A VIRTUAL DISPLAY SYSTEM FOR CONVEYING THREE-DIMENSIONAL ACOUSTIC INFORMATION

Elizabeth M. Wenzel
NASA-Ames Research Center
Moffett Field, CA

Frederic L. Wightman
University of Wisconsin
Madison, WI

Scott H. Foster
Crystal River Engineering
Groveland, CA

ABSTRACT

A three-dimensional auditory display could take advantage of intrinsic sensory abilities like localization and perceptual organization by generating dynamic, multidimensional patterns of acoustic events that convey meaning about objects in the spatial world. Applications involve any context in which the user's situational awareness is critical, particularly when visual cues are limited or absent; e.g., air traffic control or telerobotic activities in hazardous environments. Such a display would generate localized cues in a flexible and dynamic manner. Whereas this can be readily achieved with an array of real sound sources or loudspeakers, the NASA-Ames prototype maximizes flexibility and portability by synthetically generating three-dimensional sound in realtime for delivery through headphones. Psychoacoustic research 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 complete, recent studies at the University of Wisconsin have confirmed the perceptual adequacy of the basic approach to synthesis.

INTRODUCTION

With rapid advances in technology and the concomitant requirement for managing complex informational systems, an increasing amount of research has been devoted to reconfigurable interfaces like the virtual display. As computer graphics become more sophisticated, virtual displays are assuming a three-dimensional spatial organization, providing a more natural means of interaction which may improve access and manipulation of data. Recent projects, building upon Sutherland's work (1968) on binocular head-mounted displays, have taken the spatial metaphor the farthest by directly involving the operator in the data environment. The goal is to create a highly flexible and interactive simulation/telepresence environment which integrates visual, auditory, tactile, and kinesthetic cues into a complex three-dimensional, virtual world (Fisher, this volume; Furness, 1986; Brooks, 1988).

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 information channel. We believe that a three-dimensional auditory display can substantially enhance situational awareness by combining spatial and semantic information to form dynamic, multidimensional patterns of acoustic events which convey meaning about objects in the spatial world of the user. Such a display can be realized with an array of real sound sources or loudspeakers (Doll et. al., 1986). The signal-processing device being developed at NASA-Ames maximizes flexibility and portability by synthetically generating three-dimensional sound in realtime for delivery through headphones. Unlike conventional stereo, sources can be perceived outside the head at discrete distances and directions from the listener. The 3-D auditory display will be integrated with Ames' Virtual Interactive Environment Workstation (VIEW) which allows the user to explore and interact with a 360-degree synthesized or remotely-sensed world using a head-mounted, wide-angle, stereoscopic display controlled by operator position, voice, and gesture (Fisher, this volume).

SYNTHESIS OF LOCALIZED SOUND

Theory

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 hypothesis. 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,

Downloaded from pro.sagepub.com at PENNSYLVANIA STATE UNIV on September 17, 2016

Meta Platforms, Inc.
Exhibit 1104
Page 001

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, others 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.

Technique

A number of techniques for synthesizing various features of auditory spatial perception have been explored. Among these are simulation of elevation cues with comb-filtering techniques, creation of auditory "spaciousness" using reverberation effects, methods of binaural recording, structural modelling of the pinnae, and the measurement of pinna transform functions (see Blauert, 1983 for discussions of these topics). Our technique involves preserving the frequency-dependent interaural time (phase) and intensity information with precise measurements of Head-Related Transfer Functions (HRTFs); the listener-specific, direction-dependent acoustic effects imposed on an incoming signal by the outer ears.

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 semicircular arc (1.4 m radius) with eight loudspeakers mounted at 18 degree intervals of elevation (range: -36 to 90 degrees where 0 is at ear level). The arc is then re-positioned at 15-degree increments of azimuth (range: 0 to 360 degrees) in order to sample most of the spherical space. The stimulus, a 50-Hz train of 20.48-msec bursts of pseudorandom noise, is presented over one of the speakers at a level of approximately 70 dB SPL and the responses of the probe microphones to 1000 repetitions of the noise-burst are averaged (digitized with 16-bit A/D converters at a sampling rate of 50 kHz). A new pair of HRTFs or pinna transforms is then measured for each location.

Although the HRTFs are measured in the time domain, application of the Fourier Transform and linear systems theory allows analysis of the measurements in the frequency domain and the construction of pairs of digital filters for each location. Thus, the eight panels in Figure 1 exemplify the data with plots of the amplitude (relative decibels) and phase (radians) responses derived from the Fourier Transforms of the empirical measurements at the left (solid lines) and right ears (dotted lines) of a single listener. The HRTFs shown are for source locations at 0 degrees elevation and 0, 45, 135, or 270 degrees azimuth (0 is directly in front, degrees increase counter-clockwise).

