- **Hide:** Stop a primitive from being available to the user.
- **Set Contents:** Either cause a parent-child relationship to exist if this is a context primitive, or set the contents of the Text/Marker/Symbol primitive.
- **Set Attributes:** Primitives may have some presentation specific details which do not really contribute to the information, but may be aesthetically important.

Example

One common situation which could be used as an example to help explain this strategy is an application program that needs to present a pop-up menu to the user.

To accomplish this, the application would produce output primitive operators to guide appropriate modifications to the primitive feedback tree. A result of this would be the addition of a new subtree as shown in Figure 1.

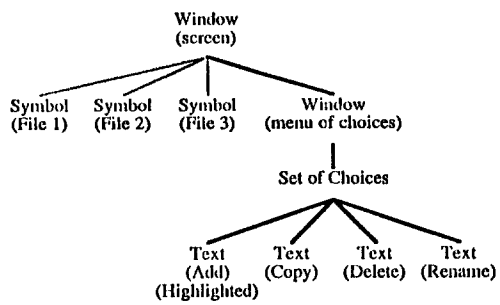

Figure 1: A new subtree is added

This primitive feedback tree would then be interpreted visually as in Figure 2, or auditorially as in Figure 3, depending on the interface manager chosen for use by the end user.

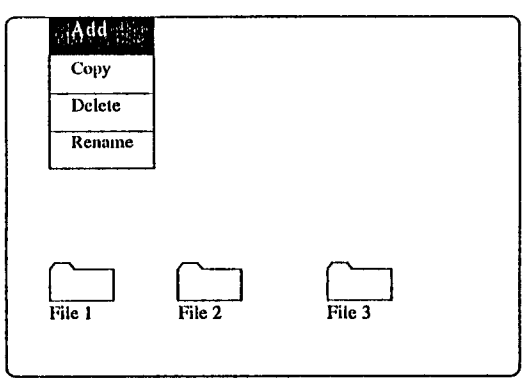

Figure 2: Visual presentation of the feedback tree

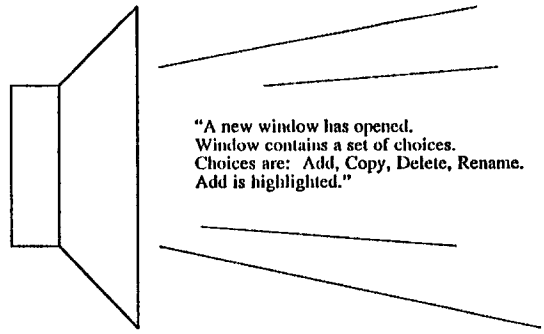

Figure 3: Auditory presentation of the feedback tree

Summary

A strategy has been proposed to allow people with disabilities the choice of feedback presentation modality. To accomplish this a set of output primitives has been developed which presents the content, context and relationships between information in an unambiguous tree structure. This structure is then interpreted by an interface manager to determine the final form of the presentation.

Acknowledgements

This project was funded by IBM Corporation as a Shared University Research Project with the University of Toronto and the Hugh MacMillan Medical Centre, Toronto, Canada.

DEVELOPMENT OF A THREE-DIMENSIONAL AUDITORY DISPLAY SYSTEM

ELIZABETH M. WENZEL, FREDERIC L. WIGHTMAN, SCOTT H. FOSTER

Abstract

We propose that the most powerful method of auditory cueing takes direct advantage of human perceptual capabilities, providing a dynamic, multidimensional pattern of events which conveys meaning about objects in the spatial world. Applications of such a three-dimensional auditory display involve any context in which the user's situational awareness is critical, particularly when visual cues are limited or absent. Examples include air traffic control displays, advanced teleconferencing environments, and monitoring telerobotic activities in hazardous situations.

Meta Platforms, Inc.
Exhibit 1042
Page 001

This type of display system requires the ability to generate localized sound cues in a flexible and dynamic manner. Whereas this can be achieved with an array of real sound sources or loudspeakers, the prototype device being developed at NASA-Ames maximizes flexibility and portability by synthetically generating three-dimensional sound cues in realtime for delivery through headphones. Unlike conventional stereo, sources can be perceived outside the head at discrete distances and directions from the listener. When completed, the device will be integrated with the Virtual Interactive Environment Workstation (VIEW), a head-mounted, wide-angle, stereoscopic display system controlled by operator position, voice, and gesture (Fisher, et. al., 1986).

Previous research in psychoacoustics suggests that perceptually-veridical localization over headphones is possible if both the direction-dependent pinna cues and the more well understood cues of interaural time and intensity are adequately synthesized. Although the realtime device is not yet finished, recent studies at the University of Wisconsin have confirmed the perceptual adequacy of the basic approach to synthesis.

