Applications of Virtual Auditory Displays

B. Shinn-Cunningham
Hear. Res. Center, Depts. of Cognitive and Neural Systems and Biomedical Engineering, Boston University, 677 Beacon St., Boston, MA 02215
E-mail: shinn@cns.bu.edu

Abstract-- Current technology makes it possible to simulate naturally-occurring spatial auditory cues quite accurately. However, the cost of such a system is not justifiable, or even desirable, for all applications. This paper surveys some of the applications currently using virtual auditory displays and some of the issues important in the design of virtual auditory displays.
***Keywords*—**virtual environments, spatial hearing

1 Introduction

Virtual auditory displays differ from other acoustic displays in that the signals reaching the listener's ears are like those that would arise from real sources around the listener. Most "virtual auditory displays" are headphone-based systems in which sounds are processed using Head-Related Transfer Functions (HRTFs) to simulate the normal auditory localization cues [1, 2]). Some systems vary the signals presented from two (or more) loudspeakers to simulate sources from different locations [3, 4]. The first approach allows precise stimulus control and can achieve very good simulations, especially if the acoustic signals change with listener movement. A loudspeaker-based system is usually less costly to implement and can be very effective, especially if the listener's head position is fixed relative to the speakers. However, the acoustic signals reaching the listener cannot be controlled with the same precision as with a headphone-based display. In some applications, the most important feature of virtual auditory displays is that the simulated sound sources are "realistic," that is, they sound like real sources external to the listener's body. In other cases, the main feature is that accurate spatial information is conveyed to a listener.

2 Application Areas

2.1 Scientific Study of Sound Localization

Historically, most studies of auditory localization have been performed using simplified acoustic cues presented over headphones or using real sound sources positioned around the listener. The first method allows the stimuli reaching the listener to be completely characterized and enables detailed study of the perceptual salience of cues such as interaural time delays (ITDs) and interaural level differences (ILDs). However, these simplified cues invariably cause sounds to be heard inside the head, rather than at locations external to the listener's body [5-7]. In contrast, studies in rooms have enabled researchers to characterize localization performance in realistic settings. Unfortunately, in these studies it is nearly impossible to determine the contributions of each of the naturally-occurring acoustic cues to subject perception and performance in real-world localization studies.

The development of virtual auditory display techniques has enabled careful scientific study of the processes governing normal auditory localization. For instance, using HRTF-based systems, researchers have begun to investigate the perceptual significance of various localization cues [8-12]. Physiological studies using HRTF-based displays have started to investigate how different spatial acoustic cues affect neural responses at different levels of the auditory pathway [13-16]. Although less precise stimulus control is achieved, some studies (e.g., studies of source movement [17]) have been performed using a speaker-based simulation.

When virtual auditory displays are used to study localization perception, the goal is to determine which naturally-occurring localization cues are important and how they influence perception and/or neural responses. Since the focus of these basic studies is to characterize the processes underlying auditory localization in normal listening environments, much of the work has focused on creating sound source simulations that cannot be discriminated from real sources. Whether or not the simulation is "perfect," HRTF-based displays allow the experimenter to methodically examine the contributions of different acoustic cues to sound localization and to gain insight into the way in which listeners perceptually integrate acoustic localization cues. Generally speaking, for scientific applications, cost is not a driving factor in choosing a virtual auditory display.

2.2 Clinical Evaluation

Clinically, spatially auditory perception can be an important factor in predicting a patient's ability to function in common, noisy listening environments [18]. Tests of spatial hearing have been suggested for evaluating everything from hearing aid function to deficits in auditory brainstem processing [19-23].

One of the greatest benefits of spatial hearing in normal-hearing individuals derives from a person's ability to selectively attend to sources from particular directions [24]. For clinical purposes, the most important information provided by auditory localization tests is whether patients can make use of spatial acoustic cues that provide great benefit to normal-hearing listeners. It is not necessary to simulate sources with perfect accuracy from all possible directions for such a test to be effective. However, it is important that the simulated sources are easy to localize (so that explaining the task to patients is simple and the goal of the task is easy to understand). It is also crucial that the test system be inexpensive and easy to use.

