Localization using nonindividualized head-related transfer functions

Elizabeth M. Wenzel
Aerospace Human Factors Research Division, NASA-Ames Research Center, Mail Stop 262-2, Moffett Field, California 94035-1000

Marianne Arruda[a] and Doris J. Kistler
Waisman Center, University of Wisconsin–Madison, Madison, Wisconsin 53705

Frederic L. Wightman
Department of Psychology and Waisman Center, University of Wisconsin–Madison, Madison, Wisconsin 53705

(Received 27 May 1992; accepted for publication 10 March 1993)

A recent development in human–computer interfaces is the virtual acoustic display, a device that synthesizes three-dimensional, spatial auditory information over headphones using digital filters constructed from head-related transfer functions (HRTFs). The utility of such a display depends on the accuracy with which listeners can localize virtual sound sources. A previous study [F. L. Wightman and D. J. Kistler, J. Acoust. Soc. Am. **85**, 868–878 (1989)] observed accurate localization by listeners for free-field sources and for virtual sources generated from the subjects' own HRTFs. In practice, measurement of the HRTFs of each potential user of a spatial auditory display may not be feasible. Thus, a critical research question is whether listeners can obtain adequate localization cues from stimuli based on nonindividualized transforms. Here, inexperienced listeners judged the apparent direction (azimuth and elevation) of wideband noisebursts presented in the free-field or over headphones; headphone stimuli were synthesized using HRTFs from a representative subject of Wightman and Kistler. When confusions were resolved, localization of virtual sources was quite accurate and comparable to the free-field sources for 12 of the 16 subjects. Of the remaining subjects, 2 showed poor elevation accuracy in both stimulus conditions, and 2 showed degraded elevation accuracy with virtual sources. Many of the listeners also showed high rates of front-back and up-down confusions that increased significantly for virtual sources compared to the free-field stimuli. These data suggest that while the interaural cues to horizontal location are robust, the spectral cues considered important for resolving location along a particular cone-of-confusion are distorted by a synthesis process that uses nonindividualized HRTFs.

PACS numbers: 43.66.Qp, 43.66.Pn, 43.66.Yw [HSC]

INTRODUCTION

A new type of interface technology, the virtual acoustic display, is being developed to enhance information transfer for applications like cockpit and air traffic control displays, advanced communications and teleconferencing systems, architecural acoustic design, virtual reality systems, and acoustic "visualization" of multidimensional data (Wenzel, 1992). The goal is to combine both directional and iconic information in a naturalistic representation of dynamic objects in the interface. Spatial auditory displays can be realized with an array of loudspeakers (Doll *et al.*, 1986; Calhoun *et al.*, 1987). An alternative approach is to generate externalized, three-dimensional sound images in real time over headphones using digital (McKinley and Ericson, 1988; Wenzel *et al.*, 1988a; Wenzel, 1992; Persterer, 1989) or analog technology (Loomis *et al.*, 1990). Headphone presentation is desirable because it enables complete control over the acoustic waveforms delivered to the two ears. Real-time presentation, coupled with a head-tracking device, allows the user to experience virtual sounds interactively; virtual sources can thus be either moving or static and can respond appropriately to the listener's head movements.

Spatial auditory displays depend critically on the user's ability to localize the various sources of information in auditory space. Consequently, the design of such a display must carefully consider the acoustic cues needed by human listeners for accurate localization. One technique for capturing some of the cues involves binaural recording with microphones placed in the ears of a manikin (Plenge, 1974; Doll *et al.*, 1986) or the ear canals of a human (Butler and Belendiuk, 1977). When such stimuli are presented over headphones, there is a realistic perception of auditory space (Plenge, 1974; Butler and Belendiuk, 1977; Doll *et al.*, 1986), particularly with recordings made in a complex, reverberent environment such as a concert hall (Blauert, 1983, p. 358). The procedure used here is closely related to binaural recording. Rather than record stimuli directly, the acoustical transfer functions, from free-field-

[a]Current address: 39 Chipaway Road, East Freetown, MA 02717.

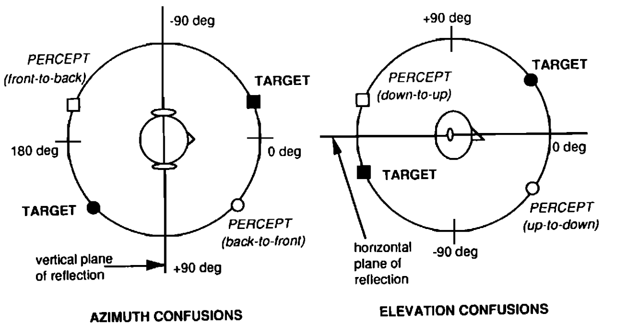

FIG. 1. Illustration of the types of confusion errors observed for location judgments in the study. Confusions of perceived azimuth, with respect to the target azimuth, are shown on the left (i.e., front-to-back confusions: Perception of a forward target in the rear hemisphere; back-to-front confusions: Perception of a rear target in the forward hemisphere). Analogously, confusions of perceived elevation are illustrated on the right (i.e., up-to-down or down-to-up confusions). Reprinted with permission from ACM Press (Wenzel *et al.*, 1991).

to-eardrum, were measured at many source positions and incorporated in digital filters which were then used to synthesize stimuli. These transfer functions, generally called head-related transfer functions (HRTFs), were measured using the technique described by Wightman and Kistler (1989a). The advantage of this technique is that it preserves the complex pattern of interaural differences over the entire spectrum of the stimulus, thus capturing the effects of filtering by the pinnae, head, shoulders, and torso.

The utility of a spatial auditory display will be compromised if the operator misjudges the apparent direction of a sound image. Judgments of the position of a sound image in the free-field are corrupted by two different kinds of error. One, which Blauert (1983) refers to as localization blur, is a small random error on the order of about 5° to 20°. Another kind of error observed in nearly all localization studies is represented by front-back "confusions." These are judgments which indicate that a source in the front hemisphere, usually near the median plane, is perceived by the listener to be in the rear hemisphere (Fig. 1). Occasionally, back-to-front confusions are also found (e.g., Oldfield and Parker, 1984a, b, 1986).

Front-back confusions probably result from ambiguities caused by the roughly spherical shape of the head and the primary role of ITDs (interaural time differences) and IIDs (interaural intensity differences) as localization cues (the so-called cone of confusion; Mills, 1972). Assuming a stationary, spherical head and symmetrically located ear canals (without pinnae), a given interaural time or intensity difference signals a range of potential sound source locations with the locus of all possible sources describing a cone (Fig. 2). Intersection of these conical surfaces with the surface of a sphere (i.e., considering sources at an arbitrary fixed distance) produces circles of constant ITD or IID that are smaller in diameter for larger values of ITD or IID. Note that the circular shapes of these iso-ITD and iso-IID contours would predict both front-back and up-down confusions. Here, we have observed confusions in elevation, with up locations heard as down, and vice versa (Fig. 1). While front-back confusions are frequently re-

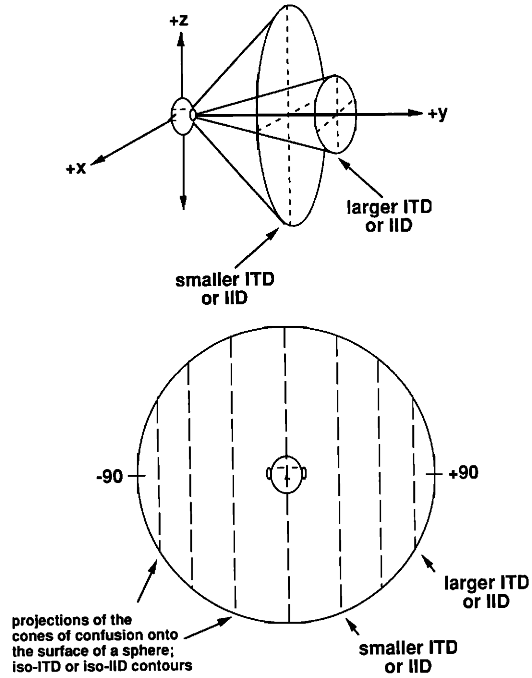

FIG. 2. Illustration of the cone-of-confusion effect for different interaural delays and intensities. Assuming a spherical head and symmetrically located ear canals (without pinnae), all sound sources lying along a conical surface would produce the same interaural time difference (ITD) and interaural intensity difference (IID). Intersection of these conical surfaces with the surface of a sphere results in circular projections corresponding to contours of constant ITD or IID. Such projections, shown in two dimensions in the drawing, are known as iso-ITD or iso-IID contours. Reprinted with permission from MIT Press (Wenzel, 1992).

ported, there has been no prior analysis of the rate of up-down confusions (although Kistler and Wightman, 1992, mention their existence).

