Localization with non-individualized virtual acoustic display cues

Elizabeth M. Wenzel

Aerospace Human Factors Research Division
NASA-Ames Research Center, Mail Stop 262-2
Moffett Field, CA 94035
(415) 604-6290
beth@eos.arc.nasa.gov

Frederic L. Wightman, Doris J. Kistler

Department of Psychology & Waisman Center
University of Wisconsin, Madison
Madison, WI 53705
(608) 263-3270
wightman%gustav.decnet@vms3.macc.wisc.edu

ABSTRACT

A recent development in advanced interface technologies is the virtual acoustic display, a system that presents three-dimensional auditory information over headphones [20]. The utility of such a display depends on the accuracy with which listeners can localize the virtual, or simulated, sound sources. Synthesis of virtual sources involves the digital filtering of stimuli using filters based on acoustic Head-Related Transfer Functions (HRTFs) measured in human ear-canals. In practise, measurement of the HRTFs of each potential user of a 3-D display may not be feasible. Thus, a critical research question is whether listeners from the general population can obtain adequate localization cues from stimuli based on non-individualized filters.

In the present study, 16 inexperienced listeners judged the apparent spatial location (azimuth and elevation) of wideband noisebursts that were presented either over loudspeakers in the free-field (an anechoic or non-reverberent environment) or over headphones. The headphone stimuli were synthesized using HRTFs from a representative subject in a previous study [23]. Localization of both free-field and virtual sources was quite accurate for 12 of the subjects, 2 showed poor elevation accuracy in both free-field and headphone conditions, and 2 showed degraded elevation accuracy only with virtual sources. High rates of confusion errors (reversals in judgements of azimuth and elevation) were also observed for some of the subjects and tended to increase for the virtual sources. In general, the data suggest that most listeners can obtain useful directional information from an auditory display without requiring the use of individually-tailored HRTFs, particularly for the dimension of azimuth. However, the high rates of confusion errors remain problematic. Several stimulus characteristics which may help to minimize these errors are discussed.

KEYWORDS: User-interface design issues, virtual acoustic displays, auditory perception & localization.

Permission to copy without fee all or part of this material is granted provided that the copies are not made or distributed for direct commercial advantage, the ACM copyright notice and the title of the publication and its date appear, and notice is given that copying is by permission of the Association for Computing Machinery. To copy otherwise, or to republish, requires a fee and/or specific permission.
© 1991 ACM 0-89791-383-3/91/0004/0351...$1.50

INTRODUCTION

Recently, a considerable amount of attention has been focused on the development of a truly three-dimensional, interactive display called the virtual interface [e.g., 3, 10]. 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., 5]. 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. Primary advantages of the binaural auditory system include the ability to monitor and identify sources of information from all possible locations, improved intelligibility of sources in noise and enhanced segregation of multiple sound sources [6], and in conjunction with other modalities, the reinforcement of information combined with a greater sense of presence or realism [19]. These features could be critical for applications that are inherently spatial, such as an air traffic control display, applications involving encoded non-speech messages, such as the acoustic "visualization" of multi-dimensional data [17] or alternative interfaces for the visually-impaired [8], and telepresence applications, such as advanced teleconferencing environments and shared electronic workspaces [11]. Thus, the combination of veridical spatial cues with good principles of iconic design could provide an extremely powerful and information-rich display which is also quite easy to use.

A three-dimensional auditory display can be realized with an array of real sound sources or loudspeakers [e.g., 7]. An alternative approach is to generate externalized, three-dimensional sound cues over headphones in realtime using digital means [e.g., 20]. Headphone presentation is desirable because it enables complete control over the acoustic waveforms delivered to the two ears, a feature which is critical to the control of apparent spatial position. The realtime capability, coupled with a head-tracking device, allows the user to experience virtual sounds interactively; virtual sources can thus be either moving or static and they can respond appropriately to the listener's head movements.

Meta Platforms, Inc.
Exhibit 1106
Page 001

The utility of a 3-D auditory display depends critically on the user's ability to localize the various sources of information in auditory space. Thus, the design of a 3-D auditory display must carefully consider the acoustic cues needed by human listeners for accurate localization. Since the time of Lord Rayleigh [18], much of the research on human sound localization has emphasized the role of two primary cues, interaural differences in time of arrival at low-frequencies and interaural differences in intensity at high frequencies. However, more recent research points to serious limitations with this view, identifying the important contribution to localization of the direction-dependent filtering which occurs when incoming sound waves interact with the outer ears or pinnae. Experiments show that spectral shaping by the pinnae is highly direction-dependent [16], that the absence of pinna cues degrades localization accuracy [14], and that pinna cues are primarily responsible for externalization or the "outside-the-head" sensation [15]. Such data suggest that perceptually-veridical localization over headphones might be possible if the spectral shaping by the pinnae as well as the interaural differences of time and intensity can be adequately synthesized.

