

INTEGRATING AUDITORY WARNINGS WITH TACTILE CUES IN MULTIMODAL DISPLAYS FOR CHALLENGING ENVIRONMENTS

Ellen C. Haas, Ph.D.

U.S. Army Research Laboratory
Multimodal Controls and Displays Laboratory
Aberdeen Proving Ground, Maryland, USA 21078
un.decided@nowhere.com

ABSTRACT

In battlefield environments of the future, auditory warnings may be integrated with tactile cues in multimodal displays. The U.S. Army is exploring the use of audio and tactile multimodal displays in applications such as the human robotic interface (HRI) to enhance Soldier performance in controlling battlefield robots. Particularly important issues in the Army HRI, as in many challenging environments, include maintaining user spatial situation awareness and providing warning signals for safety hazards. This paper will describe current research in audio and tactile display design for HRI and other applications. Best practices for integrating audio with tactile signals will be described, as well as design issues that need to be resolved.

[Keywords: auditory , tactile, multimodal]

1. INTRODUCTION

In challenging environments such as the U.S. Army battlefield, auditory warnings may be integrated with tactile cues in multimodal displays to provide information in settings where the Soldier experiences visual overload or has no access to visual displays. One such application is the human-robotic interface (HRI), a set of controls and displays that the Soldier uses to manage one or more robotic unmanned vehicles (UV's). In the U.S., every branch of the military deploys some form of UV in reconnaissance, surveillance, and intelligence operations. In addition, UVs have appeared in many civilian applications, including border and wildfire surveillance, crop dusting and crop health monitoring, and search and recovery operations [1]. Both civilian and military robotic applications have the benefit of keeping users out of harm's way in environments in which it is dangerous or impossible to work.

Both civilian and military environments present unpredictable and challenging conditions that create UV operator workload. These conditions include weather, darkness, dust, and noise. Operator workload can also be very high in cognitively demanding tasks such as individual control of one or more robots, robot sensor control and interpretation, air or ground space management, and maintaining situation awareness of the environment. In addition, Army Soldiers must maintain awareness of friendly and enemy battlefield entities. Battlefield challenges also arise from new demands for Soldier mobility; some Army systems propose that robot control operations take place in highly mobile vehicles such as High-Mobility Multipurpose Wheeled (HMMWV) "jeeps" in order to enhance

robotic command and control function and survivability [2]. In mobile environments, vehicle vibration and jolt may tax visual performance [3] and visual search [4], [5], making cues in other modalities valuable.

Early robotic systems used unimodal feedback, primarily in the visual modality. Auditory cues were developed as awareness grew that additional modalities could supplement the visual channel when it was heavily loaded. Chong, Kotoku, Ohba, Sasaki, Komoriya, and Tanie [6] examined the use of audio feedback with visual displays for multiple telerobotic operations, in which several robots were controlled by multiple remote human operators physically distant from each other. They found that by using audio and visual feedback, operators could more easily detect the possibility of collision and were able to coordinate conflicting motions between two telerobots, as compared to no audio cues. Nagai, Tsuchiya, and Kimura [7] found that audio feedback cues were a powerful tool in helping operators make decisions in simulated robotic space operations, and recommended that they would be helpful in preventing accidents during actual space operations.

Providing spatial auditory display cues can enhance UV-related tasks such as maintaining 360-degree situation awareness around a robot. Spatial audio displays permit a listener using earphones to perceive *spatialized* sounds that appear to originate at different azimuths, elevations, and distances from locations outside the head. Spatial audio displays permit sounds to be presented in different spatial locations that are meaningful to the listener, and can provide tracking information regarding object position, velocity, and trajectory beyond the field of view [8], [9], [10]. Spatial audio cues have also been shown to increase situational awareness in target search tasks using unmanned aerial vehicle displays [11].