Synthesis: Theory & Technique

The 'Duplex Theory' of human sound localization, first described by Lord Rayleigh (1907), states that the two cues primarily responsible for localization are interaural differences in time of arrival for low-frequency waveforms and interaural differences in intensity at high frequencies. Although the theory has long dominated thinking about binaural hearing, recent research points to serious limitations with this formulation. For example, it cannot account for the ability of subjects to precisely locate sounds on the median plane where interaural cues are minimal or absent (see Blauert, 1983). Similarly, when subjects listen to stimuli over headphones, they are perceived as inside the head even though interaural temporal and intensity differences appropriate to an external source location are introduced. A variety of research indicates that deficiencies of the Duplex Theory stem from an inadequate consideration of filtering by the outer ears or pinnae; Shaw (1974) demonstrated that spectral shaping by the pinnae is indeed direction-dependent, while other experiments showed that the absence of pinna cues degrades localization accuracy (Gardner & Gardner, 1973; Oldfield & Parker, 1984), and Plenge (1974) concluded that pinna cues are primarily responsible for externalization or the sensation of 'out-there-ness'. Such studies suggest that perceptually-veridical localization over headphones is possible if both the direction-dependent pinna cues and the more well understood cues of interaural time and intensity can be adequately synthesized.

A number of experimenters have been interested in simulating various features of auditory spatial perception. For example, one approach following from the work of Batteau (1968) simulates elevation with moderate success using a time-domain or comb-filtering technique (Watkins, 1978). Spatial processing has also been of interest in computer music where there is an emphasis on simulating auditory 'spaciousness', rather than veridical localization *per se*, with such cues as reverberation, distance, and Doppler shifts (e.g., Moore, 1983). Other attempts have centered on synthesizing essentially all available cues by perfecting methods of binaural recording using both stationary (see Blauert, 1983, p. 358) and head-tracked stereophony (Doll, et. al., 1986). Similarly, much of the recent work in this area has been on developing accurate techniques for modelling the pinnae (Shaw, 1974) or measuring pinna transform functions (Mehrgardt & Mellert, 1977; Wightman & Kistler, 1980; Foster, 1986).

Our technique derives from the latter approach, preserving the frequency-dependent interaural time and intensity information via precise measurement of the outer ear filtering characteristics. Figure 1 illustrates the method.

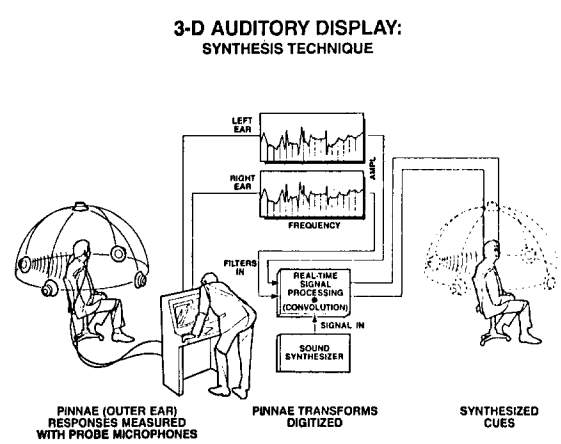

Figure 1. A conceptual illustration of the synthesis technique.

Meta Platforms, Inc.
Exhibit 1042
Page 002

Probe microphones are placed near the two eardrums of a listener seated in the anechoic chamber of the Psychoacoustics Lab at Wisconsin. The subject is at the center of a spherical array of loudspeakers with a radius of 1.4 meters and a density of 15 degrees azimuth (range: 0 to 360 degrees) and 18 degrees elevation (range: -36 to 90 degrees where 0 is at ear level). The stimulus, consisting of a 4 Hz train of 100 acoustic clicks at a level of approximately 80 dB SPL, is then presented to the subject over one of the speakers. The responses of the probe microphones to the 100 clicks are averaged, digitized at a sampling rate of 50 kHz, and stored as 16-bit numbers. A new pair of filters or pinna transforms is then measured for each location.