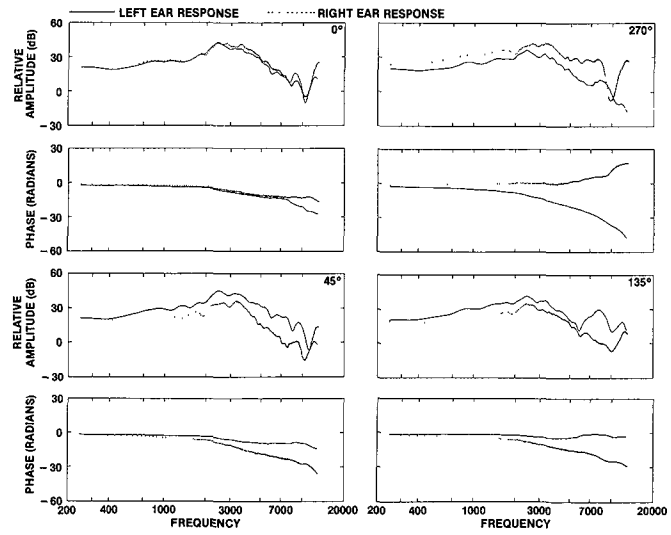

Figure 1. The eight panels plot HRTFs for a single listener for four source locations. (See text.)

A table of "location filters" is constructed from the pairs of HRTFs by first transforming them to the frequency domain, removing the spectral effects of the original loudspeakers and headphones using Fast Fourier Transform techniques, and then transforming back to the time domain. The table of corrected HRTFs acts as input to a realtime signal-processing device designed by Scott Foster and currently being prototyped at Ames. The device performs a mathematical convolution of an arbitrary signal, such as a voice, with the filter coefficients thus "placing" the signal within the perceptual 3-space of the user. The initial configuration allows up to three independent and simultaneous sources, requiring more than 100,000,000 multiply-accumulates per second. The resulting data stream is converted to an analog signal and presented over headphones.

Motion trajectories and static locations at greater resolution 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

Downloaded from pro.sagepub.com at PENNSYLVANIA STATE UNIV on September 17, 2016

Meta Platforms, Inc.
Exhibit 1104
Page 002

fixed locations 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 the approach is feasible; simple linear interpolations between locations as far apart as 60 degrees azimuth are perceptually indistinguishable from stimuli synthesized from measured coefficients. As with any system required to compute data "on the fly", the term "real-time" is a relative one. The present system is designed to have a maximal latency or directional sampling interval of 30-50 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.

PSYCHOPHYSICAL VALIDATION

Basic Synthesis Technique

The working assumption of the synthesis technique is that if, using headphones, we could produce ear canal waveforms identical to those produced by a free-field source, we would duplicate the free-field experience. The only conclusive test of this assumption must come from psychophysical studies in which free-field and synthesized, free-field listening are directly compared. A recent study at the University of Wisconsin has confirmed the perceptual adequacy of the basic approach for static sources. The stimuli were similar to those used to measure the HRTFs and subjects were blindfolded. Table 1 compares six subjects' absolute judgements of location under free-field and synthesized conditions. Note that 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 free-field sources. The correlation measures indicate that while source azimuth appears to be synthesized nearly perfectly, synthesis of source elevation is less than perfect, particularly for subjects that have difficulty judging elevation in the free-field. Similarly, the percentage of front-back reversals for synthesized stimuli is greater for all listeners.

Localization performance was also examined for regional differences, e.g., low, middle, and high elevations crossed with front, side, and back azimuths. For the sake of brevity, the data are only summarized here. As in previous free-field studies (e.g., Oldfield & Parker, 1984), localization is generally poorest at high elevations in the rear for both free-field and synthesized stimuli. Like the global data, the nine regions show a close correspondence between each subject's judgements in the free-field and synthesized conditions, although correlations are somewhat lower and front-back confusions approximately double for synthesized stimuli. Again, performance is generally better for azimuth than elevation, but with no obvious regional differences in the size and variability of errors for free-field and headphone listening. Thus, while individual differences do occur, the pattern of results across stimulus conditions is quite consistent for a given subject. That is, the frequently cited notion of "good" and "bad" localizers (Butler & Belendiuk, 1977) appears to be supported by both the global and regional analyses.

Global Measures of Localization Performance

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 localization performance for 6 subjects are compared for free-field (boldface type) and synthesized stimuli (in parentheses).

