilot stated that the auditory warnings prevented him from crashing the helicopter during his simulated mission because the speech signal called his attention to a cockpit malfunction when he did not see the accompanying visual indicator light because of his involvement in a high-demand visual task.

Several pilots stated that having a fire indication announced on the same side as the engine fire helped them assess the emergency, and that the presence of the audio signal shortened their reaction time in responding to emergencies. These results were similar to those obtained by Oving, Veltman and Bronkhorst (2004), who found that when 3-D audio cues supplemented visual warnings in cockpit traffic alert and collision avoidance systems, pilot response time was reduced and flight safety was enhanced.

Skin-Based (Haptic & Tactile) Displays

Skin-based displays (displays that interface with the user's skin) include haptic and tactile displays. Because skin-emplaced sensors can signal events, haptic and tactile displays can be used to provide information when visual or audio cues may not be available, or can supplement audio and visual displays. Like audio and visual displays, haptic and tactile displays can be used to call user attention to critical events.

Haptic displays involve the human hand and manual sensing and manipulation to generate information to the skin, allowing users to touch, feel and manipulate objects. Examples of these displays include a computer mouse and a joystick that gives force-feedback output when moved. Haptic displays can generate skin-based vector force-feedback information to indicate proprioceptive information such as body position, orientation and movement (Briggs & Srinivasan, 2002).

Tactile displays stimulate the skin to create a sense of contact, and include pressure or vibration stimulators that interact with the skin. One effective example is the vibrate function found on most pagers and cell phones (Gemperle, Ota & Siewiorek, 2001). In this example, a tactile signal is presented through a small,

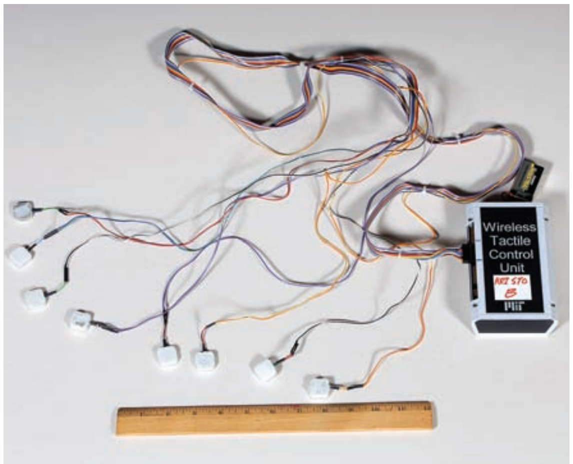

Photo 1 (right): Tactors, control unit, torso belt and forearm sleeve.

Photo 2 (below): An arm-mounted tactile display provides warning signals in a U.S. Army Research Laboratory research project.

dime-sized vibrator motor to announce the event of an incoming telephone call to cell phone user. Photo 1 shows Massachusetts Institute of Technology (MIT) pager-motor tactors, along with the MIT wireless tactile control unit, a U.S. Army Research Laboratory tactor belt, and a forearm sleeve on which the tactors can be mounted (Lockyer, 2004).

Tactile displays in particular have been used in several military applications. U.S. Army Research Laboratory researchers have successfully used tactile displays for soldier point-to-point navigation (Elliott, Redden, Krausman, et al., 2005), and military communication (Pettit, Redden & Carstens, 2006). Tactile display output was reliably detected and recognized even when soldiers performed running, jumping, climbing and crawling operations on an obstacle course (Krausman & White, 2006). Photo 2 shows a U.S. Army soldier wearing an arm-mounted tactor. Army research has shown that tactile displays can ensure that the soldier perceives signals even when s/he may be engaged in mentally or physically demanding tasks, thus enhancing system safety and effectiveness.

As with audio displays, tactile displays have been used to provide warning information regarding orientation and direction (Cholewiak & Collins, 2000) as well as user position and velocity (Rochlis & Newman, 2000), which can be helpful in reducing pilot spatial disorientation.

The Tactile Situational Awareness System (TSAS) tactile vest was developed by the U.S. Navy to reduce pilot spatial disorientation by informing them of their orientation in 3-D space (Chiasson, McGrath & Rupert, 2003). The TSAS consisted of a vest containing 20 tactors arranged into four vertical columns, with five tactors in each column. The tactors were operated in two modes—a high mode using all sensors to transmit directional information in a sequential pattern of continuous motion across the body; and a low-level mode in which only three tactors per column were activated to signal warning and alarm conditions.