Virtual auditory displays have been suggested as a cost-effective method for performing clinical tests of spatial hearing [33, 34]. An HRTF-based system has been used by Koehnke and her associates in studies of spatial perception of subjects with different hearing aids [25], the comparative localization ability of children with normal hearing and those with a history of otitis media [26], and in studies of speech intelligibility in noise [27]. Other researchers have begun to develop a low-cost, portable loudspeaker-based system to screen listeners' spatial hearing in the clinic [4].

2.3 Display of Spatial Information

A number of tasks demand that a human operator analyze and respond to spatial information. Air traffic controllers must keep track of multiple planes under their control. Pilots must navigate their craft in air and on land while tracking enemy craft, hostile targets, and other objects . Operators of a remote vehicle or a telerobot must make sense of data transmitted from a distant location, building up an internal mental model of the remote site.

Spatial information has usually been presented using visual displays (because the spatial acuity of the visual channel is much better than that of the auditory channel). Unfortunately, the visual channel is often overloaded, with operators monitoring a myriad of dials, gauges, and graphic displays [28]. Consequently, there is growing interest in using spatial auditory displays for these tasks [29, 30]. Spatial auditory displays are also being developed to present information to the blind [31-33].

In these command/control applications, the primary goal is to convey unambiguous information to the human operator. Realism, *per se*, is not useful, except to the extent that it makes the operator's task easier (i.e., reduces the workload). Conversely, spatial resolution is critical. In these applications, signal-processing schemes that could enhance the amount of information transferred to the human operator may be useful, even if the result is "unnatural."

2.4 Entertainment

Many entertainment applications use multiple speakers to simulate spatial acoustic cues. In home and movie theaters, the industry standard Dolby "Surround" technology uses five separate speakers. Two-speaker simulations are becoming increasingly common in the personal computer market. A variety of manufacturers (e.g., Qsound Labs, Spatializer Audio Laboratories, Creative Labs, Inc., etc.) sells products that use proprietary algorithms to generate spatial effects over loudspeakers. Headphone-based spatial audio systems are also available. (e.g., Virtual Listening Systems, Inc. and Sennheiser Electronics sell products that use HRTFs to simulate spatial audio for home theater applications; Microsoft Corp. has developed low-level functions to produce spatial sound over headphones for PC-based computer programs).

The most important considerations in the development of spatial audio displays for entertainment applications are the system cost and the perceptual impact of the simulation. For these applications, sources must sound "real" (in particular, they must be externalized), but precision in the simulation is not important (the perceived position of the simulated object does not have to be controlled with very much accuracy). Instead, the emphasis is on developing an inexpensive system to generate the most compelling experience possible for the end user.

3 Examples of Factors to Consider

Much of the research devoted to developing and verifying virtual auditory display technology emphasizes the subjective "realism" of the display; however, this is not the most important consideration for all applications.

In some cases, signal processing that improves "realism" actually interferes with the amount of information a listener can extract. The inclusion of a room-acoustics model to simulate early echoes and reverberation can significantly increase the perceived realism of an HRTF-based display [6, 7, 34] (in speaker-based displays, naturally-occurring reverberation is already present). This is not surprising, considering even *real* sources heard in an anechoic room can sound "unnatural" [5]. Nonetheless, echoes decrease sensitivity to changes in the primary source location. Many researchers discount the perceptual effects of echoes on source localization, citing a phenomenon known as the "precedence effect" (whereby the location of a leading sound dominates the perception of source location; e.g., see [35]). However, there is ample evidence that echoes can cause a measurable effect on perceived source location and on discriminability of the source position, particularly when the source has a slow rise time or when its onset is masked [36, 37]. For applications in which information about the location of the source is more important than the realism of the display, including echoes and reverberation may be ill advised.

In some cases, the cost of creating a "realistic" display is simply too great when weighed against the benefit gained in a particular application. In headphone-based systems, realism is enhanced with the use of individualized HRTFs, particularly in the perception of source elevation [38]. Ideally, HRTFs should also be sampled in both distance and direction at a spatial density dictated by human sensitivity. In practice, the measurement of HRTFs is a time-consuming process that requires precise measurement and careful calibration [39]. The effort involved in HRTF measurement increases dramatically when more than one distance is included and when the spatial sampling density increases. Also, as the number of HRTFs increases, so does the amount of computer storage needed. Thus, while the most "realistic" system would use individualized HRTFs that are sampled densely in both direction and distance, most systems use generic HRTFs sampled coarsely in direction at only one distance. A number of researchers have investigated various HRTF encoding schemes to ameliorate problems associated with HRTF interpolation; however, such efforts do not entirely solve these problems [13, 40-43].

