Headphone simulation of free-field listening. II: Psychophysical validation

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

Doris J. Kistler
Waisman Center, University of Wisconsin—Madison, Madison, Wisconsin 53705

(Received 14 May 1988; accepted for publication 12 October 1988)

Listeners reported the apparent spatial positions of wideband noise bursts that were presented either by loudspeakers in free field or by headphones. The headphone stimuli were digitally processed with the aim of duplicating, at a listener's eardrums, the waveforms that were produced by the free-field stimuli. The processing algorithms were based on each subject's free-field-to-eardrum transfer functions that had been measured at 144 free-field source locations. The headphone stimuli were localized by eight subjects in virtually the same positions as the corresponding free-field stimuli. However, with headphone stimuli, there were more front–back confusions, and source elevation seemed slightly less well defined. One subject's difficulty with elevation judgments, which was observed both with free-field and with headphone stimuli, was traced to distorted features of the free-field-to-eardrum transfer function.

PACS numbers: 43.66.Qp, 43.66.Pn, 43.66.Yw, 43.88.Si [WAY]

INTRODUCTION

Until recently, research on human sound localization focused on encoding and processing of two major cues, interaural differences in the time of arrival of a sound at a listener's ears (ΔT), and interaural differences in intensity (ΔI). The emphasis has shifted in the past few decades, as a result of a renewed appreciation of the role of the pinnae. It is now clear that important information about the location of a sound source is provided by the direction-dependent interactions of an incoming sound wave with the folds of the pinnae. While most of the existing research suggests that these pinna cues are important primarily for coding of source elevation (e.g., Gardner and Gardner, 1978; Hebrank and Wright, 1974a,b; Roffler and Butler, 1968; Butler, 1975), other studies argue that some aspects of azimuth coding, especially resolution of front–back confusions, are also mediated in part by pinna cues (e.g., Blauert, 1969; Oldfield and Parker, 1984b). In spite of all the recent research, however, important questions remain about the conditions in which pinna cues are important, and how those cues are processed by the auditory system.

We have learned a great deal about the processing of ΔT and ΔI cues because control of interaural time and intensity differences in an experimental setting is straightforward. Headphones are typically used to present the stimuli in experiments on ΔT and ΔI coding, thus allowing complete specification of the stimulus at each ear. Systematic manipulation of pinna cues poses more difficult technical problems, and for this reason our understanding of how pinna cues are processed has advanced relatively slowly. Some investigators have presented stimuli in free field and have attempted to modify pinna cues by filling the folds of the pinnae with putty (Gardner and Gardner, 1973; Oldfield and Parker, 1984b) or by covering the pinnae with blocks (Gardner and Gardner, 1973). Unfortunately, the utility of these techniques is limited, since they do not allow precise, systematic control over the acoustical stimulus delivered to a listener's ear. Other investigators have attempted to exploit the advantages of headphone stimulus delivery by electronically simulating pinna cues, either by processing stimuli with some electronic analog of pinna function (Bloom, 1977; Watkins, 1978) or by recording and reproducing actual pinna cues (Blauert, 1983; Butler and Belendiuk, 1977; Morimoto and Ando, 1982; Plenge, 1974; Searle *et al.*, 1975). Unfortunately, these techniques as well are of limited usefulness, since the degree to which the simulations matched the acoustics of real free-field listening conditions was never quantified. We have developed new techniques for synthesizing headphone-delivered stimuli, with which it is possible to duplicate, in a listener's ear canals, the acoustical waveforms produced by free-field sources (Wightman and Kistler, 1989). We believe that our procedures provide potential solutions to the technical problems associated with manipulation and control of pinna cues.

While headphone stimulus presentation can solve the stimulus control problem, an equally important issue relates to the perceptual equivalence of free-field and headphone listening. Briefly, if we wish to study pinna cues by simulating them with headphone-presented stimuli, we must be sure that the simulated pinna cues evoke the same percepts as real pinna cues. We feel that such a psychophysical validation constitutes the only convincing evidence that it is "localization," as opposed to some other auditory process (such as "lateralization"), which is being studied. Relatively few of the previous studies using headphone stimulus delivery included a comprehensive psychophysical evaluation of the equivalence of free-field and headphone-presented stimuli. One of the more complete studies was reported by Butler and

Belendiuk (1977). Since the focus of this study was on localization in the median plane, the listener's response alternatives in both free-field and headphone conditions were restricted to a few locations on the median plane. This restriction of response alternatives causes a problem. If, with headphone stimuli, the actual percepts were not on the median plane, the listener would have no way of indicating that fact, and thus the data might incorrectly indicate an equivalence of free-field and headphone percepts. Thus the correspondence between the results of the free-field and headphone conditions cannot be taken as conclusive evidence that pinna effects were actually mediating the listener's responses in both conditions.