Recent studies that have measured ITDs and IIDs as a function of signal frequency and position show that the situation is considerably more complicated than that portrayed in Fig. 2. However, to a first approximation, the rigid sphere model does seem to predict iso-ITD and iso-IID contours for static sources (Kuhn, 1977; Middlebrooks and Green, 1990; Middlebrooks *et al.*, 1989). Thus, while the model is not completely valid, observed patterns of iso-ITD and iso-IID contours suggest that the interaural characteristics of sounds in three-dimensional space are inherently ambiguous. In the absence of other cues, and given the primary importance of ITDs and IIDs (Wightman and Kistler, 1992; Kistler and Wightman, 1992), both front-back and up-down confusions would appear to be quite likely.

One cue thought to help in disambiguating the cones of confusion is the complex spectral shaping provided by the HRTFs. The notion is that the shape of the stimulus spectrum at one ear, not just the interaural differences, provides information about position. For example, presumably because of the orientation and shell-like structure of the pinnae, high frequencies tend to be more attenuated for sources in the rear than for sources in the front (Blauert, 1983, pp. 107–116). In addition, detailed features of these monaural cues, such as peaks and notches in the spectrum, are thought to be important for the perception of source elevation as well as front-back disambiguation (Roffler and Butler, 1968a, b; Bloom, 1977; Watkins, 1978; Humanski and Butler, 1988; Middlebrooks, 1992). Any degradation of these complex spectral cues might be expected to increase the rates of front-back and up-down confusions and decrease the ability of listeners to discriminate direction in the vertical plane.

Synthesis using individualized HRTFs would be the most likely to replicate the free-field experience for a given listener and the least likely to degrade the important spectral cues. Wightman and Kistler (1989b) compared subjects' localization judgments of static sources in the free-field with their judgments of virtual (headphone-presented) sources synthesized from the subjects' own HRTFs. They reported that, for the eight experienced listeners in their study, localization accuracy for the free-field and headphone stimuli was generally comparable. 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 spatial 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 nonindividualized 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. Wenzel *et al.*, (1988b) compared localization of real sources with stimuli synthesized from individualized and nonindividualized transfer functions. For real sources and for stimuli based on their own HRTFs, two of the subjects, "good localizers," produced accurate estimates of both azimuth and elevation, while the third, a "poor localizer," showed little ability to determine elevation in either case. Localization of stimuli based on nonindividualized HRTFs was only slightly degraded for a good localizer, as long as the HRTFs were derived from another good localizer. Large errors in judging source elevation were made by a good localizer listening to stimuli synthesized from the poor localizer's HRTFs (i.e., "listening through the ears" of the poor localizer). However, the converse was not true. After repeated trials (both with and without feedback), the poor localizer was unable to improve elevation accuracy by "listening through" a good localizer's HRTFs. These 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 and Kistler, 1989b, pp. 876–877).

The preliminary results of Wenzel *et al.* (1988b) suggest that it may be feasible to use nonindividualized transfer functions to synthesize spatial auditory display cues, as long as the HRTFs that are used come from a subject whose measurements are correlated with accurate localization performance in both free-field and headphone conditions. The present study represents a more comprehensive comparison of free-field and virtual free-field localization in two dimensions by inexperienced listeners using nonindividualized 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 and Kistler (1989b). Sixteen inexperienced subjects were tested in an attempt to assess whether the general population of listeners could readily use a virtual acoustic display.

I. METHOD

A. Subjects

Sixteen young adults (2 male, 14 female) 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.

B. 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 noise was bandpassed with a 10th-order finite impulse response (FIR) filter having cutoff frequencies at 200 Hz and 14 kHz. Further, the noise spectrum was scrambled on each experimental trial in order to minimize the contributions of stimulus familiarity and transducer idiosyncracies to the judgments 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.

The stimuli were transduced either by loudspeakers (Realistic Minimus 7) or by headphones (Sennheiser HD-430). Details of the presentation system can be found in Wightman and Kistler (1989a, b). In the free-field case, six speakers were mounted on a semicircular arc, 2.76 m in diameter, at intervals of 18° elevation (range -36 to $+54°$, 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. The subject was 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 studies by Wightman and Kistler. Details of the measurement technique and the synthesis procedure can be found in Wightman and Kistler (1989a,b). Briefly, the spectrally shaped bursts of Gaussian noise were processed independently for the left- and right-ear stimuli by cascading two FIR filter sections. The first filter section consisted of SDO's left (or right) ear HRTF for a given source position divided by SDO's headphone-to-ear-canal transfer function for the same ear. The inverse transfer function was required to remove the spectral characteristics imposed by the headphones alone when worn by SDO (see Wightman and Kistler, 1989a). The second filter section was a zero-phase bandpass filter (200 Hz to 14 kHz) used to remove processing artifacts at low and high frequencies.

Stimuli for a given subject and a given set of trials were pre-computed on a DEC Micro Vax 3500 and stored on the hard disk of an IBM PC/AT. The signals were then converted to analog form via PC-controlled, 16-bit D/A converters (Ariel DSP-16) at a rate of 50 kHz. Antialiasing filters were not used since the nearest aliased components, 36 kHz, were well beyond the audible range. The overall level of the stimuli was approximately 70 dB SPL in both the free-field and headphone conditions. Testing in both conditions was conducted in the anechoic chamber.

C. Procedure

The goal of this experiment was to compare inexperienced subjects' judgments of the apparent locations of sound sources in free-field with their judgments of locations in a virtual free-field. The virtual sources were synthesized from a "generic" set of HRTFs in the sense that they were derived from a single individual, SDO, whose localization accuracy was representative of seven of the eight subjects in the study by Wightman and Kistler (1989b).[1] The data from this subject showed good localization accuracy in both azimuth and elevation and an average rate of front-back confusions in both free-field and headphone conditions. The aim was to generate the waveforms produced by free-field stimuli at the eardrums of subject SDO and reproduce them at the eardrums of the inexperienced listeners. To the extent that each subject's headphone-to-eardrum transfer function differs from SDO's, a less faithful reproduction would result.

We wished to compare the data from this study with those from the previous experiment (Wightman and Kistler, 1989b) in which free-field performance was contrasted with headphone performance of a different group of subjects using their own HRTFs. Thus, the same absolute judgment paradigm that was used to estimate perceived spatial location in the earlier experiment was also used here. In both free-field and headphone conditions, the subject indicated the apparent spatial position of a sound source by calling out numerical estimates of azimuth and elevation (in degrees) using a modified spherical coordinate 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." The left/right and up/down distinctions were used to avoid the response confusions that sometimes occurred with the purely numerical coordinates used in the prior experiment.

TABLE I. Source positions of the stimuli used in the free-field and virtual free-field conditions. L and R refer to positions on the left and right of the vertical median plane. U and D refer to locations above and below 0° elevation, or the horizontal plane at ear level. Positions are organized according to the six regions of auditory space used in the data analysis of Figs. 6 and 7.

	Azimuth	Elevation	Azimuth	Elevation
Front-low	L45	D36	R45	D18
	L15	D18	0	0
Front-high	L15	U18	R30	U36
	L45	U18	R15	U54
	0	U36		
	L45	U54		
Sides-low	L75	0	R90	D36
	L90	0	R105	D18
Sides-high	L105	U36	R120	U36
	L105	U54		
Back-low	L150	D36	180	D36
	L135	D18	R165	0
Back-high			R135	U18
			R150	U18
			180	U54

With this paradigm, a subject's skill in using the numerical coordinate system is confounded with his/her ability to localize an acoustic source. However, since the goal of the experiment was to compare performance in the free-field and headphone conditions, we felt that this was not a serious problem. Further, subjects appear to learn the task easily and quickly produce stable judgments. Subjects were given no feedback regarding the accuracy of their judgments.