The tactile modality is also promising for providing information and warnings for robotic systems. Tactile displays use pressure or vibration stimulators that interact with the skin [12]. To provide an example of one type of tactile display, Figure 1 shows ruggedized Massachusetts Institute of Technology (MIT) pager-motor factors, along with the MIT wireless tactile control unit, a U.S. Army Research Laboratory tactor belt, and a forearm sleeve upon which the tactors can be mounted [13].

Tactile cues have been used to provide safety warning information and communicate information regarding orientation and direction [14] as well as user position and velocity [15]. Calhoun, Fontejon, Draper, Ruff and Guilfoos [16] found that tactile displays can significantly improve detection of faults in unmanned aerial vehicle teleoperation control tasks, and can serve as an effective cueing mechanism. They suggested that

tactile alerts may be advantageous in noisy task environments that require long periods of vigilance, where both audio and visual channels are taxed. Researchers from the U.S. Army Research Laboratory explored tactile cues for localization in dismounted Soldier tasks [17] [18], in a moving vehicle vibration simulator [19], and in a moving HMMWV [20]. They found that tactile displays provided better situational awareness, faster decision time, and lower workload, than the use of visual displays alone.

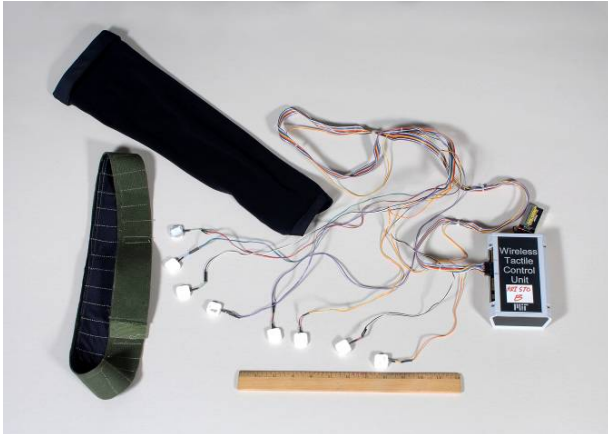


Figure 1. *Tactors, control unit, torso belt, and forearm sleeve.*

Researchers have explored the use of audio and tactile cues in HRI tasks. Gunn, Nelson, Bolia, Warm, Schumsky and Corcoran [21], and Gunn, Warm, Nelson, Bolia, Schumsky and Corcoran [22] used multimodal displays to communicate threats in an unmanned aerial vehicle (UAV) target acquisition visual search task. They found that spatial (3D) audio and tactile cues used separately enhanced target acquisition performance over no cueing. Chou, Wusheng, Wang and Tianmiao [23] designed a multimodal interface for internet-based teleoperation in which live video images, audio, and tactile force feedback information were organized and presented simultaneously. Other researchers have shown that providing additional auditory and tactile display cues can be useful in reducing HRI task difficulty [24] and creating a greater sense of operator immersion in robotic tasks [25].

2. WHY INTEGRATE AUDIO AND TACTILE DISPLAYS?

There are several advantages to integrating audio and tactile displays in challenging environments. Audio and tactile signals work well together because they have much in common. Both are useful if the user's visual field is heavily taxed (i.e., in environments that are poorly lit) or if a visual display is not available. Audio and tactile displays are effective for simple, short messages that do not need to be referred to later. Either modality is useful for mobile or stationary applications, either can be used to call for immediate response, and both can signal events in time and in space.

When used together, audio and tactile signals can supplement each other in surroundings in which variable levels of noise and vibration might mask a signal if only one modality

was used. In a noisy, high-vibration environment such as in a HMMWV, arm- or torso-mounted tactors might come into contact with a seat back, steering wheel, or dashboard, which could attenuate, mask, or change the characteristics of the tactile signal. At the same time, high levels of noise from the vehicle or communication system might mask audio signals. When used redundantly (both audio and tactile delivering the same message at the same time), audio and tactile signals would better ensure that the message is received by the user.