The work described in this article represents our attempt, through a comprehensive psychophysical validation experiment, to assess the perceptual adequacy of our new simulation techniques (Wightman and Kistler, 1989). This validation consists of direct comparisons of listeners' localization performance in free-field and headphone listening conditions.

I. METHOD

A. Subjects

Eight young adults (four male, four female) served as paid volunteers. All had normal hearing, as verified by audiometric screening at 15 dB HL, with no history of hearing problems of any kind. 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 in this experiment was a train of eight 250-ms bursts of Gaussian noise (20-ms cosine-squared onset–offset ramps), with 300 ms of silence between the bursts. The noise bursts were presented at an overall level of about 70 dB SPL. The Gaussian noise was bandpassed with a tenth-order digital FIR bandpass filter between 200 Hz and 14 kHz. The energy spectrum of the noise was shaped (differently for each stimulus) according to an algorithm that divided the spectrum into critical bands and assigned a random intensity (uniform distribution, 20-dB range) to the noise within each critical band. This trial-by-trial randomization of stimulus spectrum was used to prevent listeners from becoming familiar with specific stimulus or transducer characteristics.

The noise stimuli were presented either by loudspeaker or by headphones. In the former condition, the stimulus was routed to one of six small loudspeakers (Realistic Minimus-7). The loudspeakers were chosen to have similar response characteristics (± 5 dB from 200 Hz to 14 kHz), so no attempt was made to compensate for loudspeaker differences beyond the trial-by-trial stimulus spectral shaping described above. The loudspeakers were mounted on a semicircular steel arc, 2.76 m in diameter, the ends of which were attached to bearings directly above and below the subject's seat in an anechoic chamber. The subject was seated on an adjustable stool such that his/her head was at the center of the arc of loudspeakers. The arc could be rotated around the vertical axis, thus allowing stimulus presentation at any azimuth and at any one of six elevations. The loudspeakers were positioned at the following elevations relative to the horizontal plane passing through the subject's ears: 54, 36, 18, 0, -18, and -36 deg.

For headphone conditions, the noise bursts were transduced by Sennheiser dynamic headphones (HD-340). Each headphone stimulus was digitally processed so that it would simulate a specific free-field stimulus. This processing compensated for the characteristics of the headphones and superimposed a given subject's direction-specific outer ear characteristics (HRTF) on the stimulus (Wightman and Kistler, 1989). Production of each stimulus involved passing a shaped burst of Gaussian noise, spectrally contoured according to the algorithm described above, through two digital filters, one for the left-ear stimulus, and the other for the right-ear stimulus. Each digital filter consisted of two cascaded sections. The first was the filter described in the companion article [Eq. (4) from Wightman and Kistler, 1989], which includes the subject's HRTF for a given ear and source position and the inverse of the subject's headphone-to-ear-canal transfer function for that same ear. The HRTF and headphone transfer functions were measured according to the procedures described in the companion article (Wightman and Kistler, 1989). The second section was a zero-phase bandpass filter (200 Hz to 14 kHz) that was used to eliminate processing artifact at low and high frequencies. Finally, since the particular D/A system used to output the stimuli (Ariel DSP-16) imposed a constant 10-μs delay between left and right stimuli, a 10-μs time shift was added to the phase response of the right bandpass filter section to compensate for the delay. Stimuli were filtered in the frequency domain, using techniques based on the "overlap and add" FFT algorithm described by Stockham (1966).

Stimuli for a given subject and a given run were precomputed (using Signal Technology Inc.'s ILS software on a DEC VAX-11/750) and stored on an IBM-PC disk. They were then converted to analog form via PC-controlled 16-bit D/A converters at a 50 kHz/channel rate. No antialiasing filters were used, since the nearest aliased components were at 36 kHz, well beyond the range of hearing. Stimuli were presented at about 70 dB SPL in free field and at approximately the same level under headphones. The digital processing of the headphone stimuli preserved all the interaural level and time differences, and the slight position-to-position level differences (e.g., from front to back) that existed in free field.