TSAS testing was carried out by U.S. Navy pilots who used the vest during hover and flying operations. In addition to aviation, TSAS has been tested in underwater U.S. Navy SEAL applications and by U.S. Army pilots in an aircraft simulator as part of the Virtual Cockpit Optimization Program. These tests revealed a psychological limit for information density; pilots could not distinguish between signals containing more than two parameters of information (each parameter consisting of altitude, target location or threat location information). Thus, TSAS researchers suggest limiting tactile vests to two layers of information, such as direction and speed, or speed and rotation. The TSAS was considered successful and generated a surge of interest; it has been incorporated into the Touch Lab at MIT and into the Cutaneous Communications Lab at Princeton University (Chiasson, et al., 2003).

TSAS also has been evaluated for U.S. Navy Special Forces operations (Chiasson, et al., 2003). The TSAS vest was upgraded to present tactile directional navigation information in high-altitude, high-opening parachute operations, in ground environments and in underwater operations. Chiasson, et al. purported that the displays with tactile and visual cues resulted in better human performance than those using visual cues alone, and that superior navigational accuracy can be achieved with less mental fatigue on the operator. They suggested that a tactile display which provides eyes-free and hands-free air and ground information may allow the user to devote more time to other instruments and tasks when operating in high workload conditions, thus increasing mission safety and effectiveness.

Researchers at the TNO Human Factors Research Institute in the Netherlands designed and used tactile displays in many different military and civilian applications (van Erp, Veltman, van Veen, et al., 2003). In a recent civilian application, TNO developed a tactile display useful for automobile navigation applications. Van Erp, Mepplelink and van Veen (2002) developed a display in which small vibrators were embedded into a car seat to provide directional and navigation information to drivers. The actuators vibrated on certain sides to alert drivers when a turn was suggested, and vibrated faster the closer the car came to a turn. This display was tested in a driving simulator in which participants drove different routes through a simulated city. In this study, vehicle navigation information was presented via a visual display, a tactile display or both.

Results indicated that the addition of a tactile navigation display resulted in better performance and lower driver workload than the visual display alone. Van Erp, et al. (2002) noted that the tactile automotive display was effective in releasing other heavily loaded sensory (i.e., visual) channels, and may lead to major improvements in driver safety. This system is being adapted to motor vehicles in 2 years, and researchers predict that the haptic seat will make its debut in high-end automobiles within the next few years (Glaskin, 2004).

Haptic force-feedback displays, in which the feel of an object (such as a joystick) changes under different conditions, have been used in civilian surgical simulator applications to increase surgeon accuracy and safety. Kragic, Marayong, Li, et al. (2005) showed that haptic force feedback can be useful in retinal microsurgery training simulators, in which minimally invasive surgical tasks require micrometer-scale accuracy. In retinal microsurgery simulation, surgical tools are mounted on a robotic arm that also contains a force/torque sensor. Surgical tools are manipulated by the surgeon applying force to a handle attached to the force sensor, with the robot moving in proportion to the applied force.

To enhance surgical accuracy, Kragic, et al. (2005) recommended the use of haptic "virtual fixtures," which provide haptic force feedback to stiffen the hand-held guidance mechanism against certain directions of motion or forbidden regions of the surgical workspace. Rosenberg (1994) found that user performance can increase as much as 70% with fixture-based guidance. Visual and haptic displays also were used in the Robot Assistance Micro Surgery (RAMS) system developed by the NASA Jet Propulsion Laboratory (Charles, Das, Ohm, et al., 1997). Kennedy, Hu, Desai, et al. (2002) used haptic and visual displays in robotic cardiac surgery. Kitagawa, Dokko, Okamura, et al. (2005) found that haptic feedback significantly enhanced the execution of cardiothoracic surgery tasks requiring fine suture manipulation and knot-tying tasks.

