Auditory and Visual Facilitation: Cross-Modal Fusion of Information in Multi-Modal Displays

Captain Stephen Boyne and Nada Pavlovic
Defence Research and Development Canada
DRDC Toronto
1133 Sheppard Ave. W.
Toronto, Ontario M3M 3B9
CANADA

stephen.boyne@drdc-rddc.gc.ca
nada.pavlovic@drdc-rddc.gc.ca

Ryan Kilgore and Mark H. Chignell
Department of Mechanical and
Industrial Engineering, University of Toronto
MIE, 5 King's College Rd.
Toronto, Ontario M5S 3G8
CANADA

r.Kilgore@utoronto.ca
chignell@mie.utoronto.ca

ABSTRACT

The modern battlefield is increasingly populated with vast amounts of electronic information. The plethora of data that can be delivered needs to be filtered, interpreted, and formatted in ways that are meaningful and useful for particular tasks and situations. Cross-modal fusion of information should be helpful in optimizing battlespace interfaces to provide the maximum amount of data to the commander and in enhancing their operational picture while avoiding increasing working memory load.

In our research, carried out by researchers at the University of Toronto and Defence Research and Development Canada, we are looking at fundamental questions concerning the cross-modal fusion of information. In particular, we are focussing on how visual spatial awareness can be facilitated by presentation of auditory cues or information, and how auditory spatial awareness can be facilitated by visual information. We shall refer to instances of such facilitation as auditory facilitation, and visual facilitation, respectively.

Auditory facilitation occurs when the information about entities and spatial relations among them is coded redundantly, using the auditory modality to augment the visual. For instance, designing for auditory facilitation can be used to transmit more information to the commander and aid integration and coordination of disparate pieces of information (Flanagan et al., 1998; Nelson et al., 1998; McKinley & Ericson, 1995; Begault, 1993). There are many potential advantages of using spatialized sound in visualization platforms to aid decision-making and to enhance situational awareness (SA) of the operational picture. The short-term auditory store supplements the visual store, providing an opportunity to store greater amounts of relevant information (Wickens and Hollands, 2000). In addition, the auditory sense is omni-directional and can be used to help in orienting, and in detecting speed and accuracy (Bronkhorst et al., 1996; Begault, 1993). Redundant information presented in a different modality must be compatible with the primary modality if improved task performance and enhanced SA are to be achieved (e.g., St. John et al., 2001; Hollands et al., 2003; Tlauka et al., 2000; Shelton and McNamara, 2004; Harwood and Wickens, 1991; Aretz, 1991).

Visual facilitation occurs when visually displayed information enhances auditorily presented information. The type of visual facilitation considered in our research is where visual data representations are used to enhance awareness of target locations in spatial audio displays. This is an important application, because of

Boyne, S.; Pavlovic, N.; Kilgore, R.; Chignell, M.H. (2005) Auditory and Visual Facilitation: Cross-Modal Fusion of Information in Multi-Modal Displays. In *Visualisation and the Common Operational Picture* (pp. 19-1 – 19-4). Meeting Proceedings RTO-MP-IST-043, Paper 19. Neuilly-sur-Seine, France: RTO. Available from: http://www.rto.nato.int/abstracts.asp.

the advantages of using spatialization in audio such as increased discriminability and intelligibility of multiple sound sources (Ericson and McKinley, 1997; Drullman and Bronkhorst, 2000). The distinct location of voices in space also aids in the cognition of collaborative audio communications, in part by facilitating the task of speaker identification. An investigation of spatialized voice streams in an audioconference-style listening task (Baldis, 2001) has established that the locational separation of conferee voices increased listeners' speaker identification performance and perceived overall comprehension of conference events while simultaneously reducing perceived attention requirements for speaker identification.