C. Procedure

The aim of this experiment was to compare the apparent positions of sounds presented in free field and under headphones. Therefore, we felt that the paradigm used to quantify apparent spatial position must be the same for both free-field and headphone listening. After considerable pilot work in which we compared the strengths and weaknesses of a number of techniques (Wightman and Kistler, 1980), we chose an "absolute judgment" technique. With this procedure, a subject indicates the apparent spatial position of a sound source by calling out numerical estimates of apparent azimuth and elevation, using standard spherical coordinates. (In our previous work with this procedure, we also asked for distance estimates.) To give some examples, a sound heard directly in front would produce a response "0,0," a sound heard on the right and slightly elevated would produce "90,10," a sound heard on the left and below the horizontal plane would produce " − 90, − 10," and a sound in the rear and well elevated would produce "180, 60."

We were initially concerned that our subjects would demonstrate a wide range of skill with the absolute judgment paradigm, and that this source of variance would contaminate our results. It would then be difficult to separate individual differences in localization ability from individual differences in position estimation skill. However, for several reasons we proceeded anyway. First, our main interest was the comparison of performance in free field with performance under headphones, and both would be measured with the absolute judgment procedure. Second, our subjects appeared to learn the procedure very quickly and produced very stable judgments. Nevertheless, all subjects were given 10 h of experience in the free-field listening condition before final data were collected.

The free-field condition, which was tested first for all subjects, required subjects to estimate the apparent position of sounds delivered from 36 different positions, covering a 360-deg range of azimuths and elevations from 36 deg below the horizontal plane to 54 deg above it. The source locations were chosen from a list of 144 potential positions, which were those at which each subject's HRTFs had been measured (Wightman and Kistler, 1989). The choice was made with the aim of sampling the possible range of azimuths and elevations equally. Later in the experiment, after subjects had completed testing in both free-field and headphone conditions, a second set of 36 positions was selected and seven of the eight subjects were tested again in both free-field and headphone conditions. Table I gives the coordinates of all 72 source locations, and shows how they were divided into "low," "middle," and "high" elevations, and "front," "side," and "back" azimuths for later analysis.

At the beginning of a run in the free-field condition, subjects were blindfolded, led into the anechoic chamber, and seated at the center of the loudspeaker arc (no subject saw the inside of the anechoic chamber or the loudspeaker arrangement at any time during free-field testing). The subject was instructed to look straight ahead and not to move the head while a trial was in progress. The experimenter, who was present with the subject in the chamber in order to move the loudspeaker arc and to record the subject's responses, verified head position and stability. Each trial began with the presentation of a 15-s burst of white Gaussian noise from a loudspeaker (not one of those used for localization) mounted in front of (or, in a separate condition, behind) the subject at floor level. The purpose of this noise was to mask the sounds made by moving the loudspeaker arc, which was positioned by the experimenter during this 15-s pretrial period. When questioned later, all subjects reported that they could not detect the movement of the loudspeaker arc. After the masking noise terminated, the stimulus was presented. Recall that each stimulus consisted of eight 250-ms identical bursts of spectrally contoured noise. During a 5-s silent period immediately after termination of the stimulus, the subject called out azimuth and elevation estimates, and the experimenter entered the responses on a data sheet (no feedback was given to the subjects). A new trial began with the experimenter repositioning the loudspeaker arc according to a script shown on the data sheet. The experimenter attempted to move the arc for about the same length of time, regardless of the required azimuth. During stimulus presentation, the experimenter moved to a corner of the chamber, so as to be acoustically unobtrusive. On a given run, subjects heard a stimulus from each of the 36 locations once; the order of locations presented on each run was random. Each 36-trial run lasted about 20 min, and breaks of about 5 min were taken after each run.

TABLE I. Source positions.