Subjects were presented stimuli from 24 different source positions in the free-field (anechoic chamber) and the same 24 positions in the virtual free-field (headphone) conditions (see Table I). The source locations were chosen from the 144 positions tested by Wightman and Kistler (1989a) 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-min block. A total of 18 blocks of trials were run over a period of three days, with subjects giving 9 judgments of apparent spatial location for each of the 24 locations in both free-field and virtual free-field conditions. Blocks of trials for the free-field and headphone conditions were alternated to minimize order effects, with a 5-min break between each pair of blocks. Prior to the experimental runs, a 15-min training session was conducted which included a verbal explanation of the response coordinates and a practice block in the free-field condition. Relatively few positions were tested (24, compared to 72 in the study by Wightman and Kistler, 1989b) and minimal

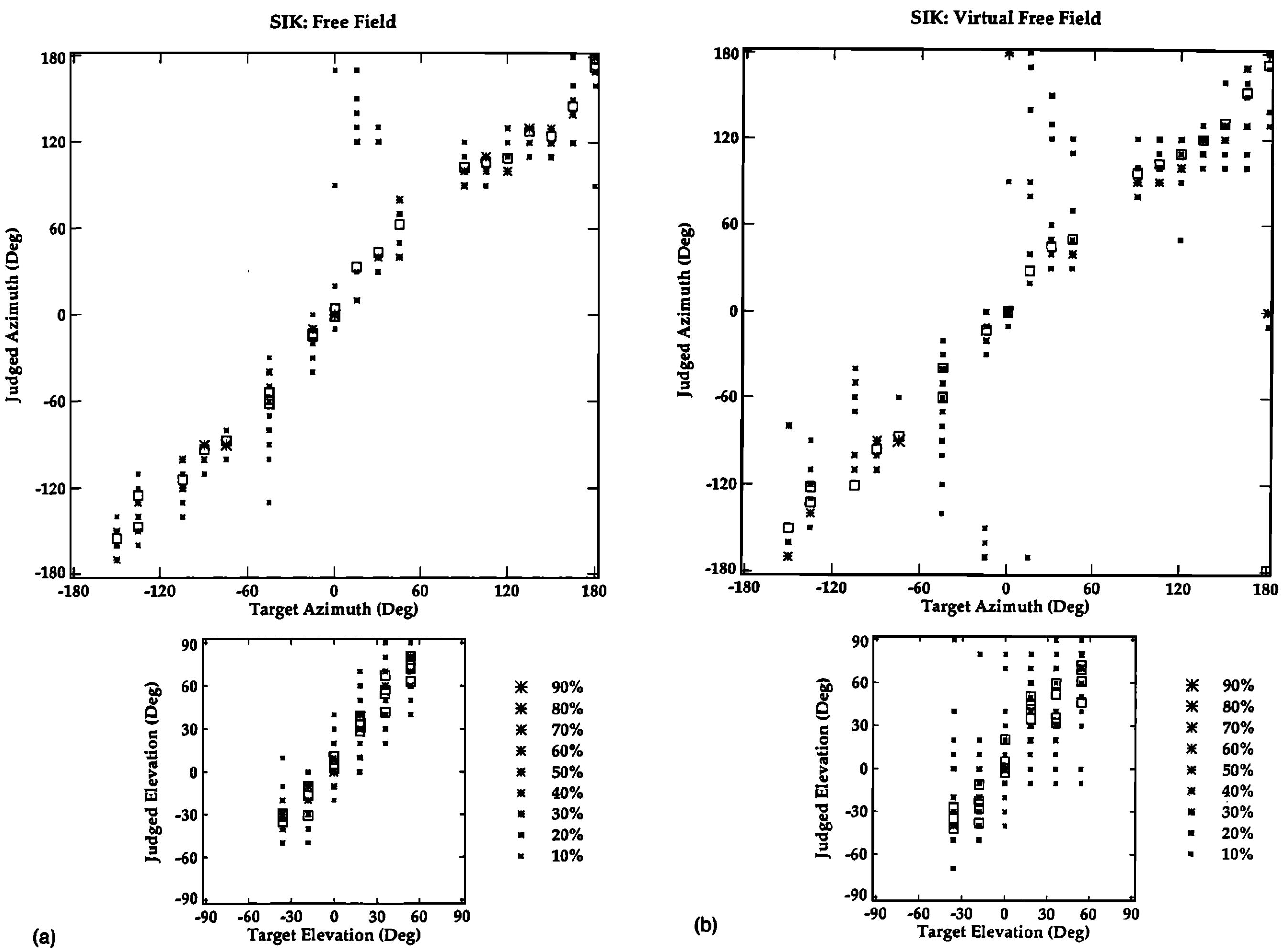

FIG. 3. Scatterplots of actual source azimuth (top panels) and actual source elevation (bottom panels) versus judged source azimuth and elevation for subject SIK in both the free-field and headphone conditions. The panels in (a) plot free-field judgments and the panels in (b) show judgments for the stimuli synthesized from nonindividualized transfer functions. Each open symbol represents the centroid of 9 judgments. The asterisks represent the subject's individual location judgments for a particular target location. The size of each asterisk indicates the number of judgments within a 10°-wide interval at that location. For example, the smallest symbol shown in the legend illustrates the case when 10% of the judgments occur within 10° of a particular response location. All 24 source positions are plotted in each panel. Thus, data from 6 different source elevations are combined in the azimuth panels and data from 18 different source azimuths are combined in the elevation insets. Note that the scale is the same in the azimuth and elevation plots. The centroid data are reprinted with permission from ACM Press (Wenzel *et al.*, 1991) and MIT Press (Wenzel, 1992).

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. During free-field presentation, the experimenter remained in the chamber in order to move the loudspeaker arc and record responses and thus was able to verify head position and stability. Each trial began with a 15-s burst of white Gaussian noise from a speaker (not used for localization in the study) mounted at floor level in front of the subject. This noise was used to mask the sounds made by repositioning the arc between trials. The subjects then heard the train of eight identical 250-ms bursts of random-spectrum noise and called out estimates of azimuth and elevation during a 5-s response interval. The experimenter recorded responses on a data sheet. A new trial began with the experimenter repositioning the arc for the next location listed on the data sheet and then moving to a corner of the chamber so as to be acoustically unobtrusive. The arc was moved for about the same amount of time between each trial, regardless of the required azimuth.

Following a free-field block, the subjects remained in the anechoic chamber but donned the headphones in preparation for a trial block with virtual stimuli. The trial sequence was the same as for the free-field condition, except that no masking noise was used between trials. The experimenter, who was now outside the chamber listening over an intercom, entered subjects' responses on a PC keyboard as they called out their estimates for each trial.