Acoustic Determinants of Performance

The presence of substantial individual differences in localization behavior suggests that there are acoustic features peculiar to each subject's HRTFs which influence performance. Thus, the use of "average" transforms, or even measurements derived from normative manikins such as the Kemar, may not be an optimum approach for simulating free-field sounds. For example, Figure 2 illustrates the between-subjects variability in the magnitude responses for a single source location. It can be seen that any straightforward averaging of these functions would tend to smooth the peaks and valleys, thus removing potentially significant features in the acoustic transforms.

Alternatively, it may be possible to enhance or avoid specific features of transforms which result in good or bad localization. The psychophysical data indicate that

Downloaded from pro.sagepub.com at PENNSYLVANIA STATE UNIV on September 17, 2016

Meta Platforms, Inc.
Exhibit 1104
Page 003

elevation is particularly difficult to judge, especially for subject SDE. A preliminary analysis of elevation coding suggests that there is an acoustic basis for this poor performance. Figure 3 plots "interaural elevation dependency" functions for four subjects' magnitude data. Interaural difference functions were first computed by dividing all leading ear HRTFs by their corresponding lagging ear HRTFs. These difference functions were then "normalized" to zero elevation by dividing all the difference functions at a given azimuth by the difference function at zero elevation and that azimuth. Finally, the normalized, elevation dependency functions were averaged over all azimuths. Thus, for each subject in Figure 3, the six functions show how interaural intensity changes for the different elevations with respect to zero elevation (the flat function). In spite of the large intersubject variabilty exemplified by Figure 2, the functions for the better localizers, SDL, SDH, and SDO, are quite similar to each other and show clear elevation dependencies. (The functions for the two subjects not shown were also similar.) SDE's, on the other hand, are different from the other subjects and show little elevation dependency. Thus, it appears that SDE's poor performance in judging elevation for both real and synthesized stimuli may be due to the lack of distinctive acoustic features in the HRTFs.

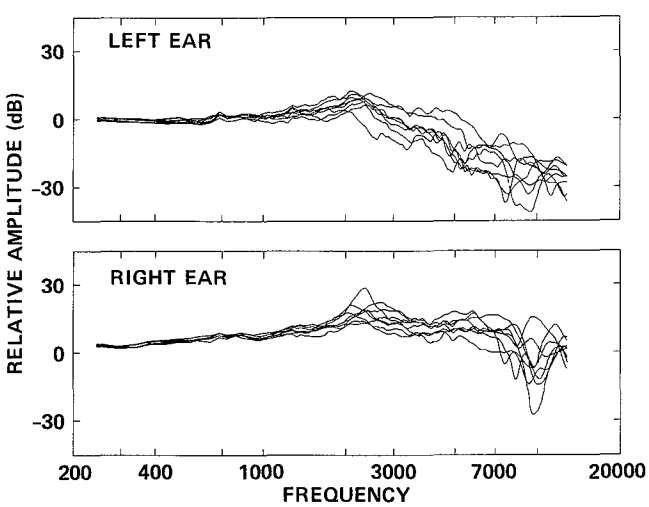

Figure 2. Magnitude responses for a single source position for 8 subjects. The left and right ears are plotted separately.

HUMAN FACTORS APPLICATIONS

There are many potential applications of a 3-D acoustic display. Such an interface will be valuable in any context 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. Examples include tracking the locations of other objects and traffic during extra-vehicular activity in space or Nap-of-the-Earth flight in the helicopter, acoustic navigation displays for the blind, advanced teleconferencing environments, and sophisticated auditory image processing for the entertainment industry (also see Fisher, this volume).