	Low		Middle		High	
	azimuth	elevation	azimuth	elevation	azimuth	elevation
Front						
	− 15	− 36	− 15	0	− 45	36
	− 45	− 36	− 45	0	0	36
	30	− 36	0	0	15	36
	45	− 36	15	0	30	36
	− 15	− 18	45	0	− 30	54
	0	− 18	− 15	18	− 45	54
	30	− 18	− 30	18	15	54
	45	− 18	− 45	18		
			0	18		
			45	18		
Side						
	90	− 36	− 75	0	− 60	36
	105	− 36	− 90	0	120	36
	− 60	− 18	− 105	0	90	36
	− 90	− 18	60	0	− 105	54
	75	− 18	90	0	75	54
	105	− 18	− 75	18	90	54
			− 90	18		
			− 120	18		
			75	18		
			105	18		
Back						
	− 135	− 36	− 135	0	− 135	36
	− 150	− 36	150	0	− 150	36
	135	− 36	165	0	165	36
	150	− 36	180	0	180	36
	180	− 36	135	18	− 150	54
	− 135	− 18	150	18	− 165	54
	− 150	− 18	165	18	180	54
	− 165	− 18				
	135	− 18				

The procedure for the headphone condition was nearly identical to that used for the free-field condition, except that the subjects heard the stimuli over headphones. To avoid the potential influence of visual cues, the subjects were blindfolded as in the free-field condition, even though they had seen the inside of the anechoic chamber during the acoustical measurement phase of the experiment, which came after free-field testing. They were also seated in the anechoic chamber during headphone testing. The trial sequence was the same as for the free-field condition, except that no masking noise was presented before each trial. After each stimulus was presented, and the subject called out azimuth and elevation estimates, the experimenter, who was outside the chamber listening over an intercom, entered the responses on a PC keyboard. As before, each run required estimates of 36 source positions, and because of the slightly faster pace, about four runs were completed in each 90-min session.

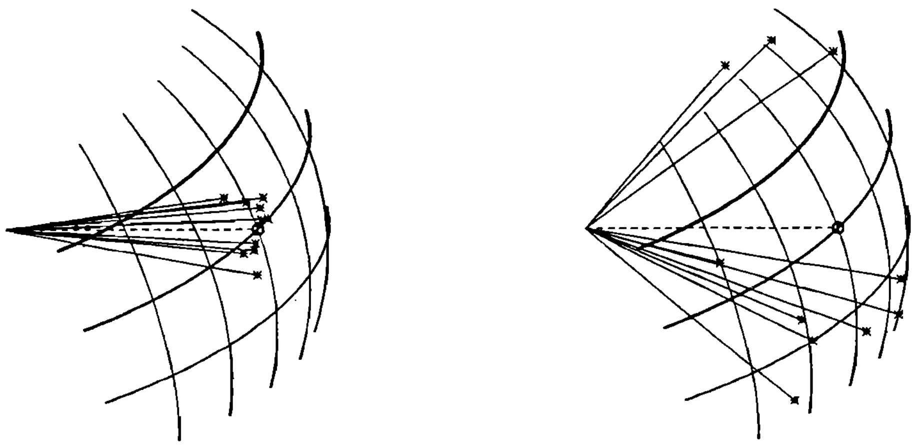
Each subject first completed six runs in the free-field condition and then, after the acoustical measurements were made, completed ten runs in the headphone condition. Next, each subject was tested in an additional six runs in the free-field condition, to evaluate learning effects. As an additional check on learning effects, seven of the eight subjects were tested in both the free-field (six runs) and headphone (six runs) conditions with an entirely new set of 36 source locations.

II. RESULTS

Before discussing the results of the main experiment, we will describe two additional conditions that we evaluated on a subset of the subjects. Both of these conditions used free-field stimulus presentation and were included as checks on certain potentially confounding aspects of our procedure. First, to evaluate the possibility that the position of the masking noise loudspeaker in the free-field conditions might bias the subjects' judgments of the positions of the other sources, we tested five of the six subjects with the masking loudspeaker moved to the rear. Data from six runs in both conditions revealed no differences. Second, as a check on the extent to which subjects' judgments might be influenced by the nature of the response scale, we required two subjects to respond in "clock time" coordinates instead of "degree" coordinates. In clock time coordinates, a source in front and level with the ears would produce a response of "12 o'clock, 3 o'clock" and a source behind and elevated would produce "6 o'clock, 1 o'clock." After a short training period with the clock time procedure, the subjects completed six test runs with both types of response. The data showed no differences between responses based on clock time coordinates and those based on degree coordinates (correlations of 0.96 or higher).