As stated above, realism is enhanced when echoes and reverberation are included in a simulation. However, the processing power needed to simulate a sound source in a reverberant room can be prohibitive. Other acoustic effects are often ignored in order to reduce the computational complexity of the acoustic simulation, including the non-uniform radiation pattern of a realistic sound source, spectral changes in a sound due to atmospheric effects, and Doppler shift of the received spectrum of moving sources.

The perceptual significance of many of these effects is not well understood; further work must be done to examine how these factors affect the realism of the display as well as how much spatial information such cues may convey.

If information transfer is of primary importance, it may even be useful to present acoustic spatial cues that are intentionally distorted so that they are perceptually more salient than are more "realistic" cues. In particular, since auditory spatial resolution is relatively poor (compared, for instance, to vision or proprioception), it may be useful to change how spatial position is encoded in the acoustic signals reaching the ears. For example, the acoustic spatial cues that would result if a listener's head were larger than normal would be exaggerated relative to more realistic spatial auditory cues [44, 45]. Encoding spatial location with these cues can improve auditory resolution to some extent [46]. However, great care must be taken when such "unnatural" encoding schemes are used. In particular, whenever a new encoding scheme is used, subjects will have to be trained to interpret the spatial encoding scheme correctly [45, 46]. Experiments investigating this type of "sensory adaptation" indicate that subjects can only adapt to linear remappings of auditory localization [47]. This means that even with extended training, subjects may never completely learn more complex encoding schemes. Finally, experiments suggest that spatial auditory resolution is limited in many cases by central, cognitive factors (such as memory or attention), rather than by peripheral sensitivity to acoustic spatial cues [46, 48]. This work underscores the impact of human limitations on the efficacy of spatial auditory displays.

4 Conclusions

While it is possible to create auditory spatial displays in which simulations are nearly indistinguishable from the real-world, such systems are very expensive. For scientific research, these high-end systems are necessary in order to allow careful examination of normal localization cues. In clinical applications, the auditory display must only be able to deliver stimuli that can distinguish listeners with normal spatial hearing from those with impaired spatial hearing. Such systems must be inexpensive and easy to use, but there is no need for a "perfect" simulation. In command and control applications, the goal is to maximize information transfer into the human operator; subjective impression (i.e., "realism") is unimportant. In these applications, both technological and perceptual issues must be considered to achieve this goal. Finally, in entertainment applications, cost is the most important factor; the precision of the display is unimportant as long as the simulation is subjectively satisfactory.

Acknowledgements-- This work was supported by AFOSR grant F49620-98-1-0108.

References

[1] F. L. Wightman and D. J. Kistler, "Headphone simulation of free-field listening. I. Stimulus synthesis," *J. Acoust. Soc. Am.*, vol. 85, pp. 858-867, 1989.

[2] S. Carlile, Virtual Auditory Space: Generation and Applications. New York: RG Landes, 1996.

[3] W. G. Gardner, "Transaural 3-D audio," MIT Media Laboratory Perceptual Computing Section, Technical Report No. 342, July 20 1995.

[4] B. G. Shinn-Cunningham, P. M. Zurek, E. R. Stutman, and R. Berkovitz, "Perception of azimuth for sources simulated using two loudspeakers in natural listening environments," 19th ARO Midwinter Meeting, St. Petersburg Beach, FL, 1996.

[5] G. Plenge, "On the differences between localization and lateralization," *J. Acoust. Soc. Am.*, vol. 56, pp. 944-951, 1974.

[6] N. I. Durlach, A. Rigapulos, X. D. Pang, W. S. Woods, A. Kulkarni, H. S. Colburn, and E. M. Wenzel, "On the externalization of auditory images," *Presence*, vol. 1, pp. 251-257, 1992.

[7] W. M. Hartmann and A. Wittenberg, "On the externalization of sound images," *J. Acoust. Soc. Am.*, vol. 99, pp. 3678-3688, 1996.