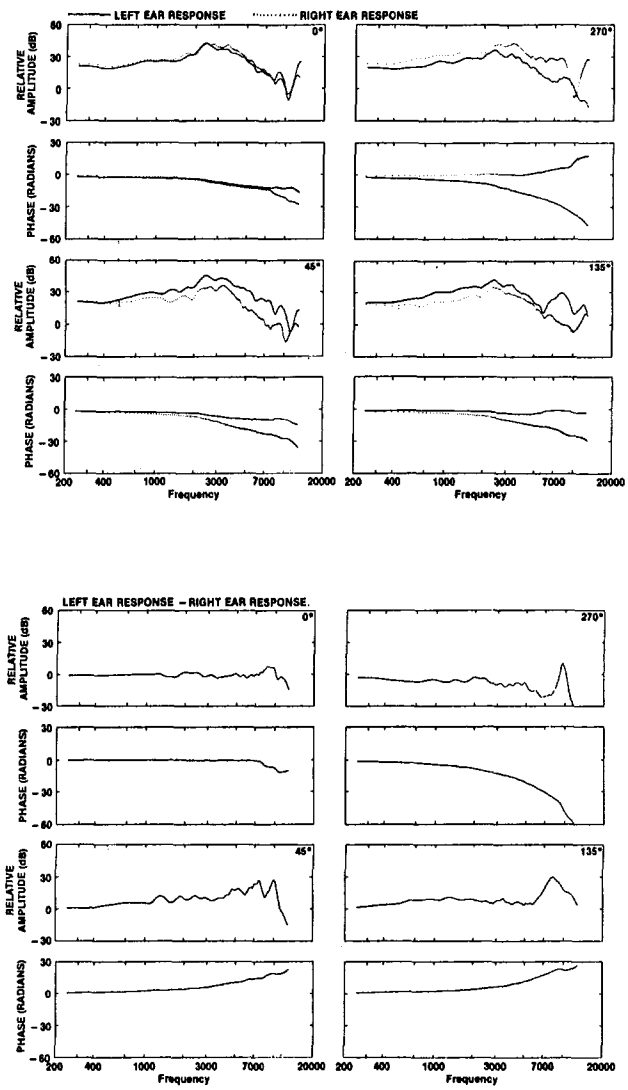

Figure 2. The panels on the top show pinna transform pairs for a single listener for four source locations. The bottom panels plot the interaural difference cues for the same locations. (See text.)

Figure 2 gives examples of empirical measurements of pairs of pinna transforms for a single listener for source locations at 0, 45, 135, and 270 degrees on the azimuthal plane (0 degrees is defined as directly in front, degrees increase counter-clockwise). The eight panels on the top plot the amplitude (relative decibels) and phase (radians) responses of the Fourier transforms for both the left (solid lines) and right ears (dotted lines) at the four locations. The eight bottom panels illustrate how the interaural difference cues change with location by plotting the difference between the left and right amplitude and phase responses as a function of frequency. Positive values indicate that the more intense and leading signal.

The table of 'location filters' derived from a complete set of measurements acts as input to a realtime signal-processing device designed by Scott Foster and currently being prototyped at Ames. The device convolves an arbitrary signal, such as a voice, with the measured filtered pairs, thus 'placing' it within the perceptual 3-space of the user. The initial configuration allows up to three independent and simultaneous sources, requiring approximately 100,000,000 multiply-accumulates per second. The resulting data stream is converted to an analog signal and presented via headphones. Prior to the actual synthesis, however, the spectral effects of the original loudspeakers and the headphones will have been removed digitally.

Motion trajectories and static locations at greater resolutions than the empirical measurements are simulated by interpolation. When integrated into VIEW, the operator's position can be monitored in realtime so that several simultaneous sources can be maintained in fixed positions or motion trajectories relative to the user. Such head-coupling should substantially enhance the simulation since previous work indicates that head movements may be important for localization (Thurlow & Runge, 1967). Informal tests at the University of Wisconsin and at Ames also suggest that this approach is feasible; simple linear interpolations between locations as far apart as 60 degrees azimuth are perceptually indistinguishable from stimuli synthesized from

Meta Platforms, Inc.
Exhibit 1042
Page 003

the measured coefficients. As with any system required to compute data 'on the fly', the term realtime is a relative one. The present system is designed to have a maximal latency or directional sampling interval of 50 msec. Recent work on the perception of auditory motion by Perrott (1982) and others indicates that this value should be acceptable since it is less than the minimum perceivable delay measured by the Minimum Audible Movement Angle for speeds of up to 360 degrees/sec.

Perceptual Verification of the Synthesis

A recent study at the University of Wisconsin has confirmed the perceptual adequacy of the basic synthesis technique. Table 1 compares the outcome of several global statistics for six subjects' absolute judgements of location under free-field (boldface type) and synthesized (in parentheses) conditions. For example, overall goodness of fit between actual and estimated source coordinates are 0.89 or better for the synthesized stimuli and 0.91 or better for real sources. The correlation measures, however, indicate that while source azimuth appears to be synthesized nearly perfectly for all listeners, synthesis of source elevation is less than perfect for some subjects. Similarly, the percentage of front-back reversals for synthesized stimuli is consistently greater for the same listeners. Thus, while individual differences do occur, the pattern of results across stimulus conditions is consistent within subjects. That is, the frequently cited notion of 'good' and 'bad' localizers appears to be supported by these data.