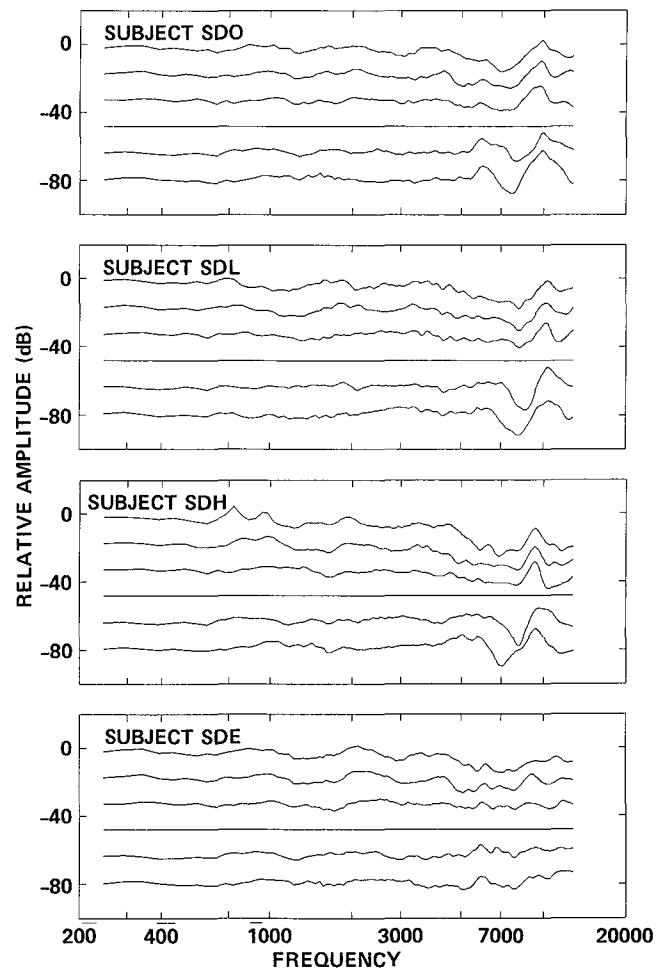

Figure 3. Interaural elevation dependency functions are plotted for 4 subjects. From top to bottom, the functions within a panel represent elevations of 54, 36, 18, 0 (the reference elevation), -18, and -36 degrees (See text.)

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 4). Localized acoustic cues can provide direct information about spatial relationships in a situation with limited field-of-view and no natural acoustic input. The operator need

Downloaded from pro.sagepub.com at PENNSYLVANIA STATE UNIV on September 17, 2016

Meta Platforms, Inc.
Exhibit 1104
Page 004

not rely solely on his/her interpretation of a three-dimensional space which has been transformed into a two-dimensional visual display or camera viewpoint. Even quite simple auditory cues, such as a sound signalling direction, distance, and finally contact with the virtual shuttle could greatly aid in supervising the remote docking operation.

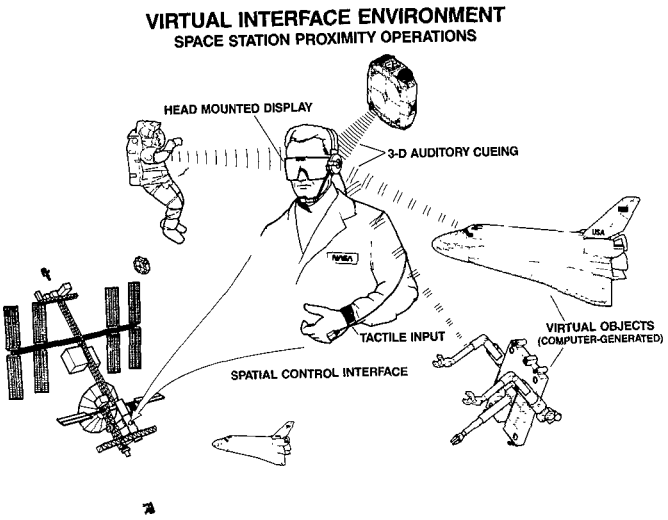

Figure 4. A possible application of 3-D auditory cues: Proximity Operations Display for the Space Station.

CONCLUSIONS

The 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. For example, relatively little is known about the nature of auditory motion perception (Perrott; Strybel, this volume), the critical cues for distance (Coleman, 1962) or the potential contribution of reverberation (Blauert, 1983) to an accurate perception of auditory space. Earlier work also 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 cues (Butler and Belendiuk, 1977). Indeed, the data reported here suggest that differences in the pinnae determine an individual's localization performance, especially for elevation.

Finding solutions to problems like listener-dependence will have a critical impact on the general utility of a three-dimensional auditory display in any application. In the past it has not been possible to adequately test many aspects of these questions 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 acoustic displays.

REFERENCES

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

Brooks, F.P. (1988) Grasping reality through illusion -- Interactive graphics serving science. Proc. CHI'88, ACM Conf. Hum. Fac. Comp. Sys., Washington, D.C., 1-11.

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.

Furness, T.A. (1986) The super cockpit and its human factors challenges. Proc. Hum. Fac. Soc., 1986 (1), 48-52.

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.

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.

Sutherland, I.E. (1968) Head-mounted three-dimensional display. Proc. Fall Joint Comp. Conf., 33, 757-764.

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.

Downloaded from pro.sagepub.com at PENNSYLVANIA STATE UNIV on September 17, 2016

Meta Platforms, Inc.
Exhibit 1104
Page 005