[8] F. Wightman, D. Kisler, and K. Andersen, "Reassessment of the role of head movements in human sound localization," *J. Acoust. Soc. Am.*, vol. 95, pp. 3003-3004, 1994.

[9] A. Kulkarni, S. K. Isabelle, and H. S. Colburn, "Human sensitivity to HRTF phase spectra," 18th ARO Midwinter Meeting, St. Petersburg, FL, 1995.

[10] B. J. May and A. Y. Huang, "Sound orientation behavior in cats. I. Localization of broadband noise," *J. Acoust. Soc. Am.*, vol. 100, pp. 1059-1069, 1996.

[11] F. L. Wightman and D. J. Kistler, "Monaural sound localization revisited," *J. Acoust. Soc. Am.*, vol. 101, pp. 1050-1063, 1997.

[12] J. C. Middlebrooks, "Spectral shape cues for sound localization," in *Binaural and Spatial Hearing in Real and Virtual Environments*, R. Gilkey and T. Anderson, Eds. New York: Erlbaum, 1997, pp. 77-98.

[13] P. W. F. Poon and J. F. Brugge, "Virtual space receptive fields of single auditory nerve fibers," *J. Neurophys.*, vol. 70, pp. 667-676, 1993.

[14] J. F. Brugge, R. A. Reale, J. E. Hind, J. C. K. Chan, A. D. Musicant, and P. W. F. Poon, "Simulation of free-field sound sources and its application to studies of cortical mechanisms of sound localization in the cat," *Hear. Res.*, vol. 73, pp. 67-84, 1994.

[15] S. Carlile and A. J. King, "Monaural and binaural spectrum level cues in the ferret: Acoustics and the neural representation of auditory space," *J. Neurophys.*, vol. 71, pp. 785-801, 1994.

[16] E. D. Young, J. J. Rice, G. A. Spirou, I. Nelken, and R. A. Conely, "Head-related transfer functions in cat: Neural representation and the effects of pinna movement," in *Binaural and Spatial Hearing in Real and Virtual Environments*, R. Gilkey and T. Anderson, Eds. New York: Erlbaum, 1997, pp. 475-498.

[17] D. W. Grantham, "Detection and discrimination of simulated motion of auditory targets in the horizontal plane," *J. Acoust. Soc. Am.*, vol. 79, pp. 1939-1949,