Global Measures of Localization Performance Comparison of Free-Field vs. Synthesized Stimuli								
Subject ID	Goodness of Fit		Azimuth Correlation		Elevation Correlation		% Front-Back Reversals	
SDE	.91	(.89)	.993	(.992)	.69	(.50)	13.4	(20.6)
SDH	.95	(.95)	.991	(.983)	.94	(.86)	4.9	(10.3)
SDL	.97	(.95)	.995	(.991)	.92	(.91)	6.9	(13.6)
SDM	.98	(.96)	.995	(.996)	.95	(.94)	4.6	(9.4)
SDO	.97	(.97)	.996	(.996)	.97	(.95)	3.0	(14.7)
SDP	.99	(.97)	.999	(.997)	.98	(.92)	3.7	(5.6)

Table 1. Measures of free-field performance are in boldface type and measures of performance during synthesized conditions are in parentheses.

Human Factors Applications

The prototype auditory display being developed at the Aerospace Human Factors Research Division of NASA-Ames Research Center will form part of the Virtual Interactive Environment Workstation (VIEW), a head-mounted, wide-angle, stereoscopic display system controlled by operator position, voice, and gesture.

The potential applications of a 3-D acoustic display are many. Such an interface will be valuable in any situation where the user's awareness of his spatial surroundings is important. This information is particularly critical under high workload, especially when visual cues are degraded or absent and direct sensation of the auditory world may not be possible or desirable.

One application suggested for the Space Station is the use of a spatial auditory display during Proximity Operations for monitoring traffic in the vicinity of the station and controlling docking with the shuttle (Figure 3). Localized acoustic cues can provide direct information about spatial positions and relationships in a situation with limited field-of-view. The operator need not rely solely on his/her interpretation of a three-dimensional space which has been transformed into a two-dimensional visual display.

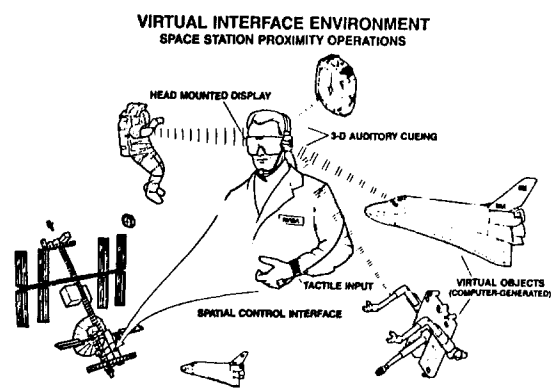

Figure 3. A possible application of 3-D auditory cues: Proximity Operations Display for the Space Station.

When integrated with VIEW, the 3-D auditory display will be coupled to the visual and gestural subsystems via realtime position-tracking

Meta Platforms, Inc.
Exhibit 1042
Page 004

so that simultaneous sound sources can be maintained in fixed positions or motion trajectories relative to the user. Thus, the display can provide an additional information channel while substantially augmenting the operator's situational awareness during activities such as supervision and control of telerobotic operations in hazardous environments (Figure 4). Even quite simple auditory cues, such as a sound signalling direction, distance, and finally contact with a virtual object could greatly aid the user in manipulating the remote world.

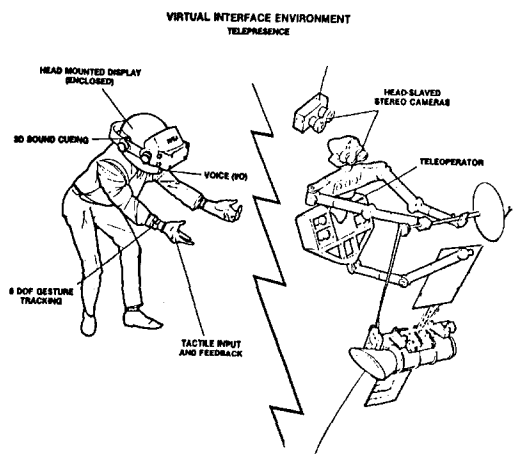

Figure 4. 3-D auditory displays can provide an additional information channel and augment situational awareness during telerobotic operations.