One technique for capturing both pinnae and interaural difference cues involves binaural recording with microphones placed in the ears of a manikin [15] or the ear-canals of a human [4]. When stimuli recorded this way are presented over headphones, there is an immediate and veridical perception of 3-D auditory space [2, 4, 15]. Our procedure is closely related to binaural recording. Rather than record stimuli directly, we measure the acoustical transfer functions, from free-field-to-eardrum, at many source positions, and use these transfer functions as the basis of filters with which we synthesize stimuli. The transfer functions are generally called Head-Related Transfer Functions (HRTFs) and are measured in the ear-canals of individual subjects [22].

The working assumption of the synthesis technique is that if, using headphones, one can produce ear-canal waveforms identical to those produced by a free-field source, the free-field experience will be duplicated. The only conclusive test of this assumption must come from psychophysical studies in which the perception of real and synthesized stimuli are directly compared. Wightman & Kistler [23] compared subjects' localization judgements of static sources in the free-field with their judgements of virtual (headphone-presented) sources synthesized from the subjects' own HRTFs. Presumably, synthesis using individualized HRTFs would be the most likely to replicate the free-field experience for a given listener. In general, Wightman & Kistler reported that localization accuracy for the free-field and headphone stimuli was comparable for the experienced listeners of their study. However, some minor degradation of localization accuracy was observed with headphone-presented sounds. The rate of front-back confusions increased from about 6 to 11% and source elevation appeared to be less well-defined for virtual sources.

If virtual sources are to be used in a general-purpose 3-D auditory display, it may not be feasible to measure the HRTFs from each potential listener. It may also be the case that the user of such a display may not have the opportunity for extensive training. Thus, a critical issue for the design of virtual acoustic displays is the degree to which the general population of listeners can obtain adequate localization cues from stimuli based on non-individualized HRTFs.

There is some evidence that the ability to localize, and especially to determine the elevation of a sound source, appears to depend on the use of acoustical cues provided by one's own ears [21]. Wenzel, et. al. [21] compared localization of real sources for 3 subjects with stimuli synthesized from individualized and non-individualized transfer functions. Localization with the non-individualized transforms was only somewhat degraded compared to performance with both free-field sources and individualized HRTFs, as long as the transforms were derived from what one might call a "good" localizer (e.g., subject SDO). Further, when performance degradations did occur, they were primarily seen in the subjects' elevation judgements for all stimulus conditions. The individual differences in performance appeared to be correlated with the presence or absence of elevation-dependent, acoustical features in the 5 to 10 kHz region of the subjects' HRTFs (see also Wightman & Kistler [23], Figure 10).

The preliminary results of Wenzel, et. al. [21] suggest that it may be feasible to use non-individualized transfer functions to synthesize 3-D auditory display cues, so long as the HRTFs that are used come from a subject whose measurements have been "behaviorally-calibrated" and are thus correlated with known perceptual ability in both free-field and headphone conditions. The present study represents a more comprehensive comparison of free-field and virtual free-field localization by inexperienced listeners using non-individualized transfer functions. The virtual free-field stimuli were generated digitally using HRTFs measured in the ear canals of a representative subject, SDO, from the study by Wightman & Kistler [23]. A relatively large number of inexperienced subjects were tested in an attempt to assess whether the general population of listeners could readily use a virtual acoustic display.

METHOD

Subjects

Sixteen young adults served as paid volunteers. All had normal hearing, verified by standard audiometric screening at 15 dB HL, and reported no history of hearing problems. None of the subjects had any previous experience in psychoacoustical experiments, and all were naive regarding the purpose of the experiment.

Stimuli

The basic stimulus consisted of a train of eight, 250-ms bursts of Gaussian noise with 20-ms, cosine-squared ramps at onset and offset and 300-ms of silence between the bursts. The overall level of the train was approximately 70 dB SPL in both the free-field and headphone conditions. The noise was bandpassed between 200 Hz and 14 kHz and its

Meta Platforms, Inc.
Exhibit 1106
Page 002

spectrum was scrambled on each experimental trial in order to minimize the contributions of stimulus familiarity and transducer idiosyncracies to the judgements of location. The scrambling algorithm divided the spectrum into adjacent critical bands and assigned a random intensity level (uniform distribution, 20-dB range) to the noise in each band. Stimuli for a given subject and a given set of trials were pre-computed and converted to analog form via 16-bit D/A converters at a rate of 50 kHz.

The stimuli were transduced either by loudspeakers or by headphones. Details of the presentation system can be found in [22, 23]. In the free-field case, six speakers were mounted on a semicircular arc, 2.76 m in diameter, at intervals of 18 degrees elevation (range -36 to +54 deg., where 0 is the horizontal plane passing through the subjects' ears). The arc was located in an anechoic chamber and could be rotated around the vertical axis, allowing stimulus presentation at any azimuth, with the subject seated on an adjustable stool such that his/her head was at the center of the arc.

For the headphone conditions, each stimulus was digitally-processed so that it would simulate a particular free-field location. The processing was based on the direction-specific, outer ear characteristics (HRTFs) measured for subject SDO in the study by Wightman & Kistler. Briefly, HRTFs are estimated by deconvolving the loudspeakers, test stimulus, microphone, and headphone responses from the recordings made with probe microphones. Further details of the measurement technique, synthesis procedure, and psychophysical task can be found in [22, 23].

