

VIRTUAL ACOUSTIC DISPLAYS

Elizabeth M. Wenzel
NASA Ames Research Center
Moffett Field, California

As with most research in information displays, virtual displays have generally emphasized visual information. Many investigators, however, have pointed out the importance of the auditory system as an alternative or supplementary information channel (e.g., Deatherage, 1972; Doll, et. al., 1986; Patterson, 1982; Caver, 1986). A three-dimensional auditory display can potentially enhance information transfer by combining directional and iconic information in a quite naturalistic representation of dynamic objects in the interface. Borrowing a term from Gaver (1986), an obvious aspect of "everyday listening" is the fact that we live and listen in a three-dimensional world. Indeed, a primary advantage of the auditory system is that it allows us to monitor and identify sources of information from all possible locations, not just the direction of gaze. This feature would be especially useful in an application that is inherently spatial, such as an air traffic control display for the tower or cockpit. A further advantage of the binaural system, often referred to as the "cocktail party effect" (Cherry, 1953), is that it improves the intelligibility of sources in noise and assists in the segregation of multiple sound sources. This effect could be critical in applications involving encoded nonspeech messages as in scientific "visualization," the acoustic representation of multi-dimensional data (e.g., Bly, 1982) and the development of alternative interfaces for the visually-impaired (Edwards, 1989; Loomis, et. al., 1990). Another aspect of auditory spatial cues is that, in conjunction with other modalities, it can act as a potentiator of information in the display. For example, visual and auditory cues together can reinforce the information content of the display and provide a greater sense of presence or realism in a manner not readily achievable by either modality alone (Colquhoun, 1975; Warren, et. al., 1981; O'Leary and Rhodes, 1984). This phenomenon will be particularly useful in telepresence applications, such as advanced teleconferencing environments, shared electronic workspaces, and monitoring telerobotic activities in remote or hazardous situations. Thus, the combination of direct spatial cues with good principles of iconic design could provide an extremely powerful and information-rich display which is also quite easy to use.

This type of display could be realized with an array of real sound sources or loudspeakers (Doll, et. al., 1986; Calhoun, et. al., 1987). An alternative approach, recently developed at NASA-Ames, generates externalized, three-dimensional sound cues over headphones in realtime using digital signal-processing (Wenzel, et. al., 1988a). Here, the synthesis technique involves the digital generation of stimuli using Head-Related Transfer Functions (HRTFs) measured in the two ear-canals of individual subjects (see Wightman and Kistler, 1989a). Up to four moving or static sources can be simulated in a head-stable environment by digital filtering of arbitrary signals with the appropriate HRTFs. This type of presentation system is desirable because it allows complete control over the acoustic waveforms delivered to the two ears and the ability to interact dynamically with the virtual display. Other similar approaches include an analog system developed by Loomis, et. al. (1990) and digital systems which make use of transforms derived from normative mannikins and simulations of room acoustics (Genuit, 1986; Posselt, et. al., 1986; Persterer, 1989; Lehnert and Blauert, 1989).

Such an interface also requires the careful psychophysical evaluation of listeners' ability to accurately localize the virtual or synthetic sound sources. For example, a recent study by Wightman and Kistler (1988b) confirmed the perceptual adequacy of the basic technique for static sources; source azimuth was synthesized nearly perfectly for all listeners while source elevation was somewhat less well-defined in the headphone conditions.

From an applied standpoint, measurement of each potential listener's HRTFs may not be possible in practice. It may also be the case that the user of such a display will not have the opportunity for extensive training. Thus, a critical research issue for virtual acoustic displays is the degree to which the general population of listeners can obtain adequate localization cues from stimuli based on non-individualized transforms. Preliminary data (Wenzel, et. al., 1988b) suggest that using non-listener-specific transforms to achieve synthesis of localized cues is at least feasible.

For experienced listeners, localization performance was only slightly degraded compared to a subject's inherent ability, even for the less robust elevation cues, as long as the transforms were derived from what one might call a "good" localizer. Further, the fact that individual differences in performance, particularly for elevation, could be traced to acoustical idiosyncracies in the stimulus suggests that it may eventually be possible to create a set or "universal transforms" by appropriate averaging (Genuit, 1986) and data reduction techniques (e.g., principal components analysis), or perhaps even enhancing the spectra of empirically-derived transfer functions (Durlach and Pang, 1986).

Alternatively, even inexperienced listeners may be able to adapt to a particular set of HRTFs as long as they provide adequate cues for localization. A reasonable approach is to use the HRTFs from a subject whose measurements have been "behaviorally-calibrated" and are thus correlated with known perceptual ability in both free-field and headphone conditions. In a recently completed study, sixteen inexperienced listeners judged the apparent spatial location of sources presented over loudspeakers in the free-field or over headphones. The headphone stimuli were generated digitally using HRTFs measured in the ear canals of a representative subject (a "good localizer") from Wightman and Kistler (1988a,b). For twelve of the subjects, localization performance was quite good, with judgements for the non-individualized stimuli nearly identical to those in the free-field.