Combined Audio & Skin-Based Displays

The preceding discussion has described how audio, tactile and haptic displays can be used separately to improve system safety. Researchers have also used visual, audio, and haptic or tactile displays in combination. At the U.S. Army Research Laboratory, Haas and Stachowiak (2007) conducted field studies to determine whether tactile and 3-D audio technologies effectively convey information in moving vehicles that contain relatively high levels of vibration and jolt. Haas and Stachowiak mounted spatial audio, tactile, and combined audio and tactile displays in a highly mobile (high-mobility) multipurpose wheeled vehicle to determine to what extent vehicle travel over gravel roads and cross-country terrain affects the use of audio, tactile, and combined audio and tactile cues in several robotics-oriented tasks. The combined audio and tactile displays provided the same (redundant) information simultaneously.

Emerging results indicate that audio and tactile displays are useful even in demanding cross-country terrain conditions, and that tactile displays and combined audio and tactile displays are best at reducing operator mental workload, thus making robotic operations safer. However, these researchers noted that poorly implemented cues of either modality (e.g., audio and tactile cues that cannot be detected) are less effective.

Because skin-emplaced sensors can signal events, haptic and tactile displays can be used to provide information when visual or audio cues may not be available, or can supplement audio and visual displays. Like audio and visual displays, haptic and tactile displays can be used to call user attention to critical events.

In a civilian safety-related application, Ng, Man, Fels, et al. (2005) used surgeon forearm-mounted tactors and audio signals to provide information on operating room physiological variables in a simulated operating room. They compared tactile with audio and audio/tactile alarms in a physiological monitoring system and found that the tactile alarm was as easy to learn and had a higher alarm identification rate than an auditory alarm alone. They also found that tactile display used alone provided greater alarm identification accuracy when compared to a combination tactile and auditory alarm.

Other civilian applications that use combinations of audio and tactile displays are web-based teleoperation systems in which remote objects such as robotic arms or mobile robots are controlled through an Internet network by means of a browser. Elhajj, Xi, Fung, et al. (2003) used video, audio and haptic information to perform real-time robotic teleoperation via the Internet. Chou and Wang (2001) designed a multimodal interface for Internet-based teleoperation, in which live video images, audio and force information were organized and presented in a predictive display. They found that presenting multimodal information redundantly reduced operator mental workload, which could make robotic operations safer.

Practical Considerations for the Use of Spatial Audio & Skin-Based Displays

The spatial audio and skin-based systems described here are not yet available to the public, but they may be adapted in some form in commercial applications, such as the haptic automobile seat (van Erp, et al., 2002). Realistic cost information will be available when the displays are commercially available.

However, it is possible to define user characteristics for designers who want to obtain the most effective use of these systems. Spatial audio and skin-based systems do not have vision requirements since they are designed to be used in environments where the user's visual field is heavily taxed or where a visual display is not available. Audio and

Military and civilian research and application have demonstrated the value of multimodal interfaces for enhancing system safety. The addition of auditory, haptic and tactile displays to traditional visual displays has been valuable in mitigating the effects of spatial disorientation and in presenting efficient location-based warnings.

skin-based displays are useful for all ages and body types. As with visual displays, if audio or skin-based signals are designed to be intuitive and easily understood, little training is needed to use them. Although fighter pilots and surgeons are described in this article as potential users, spatial audio and skin-based displays are designed to be usable by people of all education levels.

Spatial audio displays have some limitations. Because audio systems require two working ears to hear spatial audio cues, users with limited hearing in one or both ears should not use these systems. However, hearing loss alone (as defined by a measured audiometric threshold) may not limit users' ability to correctly hear spatial signals. Users with hearing impairment who have normal or near normal performance on tests of spatial hearing (binaural detection and localization) can show performance benefits from the use of amplification (i.e., hearing aids) when they use spatial audio displays, while those who do not have good spatial hearing will find no benefit from amplification (Koehnke & Besing, 1997).

Skin-based displays also have some limitations. In general, those who experience skin numbness or irritation should avoid the use of skin-based elements at those locations. In addition, large individual differences exist with respect to pain and annoyance related to vibration felt on the skin, both between people and over life span, but this can be remedied by providing the user a control to adjust display element intensity (van Erp, 2002). Tactile display limitations may preclude their use in some industrial environments. For safety reasons, tactors and skin-based display elements should be used only in applications where they do not come into contact with equipment, barriers or obstacles; such contact not only could harm the user, but also could mask, attenuate or change the characteristics of the tactile signal.