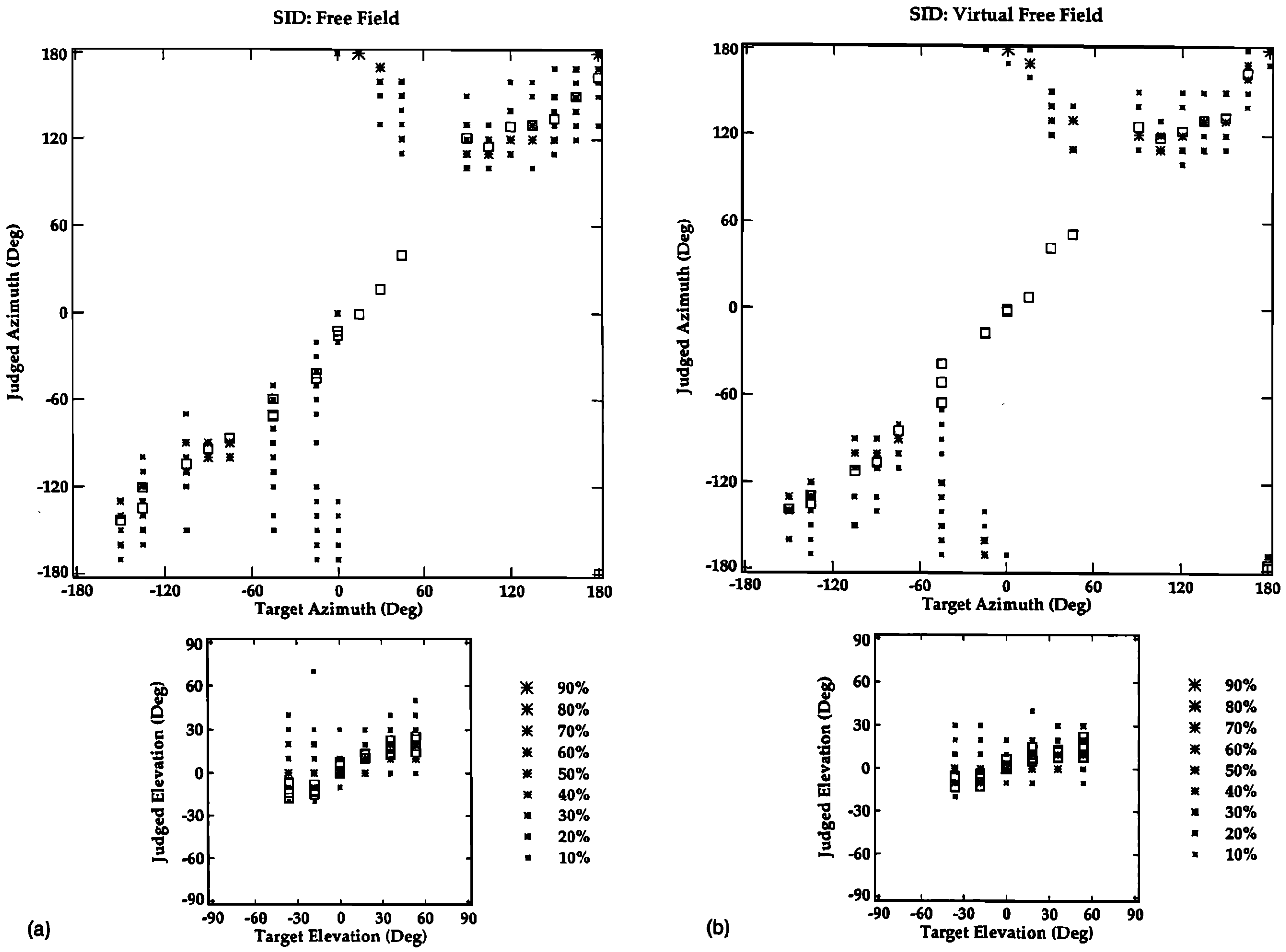

FIG. 4. Same as Fig. 3, but for subject SID.

II. RESULTS

A. Localization error and data analysis

Since both the stimuli and responses are represented by points on the surface of a sphere (distance remained constant in this experiment), spherical statistics were used to characterize the psychophysical data (Fisher *et al.*, 1987; see also Wightman and Kistler, 1989b, for a discussion of spherical statistics applied to localization data). The spherical statistics computed here, average angle of error, judgment centroid, and inverse kappa (K^{-1}), also allow comparison to the data of Wightman and Kistler (1989b) which was a primary goal of the present study.

The average angle of error is the mean of the unsigned angles between each judgment vector and the vector from the origin to the actual (or synthesized) target location. The judgment centroid is a unit-length vector with the same direction as the resultant, the vector sum of all the unit-length judgment vectors. The direction of the centroid can be thought of as the "average direction" of a set of judgments from the origin, the center of the subject's head. K is estimated from the length of the resultant vector and is proportional to the dispersion of the judgments around the centroid. Generally, the parameter K^{-1} is reported since the inverse value varies with dispersion in the same manner as a variance estimate. For the range of values observed here, a K^{-1} of 0.01 corresponds to a 95% confidence angle of 4.9°, a K^{-1} of 0.10 corresponds to a confidence angle of 16.0°, and a K^{-1} of 0.19 corresponds to a confidence angle of 23.1° (number of trials=9, as used here) with respect to the centroid estimated for a particular target location. The confidence angle is conceptually analogous to the 95% confidence interval used in standard statistical techniques. It describes a conical section of a sphere, surrounding the centroid estimated for a particular target location, which includes the population centroid of the sampling distribution of judgment vectors with a probability of 0.95 (Fisher *et al.*, 1987, p. 131, Eq. 5.34). Thus, if K^{-1} is 0.10 for each of two neighboring centroids, the precison of the estimates is such that the centroids must be at least 32° apart to be distinguishable as the centroids of two different sample populations of judgment vectors.

As noted above, a feature of localization data observed in nearly all such studies is the presence of front-back confusions. In this experiment we also observed confusions in elevation, with up locations heard as down, and vice versa. 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 (i.e., the responses are coded

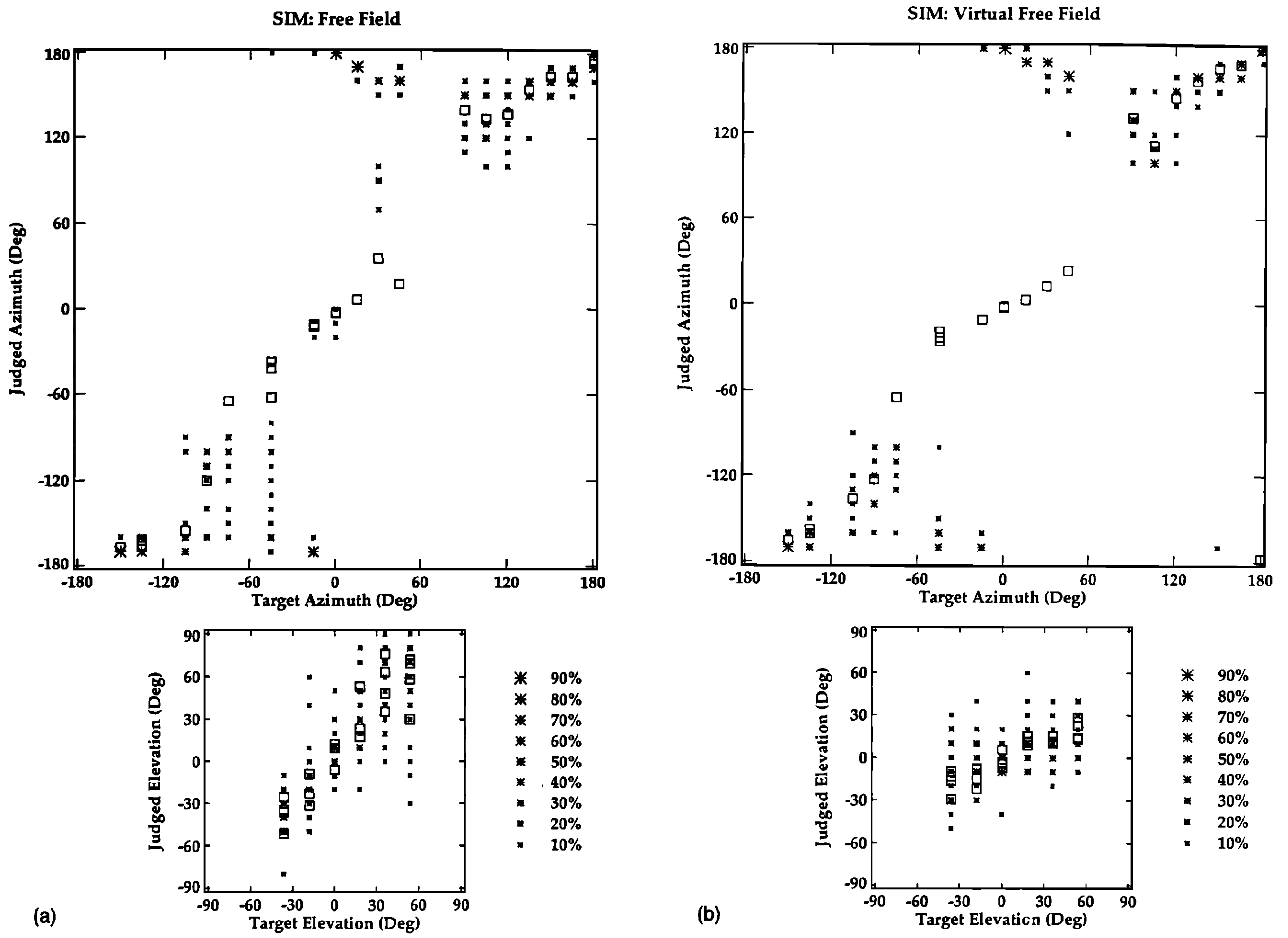

FIG. 5. Same as Fig. 3, but for subject SIM.

as if they had indicated the correct hemisphere; Wightman and Kistler, 1989b) or simply not included (Makous and Middlebrooks, 1990) when computing descriptive statistics. Otherwise, estimates of error would be greatly inflated. 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 their rates in each condition.