Procedure

The goal of this experiment was to compare the inexperienced subjects judgements of the apparent locations of sound sources in free-field with their judgements of locations in a virtual free-field. The virtual sources were synthesized from a "generic" set of HRTFs in the sense that they derived from a single individual, SDO, whose localization accuracy was representative of seven of the eight subjects in the study by Wightman and Kistler [23]. The data from this subject showed good localization accuracy in both azimuth and elevation and an average (low) rate of front-back confusions in both free-field and headphone conditions.

Further, we hoped to compare the data from this study with those from the previous experiment in which free-field performance was contrasted with headphone performance with each subject using his/her own HRTFs. Thus, the same "absolute judgement" paradigm that was used to estimate perceived spatial location in the Wightman & Kistler [23] experiment was also used here. In both free-field and headphone conditions, the subject indicates the apparent spatial position of a sound source by calling out numerical estimates of azimuth and elevation (in degrees) using a modified spherical co-ordinate system. For example, a sound heard directly in front would produce a response of "0, 0", a sound heard directly to the left and somewhat elevated might produce "left 90, up 15", while one far to the rear on the right and below might produce "right 170, down 30".

Subjects heard stimuli from 24 different source positions in the free-field (anechoic chamber) and the same 24 positions in the virtual free-field (headphone) conditions. The source locations were chosen from the 144 positions tested by Wightman & Kistler [22] with the aim of sampling the possible range of azimuths and elevations equally. Each of the 24 positions was heard once in a single block of trials with a different randomized order for each 10-minute block. A total of 18 blocks of trials were run over a period of two days, with subjects giving 9 judgements of apparent spatial location for each of the 24 locations. Blocks of trials for the free-field and headphone conditions were alternated in order to minimize order effects. Prior to the experimental runs, a 15-minute training session was conducted which included a verbal explanation of the response co-ordinates and a practice block in the free-field condition. Relatively few positions were tested and minimal training given because it was important for the purposes of this study that subjects' performance reflect that of listeners inexperienced with localizing free-field or virtual free-field stimuli.

At the beginning of each session, subjects were blindfolded, led into the anechoic chamber, and seated at the center of the loudspeaker arc. Subjects never saw the inside of the anechoic chamber at any time during the study, whether in the free-field or headphone conditions. The subject was instructed to orient straight ahead and not move the head while a trial was in progress. Head position and stability was verified visually by the experimenter. For each trial, the subjects heard the train of eight identical 250-ms bursts of spectrally-shaped noise, and called out estimates of azimuth and elevation during a 5-s response interval. The experimenter recorded their responses and no feedback was given. Following a free-field block, the subjects remained in the anechoic chamber but donned the headphones in preparation for the virtual free-field conditions. The trial sequence was essentially the same as for the free-field condition.

RESULTS

Analysis of the results of a localization study such as the one described here is complicated by two factors. One is the fact that the stimuli and responses are represented by points in three-dimensional space; in particular, as points on the surface of a unit-sphere since distance remained constant in this experiment. For these spherically-organized data, the usual statistics of means and variances are potentially misleading. For example, an azimuth error of 15 degrees on the horizontal plane is much larger in terms of absolute distance than a 15-degree error at an elevation of 54 degrees. Thus, it is more appropriate to apply the techniques of spherical statistics to characterize these psychophysical data [9]. The spherical statistic reported here, the judgement centroid, is a unit-length vector with the same direction as the resultant, the vector sum of all the unit-length judgement vectors. The direction of the centroid can be thought of as the "average direction" of a set of judgements from the origin, the subject's position.

Meta Platforms, Inc.
Exhibit 1106
Page 003

A second complication in the analysis of localization data, observed in nearly all such studies, is the presence of front-back "confusions". These are responses which indicate that a source in the front hemisphere, usually near the median plane, was perceived to be in the rear hemisphere. Occasionally, the reverse situation is also heard. Further, in this experiment we also observed confusions in elevation, with up locations heard as down, and vice versa. These types of azimuth and elevation errors are illustrated in Figure 1. It is difficult to know how to treat these types of errors fairly. Since the confusion rate is often low, confusions have generally been resolved when computing descriptive statistics (i.e., the responses are coded as if they had indicated the correct hemisphere). Otherwise, estimates of error would be greatly inflated. On the other hand, if we assume that subjects' responses correctly reflect their perception, resolving such confusions could be misleading. Thus, the rate of confusions is usually reported as a separate statistic. Since we were primarily interested in comparing free-field and headphone conditions, we elected to resolve both front-back and up-down confusions and report the rates for each type of confusion. If the angle between the target and judged location was made smaller by reflecting the judgement about the vertical plane passing through the subject's ears (or, in the case of up-down confusions, the horizontal plane passing through the ears), the judgement was coded in reflected form and the confusion count was increased.

Judgement centroids were computed separately for each of the 24 positions in both the free-field and headphone conditions, for each of the 16 subjects. The left panels of Figures 2, 3, and 4 show the relation between target and response azimuth (for all stimulus elevations) and the panels on the right show the relation between target and response elevation (for all stimulus azimuths) in both the free-field and virtual source conditions. Because of space limitations, only three types of plots are shown which represent the three response patterns observed in the experiment. Figure 2 plots centroids for 12 subjects who behaved similarly, while Figures 3 and 4 plot data from individual subjects who reflect the two remaining behavior patterns.