We have developed and implemented a collaborative virtual environment, named the Vocal Village (www.vocalvillage.net), to capitalize upon these known benefits of spatialized audio. The Vocal Village is a client/server based VoIP system that accommodates spatialized audio communication in a flexible, customizable audio space, where voice locations are monitored and controlled via graphical user interface. Our recent investigations of this audio space have demonstrated that providing listeners with the ability to control the apparent location of incoming voice streams (personalization) using a visual interface (Visual facilitation) further decreases the perception of both the difficulty associated with, and the amount of required attention for, speaker identification in a multi-talker environment (Kilgore et. al., 2003).

As an example of the use of visual facilitation, interactions within the Vocal Village are augmented with a visual display of participant names, apparent audio locations, relative volume settings, and position within the virtual space of the conference. In military applications, similar uses of additional visual information could be aid in disambiguating, elaborating, or qualifying incoming audio messages, and in creating flexible voice collaboration systems where cues as to relative location are embedded in the audio information.

However, while spatial audio GUIs would be intended to aid in maintaining users' awareness of a complex audio space and aid in speaker identification, the use of such visualization may inadvertently introduce detrimental performance effects by inappropriately altering users' perception of spatial auditory events. This alteration could be the result of inter-sensory effects caused by perceptually dominant visual stimuli. A powerful example of inter-sensory bias is the ventriloquism effect, by which a visual event is able to 'capture' the location of a non-collocated auditory event – such as the moving mouth of a dummy is able to visually capture apparent location of a ventriloquist's voice (Bertelson and Radeau, 1981), (Warren et. al., 1981). Another example would be the result of a perceptual 'mismatch' between the apparent spatial location of an audio event and the location of its related visual depiction. Such non-collocated audio/visual events are problematic as they may actually result in decreased electrophysiological responses to audio/visual stimuli (Calvert et. al., 2001) and often cause a 'disconcerting' experience for listeners, who have difficulty interpreting conflicting cues.

The full version of this paper will report on three studies of cross-modal fusion that we are carrying out. The first study is concerned with the visual facilitation of audio in voice collaboration and is motivated by the recent finding that listeners become more sensitive to disparate audio and visual cues as the screen size of the visual interface is reduced (Walker and Brewster, 2001). This finding is relevant to addressing portability needs of soldiers in the field where large audio spaces are being condensed into small visual depictions (such as PDA or wrist-mounted displays).

We are also carrying out two studies of auditory facilitation, one concerned with urban warfare as the application, and the other concerned with search and rescue operations. The Canadian forces are in the process of acquiring or developing a number of systems to provide information to the front line soldier. One of the purposes behind the current research is to determine the relative value of these different technologies to better guide acquisition and deployment strategies. 3D audio is being investigated for a variety of different

applications including navigation displays and radio communications. As discussed above the use of redundant 3D audio information provides a number of advantages to the user. In addition to the issues of the compatibility of the redundant information provided 3D audio presents the additional complication of being available in widely different levels of fidelity, ranging from simple stereo, or left/right cuing, to almost perfect replications of the spatialized aspects of naturally occurring sounds.

In one recent research study we are examining the effects of the level of fidelity and orientation on auditory facilitation by examining three different cases, simple stereo, a generic spatialized audio and a free field condition. The first two conditions are oriented to the soldier's direction of travel; the last is based on head orientation. We have assessed the impact of the different displays on a soldier's ability to develop situation awareness and perform his mission in a virtual built-up area.

Our goal in this research is to provide science-based guidance for the use of cross-modal fusion of information in multi-modal displays. By developing methods of auditory and visual facilitation that are compatible with the properties of human perception and cognition it should be possible to transfer more information to the user/soldier without compromising attentional resources or increasing working memory load.

REFERENCES

Aretz, A.J. (1991). The design of electronic map displays. *Human Factors,* 33(1), 85-101.

Baldis, Jessica. "Effects of spatial audio on memory, comprehension, and preference during desktop conferences," (CHI 2001) CHI Letters, Vol. 3, No. ACM Press, 2001, pp. 166-173.

Begault, D.R. (1993). Head-up auditory displays for traffic collision avoidance system advisories: A preliminary investigation. *Human Factors,* Vol. 35(4), pp. 707-717.