In general, these data suggest that most listeners can obtain useful directional information from an auditory display without requiring the use of individually-tailored HRTFs. However, a caveat is important here. The results described above are based on analyses in which errors due to front/back confusions were resolved. For free-field versus simulated free-field stimuli, experienced listeners exhibit front/back confusion rates of about 5 vs. 10 % and inexperienced listeners show average rates of about 22 vs. 34 %. Although the reason for such confusions is not completely understood, they are probably due in large part to the static nature of the stimulus and the ambiguity resulting from the so-called cone of confusion (see Blauert, 1983). Several stimulus characteristics may help to minimize these errors. For example, the addition of dynamic cues correlated with head-motion and well-controlled environmental cues derived from models of room acoustics may improve the ability to resolve these ambiguities.

REFERENCES

- Blanert, J. (1983) *Spatial Hearing*. The MIT Press: Cambridge, MA.
- Bly, S. (1982) *Sound and computer information presentation*. Unpublished doctoral thesis (UCCRL-53282) Lawrence Livermore National Laboratory and University of California, Davis, CA.
- Calhoun, G. L., Valencia, G., and Furness, T. A. III (1987) Three-dimensional auditory cue simulation for crew station design/evaluation. *Proc. Hum. Fac. Soc.*, 31, 1398-1402.
- Cherry, E. C. (1953) Some experiments on the recognition of speech with one and two ears. *J. Acoust. Soc. Am.*, 22, 61-62.
- Colquhoun, W. P. (1975) Evaluation of auditory, visual, and dual-mode displays for prolonged sonar monitoring in repeated sessions. *Hum. Fact.*, 17, 425-437.
- Deatherage, B. H. (1972) Auditory and other sensory forms of information presentation. In H.P. Van Cott and R.G. Kincade (Eds.), *Human Engineering Guide to Equipment Design*, (rev. ed.), Washington, DC: U.S. Government Printing Office, 123-160.
- Doll, T. J., Gerth, J. M., Engelman, W. R. and Folds, D. J. (1986) Development of simulated directional audio for cockpit applications. USAF Report No. AAMRL-TR-86-014.
- Durlach, N. I. and Pang, X. D. (1986) Interaural magnification. *J. Acoust. Soc. Am.*, 80, 1849-1850.
- Edwards, A. D. N. (1989) Soundtrack: An auditory interface for blind users. *Hum. Comp. Interact.*, 4, 45-66.
- Gaver, W. (1986) Auditory icons: Using sound in computer interfaces. *Hum. Comp. Interact.*, 2, 167-177.
- Genuit, K. (1986) A description of the human outer ear transfer function by elements of communication theory. *Proc. 12th ICA (Toronto)*, Paper B6-8.
- Lehnert, H. and Blauert, J. (1989) A concept for binaural room simulation. *ASSP Workshop on Applications of Signal Processing to Audio and Acoustics*, New Paltz, NY.
- Loomis, J. M., Herbert, C., and Cicinelli, J. G. (1989) Active localization of virtual sound sources. Submitted to *J. Acoust. Soc. Am.*
- O'Leary, A. and Rhodes, G. (1984) Cross-modal effects on visual and auditory object perception. *Perc. & Psychophys.*, 35, 565-569.
- Patterson, R. R. (1982) Guidelines for Auditory Warning Systems on Civil Aircraft. Civil Aviation Authority Paper No. 82017, London.

- Persterer, A. (1989) A very high performance digital audio signal processing system. *ASSP Workshop on Applications of Signal Processing to Audio and Acoustics*, New Paltz, NY.
- Posselt, C., Schroter, J., Opitz, M., Divenyi, P., and Blauert, J. (1986) Generation of binaural signals for research and home entertainment. *Proc. 12th ICA (Toronto)*, Paper B1-6.
- Warren, D. H., Welch, R. B., and McCarthy, T. J. (1981) The role of visual-auditory "compellingness" in the ventriloquism effect: Implications for transitivity among the spatial senses. *Perc. & Psychophys.*, 30, 557-564.
- Wenzel, E. M., Wightman, F. L., and Foster, S. H. (1988a) A virtual display system for conveying three-dimensional acoustic information. *Proc. Hum. Fac. Soc.*, 32, 86-90.
- Wenzel, E. M., Wightman, F. L., Kistler, D. J., and Foster, S. H. (1988b) Acoustic origins of individual differences in sound localization behavior. *J. Acoust. Soc. Amer.*, 84, 579.
- Wightman, F. L. and Kistler, D. J. (1989a) Headphone simulation of free-field listening I: stimulus synthesis. *J. Acoust. Soc. Amer.*, 85, 858-867.
- Wightman, F. L. and Kistler, D. J. (1989b) Headphone simulation of free-field listening II: psychophysical validation. *J. Acoust. Soc. Amer.*, 85, 868-878.