Table 1 shows the total percentage of azimuth confusions (both front-to-back and back-to-front, collapsed across all elevations) and elevation confusions (both up-to-down and down-to-up, collapsed across all azimuths) for the free-field and headphone conditions, for each subject. Chi-square values, which reflect whether a significant difference in confusion rates occurred between the the free-field and headphone conditions, are also given. Table 2 gives a breakdown of the confusion percentages by hemisphere.

The data reveal several important features of localization behavior by inexperienced subjects. One is the presence of individual differences evidenced in both the judgements of source identification and the confusion rates. For example, three types of response patterns are shown in the plots of resolved location judgements. An example of the first pattern can be seen in Figure 2. Here, localization performance for 12 subjects who behaved similarly is represented by the centroids of their individual centroid estimates; resolved judgements of the apparent position of

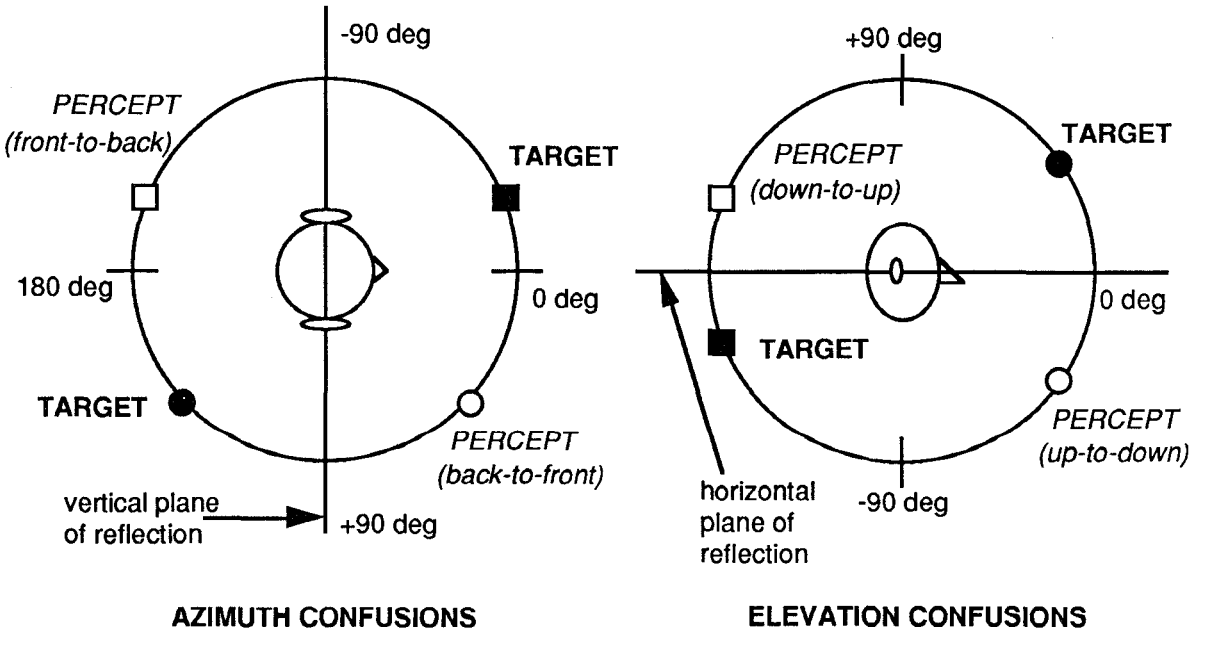

Figure 1. Illustration of the types of confusion errors observed for location judgements in the study. Reversals of perceived azimuth, with respect to the target azimuth, are shown on the left (i.e., front-to-back confusions: perception of a forward target in the rear hemisphere; back-to-front confusions: perception of a rear target in the forward hemisphere). Analogously, elevation reversals are illustrated on the right (i.e., up-to-down or down-to-up confusions).

