Shinn-Cunningham, BG (2002). "Spatial auditory displays," to appear in *International Encyclopedia of Ergonomics and Human Factors, 2nd edition,* W. Karwowski, Ed. London: Taylor and Francis, Ltd.

Spatial Auditory Displays

Barbara G. Shinn-Cunningham

Departments of Cognitive and Neural Systems and Biomedical Engineering
Boston University
Boston, MA 02215
phone: 617-353-5764
fax: 617-353-7755
email: shinn@cns.bu.edu

In command and control applications, complex spatial data must be presented to the human operator. Examples of this kind of application include presenting information to air traffic controllers, fighter pilots, and operators of remote exploration vehicles. For these applications, the most important factor is the amount of information that the operator can extract from the display. Realism of the display is not critical, except to the extent that unrealistic displays may increase the workload on the operator. For some of these applications, it is not important to accurately present 3-dimensional spatial information over the auditory channel; one or two spatial dimensions may meet the requirements for a particular task. Expense of the display is also not a critical consideration for command and control applications. Headphone-based displays are usually employed for these systems.

Blind users are increasingly making use of spatial auditory displays. For instance, spatial auditory displays are used in navigational aids or as a substitute for graphical computer interfaces. As with command and control applications, realism of the spatial auditory display is not important by itself with these applications; what matters is the amount of spatial information conveyed to the user. Relatively inexpensive and robust systems are generally more appropriate for this kind of use than are extremely costly systems that require extensive calibration and maintenance. As with command and control applications, headphone-based systems are generally preferred for use as sensory aids.

One of the largest and most rapidly growing application areas for spatial auditory displays is in the entertainment industry. Movie theaters employ displays with multiple speakers to elicit the impression of sources moving around the listener. Computer games increasingly rely on stereo speakers to simulate sources at various locations. For these applications, subjective realism is the most important goal, but the accuracy and resolution achieved by the display is not as important. As a result, speaker-based techniques are cost effective: the subjective realism of the display is robust, even though objectively these displays are not particularly accurate.

3. Acoustic cues for sound position

The sounds that reach a listener's two ears are used to determine both the source content (what the source is) and location (where the source is) as well as acoustic attributes of the room. The sounds reaching the ears depend on the content of the sound source, the position of the source relative to the listener, and the listening environment. In most environments, sound from the source reaches the listener both directly and via reflections off walls, floors, and other objects. Sound location is determined by the specific sound attributes (cues) that change with source position.

Differences in the time of arrival of the sound at the left and right ears (i.e., interaural time differences, or ITDs) are the main cue indicating the laterality (left/right location) of a source. Interaural intensity differences (IIDs) also contribute to the perceived laterality of a source. The spectral content of the direct sound reaching the listener helps to indicate whether a sound is in front of, behind, above, or below a listener. Such spectral cues arise from direction filtering due to the external ears (pinnae). Individual differences in the pinnae are important for accurate perception of up/down and front/back.

Overall intensity of the signals reaching the ears conveys distance information for sources familiar to the listener. In a given reverberant environment, the direct-to-reverberant energy ratio provides an absolute measure of source distance.

As a listener moves relative to a sound source, spatial cues change accordingly. These dynamic changes help to disambiguate possible source positions that would otherwise give rise to very similar cues. For instance, sources directly in front or directly behind both cause near zero ITDs and IIDs; however, a leftward rotation of the head results in either ITDs and IIDs favoring the right ear (for a source in front) or the left ear (for a source behind).

4. Speaker-based simulations

In speaker-based simulations, two or more loudspeakers are used to control the total acoustic signals reaching the listener. The signals presented from each speaker are calculated so that the total signals reaching the eardrums of the listener approximate normal spatial auditory cues. In particular, the speaker signals must be controlled such that they cancel and sum differently at the two ears to generate appropriate ITD, IID, spectral, and level cues.

Although it is difficult to precisely control the apparent location of a simulated source using speaker-based techniques, even the simplest versions of such an approach can cause the apparent location of a "phantom" source to change. For instance, it is relatively easy to mimic the gross spatial cues of sources at lateral positions between the two speakers. However, it is more difficult to simulate sources outside the range of speaker locations or control the perceived up/down front/back direction of the source. In order to simulate this spatial dimension, other spatial cues (particularly spectral cues) must be simulated correctly. However, computations become increasingly more complex and less stable as these cues are included. As a result, most free-field systems do not try to control spectral cues with precision.