The algorithm for resolving confusions used here treats each judgment identically. If the angle between the target and judged location is made smaller by reflecting the judgment 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), then the judgment is coded in reflected form and the confusion count is increased. This procedure probably results in overestimates of confusion rates and underestimates of error for positions near 90° in azimuth and 0° in elevation. Such biases should not affect the comparison of the free-field and headphone conditions.

Judgment centroids, average angles of error, K^{-1}, and confusion rates were computed separately for each of the 24 positions in both the free-field and headphone conditions, for each of the 16 subjects. Figures 3–5 show the relation between target and response azimuth (for all stimulus elevations) and between target and response elevation (for all stimulus azimuths) for three representative subjects in both the free-field and virtual source conditions. The centroid data alone have been previously reported for these three subjects (Wenzel *et al.*, 1991; Wenzel, 1992). However, given the high rates of confusions observed for some of our subjects and the fact that centroids are necessarily based on data which have had confusions resolved, we felt that plotting only centroids can be misleading without providing the context of the overall pattern of unresolved location judgments. Thus here, centroids (open squares) are shown superimposed over the asterisk symbols which represent all of the subjects' individual location judgments for each target location. As illustrated in the legend, the size of the asterisk represents the number of judgments within a 10°-wide interval at that location. For example, the smallest symbol shown in the legend illustrates the case when 10% of the judgments occur within 10° of a particular response location. Perfect correlations between target positions and response judgments corre-

TABLE II. Goodness of fit estimates computed for both azimuth and elevation (global), azimuth only, and elevation only judgments. Free-field data (FF) are in **boldface type** and data for the headphone stimuli (HP) are in plain text.

	Goodness-of-fit					
	Global		Azimuth		Elevation	
Subject	FF	HP	FF	HP	FF	HP
SEY	**0.760**	0.521	**0.716**	0.529	**0.853**	0.511
SHT	**0.814**	0.677	**0.803**	0.633	**0.865**	0.760
SHU	**0.868**	0.604	**0.892**	0.594	**0.892**	0.655
SHW	**0.804**	0.674	**0.772**	0.687	**0.865**	0.500
SHY	**0.856**	0.696	**0.850**	0.711	**0.913**	0.690
SHZ	**0.816**	0.623	**0.830**	0.665	**0.908**	0.576
SIB	**0.634**	0.591	**0.581**	0.572	**0.817**	0.669
SIC	**0.847**	0.676	**0.884**	0.765	**0.841**	0.334
SIF	**0.872**	0.581	**0.881**	0.638	**0.876**	0.395
SIG	**0.741**	0.627	**0.728**	0.597	**0.721**	0.696
SIH	**0.891**	0.760	**0.886**	0.789	**0.934**	0.541
SIK	**0.912**	0.675	**0.910**	0.642	**0.935**	0.765
SID	**0.596**	0.570	**0.646**	0.604	**0.400**	0.493
SIE	**0.559**	0.588	**0.533**	0.554	**0.687**	0.648
SII	**0.859**	0.632	**0.876**	0.673	**0.802**	0.475
SIM	**0.699**	0.558	**0.652**	0.582	**0.850**	0.374

spond to a diagonal slope of $+1.0$ on these graphs. On the other hand, two short negative diagonals (slopes of -1.0) running from target-response coordinates of -180,0 to 0, -180 and 0,180 to 180,0 (-90,0 to 0,-90 and 0,90 to 90,0 for elevation) correspond to regions where the different types of confusions would fall. To facilitate plotting, target and response positions in the left and down hemispheres are shown as negative numbers.

As a supplement to the representative data of Figures 3–5, Table II provides quantitative estimates of localization performance for each of the 16 subjects. Goodness of fit (GOF) estimates (see Wightman and Kistler, 1989b, p. 872) were computed for the unresolved judgments versus the target locations for both azimuth and elevation (global), azimuth only, and elevation only judgments.

Figures 6 and 7 show regional averages for each subject for the average angle of error, K^{-1}, and confusion rates (based on the number of possible azimuth or elevation confusions in each region). The six regions of auditory space (see Table I) were defined by partitioning azimuth into three groups, front (L45 to R45), side (L60 to L120, R60 to R120), and back (L135 to R135); and partitioning elevation into two groups, low (D36 to 0) and high (U18 to U54). Parameter values for the individual subjects are plotted as open squares while data averaged across all subjects are indicated by solid squares. For purposes of comparison, regional means from the study by Wightman and Kistler (1989b) are indicated for their nominally equivalent regions (open circles); i.e., their categories including middle elevations (0°, -18°, $+18$°) are not plotted in Figs. 6 and 7.

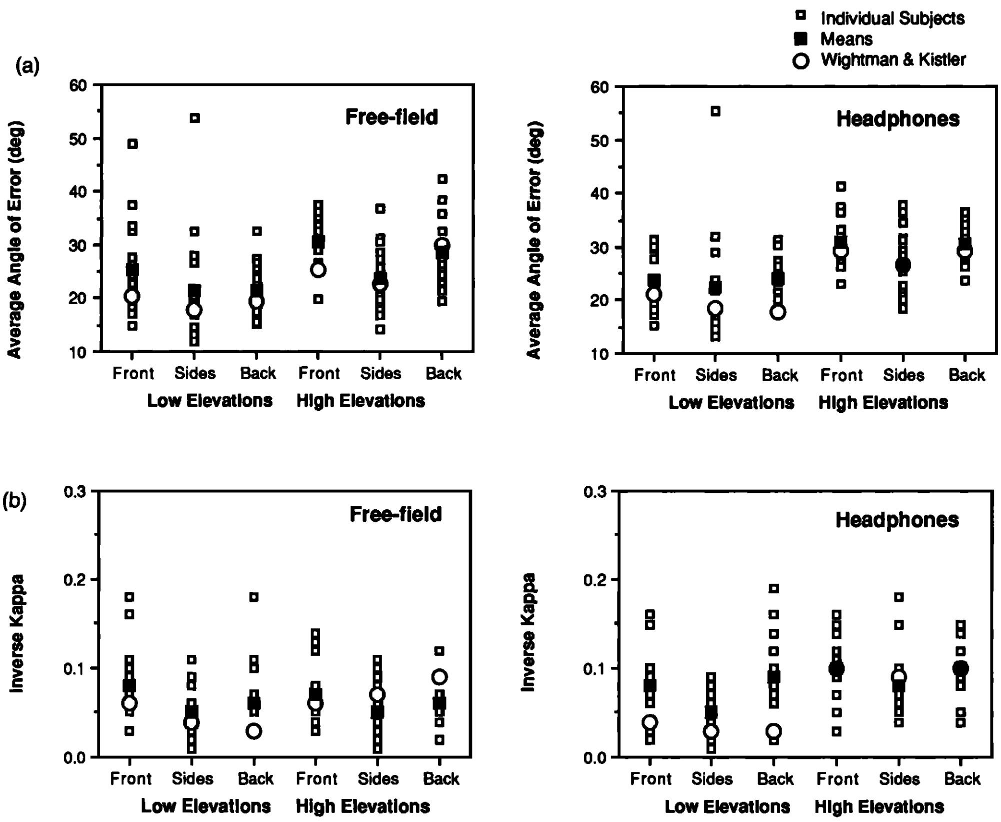

FIG. 6. Regional averages for the average angle of error (a) and K^{-1} (b) are shown for the six regions of auditory space outlined in Table I. Parameter values for the 16 individual subjects are plotted as open squares (with identical means superimposed) while data averaged across all subjects are indicated by solid squares. For purposes of comparison, regional means from the study by Wightman and Kistler (1989b) are indicated for their nominally equivalent regions (open circles).