Meta Platforms, Inc.
Exhibit 1106
Page 004

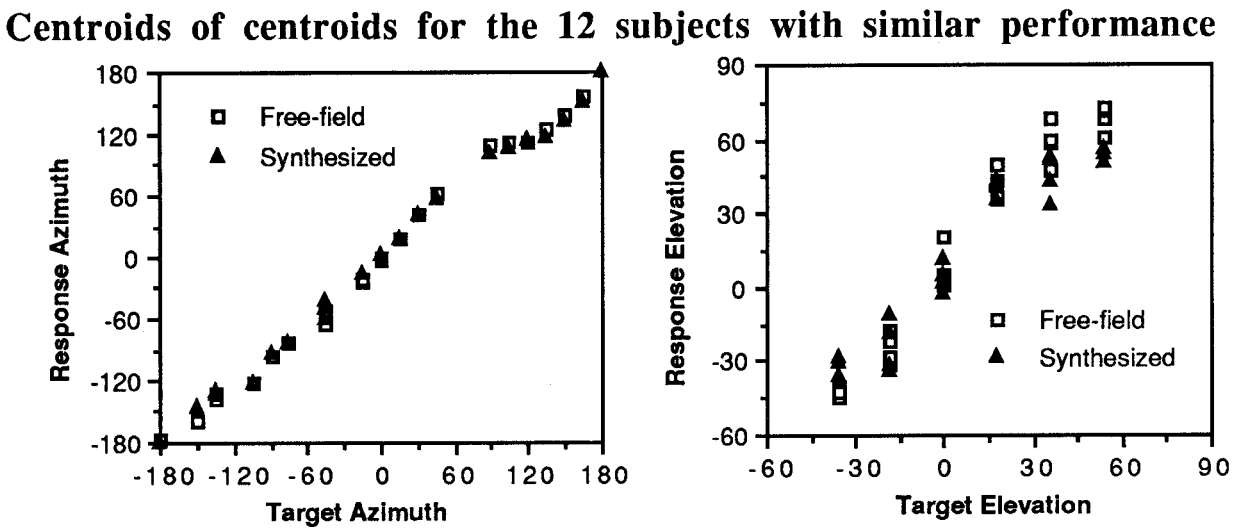

Figure 2. Scatterplots of centroids based on 12 individual subjects' centroids originally estimated from 9 judgements per target. Actual source locations versus judged source locations in both the free-field (open symbols) and headphone (solid symbols) conditions are plotted. The panel on the left plots azimuth judgements and the panel on the right shows elevation judgements. All 24 source positions are plotted in each panel. Thus, data from 6 different source elevations are combined in the azimuth panel and data from 18 different source azimuths are combined in the elevation panel. Sources at "left" azimuths and "down" elevations are shown as negative numbers.

virtual sources are nearly identical to resolved judgements of comparable stimuli in the free-field for both azimuth and elevation. Of the remaining subjects, all show comparable performance in azimuth judgements for both the free-field and headphone conditions (left panels of Figures 3 and 4). However, SID and SIE (see SID in Figure 3) show poor localization in elevation for both free-field and virtual free-field conditions, while SII and SIM (see SIM in Figure 4) only show a degradation in elevation accuracy with virtual sources. Thus, as in the experiment by Wightman & Kistler [23], differences among subjects emerge most clearly in the judgements of source elevation.

Perhaps the most striking feature of the data for these inexperienced listeners is the relatively high rate of azimuth confusions for both free-field and headphone conditions. Reversals also tend to increase for virtual sources when compared to free-field conditions; azimuth confusion rates increased significantly for 11 of the subjects' judgements (Table 1). For the 8 experienced subjects of the Wightman & Kistler [23] study, front-back confusion rates for the free-field ranged from 3 to 12% (mean: 5.6%) and approximately doubled under headphone conditions using the subjects' own HRTFs. Here, half of the subjects also showed relatively low free-field confusion rates (mean: 6.5%, range: 2 to 10%) which increased, on the average, by a factor of 3.8 for virtual sources synthesized from SDO's HRTFs. The other 8 subjects had much higher free-field confusion rates for azimuth judgements (mean: 32.2%, range: 20 to 43%) with an average increase of 1.26 for the headphone conditions. The breakdown of azimuth confusions shown in Table 2 indicates that the front-back confusion rates are due largely to reversals of front positions to the rear, a phenomenon which has been frequently reported by many studies of free-field localization in the median plane [see 2].

A few recent studies have also examined azimuth reversal rates. For example, the extensive free-field localization studies conducted by Oldfield & Parker [13, 14] report average rates of 3.4%, and 26% for normal free-field and pinnae-occluded conditions, respectively. Similarly, Makous & Middlebrooks [12] report an average rate of 6% (range 2 to 10%) for free-field localization. Recently, Asano, et. al. [1] examined localization of virtual sources in the median plane for two highly-practised subjects listening to sources synthesized from either their own or the other subject's transfer functions. One subject showed low reversal rates (1.9 and 0.9%) for both synthesis conditions while the other showed 7.4% and 14.8% for stimuli synthesized from his own and the other's HRTFs, respectively. These authors also examined performance for sources synthesized from simplified HRTFs (i.e., transfer functions modelled by different orders of filters derived from an auto-regressive moving-average model). The rate of azimuth confusions increased substantially (up to about 18 to 49%) as the modeled HRTFs were simplified, with the majority reversing to the rear for the smallest-order filters. Curiously, one of the subjects showed consistently fewer reversals when listening with the other subject's HRTFs.