Multimodal displays can use different modalities to provide the user with multiple dimensions of information, when the use of one modality would constrain the total amount of information that could be communicated. Tactile displays are limited to incorporating temporal (rhythm), spatial, and a small range of frequencies to communicate two or three different dimensions of information at most [26], [27], [28]. Auditory cues have a large range of frequency, temporal and spatial cues, and can use evocative cues such as icons, earcons, and speech cues to communicate several different dimensions of information. Tactile displays are very effective at communicating spatial location when mounted on sites such as the torso or arm, although resolution is limited to the allowable spacing between tactors [29]. Spatial audio displays are also effective at communicating spatial location, but localization accuracy is constrained by front-back confusion of audio signals (when a listener perceives that a sound source in the frontal hemifield seems to occur in the rear hemifield, or vice-versa). Front-back confusion can be reduced significantly with the use of a headtracking device [30]. However, tactile displays also require head- or body-trackers to provide tracking information regarding the position of objects or events of interest in the environment.

3. GUIDELINES AND BEST PRACTICES FOR INTEGRATING AUDIO WITH TACTILE WARNINGS IN MULTIMODAL DISPLAYS

Relevant guidelines for designing multimodal displays include ISO 14915, Part 3 [31], Sutcliffe [32], and Sarter [33]. General guidelines state that multimodal displays should incorporate manageable information loading by using signals that provide only information that is necessary. Signals should also be consistent, and incorporate redundancy whenever possible. Selection of signal dimensions and encoding should exploit learned or natural relationships as much as possible, and both auditory and tactile signals should be easily discernable from other audio and vibrational events in the environment. In addition, both audio and tactile signals should avoid conflict with previously used signals in terms of meaning or characteristics.

Although there are many approaches to integrating audio and tactile cues efficiently to provide information, a knowledge of the strengths and limitations of each modality is necessary to successfully integrate audio signals with tactile in multimodal displays. There are several different strategies for integrating audio and tactile cues in multimodal displays. Three such strategies involve the use of redundant, independent, and complementary information. Redundant multimodal displays use different modalities to present the same information at the same time (e.g., both auditory and tactile modalities signal the same warning). Redundant displays are useful for presenting

important information in challenging environments where information might otherwise be lost if one signal were masked. Independent displays use different modalities to present different information at different times (e.g., tactile signals to warn of obstacles and their locations, auditory signals to warn of safety hazards and their locations). Independent displays are useful for environments contain a great number of signals; the use of different modalities for different signals might reduce user confusion. Complementary displays use different modalities to present different aspects of the same signal (e.g., auditory cues signal warning functions and tactile signals denote their spatial locations). Complementary displays allow the different modalities to play to their strengths; evocative audio signals could be used to describe a large number of different signal functions, while tactile signals could describe spatial location. However, independent and complementary displays have no signal redundancy, so signal information could be lost if either audio or tactile information is masked.

Auditory and tactile output must have synchrony to attain the user's assumption of unity (the perception that the audio and tactile events are linked to the same distal object or event). It has been postulated that the greater the number of a certain group of properties shared between two modalities, the stronger the observer's unity assumption [34]. Among these properties are spatial location, motion, and temporal patterning or rate, all of which would be impacted by temporal synchrony or asynchrony in a multimodal display [35]. Detectable perceptual threshold value studies indicate that synchronization between auditory and tactile modalities should be 25 ms or less [35], [36].

Multimodal display design should also involve an awareness of potential crossmodal links in attention due to the use of multiple modalities. Crossmodal attentional effects include modality shifting, modality expectation, and crossmodal spatial links.

Sarter [33] noted that modality shifting effects have been demonstrated in numerous psychological and neurophysiological laboratory studies. The modality shifting effect involves the user's limitations in shifting attention from one modality to another. Shifting from signals presented in one modality to those presented in another might slow participant response time to signals, depending on which modality is used more often. Spence, Pavani and Driver [37] found that it appears to be particularly difficult and time-consuming to shift attention to the visual or auditory channel away from rare events that are presented in the tactile modality.