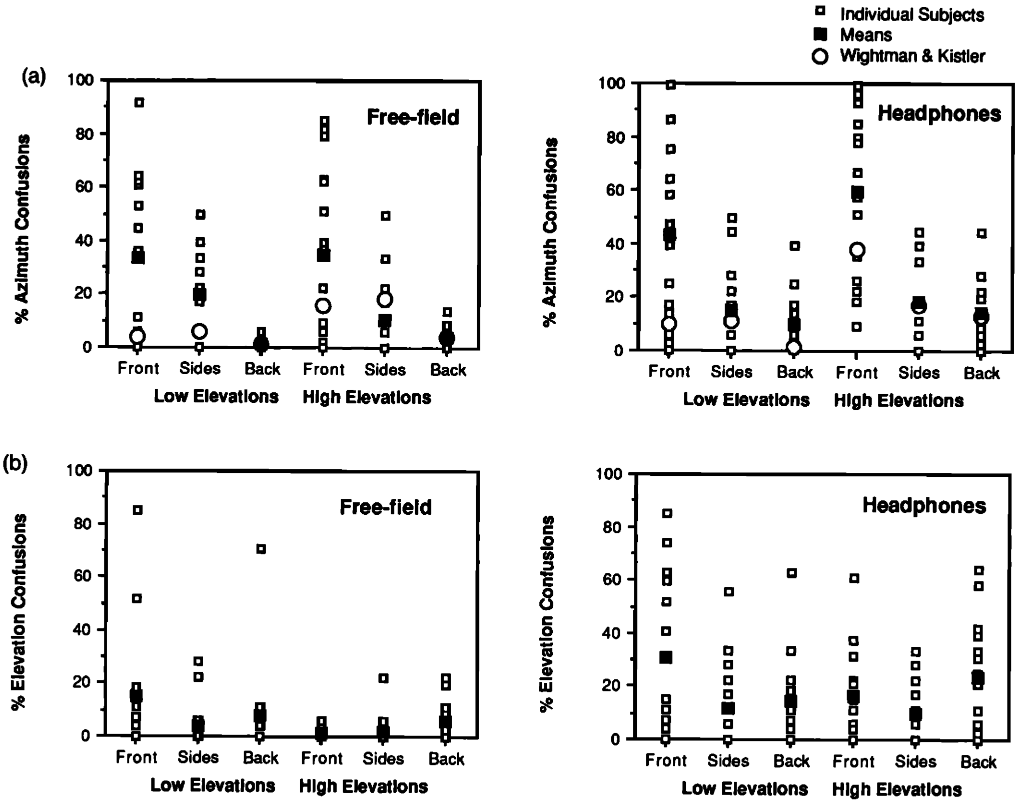

FIG. 7. Same as Fig. 6, except that the percentages of front-back (a) and up-down (b) confusions are plotted. Confusion rates are based on the number of possible front-back or up-down confusions within each region. Wightman and Kistler (1989b) did not analyze confusions in elevation so these means are missing in (b).

Table III shows the total percentage of front-back confusions (collapsed across all elevations) and up-down confusions (collapsed across all azimuths) for the free-field and headphone conditions, for each subject. Chi-square values, quantifying the difference in confusion rates that occurred between the the free-field and headphone conditions, are also given. Table IV gives a breakdown of the confusion percentages by hemisphere.

B. Individual differences

One noteworthy feature of the data reported here is the presence of pronounced individual differences. The centroid data in the top panels of Figs. 3–5 indicate that all 16 subjects produced accurate judgments of azimuth for both free-field and virtual sources. However, three very different response patterns were observed for elevation centroids. Examples of the first can be seen in the bottom panels of Fig. 3(a)–(b). Similar to this representative subject, 12 of the 16 subjects showed good elevation accuracy for resolved judgments in both free-field and virtual-source conditions. On the other hand, two of the 16 subjects (SID and SIE) showed poor elevation performance in both free-field and headphone conditions [Fig. 4(a)–(b)], while the remaining two subjects (SII and SIM) exhibited good elevation performance in the free-field but poor elevation accuracy in virtual free-field conditions [Fig. 5(a)–(b)].

As might be expected, the regional data of Fig. 6(a)–(b) suggest that the variability of judgments from our inexperienced subjects is higher than that from the more experienced listeners of Wightman and Kistler. Average error angles are often larger and dispersions are generally greater, particularly for targets at low elevations. However, the overall pattern of regional variations is comparable to Wightman and Kistler (1989b) and qualitatively similar to the observations of Oldfield and Parker (1984a). That is, localization acuity tends to be worst for rear locations at high elevations.

It is apparent that misleading interpretations of the data could result when centroids based on resolved data are used to summarize judgments. For example, while the centroid data are quite monotonic and well behaved in Figs. 3–5, the individual judgments for the three subjects' azimuth data tend to show concentrations of responses near +/−180° for targets near 0° which reflect the presence of front-back confusions. Some degree of response clustering reflecting up-down confusions also appears in the elevation judgments, although the overall pattern of the data suggests either increased variability or gross compression of elevation judgments rather than obvious mirror-image confusions. Similarly, the goodness of fit statistics for all 16 subjects shown in Table II reflect the results typified by Figs. 3–5. In general, free-field GOFs are superior to the virtual free-field data with the lowest values corresponding to the subjects and stimulus conditions that also had the highest front-back and/or up-down confusion rates. It is still useful, however, to assess rates of front-back

TABLE III. Percentages of front-back and up-down confusions for free-field and headphone listening are shown based on the total number of possible confusions. Chi-square statistics are also indicated for those subjects which showed a significant difference in front-back and/or up-down confusions for the free-field versus headphone conditions. (* $\alpha = < 0.5$; ** $\alpha < 0.01$; + indicates a continuity correction was applied to the chi-square.) Free-field data are in **boldface type** (FF) and data for the headphone stimuli are in plain text (HP). Adapted with permission from ACM Press (Wenzel *et al.*, 1991).

	% Front-back confusions			% Up-down confusions		
	FF	HP	X^2	FF	HP	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	**35**	40	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	5.89*
SIM	**43**	50	NS	**4**	32	44.8**
Mean	**19**	31	7.7**	**6**	18	14.0**

TABLE IV. Breakdown of confusion percentages by hemisphere for the free-field and headphone conditions. Percentages are based on the total number of possible confusions. Thus, a value of 50% indicates all judgments are reversed in that category. Free-field data (**FF**) are in **boldface type** and data for the headphone stimuli (HP) are in plain text. Overall means and unbiased estimates of the standard deviations are also shown. Adapted with permission from ACM Press (Wenzel *et al.*, 1991).

	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	**0**	4	**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	**2.2**	5.0	**6.4**	6.3	**2.1**	8.8

and up-down confusions, particularly with a view toward comparison with previous studies in the literature.

C. Front-back confusions

Examination of the unresolved location judgments (Figs. 3–5) and the regional analysis of azimuth confusions [Fig. 7(a); Table IV] indicates that the front-back confusion rates consist largely of apparent reversals of front positions to the rear, a phenomenon that has been frequently reported in studies of localization in the median plane (see Blauert, 1983) and appears in the regional analysis of Wightman and Kistler (1989b; p. 873). Some of our inexperienced listeners made an unusually large number of front-back confusions in both free-field and headphone conditions. Half of the subjects produced free-field, front-back confusion rates (mean: 6.5%, range: 2% to 10%) similar to Wightman and Kistler's experienced subjects. These rates approximately quadrupled with virtual sources synthesized from nonindividualized HRTFs compared to the doubling of rates seen in Wightman and Kistler (1989b) for stimuli synthesized from the subjects' own HRTFs. This quadrupling was a significant increase for each of the eight affected subjects (Table III). The other half of the subjects had much higher free-field confusion rates (mean: 32%, range: 20% to 43%) and only three of these produced a significant increase with virtual sources (Table III). The high confusion rates for these subjects may be due primarily to inexperience with the task and perceptual ambiguities of the anechoic stimuli, rather than to the simulation procedure *per se*.