To our knowledge, no previous studies have analyzed confusions in elevation. The rates observed here suggest that this type of reversal is less common; 14 of the 16 subjects showed free-field rates which were less than 10%. As with azimuth, reversals of elevation also increased for the virtual sources; compared to free-field conditions, confusion rates increased significantly for 14 of the

Meta Platforms, Inc.
Exhibit 1106
Page 005

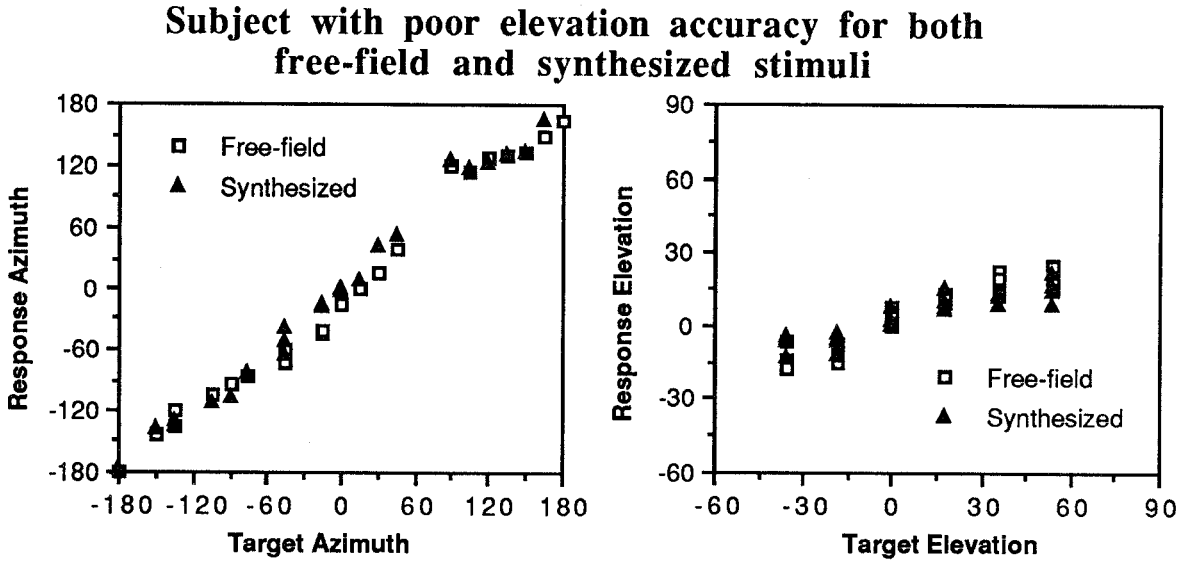

Figure 3. Same as Figure 2 except that data for an individual subject, SID, are plotted.

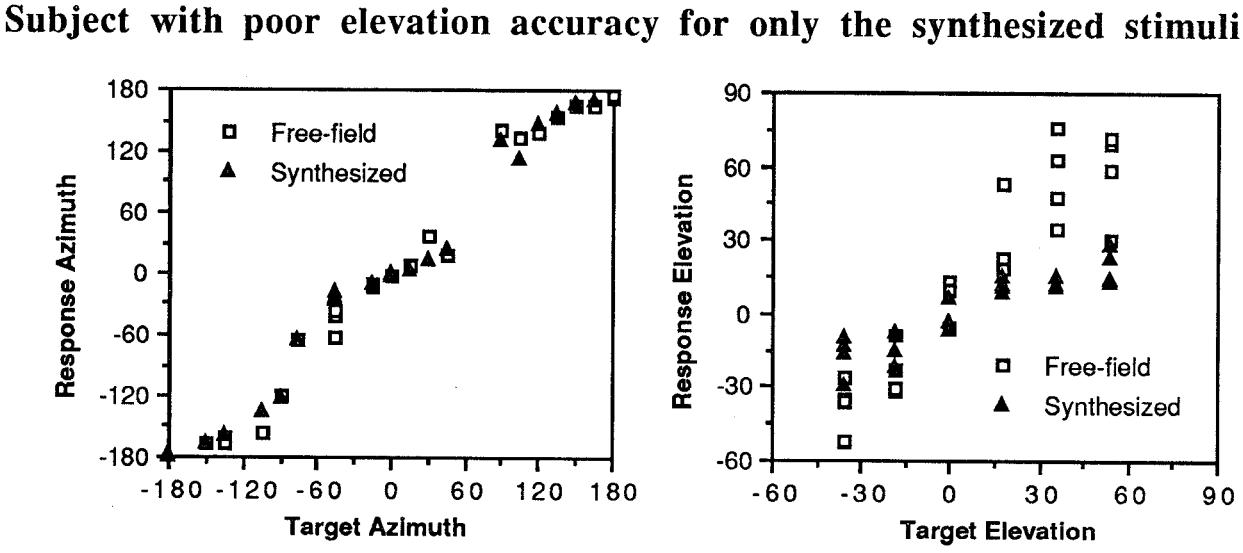

Figure 4. Same as Figure 2 except that data for an individual subject, SIM, are plotted.

subjects' elevation judgements (Table 1) with an average increase of a factor of 6.9. Unlike azimuth confusions, the breakdown of elevation confusions in Table 2 does not indicate an overall response bias toward either hemisphere.

DISCUSSION

The data indicate that the attempt to simulate free-field localization for inexperienced subjects with non-individualized transfer functions was largely successful for positions in azimuth when the criterion is the comparison to a subject's own free-field performance. When confusions are resolved, virtual sources synthesized from a "good" localizer's HRTFs are judged to have the same spatial locations as sources presented in the free-field for the limited range of stmuli investigated here. That is, the target-response plots imply strong correlations between the judged and actual source locations with a close correspondence between free-field and virtual sources. Similar to Wightman & Kistler's [23] observations using individualized HRTFs, when performance degradations do occur with virtual stimuli they are in the dimension of elevation. In fact, although judgement of elevation was poor for 4 of our subjects, only 2 showed a discrepancy between free-field and headphone conditions. Azimuth judgements, on the other hand, appear to be uniformly robust when compared with free-field performance in both experiments.