Voice Input Controls

Conventional controls in military and civilian applications consist of keyboard, mouse and joystick. However, voice input, also known as ASR, is a viable substitute for conventional controls, especially in environments where manual control is difficult (e.g., in a moving vehicle) or when both of the operator's hands are busy. Voice input controls allow a user to use speech to command a system.

Numerous military applications employ voice input controls. Draper, Calhoun, Ruff, et al. (2003) compared the utility of manual and speech input for several unmanned aerial vehicle (UAV) control tasks. Participants were asked to manually fly a high-fidelity simulated UAV and concurrently perform a series of data entry tasks using either manual (push button) or voice commands. Results showed that voice input not only made participant data entry faster and more accurate, but also improved their performance of the UAV flight and navigation tasks. Speech commands were advantageous because they required fewer steps to complete; numerous sequential button presses could be replaced by one voice command. This ability to consolidate multiple commands into single "macro" commands can reduce operator fatigue during demanding procedures.

Speech displays also have been useful in civilian and military control tasks where both of the operator's hands are busy. The Robonaut, a NASA/Defense Advanced Research Projects Agency mobile humanoid robot, allows a user to continuously employ both hands and arms to remotely control the robot's dexterous five-finger hands when performing various space orbital and planetary operation tasks (Goza, Ambrose, Diftler, et al., 2004). Use of the hand control could cause operator physical fatigue, especially during demanding operations. Goza, et al. found that voice commands reduced fatigue by permitting the operator to issue controls without having to continuously employ both hands and arms.

Voice commands have been used in many civilian applications, including dictation, medical transcription, mobile telephony, foreign language translation, control of home climate and security systems, and voice interactive telephone applications (e.g., using spoken commands over the telephone to find flight time and status).

In a medical application, Downs (1998) described a prototype neurosurgical system in which a remote expert surgeon could control a microscope by means of voice commands. Rininsland (1999) described the Advanced Robotics and Telemanipulator System for Minimally Invasive Surgery (ARTEMIS), in which a surgeon successfully used voice commands to guide an endoscopic tool at the end of a robotic arm while planning, training and performing minimally invasive surgical procedures. Freeing the hands, especially under high manual workload conditions, helps makes surgery safer.

In an industrial application, the ROBTET, a robotic system for maintaining live power lines, allows users to issue voice commands while using both hands to perform other control tasks (Penin, Aracil, Ferre, et al., 1998). Field tests indicate that time to perform teleoperation maintenance tasks such as changing insulator string was similar to the time spent by experienced users without teleoperation.

Several prototype speech systems have been suggested for industrial applications. Ressler, Antonishek, Wang, et al. (2001) designed a virtual reality collaborative engineering environment designed for applications such as building construction or manufacturing cells. Collaborating engineers could discuss interactions of objects such as vehicles or buildings by using voice controls to move synthetic representations around a virtual environment. Another application, the Multilingual and Mobile Maintenance Man, is a mobile maintenance speech recognition interface designed to provide information to maintenance crews using several different input modalities (including speech recognition) in multiple languages (Nyrkko, Carlson, Keijola, et al., 2007).

Practical Considerations for the Use of Voice Control

Practical considerations for voice control systems involve availability, cost and system limitations. Some of the systems described are prototypes built to test design concepts and are not currently available to the public. However, several civilian speech recognition applications are commercially available. A one-user ASR dictation system can cost less than $100, while high-end products such as a corporate telephone interactive voice response system can cost as much as $100,000, but can serve hundreds or even thousands of callers simultaneously (Linthicum, 1996).

Voice control system use is not limited by user vision, age or level of education. ASR systems do not have minimal vision requirements because they are designed to be used where manual controls or even visual displays may not be available. As with any type of user interface, if the user input is designed to be intuitive and easily remembered, little training will be needed to learn voice commands. Depending on the application, ASR systems are designed to be used by people of all ages and education levels; most speech recognition systems can accommodate the voices of users ranging in age from 15 to 70 years (Wilpon & Jacobsen 1996).