Modality expectation effects have also been demonstrated in numerous laboratory studies listed in a review article by Sarter [33]. Modality expectations are formed based on the observed frequency or the perceived importance of a cue in a particular modality. Sarter described studies that indicate that expecting a cue to appear in a certain modality leads to an enhanced readiness to detect and discriminate information in that sensory channel, and may lead to increased response time to cues in an unexpected modality. The effects of modality expectation appear to be somewhat less pronounced for tactile than for auditory cues [38].

Crossmodal spatial link effects have also been found to affect attention to signals in different modalities. Crossmodal spatial link effects involve deliberate shifts of attention to a particular location. Sarter [33] stated that concurrent stimuli of

different modalities in the same spatial location can greatly (and non-linearly) facilitate user response, while concurrent stimulation by different modalities at different locations can lead to non-linear suppression of user attentional response. Sarter noted that crossmodal spatial attentional links may be inadvertently evoked if the location of multimodal information presentation is not carefully controlled.

4. CONCLUSIONS AND ISSUES FOR FUTURE RESEARCH

The battlefield display of the future may integrate audio with tactile cues. The U.S. Army has focused on reducing operator workload and enhancing situation awareness and Soldier performance in their HRI interface by using multimodal displays. Multiple modalities can be advantageous; when used together, audio and tactile signals can supplement each other in demanding surroundings where variable levels of noise and vibration might mask cues in only one modality. Further, multimodal displays can use different multimodal design strategies (independent, redundant and complementary information) to provide the user with multiple dimensions of information. Although relevant guidelines exist for the design of multimodal displays, close attention must be paid to factors such as signal synchrony and crossmodal links in attention, to reduce delays in user response time and accuracy that may arise from the use of multiple modalities.

In summarizing crossmodal shift effects (modality shifting, modality expectations, and crossmodal spatial linking), Sarter [33] observed that shift effects are based on laboratory studies in which absolute effect sizes are small, and levels of user workload are low. She noted that the reaction time decrements described in the laboratory studies may turn out to be larger in more complex environments, and may be associated with increased error rates. Research is being conducted at the U.S. Army Research Laboratory (ARL) to determine whether attentional shift effects exist in more complex and demanding environments that contain variable levels of workload.

Although multimodal design strategies were described, (independent, redundant, and complementary displays), there is a lack of research comparing human performance associated with each strategy. These effects should also be tested in the laboratory as well as in demanding field environments. The ARL is conducting research in this area.

Few researchers have explored the use of coded tactile cues to efficiently communicate multiple (two or more) dimensions of information. Although not an auditory design issue, the design of the tactile cues can influence the overall effectiveness of the multimodal display. As previously noted, the small quantity of tactile research indicates that tactile displays can effectively incorporate temporal (rhythm), location, and limited range of frequencies to communicate two or three dimensions of information. The ARL is exploring different tactile coding strategies, and will integrate tactile cues with audio signals in multimodal displays used in laboratory and field environments. Future papers will describe the results of their research.