Meta Platforms, Inc.
Exhibit 1106
Page 006

ID	% Azimuth Confusions Free-Field	Headphones	X^2	% Elevation Confusions Free-Field	Headphones	X^2
SEY	**35**	47	5.5 *	**9**	24	15.7 **
SHT	**20**	34	10.1**	**4**	16	14.9**
SHU	**6**	36	53.5**	**3**	17	18.4**+
SHW	**20**	24	NS	**4**	27	42.8**+
SHY	**10**	18	6.1*	**1**	10	14.2**+
SHZ	**6**	23	22.1**	**1**	16	25.2**+
SIB	**38**	43	NS	**5**	16	10.9**
SIC	**8**	19	10.4**	**6**	27	27.8**
SIF	**7**	26	26.7**	**3**	29	44.8**
SIG	**36**	39	NS	**6**	8	NS
SIH	**2**	10	9.8**+	**1**	16	25.0**+
SIK	**7**	23	20.7**	**1**	7	7.0**+
SID	**33**	45	6.4*	**26**	17	4.3*
SIE	**33**	36	NS	**11**	18	4.3*
SII	**6**	27	32.3**	**7**	15	5.9*
SIM	**43**	50	NS	**4**	32	44.8**
Mean	**19**	31	7.7**	**6**	18	14.0**

Table 1. Percentages of front-back (azimuth) and up-down (elevation) confusions for free-field and headphone listening. Chi-square statistics are also indicated for those subjects which showed a significant difference in azimuth and/or elevation confusions for the free-field versus headphone conditions. (* $\alpha = < .05$; ** $\alpha < .01$; + indicates a continuity correction was applied to the chi-square). Free-field data are in **boldface type** and data for the synthesized stimuli are in plain text.

	% Azimuth Breakdown				% Elevation Breakdown			
	Front to Back		Back to Front		Down to Up		Up to Down	
ID	**FF**	HP	**FF**	HP	**FF**	HP	**FF**	HP
SEY	**29**	35	**6**	12	**3**	4	**6**	20
SHT	**18**	30	**2**	4	**4**	15	**0**	1
SHU	**4**	33	**2**	3	**2**	15	**1**	2
SHW	**20**	18	**0**	6	**1**	2	**3**	25
SHY	**9**	9	**1**	9	**1**	7	**0**	3
SHZ	**4**	15	**2**	8	**0**	1	**1**	15
SIB	**36**	42	**2**	1	**3**	11	**2**	5
SIC	**1**	3	**7**	16	**1**	10	**5**	17
SIF	**2**	12	**5**	14	**2**	1	**1**	28
SIG	**35**	36	**1**	3	**4**	1	**2**	7
SIH	**0**	5	**2**	5	**1**	6	**0**	10
SIK	**7**	14	**0**	9	**1**	5	**0**	2
SID	**32**	45	**1**	0	**26**	13	**0**	4
SIE	**33**	36	**0**	0	**10**	17	**1**	1
SII	**4**	18	**2**	9	**1**	0	**6**	15
SIM	**43**	50	**0**	0	**3**	18	**1**	14
Mean	**17**	25	**2**	6	**4**	8	**2**	11
S.D.	**15.1**	15.0	**4.6**	5.1	**6.3**	6.3	**2.1**	8.8

Table 2. Breakdown of confusion percentages by hemisphere for the free-field and headphone conditions. Percentages for the free-field (**FF; boldfaced type**) and headphone (HP; plain text) stimuli are shown. Percentages were based on the total number of possible azimuth or elevation judgements. Thus, a value of 50% implies all judgements were reversed in that category.

Meta Platforms, Inc.
Exhibit 1106
Page 007

As noted above, the most remarkable difference between the studies is in the high rate of front-back confusions. For free-field versus simulated free-field stimuli, the experienced listeners exhibit azimuth confusion rates of about 6 vs. 11% [23] while our inexperienced listeners show average rates of about 19 vs. 31%. In both studies these confusions are largely due to reversals to the rear. Further, although the average rates are lower, confusions in elevation are also a potential source of increased error in the simulation of localized sources.

In general, the data suggest that most listeners can obtain useful directional information from an auditory display without requiring the use of individually-tailored HRTFs, particularly for the dimension of azimuth. However, the high rates of confusion errors remain problematic. Comparison of these results to the study in which free-field listening is contrasted with localization of sources synthesized from a subject's own HRTFs [23], suggests that the use of non-individualized transforms primarily results in an increase in the rate of azimuth confusions, particularly front-to-rear confusions. Note, though, that the existence of free-field confusions indicates that these reversals are not strictly the result of the simulation. It is possible, as Asano, et. al. [1] have claimed, that these errors diminish as subjects adapt to the unusual listening conditions provided by static anechoic sources, whether real or simulated. The difference in free-field confusion rates between the inexperienced listeners of this experiment and the more experienced subjects of Wightman & Kistler tend to support this view. Perhaps the higher rates seen for virtual sources are at least partially the result of unfamiliar stimulus features contributed by the synthesis process which can eventually be overcome by adaptation.