A few other recent studies have also reported front-back confusion rates that appear to be affected by experimental conditions that, like stimuli based on nonindividualized HRTFs, could be interpreted as producing varying degrees of disruption of the spectral pinna cues. For example, the extensive free-field localization studies conducted by Oldfield and Parker (1984a, b; 1986) report average rates of 3.4%, 12.5%, and 26% for normal free-field, monaural, and pinnae-occluded localization conditions, respectively. Makous and Middlebrooks (1990) report low rates (average 6%; range 2% to 10%) for free-field localization of stimuli bandpassed at higher frequencies (1.8–16 kHz) where pinna cues are known to be most effective. Conversely, Begault and Wenzel (in press) observed high front-back confusion rates (average 29%) for virtual sources using speech stimuli with relatively little energy beyond 5 kHz. Recently, Asano *et al.* (1990) examined localization of virtual sources in the median plane for two highly practiced subjects listening to sources synthesized from either their own or the other subject's transfer functions. Although one subject showed low confusion rates (1.9% and 0.9%) for both synthesis conditions, the other showed rates of 7.4% and 14.8% for stimuli synthesized from his own and the other's HRTFs, respectively.

Another factor that may have disrupted spectral cues and resulted in the high confusion rates observed here was the use of spectrally scrambled stimuli. The studies described above that demonstrated low confusion rates all used signals with relatively stationary spectra such as white (Oldfield and Parker, 1984a; Asano *et al.*, 1990) or even frozen noise (Makous and Middlebrooks, 1990). On the other hand Wightman and Kistler (1989a), who also used spectrally scrambled stimuli, reported relatively low confusion rates.

D. Up-down confusions

To our knowledge, no previous studies have analyzed confusions in elevation. The rates observed here suggest that this type of confusion is less common; 14 of the 16 subjects showed free-field rates which were less than 10%. As with front-back confusions, up-down confusions increased dramatically with virtual sources; compared to free-field conditions, rates increased significantly for all but one subject's elevation judgments (Table III), with an average increase of about a factor of 7. Unlike front-back confusions, the regional breakdown of up-down confusions [Fig. 7(b); Table IV] does not indicate an overall response bias toward either the upper or lower hemispheres. As noted above, it is possible that the rate of up-down confusions was overestimated due to relatively large errors in acuity near 0° that were classed as reversals (e.g., error angles greater than 18° for targets at +/−18°). This notion is supported by the fact that nearly half of the up-down confusions occurred at targets of +/−18° (44% vs 49% for free-field and headphone conditions, respectively).

Some of the confusion errors reported as front-back and up-down confusions in Fig. 7 and Tables III and IV were actually the result of a combination of confusion errors. That is, occasionally subjects heard a target in the lower-front quadrant as a source in the upper-rear quadrant (most common), a target in the upper-front quadrant as a source in the lower-rear quadrant, and so on. Overall, 14% and 29% of front-back confusions (free-field and headphone conditions, respectively) and about half (55%) of the up-down confusions were the result of combined confusion errors. Although Makous and Middlebrooks (1990) did not specifically analyze up-down confusions, they did note that elevation errors were largest for trials in which there was a front-back confusion. The observation of confusions corresponding to multiple positions along iso-ITD or iso-IID contours underscores the important role that cone-of-confusion effects must play in the localization of these static, anechoic stimuli.

III. DISCUSSION

The centroid data indicate that the attempt to simulate free-field localization for inexperienced subjects with nonindividualized transfer functions was largely successful. That is, with confusions *resolved*, there were high correlations between the judged and actual source locations and a close correspondence between free-field and virtual-source conditions. When performance degradations did occur with virtual stimuli they were in the dimension of elevation. In fact, although judgment of elevation was poor for four of the subjects, only two of these showed a discrepancy between free-field and headphone conditions. Azimuth perception, on the other hand, appeared to be accurate in both free-field and virtual-source conditions. As in Wightman and Kistler (1989b), there were no clearcut differences in the variability of judgments between free-field and virtual-source conditions as indicated by average angles of error and K^{-1}. There is some suggestion, however, of an overall lower level of performance in this experiment that could be attributed to inexperience. Compared to the study by Wightman and Kistler, average error angles, K^{-1}'s, and confusion rates all tended to be higher for both free-field and virtual source conditions.

Such data suggest that many listeners can obtain useful directional information from an auditory display, particularly for the horizontal dimension, without requiring the use of individually tailored HRTFs. However, the high rates of confusion errors remain a problem. Comparison of these results to the study of localization of sources synthesized from a subject's own HRTFs (Wightman and Kistler, 1989b), suggests that the use of nonindividualized HRTFs results primarily in an increase in the rate of front-back confusions. Note that the existence of free-field reversals indicates that these confusions are not strictly the result of the simulation. Rather, as discussed earlier, they are probably caused by the inherent ambiguities in the stimuli revealed by measurements of iso-ITD and iso-IID contours (Kuhn, 1977; Middlebrooks *et al.*, 1989; Middlebrooks and Green, 1990).

It is possible, as Asano *et al.* (1990) have claimed, that confusions tend to diminish as subjects gain experience with the impoverished stimulus conditions provided by static anechoic sources, whether real or simulated. That is, even without feedback, subjects may learn to make increasingly fine discriminations of location-dependent spectral differences that eventually allow them to reliably resolve the ambiguous interaural cues. The higher overall confusion rates of our inexperienced listeners compared to the more experienced subjects of Wightman and Kistler tend to support this view. It may be that some form of adaptation and/or task-dependent training will usually be required to take full advantage of a virtual acoustic display using static sources. Cone-of-confusion effects alone cannot explain a front-to-back response bias, however, 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 cannot be seen.

In general, the results suggest that interaural difference cues are readily synthesized under headphones while the spectral details responsible for determining elevation and disambiguating front-back locations tend to suffer during a synthesis process that uses nonindividualized transforms. From the listener's perspective, both the relatively small errors introduced by the synthesis process *per se* and the larger discrepancies resulting from the use of nonindividualized pinna cues may simply exacerbate cone-of-confusion effects by adding "jitter" to the rather subtle spectral features of the HRTFs, particularly with respect to the cues that subjects are used to hearing. Such effects could act as a noise process which effectively reduces the relative size or shifts the relative location of spectral peaks and valleys, resulting in increased reversal rates and reduced elevation accuracy because of a mismatch between detected and expected spectral patterns. In a related experiment, Asano *et al.* (1990) examined performance for virtual sources synthesized from simplified HRTFs (i.e., transfer functions modeled by different orders of filters derived from an auto-regressive moving-average model). The rate of front-back confusions increased substantially (from about 1%–15% to about 18%–49%) as the modeled HRTFs were simplified, with the majority reversing to the rear for the lowest-order filters. Such data are consistent with the present study if one assumes that, in terms of the presentation of unfamiliar spectral cues, simplification of HRTFs is akin to listening through nonindividualized transforms.

In a practical display, it may be possible to minimize confusion errors by increasing the complexity of the simulation. The addition of visual cues, dynamic cues correlated with head-motion, and more complex environmental cues derived from models of room acoustics may improve the ability to resolve these ambiguities. For example, some studies have shown that allowing or inducing head motion improves localization ability by substantially reducing the rate of reversals (Burger, 1958; Thurlow and Runge, 1967; Fisher and Freedman, 1968). With head motion, a listener could potentially disambiguate front-back locations by tracking changes in the magnitude of the interaural cues over time; for a given lateral head movement, ITDs and IIDs for sources in the front will change in the opposite direction compared to sources in the rear (Wallach, 1939; 1940).

The use of familiar signals combined with cues that provide a sense of distance and environmental context, such as the characteristics specific to enclosed spaces, may also help to reduce front-back confusions. Just as we come to learn the characteristics of a particular room or concert hall, the localization of virtual sounds may improve if the listener is allowed to become familiar with sources as they interact in a particular artificial acoustic environment. For example, simulation of an asymmetric room might aid the listener in distinguishing front from rear sources by strengthening spectral or timbral differences between the front and rear sections of the room.