While one can affect the perceived position of phantom sources using free-field simulation techniques, the approach is not particularly robust. In particular, accurate simulation depends upon knowing the location and orientation of the listener relative to the loudspeakers. Generally, the head must be within a relatively small "sweet spot" for the simulation to be accurate. Alternatively, one can employ a head-tracker and can compute the signals presented from each speaker in real time based on the current position and orientation of the listener. While this approach can increase the working area, it also greatly increases the complexity and expense of the system.

Elementary free-field simulation techniques are employed in many consumer-market products (such as computer games and stereo recordings) because they are so simple and inexpensive to implement. However, the precision of such simulations is not adequate for many other applications. The technical difficulties inherent in more advanced free-field systems makes them less attractive for applications in which the spatial information displayed in the system must be precise.

5. Headphone simulations

Headphone-based systems rely on Head-Related Transfer Functions (HRTFs) in order to simulate a source in space. For each spatial location relative to the listener, a pair of HRTF filters (one for the left ear and one for the right ear) describes how an arbitrary sound source is changed as it propagates through space and impinges on the listener. Theoretically, a stereo signal generated from the appropriate HRTF pair will be identical to the signals reaching the ears from a sound source at the desired position in space and will include all ITD, IID, and other spatial cues that are present in a natural sound.

Although an HRTF simulation should yield stimuli that are perceptually indistinguishable from natural experience, a number of technical hurdles limit the accuracy of HRTF techniques. For instance, measurement of HRTFs is a difficult, time-consuming process. While individual differences in HRTFs are critical for some aspects of sound source localization (e.g., for distinguishing front/back and up/down), most systems employ a standard set of HRTFs that are not matched to the individual. Calibration of the headphone system can also be problematic; the exact positioning of the headphones over the ears can have a noticeable impact on the signals reaching the eardrum. Storage requirements of HRTFs may limit the number of discrete HRTF pairs that can be the system is carefully calibrated and individual HRTFs are used. While imperfect, HRTF-based systems allow the signals at the ears to be controlled with much greater precision than do speaker-based simulations.

6. Summary

The cost of creating a natural, realistic simulation is not justifiable, or even desirable, for all applications; instead, the optimal design must take into account the goals of the display device. It is relatively easy to generate gross binaural cues that match normal experience (either with speaker- or headphone-based approaches), and for many applications this is sufficient. For other applications, it may be necessary to calibrate the display to the individual listener in order to accurately control more than the apparent laterality of a sound source. For some applications, subjective realism is the overriding goal, but for others, the main measure of the effectiveness of the display is the amount of spatial information that a listener can extract. Ultimately, the most efficient design for a spatial auditory display depends on weighing the cost and benefits of each design, taking into account financial, perceptual, and technological constraints.

References

BEGAULT, D. R. (1994). 3-D Sound for Virtual Reality and Multimedia. London, Academic Press.

BLAUERT, J. (1997). Spatial Hearing (2e). Cambridge, MA, MIT Press.

CARLILE, S. (1996). Virtual Auditory Space: Generation and Applications. New York, RG Landes.

MIDDLEBROOKS, J. C. and GREEN, D. M. (1991). "Sound localization by human listeners." Annual Review of Psychology **42**: 135-159.

SHINN-CUNNINGHAM, B. G., LEHNERT, H., KRAMER, G., WENZEL, E.M., and DURLACH, N. I., 1997, Auditory Displays. Binaural and Spatial Hearing in Real and Virtual Environ349 12A a6w f0 12 s]135aNeal., WENZEL, E.M.D 0rsonET 177 277.7Erlbaum47 0 TD -0.01T 90wA4

Index Words: auditory localization, spatial displays, binaural hearing

Author:

Barbara G. Shinn-Cunningham, Departments of Cognitive and Neural Systems and Biomedical Engineering, Boston University, Boston, MA 02215, ph: 617-353-5764, fax: 617-353-7755, email: shinn@cns.bu.edu