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 2]. Assuming a stationary, spherical model of the head, a given interaural time difference correlates ambiguously with the direction of a sound source, with a conical shell describing the locus of all possible sources. However, cone-of-confusion effects alone cannot explain a front-to-back response bias, and it may be that visual dominance plays a substantial role in auditory localization. That is, given an ambiguous acoustic stimulus in the absence of an obvious visual correlate, it may be that the perceptual system resolves the ambiguity with a heuristic that assumes the source is behind where it can't be seen.

Several stimulus characteristics may help to minimize these errors. For example, the addition of visual cues, 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. A related problem in synthesizing veridical acoustic images is the fact that such stimuli sometimes fail to externalize, particularly when the signals are unfamiliar (e.g., the spectrally-scrambled noisebursts used here) and simulated from anechoic measurements of HRTFs. Thus additional cues which provide a sense of distance and environmental context, such as the ratio of direct to reflected energy and other characteristics specific to particular enclosed spaces, may also enhance the externalization of images [15]. However, the specific parameters used in such a model must be investigated carefully if localization accuracy is to remain intact. For example, Blauert [2] reports that the spatial image of a sound source grows larger and increasingly diffuse with increasing distance in a reverberant environment, a phenomenon which may tend to interfere with the ability to judge the direction of the source.

The simulation techniques investigated here provide both a means of implementing a virtual acoustic display and the ability to study features of human sound localization that were previously inaccessable due to a lack of control over the stimuli. The availability of realtime control systems [e.g., 20] further expand the scope of the research, allowing the study of dynamic, intersensory aspects of localization which may do much toward alleviating the problems encountered in producing the reliable and veridical perception which is critical for applied contexts.

REFERENCES

1. Asano, F., Suzuki, Y., & Sone, T. Role of spectral cues in median plane localization. *J. Acoust. Soc. Am.*, 88 (1990), 159-168.

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

3. Brooks, F.P. Grasping reality through illusion -- Interactive graphics serving science. *Proc. CHI'88,* (Washington, DC). ACM, New York, 1988, 1-11.

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

5. Buxton, W., Gaver, W., & Bly, S. *The use of non-speech audio at the interface.* Tutorial #10, CHI'89, ACM , New York, 1989.

6. Cherry, E.C. Some experiments on the recognition of speech with one and two ears. *J. Acoust. Soc. Am.*, 22 (1953), 61-62.

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

8. Edwards, A.D.N. Soundtrack: An auditory interface for blind users. *Hum. Comp. Interact.*, 4 (1989), 45-66.

9. Fisher, N.I., Lewis, T., & Embleton, B.J.J. *Statistical Analysis of Spherical Data.* Cambridge U. P.: Cambridge, United Kingdom, 1987.

Meta Platforms, Inc.
Exhibit 1106
Page 008

10. Fisher, S.S., Wenzel, E.M., Coler, C. & McGreevy, M.W. Virtual interface environment workstations. *Proc. Hum. Fac. Soc.*, 32 (1988), 91-95.

11. Ludwig, L., Pincever, N. & Cohen, M. Extending the notion of a window system to audio. *Computer*, (1990), 66-72.

12. Makous, J.C., & Middlebrooks, J.C. Two-dimensional sound localization by human listeners. *J. Acoust. Soc. Am.*, 87 (1990), 2188-2200.

13. Oldfield, S.R. & Parker, S.P.A. Acuity of sound localisation: a topography of auditory space. I. Normal hearing conditions. *Perc.*, 13 (1984), 581-600.

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

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

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

17. Smith, S., Bergeron. R.D., & Grinstein, G.G. Stereophonic and surface sound generation for exploratory data analysis. *Proc. CHI'90*, (Seattle, WA). ACM, New York, 1990, 125-132.

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

19. Warren, D.H., Welch, R.B., & McCarthy, T.J. The role of visual-auditory "compellingness" in the ventriloquism effect: Implications for transitivity among the spatial senses. *Perc. & Psychophys.*, 30 (1981), 557-564.

20. Wenzel, E.M., Wightman, F.L., & Foster, S.H. A virtual display system for conveying three-dimensional acoustic information. *Proc. Hum. Fac. Soc.*, 32 (1988), 86-90.

21. Wenzel, E.M., Wightman, F.L., Kistler, D.J., & Foster, S.H. Acoustic origins of individual differences in sound localization behavior. *J. Acoust. Soc. Amer.*, 84 (1988), S79.

22. Wightman, F.L. & Kistler, D.J. Headphone simulation of free-field listening I: stimulus synthesis. *J. Acoust. Soc. Amer.*,85 (1989), 858-867.

23. Wightman, F.L. & Kistler, D.J. Headphone simulation of free-field listening II: psychophysical validation. *J. Acoust. Soc. Amer.*, 85 (1989), 868-878.

Meta Platforms, Inc.
Exhibit 1106
Page 009