The specific parameters used in such models must be investigated carefully if localization accuracy is to remain intact. von Békésy (1960) 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. In the context of a display environment, it may be possible to create an artificial acoustic world which has been optimized for both localization and externalization of sources based on a careful parametric evaluation of this trade-off. Allowing users to interact with and learn its characteristics in a dynamic, head-tracked display (Wenzel *et al.*, 1988a; Wenzel, 1992; Foster, Wenzel, and Taylor, 1991) might then further improve the simulation and mitigate many of the errors we have observed for static sources.

IV. CONCLUSIONS

Overall, the results suggest that the interaural difference cues are readily synthesized under headphones while the spectral details responsible for determining elevation and disambiguating front-back locations are distorted by a synthesis process that uses nonindividualized HRTFs. However, it would appear that many listeners are able to obtain at least some useful directional information from an auditory display without requiring the use of individually tailored HRTFs. For 14 of the 16 subjects, virtual sources synthesized from a "good" localizer's HRTFs are judged to have the same (resolved) spatial locations as sources presented in the free-field for the limited range of stimuli investigated here. Comparison of these results with those reported by Wightman and Kistler (1989b), in which free-field listening was compared with localization of sources synthesized from subjects' own HRTFs suggests that the use of nonindividualized transforms primarily results in an increase in the rate of front-to-back confusions. It is possible that both this effect and the slight overall degradation in performance (compared to Wightman and Kistler) are due to the subjects' inexperience with the task, and may be mitigated by further training.

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 real-time control systems (e.g., Wenzel *et al.*, 1988a; Wenzel, 1992) 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 that is critical for applied contexts.

ACKNOWLEDGMENTS

This work was supported by a Co-operative Research Agreement with NASA–Ames Research Center and by the Federal Aviation Administration. The authors gratefully acknowledge the invaluable assistance of Douglas Swiggum during the technical phases of the study. As indicated, portions of the data were presented at the CHI'91 ACM Conference on Computer–Human Interaction, New Orleans, LA and reported in the proceedings (Wenzel *et al.* 1991) and also in Wenzel (1992).

[1]As Middlebrooks *et al.* (1989) and Middlebrooks and Green (1992) have noted, HRTF characteristics vary systematically with the relative size of subjects, although the behavioral consequences of these differences are not clear. While physical sizes were not recorded here, 14 of the 16 subjects tested were women and thus, to the extent that women tend to be smaller than men, the subjects probably tended to be smaller than the general population. Similarly, SDO was a small, slight woman of about 5 ft in height. Thus, while there was no doubt some mismatch in HRTF characteristics, the discrepancy was probably not as large as if the subjects tested here had all been male. SDO's HRTFs were chosen primarily

on the basis of behavioral performance data; except for the criterion of normal hearing, the subjects of this study were chosen randomly.

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

Begault, D. R., and Wenzel, E. M. (**in press**). "Headphone localization of speech," Human Factors (to be published).

Bekesy, G. von (**1960**). *Experiments in Hearing* (McGraw-Hill, New York).

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

Bloom, P. J. (**1977**). "Creating source elevation illusions by spectral manipulation," J. Audio. Eng. Soc. **25**, 560–565.

Burger, J. F. (**1958**). "Front-back discrimintation of the hearing system," Acustica **8**, 301–302.

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

Calhoun, G. L., Valencia, G., and Furness, T. A., III (**1987**). "Three-dimensional auditory cue simulation for crew station design/evaluation," Proc. Hum. Fac. Soc. **31**, 1398–1402.

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

Fisher, H., and Freedman, S. J. (**1968**). "The role of the pinna in auditory localization," J. Aud. Res. **8**, 15–26.

Fisher, N. I., Lewis, T., and Embleton, B. J. J. (**1987**). *Statistical Analysis of Spherical Data* (Cambridge U.P., Cambridge, United Kingdom).

Foster, S. H., Wenzel, E. M., and Taylor, R. M. (**1991**). "Real-time synthesis of complex acoustic environments," ASSP (IEEE) Workshop on Applications of Signal Processing to Audio and Acoustics, New Paltz, NY.

Humanski, R. A., and Butler, R. A. (**1988**). "The contribution of the near and far ear toward localization of sound sources on the median plane," J. Acoust. Soc. Am. **83**, 2300–2310.

Kistler, D. K., and Wightman, F. L. (**1992**). "A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction," J. Acoust. Soc. Am. **91**, 1637–1647.

Kuhn, G. F. (**1977**). "Model for the interaural time differences in the azimuthal plane," J. Acoust. Soc. Am. **62**, 157–167.

Loomis, J. M., Hebert, C., and Cicinelli, J. G. (**1990**). "Active localization of virtual sounds," J. Acoust. Soc. Am. **88**, 1757–1764.

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

McKinley, R. L., and Ericson, M. A. (**1988**). "Digital synthesis of binaural auditory localization azimuth cues using headphones," J. Acoust. Soc. Am. **83**, S18.

Middlebrooks, J. C., Makous, J. C., and Green, D. M. (**1989**). "Directional sensitivity of sound-pressure levels in the human ear canal," J. Acoust. Soc. Am. **86**, 89–108.

Middlebrooks, J. C., and Green, D. M. (**1990**). "Directional dependence of interaural envelope delays," J. Acoust. Soc. Am. **87**, 2149–2162.

Middlebrooks, J. C. (**1992**). "Narrow-band sound localization related to external ear acoustics," J. Acoust. Soc. Am. **92**, 2607–2624.

Mills, W. (**1972**). "Auditory localization," in *Foundations of Modern Auditory Theory*, edited by J. V. Tobias (Academic, New York).

Oldfield, S. R., and Parker, S. P. A. (**1984a**). "Acuity of sound localisation: a topography of auditory space. I. Normal hearing conditions," Perception **13**, 601–617.

Oldfield, S. R., and Parker, S. P. A. (**1984b**). "Acuity of sound localisation: a topography of auditory space. II. Pinna cues absent," Perception **13**, 601–617.

Oldfield, S. R., and Parker, S. P. A. (**1986**). "Acuity of sound localisation: a topography of auditory space. III. Monaural hearing conditions," Perception **15**, 67–81.

Persterer, A. (**1989**). "A very high performance digital audio signal processing system," ASSP (IEEE) Workshop on Applications of Signal Processing to Audio & Acoustics, New Paltz, NY.

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

Roffler, S. K., and Butler, R. A. (**1968a**). "Factors that influence the localization of sound in the vertical plane," J. Acoust. Soc. Am. **43**, 1255–1259.

Roffler, S. K., and Butler, R. A. (**1968b**). "Localization of tonal stimuli in the vertical plane," J. Acoust. Soc. Am. **43**, 1260–1266.

Thurlow, W. R., and Runge, P. S. (**1967**). "Effects of induced head movements on localization of direction of sound sources," J. Acoust. Soc. Am. **42**, 480–488.

Wallach, H. (**1939**). "On sound localization," J. Acoust. Soc. Am. **10**, 270–274.

Wallach, H. (**1940**). "The role of head movements and vestibular and visual cues in sound localization," J. Exp. Psychol. **27**, 339–368.

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

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

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

Wenzel, E. M., Wightman, F. L., and Kistler, D. J. (**1991**). "Localization with nonindividualized virtual acoustic display cues," Proc. CHI'91, ACM Conf. Hum. Fac. Comp. Sys., New Orleans, LA, 351–359.

Wenzel, E. M. (**1992**). "Localization in virtual acoustic displays," Presence **1**, 80–107.

Wightman, F. L., and Kistler, D. J. (**1989a**). "Headphone simulation of free-field listening I: Stimulus synthesis," J. Acoust. Soc. Am. **85**, 858–867.

Wightman, F. L., and Kistler, D. J. (**1989b**). "Headphone simulation of free-field listening II: Psychophysical validation," J. Acoust. Soc. Am. **85**, 868–878.

Wightman, F. L., and Kistler, D. J. (**1992**). "The dominant role of low-frequency interaural time differences in sound localization," J. Acoust. Soc. Am. **91**, 1648–1661.