Voice input controls have limitations that can affect system effectiveness. Differences in user gender and dialect have been shown to impact speech recognizer accuracy. ASR systems may be less adept in recognizing female speakers (Tashakkori & Bowers, 2003; Rodger & Pendharkar, 2004). Dialects (such as deep southern U.S. speech) can make ASR average word error rate significantly higher than with general American speech as well (Picone, 1990). Dialect differences also have been shown to affect performance in Japanese (Kudo, et al., 2006) and Arabic ASR systems (Kirchoff & Vegyri, 2004).

In addition, ASR efficiency can be degraded when any difference exists between the conditions in which the speech system has been trained and the conditions in which it is finally used. Several factors can cause ASR system error, including changes in environmental noise, changes in user speaking rate due to stress, and changes in task difficulty (Junqua, 2000). Each of these factors may limit the use of ASR in industrial environments. Many approaches have been taken to reduce ASR error, especially in high-noise environments. These approaches include the use of microphones that can attenuate background noise, and the use of algorithms to characterize and estimate frequently changing noise conditions (Junqua).

References

Benson, A. (2003). Spatial disorientation: General aspects. In J. Ernsting, A. Nicholson & D. Rainford (Eds.), *Aviation medicine.* Oxford, U.K.: Butterworth Heinemann. 419-436.

Braithwaite, M.G., Durnford, S.J., Crowley, J.S., et al. (1998, Nov.) Spatial disorientation in U.S. Army rotary-wing operations. *Aviation Space and Environmental Medicine, 69*(11), 1031-1037.

Briggs, S.J. & Srinivasan, M.A. (2002). Haptic interfaces. In K.M. Stanney (Ed.), *Handbook of virtual environments: Design, implementation and applications.* Mahwah, NJ: Lawrence Erlbaum Associates.

Charles, S., Das, H., Ohm, T., et al. (1997) Dexterity-enhanced telerobotic microsurgery. *Proceedings of the International Conference on Advanced Robotics, USA.* 5-10.

Chiasson, J., McGrath, B.J. & Rupert, A.H. (2003). *Enhanced situation awareness in sea, air and land environments.* Pensacola, FL: Naval Aerospace Medical Research Lab.

Cholewiak, R.W., & Collins, A.A. (2000). The generation of vibrotactile patterns on a linear array: Influences of body site, time and presentation mode. *Perception and Psychophysics, 62,* 1220-1235.

Chou, W., & Wang, T. (2001). The design of multimodal human-machine interface for teleoperation. *Systems, Man and Cybernetics, 5,* 3187-3192.

Collins, H.L. & Harrison, G. (1995, Nov./Dec.). Spatial disorientation episodes among F-15-C pilots during operation desert storm. *Journal of Vestibular Research, 5*(6), 405-410.

Downs, H. (1998). *Integrated remote neurosurgical system.* Charlottesville, VA: Virginia Neurological Institute.

Draper, M., Calhoun, G., Ruff, H., et al. (2003). Manual versus speech input for unmanned aerial vehicle control station operations. *Proceedings of the Human Factors and Ergonomics Society 47th Annual Meeting, USA.* 109-113.

Elhajj, I., Xi, N., Fung, W.K., et al. (2003). Supermedia-enhanced Internet-based telerobotics. *Proceedings of the Institution of Electrical Engineers, 91*(3), 396-421.

Elliott, L.R., Redden, E., Krausman, A., et al. (2005). Multimodal displays to support Army infantry decision making and performance. *Proceedings of the 7th International Naturalistic Decision Making Conference, Amsterdam: The Netherlands.*

Endsley, M.R. & Rosiles, S.A. (1995). Vertical auditory localization for spatial orientation. *Proceedings of the Human Factors and Ergonomics Society 39th Annual Meeting, USA.* 55-59.

Gemperle, F., Ota N. & Siewiorek, D. (2001). Design of a wearable tactile display. *Proceedings of the 5th IEEE International Symposium on Wearable Computers, Zurich, Switzerland.* 5-12.

Glaskin, M. (2004, May 14). Driving by the seat of your pants. Commentary. *The Engineer.*

Goza, S.M., Ambrose, R.O., Diftler, M.A., et al. (2004). Telepresence control of the NASA/DARPA Robonaut on a mobility platform. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Vienna, Austria.* 623-629.

Haas, E.C. (1998). Can 3-D auditory warnings enhance helicopter cockpit safety? *Proceedings of the Human Factors and Ergonomics Society 42nd Annual Meeting, USA.* 1117-1121.