1986.
[18] R. Hausler, H. S. Colburn, and E. Marr, "Sound localization in subjects with impaired hearing," *Acta Oto-Laryngologica*, vol. Supplement 400, 1983.
[19] A. Markides, Binaural Hearing Aids. New York: Academic Press, 1977.
[20] G. R. Bienvenue and B. M. Siegenthaler, "A clinical procedure for evaluating auditory localization," *J. Speech Hear. Disorders*, vol. 39, pp. 469-477, 1974.
[21] D. Byrne, W. Noble, and B. Glauerdt, "Effects of earmold type on ability to locate sounds when wearing hearing aids," *Ear and Hearing*, vol. 17, pp. 218-228, 1996.
[22] W. Noble, D. Byrne, and B. Lepager, "Effects on sound localization of configuration and type of hearing impairment," *J. Acoust. Soc. Am.*, vol. 95, pp. 992-1005, 1994.
[23] M. Furst, R. A. Levine, A. D. Korczyn, B. C. Fullerton, R. Tadmor, and D. Algom, "Brainstem lesions and click lateralization in patients with multiple sclerosis," *Hear. Res.*, vol. 82, pp. 109-124, 1995.
[24] N. I. Durlach and H. S. Colburn, "Binaural phenomena," in *Handbook of Perception, vol. 4*, E. C. Carterette and M. P. Friedman, Eds. New York: Academic Press, 1978, pp. 365-466.
[25] J. Koehnke and J. Besing, "Binaural performance in listeners with impaired hearing: Aided and unaided results," in *Binaural and Spatial Hearing in Real and Virtual Environments*, R. Gilkey and T. Anderson, Eds. New York: Erlbaum, 1997, pp. 725-752.
[26] J. M. Besing and J. Koehnke, "A test of virtual auditory localization," *Ear and Hearing*, vol. 16, pp. 220-229, 1995.
[27] J. Koehnke and J. M. Besing, "A procedure for testing speech intelligibility in a virtual listening environment," *Ear and Hearing*, vol. 17, pp. 211-217, 1996.
[28] R. L. McKinley, M. A. Ericson, and W. R. D'Angelo, "3-Dimensional auditory displays: Development, applications, and performance," *Aviation, Space, Environ. Med.*, May 1994, pp. A31-A38, 1994.
[29] D. R. Begault, "Virtual acoustic displays for teleconferencing: Intelligibility advantage for "telephone grade" audio," 95th AES Convention, February 25-28, 1995, Paris, France, 1995.
[30] R. L. McKinley and M. A. Ericson, "Flight demonstration of a 3-D auditory display," in *Binaural and Spatial Hearing in Real and Virtual Environments*, R. Gilkey and T. Anderson, Eds. New York: Erlbaum, 1997, pp. 665-682.
[31] B. Arons, "Hyperspeech: Navigating in speech-only hypermedia," Hypertext '91, San Antonio, TX, 1991.
[32] T. V. Raman, "Audio system for technical readings," New York: Cornell University, 1994.
[33] F. James, "Presenting HTML structure in audio: User satisfaction with audio hypertext," ICAD, Xerox Parc, 1996.
[34] D. R. Begault, "Perceptual effects of synthetic reverberation on three-dimensional audio systems," *J. Audio Eng. Soc.*, vol. 40, pp. 895-904, 1991.
[35] P. M.. Zurek, "The Precedence Effect," in *Directional Hearing*, W. A. Yost and G. Gourevitch, Eds. New York: Springer-Verlag, 1987, pp. 85-105.
[36] B. G. Shinn-Cunningham, P. M. Zurek, and N. I. Durlach, "Adjustment and discrimination measurements of the precedence effect," *J. Acoust. Soc. Am.*, vol. 93, pp. 2923-2932, 1993.
[37] W. M. Hartmann, "Listening in a room and the precedence effect," in *Binaural and Spatial Hearing in Real and Virtual Environments*, R. Gilkey and T. Anderson, Eds. New York: Erlbaum, 1997, pp. 191-210.
[38] E. M. Wenzel, M. Arruda, D. J. Kistler, and F. L. Wightman, "Localization using nonindividualized head-related transfer functions," *J. Acoust. Soc. Am.*, vol. 94, pp. 111-123, 1993.
[39] S. Carlile and D. Pralong, "Recording the head related transfer function (HRTF): The essential link in high fidelity three dimensional audio displays," Australian Airforce Conference, 1995.
[40] D. J. Kistler and F. L. Wightman, "A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction," *J. Acoust. Soc. Am.*, vol. 91, pp. 1637-1647, 1991.
[41] R. L. Jenison and K. Fissell, "A spherical basis function neural network for modeling auditory space," *Neural Comp.*, vol. 8, pp. 115-128, 1995.
[42] J. Chen, B. D. Van Veen, and K. E. Hecox, "A spatial feature extraction and regularization model for the head-related transfer-function," *J. Acoust. Soc. Am.*, vol. 97, pp. 439-952, 1995.
[43] M. A. Bloomer, P. R. Runkle, and G. H. Wakefield, "Pole-zero models of head-related and directional transfer functions," 18th ARO Midwinter Meeting, St. Petersburg, Florida, 1995.
[44] N. I. Durlach and X. D. Pang, "Interaural magnification," *J. Acoust. Soc. Am.*, vol. 80, pp. 1849- 1850, 1986.
[45] N. I. Durlach, B. G. Shinn-Cunningham, and R. M. Held, "Supernormal auditory localization. I. General background," *Presence*, vol. 2, pp. 89-103, 1993.
[46] B. G. Shinn-Cunningham, N. I. Durlach, and R. M. Held, "Adapting to supernormal auditory localization cues I: Bias and resolution," *J. Acoust. Soc. Am.*, (in press), 1998.
[47] B. G. Shinn-Cunningham, N. I. Durlach, and R. M. Held, "Adapting to supernormal auditory localization cues II: Changes in mean response," *J. Acoust. Soc. Am.*, (in press), 1998.
[48] B. G. Shinn-Cunningham, "A decision-theory model of auditory adaptation," *Perception and Psychophysics*, vol. (submitted), 1998.