Bertelson, Paul and Monique Radeau, "Cross-modal bias and perceptual fusion with auditory-visual spatial discordance," Perception and Psychophysics, 1981, 29 (6), pp. 578-584.

Bronkhorst, A.W., Veltman, J.A. and Breda, L. van (1996). Application of a three-dimensional auditory display in a flight task. *Human Factors,* Vol. 38(1), pp. 23-33.

Calvert, A. Gemma, Peter C. Hansen, Susan D. Iversen, and Michael J. Brammer, "Detection of audio-visual integration sites in humans by application of electrophysiological criteria to the BOLD effect," NeuroImage 14, 2001, pp. 427-438.

Drullman, Rob and Adelbert W. Bronkhorst, "Multichannel speech intelligibility and talker recognition using monaural, binaural, and three-dimensional auditory presentation," J. Acoust. Soc. Am. 107(4), April 2000, pp. 2224-2235.

Ericson, M.A., and R.L. McKinley, "The intelligibility of multiple talkers separated spatially in noise," in **Binaural and Spatial Hearing in Real and Virtual Environments**, Gilkey, Robert H. and Timothy R. Anderson Eds., NJ, Lawrence Erlbaum Associates, 1997, pp. 701-724.

Flanagan, P., McAnally, K.I., Martin, R.L., Meehan, J.W. and Oldfield, S.R. (1998). Aurally and visually guided visual search in a virtual environment. *Human Factors,* Vol. 40(3), pp. 461-468.

Harwood, K. and Wickens, C.D. (1991). Frames of reference for helicopter electronic maps: the relevance of spatial cognition and componential analysis. *The International Journal of Aviation Psychology,* 1(1), 5-23.

Hollands, J.G., Ivanovic, N., & Enomoto, Y. (2003). Visual momentum and task switching with 2D & 3D displays of geographic terrain. *Proceedings of the Human Factors and Ergonomics Society - 47th Annual Meeting* (pp. 1620-1624). Santa Monica, CA: Human Factors and Ergonomics Society.

Kilgore, Ryan M., Mark Chignell and Paul W. Smith, "Spatialized Audioconferencing: what are the benefits?" Proceedings of the 2003 conference of the Centre for Advanced Studies conference on Collaborative research, 2003, pp. 111-120.

McKinley, R.L. and Ericson, M.A. (1997). Flight demonstration of a 3-D auditory display. In Binaural and spatial hearing in real and virtual environments. (Eds.) Gilkey, R. H. and Anderson, T. R. Hillsdale, NJ, England: Lawrence Erlbaum Associates, Inc. pp. 683-699.

Nelson, W.T., Hettinger, L.J., Cunningham, J.A., Brickman, B.J., Haas, M.W. and McKinley, R.L. (1998). Effects of localized auditory information on visual target detection performance using a helmet-mounted display. *Human Factors,* Vol. 40(3), pp. 452-460.

Shelton, A.L. and McNamara, T.P. (2004). Orientation and perspective dependence in route and survey learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition.* 30(1) 158-170.

St. John, M., Cowen, M.B., Smallman, H.S. and Oonk, H.M. (2001). The use of 2D and 3D displays for shape-understanding versus relative-position tasks. *Human Factors, 43,* 79-98.

Tlauka, M., Stantos, D. and McKenna F.P. (2000). Dual displays. *Ergonomics,* 43(6), pp. 764-770.

Walker, Ashley and Stephen Brewster, "'Sitting too close to the screen can be bad for your ears': a study of audio-visual location discrepancy detection under different visual projections," Proceedings of the 2001 International Conference on Auditory Display, 2001, pp. 86-89.

Warren, David H., Robert Welch and Timothy J. McCarthy. "The role of visual-auditory 'compellingness' in the ventriloquism effect: implications for transitivity among the spatial senses", Perception and Psychophysics, 1981, 30(6), 557-564.

Wickens, C.D. and Hollands, J.G. (2000). Engineering Psychology and Human Performance, Third Edition. Upper Saddle River, NJ: Prentice Hall Inc.