Haas, E.C. (2007). Integrating auditory warnings with tactile cues in multimodal displays for challenging environments. *Proceedings of the 13th International Conference on Auditory Display, Montréal, Canada.*

Haas, E.C. & Stachowiak, C. (2007). Multimodal displays to enhance human robot interaction on-the-move. *Proceedings of the Performance Metrics for Intelligent Systems (PerMIS) Workshop, USA.*

Junqua, J. (2000). *Robust speech recognition in embedded systems and PC applications.* Norwell, MA: Kluwer Academic Publishers.

Kennedy, C.W., Hu, T., Desai, J.P., et al. (2002). A novel approach to robotic cardiac surgery using haptics and vision. *Cardiovascular Engineering: An International Journal, 2*(1), 15-22.

Kirchoff, K. & Vegyri, D. (2004). Cross-dialectal acoustic data sharing for Arabic speech recognition. *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, Montreal, Canada.* 765-768.

Kitagawa, M., Dokko, D., Okamura, A.M., et al. (2005). Effect of sensory substitution on suture-manipulation forces for robotic surgical systems. *Journal of Thoracic and Cardiovascular Surgery, 129,* 151-158.

(References continued on page 38)

A Look into the Future

An extensive literature search indicates that spatial audio and skin-based displays have not yet been adapted by industry, perhaps because power requirements and component size do not allow user mobility. However, that does not imply that these devices are not useful in industrial applications; either modality may be useful for some types of stationary applications where there are no barriers or obstacles with which the tactors can come into contact. Haas (2007) noted that spatial audio and skin-based displays may be effective in environments where users are visually taxed (such as in spaces that are poorly lit or have limited visibility), or if a visual display is not available. Both audio and skin-based displays are effective in providing warnings that are simple and short and do not need to be referred to later.

ASR presents several opportunities for use in industry. Researchers and industry analysts have suggested that mobile service operations are an important growth area for speech recognition applications (Manufacturing and Logistics Information Technology Unlimited, 2007). In these operations, field workers use ASR to input multiple transactions through a mobile handset in real-time. One such application allows users to create, update and close various work orders and safety reports. In another application, speech systems are used to guide workers in order-picking operations, allowing them to keep their hands free. Mobile service operations have been found to help some corporations achieve cost reduction and productivity improvements (Manufacturing and Logistics Information Technology Unlimited).

Military and civilian research and application have demonstrated the value of multimodal interfaces for enhancing system safety. The addition of auditory, haptic and tactile displays to traditional visual displays has been valuable in mitigating the effects of spatial disorientation and in presenting efficient location-based warnings, although poorly implemented cues in any modality will not be as effective. Voice input controls have been useful when user hands are busy or when manual control use is difficult. As long as humans monitor displays and operate controls, multimodal control and display technologies can be used effectively to enhance soldier and civilian system safety. ■

References *(continued from page 37)*

Koehnke, J. & Besing, J. (1997). Binaural performance in listeners with impaired hearing: Aided and unaided results. In R.H. Gilkey and T.A. Anderson (Eds.), *Binaural and spatial hearing in real and virtual environments.* Mahwah, NJ: Lawrence Erlbaum Associates. 725-751.

Kragic, D., Marayong, P., Li, M., et al. (2005). Human-machine collaborative systems for microsurgical applications. *The International Journal of Robotics Research, 24*(9), 731-741.

Krausman, A.S. & White, T.L. (2006). *Tactile displays and detectability of vibrotactile patterns while performing combat assault maneuvers* (Technical Report Nos. AD-A459260 & ARL-TR-3998). Aberdeen Proving Ground, MD: U.S. Army Research Laboratory.

Kudo, I., Nakama, T., Watanabe, T., et al. (1996). Data collection of Japanese dialects and its influence into speech recognition. *Proceedings of the 4th International Conference on Spoken Language Processing, Sydney, Australia.* 2021-2024.

Linthicum, D. (1996). Make voice response sing. *Byte.com.* Retrieved May 18, 2007, from http://www.byte.com/art/9605/sec7/art1.htm.

Lockyer, B. (2004). *Operation manual for the MIT wireless tactile control unit.* Cambridge, MA: Massachusetts Institute of Technology.