5. REFERENCES

- [1] Cooke, N. (2006). Preface: Why Human Factors of “Unmanned Systems?”, in N.J. Cooke, H.L. Pringle, H.K. Pedersen, O. Conner (Eds.), Human Factors of Remotely Operated Vehicles. Elsevier, Amsterdam.
- [2] Emmerman, P.J., Grills, J.P., and Movva, U.Y. Challenges to Agentization of the Battlefield. Proceedings of the International Command and Control Research and Technology Symposium. U.S. Department of Defense.
- [3] Ishitake, T., Ando, H., Miyazaki, Y., and Matoba, F. changes of visual performance induced by exposure to whole-body vibration. Kurume Medical Journal, 45(1), 59-62.
- [4] Griffin, M.J., and Lewis, C.H. (1978). A review of the effects of vibration on visual acuity and continuous manual control, part I: Visual acuity. Journal of Sound and Vibration, 56: 383-413.
- [5] Dennis, J.P.(1965). Some effects of vibration upon visual performance. Journal of Applied Psychology, 49:245-252.
- [6] Chong, N.Y, Kotoku, T., Ohba, K., Sasaki, H., Komoriya, K., and K. Tanie, K. (2002) “Multioperator teleoperation of multirobot systems with time delay. II – Testbed description,” Presence: Teleoperators and Virtual Environments, vol. 11, no. 3, pp. 292-303.
- [7] Nagai, Y., Tsuchiya, S., and Kimura, S., (2000). A study on an audio feedback system for the operation of Engineering Test Satellite VII – Effects on telemetry monitor, in Proceedings of the International Symposium on Space Technology and Science, pp. 1226-1231.
- [8] Perrott, D.R., Sadralodabai, T., Saberi, K., and Strybel, T.Z. (1991). Aurally aided visual search in the central visual field; effects of visual load and visual enhancement of the target. Human Factors, 33(4), 389-400.
- [9] Elias, B. (1996). The effects of spatial auditory preview on dynamic visual search performance. Proceedings of the Human Factors and Ergonomics Society 40th Annual Meeting, PP. 1227-1231.
- [10] Fujawa, G.E., and Strybel, T.Z. (1997). The effects of cue informativeness and signal amplitude on auditory spatial facilitation of visual performance. Proceedings of the Human Factors and Ergonomics Society 41st Annual Conference, pp. 556-560.
- [11] Simpson, B. D., Bolia, R. S., and Draper, M. H. (2004). Spatial auditory display concepts supporting situation awareness for operators of unmanned aerial vehicles. In D. A. Vincenzi, M. Mouloua, and P. A. Hancock (Eds.), Vol. I.- Human performance, Situational Awareness, and Automation: Current Research and Trends (pp. 61-65). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- [12] Gemperle, F., Ota, N., and Siewiorek, D. (2001). Design of a wearable tactile display. Proceedings of the 5th IEEE International Symposium on Wearable Computers, 5-12.
- [13] Lockyer, B. (2004). Operation manual for the MIT wireless tactile control unit. Cambridge, Massachusetts: Massachusetts Institute of Technology.
- [14] Cholewiak, R. W., and 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.
- [15] Rochlis, J. L., and Newman, D. J. (2000). A tactile display for international space station (ISS) extravehicular activity (EVA). Aviation and Space Environmental Medicine, 71, 571-578.
- [16] Calhoun, G., Fontejon, J., Draper, M., Ruff, H., and Guilfoos, B. (2004). Tactile versus aural redundant alert cues for UAV control applications. Proceedings of the Human Factors and Ergonomics Society 48th Meeting (pp. 137-141).
- [17] Turner, D., and Redden, E.A. (2006). Localization of tactile belt signals as a function of tactor operating characteristics. Unpublished manuscript, U.S. Army Research Laboratory, Aberdeen Proving Ground, Maryland.
- [18] Elliott, L. R., Redden, E.S., Pettit, R., Carstens, C., and Van Erp, J. (2006). Tactile Guidance for Land Navigation. Manuscript submitted for publication, U.S. Army Research Laboratory, Aberdeen Proving Ground, MD.
- [19] White, T.W., and Krausman, A. (2006). Detectability of vibrotactile signals in moving vehicles. Unpublished manuscript, U.S. Army Research Laboratory, Aberdeen Proving Ground, Maryland.
- [20] Haas, E.C., Stachowiak, C., White, T., Pillalamarri, K., and Feng, T. (2007). The Effect of Vehicle Motion on the Integration of Audio and Tactile Displays in Robotic System TCU Displays. Unpublished manuscript, U.S. Army Research Laboratory, Aberdeen Proving Ground, MD.
- [21] Gunn, D.V., Nelson, W.T., Bolia, R.S., Warm, J.S., Schumsky, D.A., and Corcoran, K.J. (2002). Target Acquisition with UAVs: Vigilance Displays and Advanced Cueing Interfaces. Proceedings of the Human Factors and Ergonomics Society 46th Annual Meeting, 1541-1545.
- [22] Gunn, D.V., Warm, J.S., Nelson, W.T., Bolia, R.S., Schumsky, D.A., and Corcoran, K.J. (2005). Target acquisition with UAVs: Vigilance displays and advance cueing interfaces. Human Factors, 47(3), 488-497.