Future Research Directions

The initial psychophysical data indicate that the technique provides a successful first attempt at synthesizing localized sound, particularly for source azimuth. However, many factors not directly addressed here will need to be considered in producing a veridical percept. Earlier work, for example, suggests that localization may be influenced by the nature of accompanying visual cues (Gardner, 1968), the listener's familiarity with the sounds to be localized (Coleman, 1962), and individual differences in pinna transforms (Butler and Belendiuk, 1977). Also, relatively little is known about the nature of the critical cues for distance (Coleman, 1962) or the potential contribution of reverberation (Blauert, 1983) to an accurate perception of auditory space. In the past it has not been possible to adequately test many aspects of these problems simply because it was technically too difficult to put the stimuli under direct experimental control. The realtime signal-processing device under development at Ames should prove a very useful tool for examining some of these issues as well as furnish the basic technology for sophisticated auditory displays.

References

Batteau, D.W. (1968) Listening with the naked ear. In *The Neuropsychology of Spatially Oriented Behavior,* S.J. Freedman (Ed.), Dorsey Press, Homewood, IL.

Blauert, J. (1983) *Spatial Hearing: The Psychophysics of Human Sound Localization.* MIT Press: Cambridge, MA.

Butler, R.A. & Belendiuk, K. (1977) Spectral cues utilized in the localization of sound in the median sagittal plane. *J. Acoust. Soc. Am.,* 61, 1264-1269.

Coleman, P.D. (1962) Failure to localize the source distance of an unfamiliar sound. *J. Acoust. Soc. Am.,* 34, 345-346.

Doll, T.J., Gerth, J.M., Engelman, W.R. & Folds, D.J. (1986) Development of simulated directional audio for cockpit applications. USAF Report No. AAMRL-TR-86-014.

Fisher, S.S., McGreevy, M., Humphries, J., & Robinett, W. (1986) Virtual Environment Display System. *ACM Workshop Interact. 3D Graph.,* Chapel Hill, N.C.

Foster, S.H. (1986) Impulse response measurement using Golay codes. *Proc. Int. Conf. Acoust. Speech. Sig. Proc.*

Gardner, M.B. (1968) Proximity image effect in sound localization. *J. Acoust. Soc. Am.,* 43, 163.

Gardner, M.B. & Gardner, R.S. (1973) Problem of localization in the median plane: Effect of pinnae cavity occlusion. *J. Acoust. Soc. Am.,* 53, 400-408.

Mehrgardt, S. & Mellert, V. (1977) Transformation characteristics of the external human ear. *J. Acoust. Soc. Am.,* 61, 1567-1576.

Moore, F.R. (1983) A general model for spatial processing of sounds. *Comp. Mus. J.,* 7, 6-15.

Meta Platforms, Inc.
Exhibit 1042
Page 005

Oldfield, S.R. & Parker, S.P.A. (1984) Acuity of sound localisation: a topography of auditory space. II. Pinna cues absent. *Perc.*, 13, 601-617.

Perrott, D.R (1982) Studies in the perception of auditory motion. In *Localization of Sound: Theory and Applications,* R.W. Gatehouse (Ed.), Amphora Press: Groton, CN, 169-193.

Plenge, G. (1974) On the difference between localization and lateralization. *J. Acoust. Soc. Am.*, 56, 944-951.

Lord Rayleigh [Strutt, J.W.] (1907) On our perception of sound direction. *Phil. Mag.*, 13, 214-232.

Shaw, E.A.G. (1974) The external ear. In *Handbook of Sensory Physiology, Vol. V/1, Auditory System,* W.D. Keidel & W.D. Neff (Eds.), Springer-Verlag: New York.

Thurlow, W.R. & Runge, P.S. (1967) Effect of induced head movements on localization of direction of sounds. *J. Acoust. Soc. Am.*, 42, 480-488.

Watkins, A.J. (1978) Psychoacoustical aspects of synthesized vertical locale cues. *J. Acoust. Soc. Am.*, 63, 1152-1165.

Wightman, F.L., & Kistler, D.L. (1980) A New "Look" at Auditory Space Perception. In *Psychophysical, Physiological, and Behavioural Studies in Hearing,* G. van den Brink & F.A. Bilsen (Eds.), Delft University Press, The Netherlands.

About the Authors

Elizabeth M. Wenzel is a Research Psychologist in the Aerospace Human Factors Research Division at NASA-Ames Research Center. She received her Ph.D. in Cognitive Psychology, with an emphasis in auditory perception, from U.C., Berkeley in 1984.

Frederic L. Wightman is a professor in the Department of Psychology at the University of Wisconsin, Madison. He also holds a joint research appointment in psychoacoustics at the Waisman Center.

Scott H. Foster is an electrical engineer with over 10 years of experience in the design of audio signal processing hardware and software. He received his B.S. in Mathematics from M.I.T. in 1976.

Meta Platforms, Inc.
Exhibit 1042
Page 006