Manufacturing and Logistics Information Technology Unlimited. (2007, Mar. 7). Speech recognition technology in the enterprise moves beyond customer service applications. Hertfordshire, U.K.: Author. Retrieved May 10, 2007, from http://www.logisticsit.com/absolutenm/templates/article-voice.aspx?articleid=2851&zoneid=39.

Ng, J.Y.C., Man, J.C.F., Fels, S., et al. (2005). An evaluation of a vibro-tactile display prototype for physiological monitoring. *Anesthesia and Analgesia, 101*(6), 1719-1724.

Nyrkko, S., Carlson, L., Keijola, M., et al. (2007). Ontology-based knowledge in interactive maintenance guide. *Proceedings of the 40th Hawaii International Conference on System Sciences, USA.* 1-10.

Oving, A.B., Veltman, J.A. & Bronkhorst, A.W. (2004). Effectiveness of 3-D audio for warnings in the cockpit. *International Journal of Aviation Psychology, 14*(3), 257-276.

Penin, L.F., Aracil, R., Ferre, M., et al. (1998). Telerobotic system for live power lines maintenance: ROBTET. *Proceedings of the 1998 IEEE International Conference on Robotics and Automation, Leuven, Belgium.* 2110-2115.

Pettit, R.A., Redden, E.S. & Carstens, C.C. (2006). *A comparison of hand and arm signals to a covert tactile communication system in a dynamic environment* (Technical Report A363354). Aberdeen Proving Ground, MD: U.S. Army Research Laboratory.

Picone, J. (1990). The demographics of speaker independent digit recognition. *Proceedings of the International Conference on Acoustics, Speech and Signal Processing (ICASSP), USA.* 105-108.

Ressler, S., Antonishek, B., Wang, Q., et al. (2001, Feb. 19-22). Integrating active tangible devices with a synthetic environment for collaborative engineering. *Proceedings of the 6th International Conference on 3D Web Technology, Paderborn, Germany.* 93-100.

Rininsland, H. (1999). ARTEMIS: A telemanipulator for cardiac surgery. *European Journal of Cardio-Thoracic Surgery, 16*(2), S106-S111.

Rochlis J.L. & Newman, D.J. (2000). A tactile display for International Space Station extravehicular activity. *Aviation and Space Environmental Medicine, 71*, 571-578.

Rodger, J.A. & Pendharkar, P.C. (2004). A field study of the impact of gender and user's technical experience on the performance of voice-activated medical tracking application. *International Journal of Human-Computer Studies, 60,* 5-6, 529-544.

Rosenberg, L. (1994). Virtual fixtures: Perceptual overlays enhance operator performance in telepresence tasks. *Proceedings of SPIE-2057.* 10-21.

Shilling R.D. & Shinn-Cunningham, B. (2002). Virtual auditory displays. In K. Stanney (Ed.) *Handbook of virtual environments.* Mahwah, NJ: Erlbaum. 65-92.

Tashakkori, R. & Bowers, C. (2003). Similarity analysis of voice signals using wavelets with dynamic time warping. *Proceedings of SPIE-5102.* 168-177.

van Erp, J.B. (2002). Guidelines for the use of vibro-tactile displays in human computer interaction. In S.A. Wall (Ed.), *Eurohaptics 2002: Conference proceedings.* Paris, France: EuroHaptics Society. 18-22.

van Erp, J.B., Mepplelink, R. & van Veen, H.A. (2002). *A touchy car* (DTIC TNO-TM-02-B001). Netherlands: TNO Human Factors Research Institute.

van Erp, J.B., Veltman, J.A., van Veen, H.A. & Oving, A.B. (2003). *Tactile torso display as countermeasure to reduce night vision goggles induced drift* (DTIC TNO X5-X5). Netherlands: Human Factors Research Institute.

Veronneau, S.J. & Evans, R.H. (2004). Spatial disorientation mishap classification. In F. Previc and W. Ercoline (Eds.), *Spatial disorientation in aviation.* Reston, VA: American Institute of Aeronautics and Astronautics. 197-241.

Wipon, J.G. & Jacobsen, C.N. (1996). A study of speech recognition for children and elderly. *Proceedings of ICASSP, USA.*