- [23] Chou, Wusheng, Wang and Tianmiao (2001). The design of multimodal human-machine interface for teleoperation. Proceedings of the IEEE International Conference of systems, Man and Cybernetics, Volume 5, 2187-3192.
- [24] Lathan, C.E., and Tracey, M. (2002). The effects of operator spatial perception and sensory feedback on human-robot teleoperation performance. Presence: Teleoperators and Virtual Environments, 11(4), 368-377.
- [25] Burdea, G., Richard, P., and Coiffet, P. (1996). Multimodality virtual reality: Input-output devices, systems integration, and human factors. International Journal of Human-computer Interaction: Special Issues of Human-Virtual Environment Interaction, 8(1), 5-24.
- [26] Brown, L.M., Brewster, S.A., and Purchase, H.C. (2006). Multidimensional tactons for non-visual information presentation in mobile devices. MobileHCI'06, 231-238.
- [27] Hoggan, E., and Brewster, S. (2006). Crossmodal spatial location: Initial experiments. NordiCHI 2660: Changing Roles, 469-472.
- [28] Brewster, S. A., and Brown, L.M. (2004). Tactons: Structured tactile messages for non-visual information display. In A. Cockburn (Ed.), Conferences in research and Practice in Information Technology.
- [29] Van Erp, J.B. (2002). Guidelines for the Use of Vibro-Tactile Displays in Human Computer Interaction, in Eurohaptics 2002: Conference Proceedings. Ed by S.A. Wall et. al., (Eds.) Edinburgh, 2002, 18-22.
- [30] Begault, D.R., Wenzel, E.M., and Anderson, M.R., (2001). Direct comparison of the impact of head tracking, reverberation, and individualized head-related transfer functions on the spatial perception of a virtual sound source. Journal of the Audio Engineering Society, 49, 904-915.
- [31] ISO, 1998 ISO (International Organization for Standardization), (1998). ISO 14915: Multimedia user interface design software ergonomic requirements, Part 3: Media combination and selection, author.
- [32] Sutcliffe, A. (2003). Multimedia and virtual reality: Designing multisensory user interfaces. Lawrence Erlbaum Associates, Mahwah, New Jersey.
- [33] Sarter, N.B., (2006). Multimodal information presentation: Design guidance and research challenges (2006). In A.M Bisantz (Ed) Cognitive Engineering Insights for Human Performance and Decision Making, *International Journal of Industrial Ergonomics*, 439-445.
- [34] Welch, R.B. (1999). Meaning, attention, and the unity assumption in the intersensory bias of spatial and temporal perceptions. In G. Ascherleben, t. Bachmann, and J. Musseler (Eds.), Cognitive Contributions to the Perception of Spatial and Temporal Events. Elsevier, Amsterdam, 1999, 371-387.
- [35] Adelstein, B.D., Begault, D.R., Anderson, M.R., and Wenzel, E.M. (2003) Sensitivity to Haptic-Audio Asynchrony. International Conference on Multimodal Interfaces (ICMI)'03, American Association for Computing Machinery, 73-76.
- [36] Altinsoy, M.E. (2003). Perceptual aspects of auditory-tactile asynchrony. Proceedings of the Tenth International Congress of Sound and Vibration, Stockholm, Sweden.
- [37] Spence, C., Pavani, F., and Driver, J. (2000). Crossmodal links between vision and touch in covert endogenous spatial attention. Journal of Experimental Psychology: Human Perception and Performance, 26, 1298-1319.
- [38] Boulter, L.R. (1977). Attention and reaction times to signals of uncertain modality. Journal of Experimental Psychology: Human Perception and Performance, 3, 379-388.