Analysis of the results of a localization experiment of this sort is complicated by the fact that the stimuli and responses are represented by points in three-dimensional space (in our case, points on the surface of a sphere, since distance was constant). For spherically organized data, the usual statistics (mean and variance) are either inappropriate or potentially misleading. For example, an azimuth error of 30 deg for a source on the horizontal plane is much larger in an absolute distance sense than a 30-deg azimuth error for a point at 60-deg elevation. This particular problem, coupled with the fact that azimuth error and elevation error are almost certainly not independent, makes it difficult to interpret such summary statistics as average azimuth error, collapsed across all elevations or average elevation error collapsed across all azimuths (Oldfield and Parker, 1984a). We have, therefore, borrowed some techniques from the field of spherical statistics (Fisher *et al.*, 1987; Watson, 1983) in order to analyze our data.

We used the following descriptive spherical statistics to characterize the psychophysical data: the average angle of error, the judgment centroid, and κ^{-1}. 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 position. The judgment centroid is a unit-length vector with the same direction as the resultant vector, 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 of the data space (the subject's position). Note that the length of the resultant is determined by the dispersion of the judgments; judgments concentrated around the centroid would produce a long resultant and scattered judgments, a short



(a)

(b)

FIG. 1. Illustration of the meaning of κ^{-1}, a measure of the dispersion of spherical data. In each panel, the points (*) represent hypothetical judgments of sound source position. The actual source position is indicated by a circle (o). The vertical lines (longitude markers) are 15 deg apart, and the horizontal lines (latitude markers) are 18 deg apart. The lines are drawn from the hypothetical subject's position (the origin of the sphere) to each data point (the dashed line) is to the target position. (a) Data with a κ^{-1} of 0.01, representative of the small dispersions observed in our data. (b) Data with a κ^{-1} of 0.18, representative of the large dispersions we occasionally observed in our data. Our data include a few values of κ^{-1} larger than those shown. Those occurred only when a clear outlier contaminated the data set.

resultant. In theory, if each of N judgment vectors has unit length, the length of the resultant could vary between N (no dispersion) and 0 (uniform distribution of judgments around the sphere). The parameter κ, which is estimated from the length of the resultant, is the preferred index of the dispersion of spherical data (Fisher *et al.*, 1987; Watson, 1983). It is one of the parameters of nearly all of the probability distributions that are used as models for spherically arranged data (see the Appendix of this article).

An unbiased estimate of κ for small (less than 16) samples is given by (Fisher *et al.*, 1987)

$$\kappa' = (N-1)^2/N(N-R),$$

where N is the number of data points and R is the length of the resultant vector. Note that as the data become more concentrated on the surface of the sphere, or as R approaches N, κ increases (to an infinite limit.) It is customary to use estimates of κ^{-1} as indices of dispersion. The two data sets shown in Fig. 1 illustrate estimated values of κ^{-1} that encompass most of the values of κ^{-1} we estimated (using the formula shown above) in our experiment.

One final peculiarity of localization data that must be addressed is the issue of front–back "confusions." These are responses, observed in nearly all localization studies, which indicate that a front-hemisphere source was perceived in the rear hemisphere, or, less frequently, the reverse. These responses are typically observed for sources near or on the median plane. It is difficult to know how to treat these apparent confusions fairly. Since the confusions are relatively infrequent, it has been common practice to resolve them (i.e., code the response as if it indicated the correct hemisphere). Otherwise, error indices would be greatly inflated. Of course, if we assume the responses correctly reflect the subject's perception, resolution of the confusions may be misleading. Since our primary motivation is to compare free-field and headphone conditions, we elected to resolve all apparent confusions, and report the incidence of confusions in each condition. Our algorithm for resolving front–back confusions treats each judgment identically, regardless of target azimuth. If the angle between the target and the judgment is made smaller by reflecting the judgment about the vertical plane passing through the subject's ears, then the judgment is entered in reflected form, and the confusion count is increased by one. This procedure undoubtedly produces a slight overestimate of confusion rate and an underestimate of error rate for target positions near ± 90 deg. However, the comparison of free-field and headphone conditions should not be affected by these biases.

The free-field data discussed below include the data from the additional six runs that were collected after the first headphone condition had been completed. The additional six runs produced data that were indistinguishable from the data obtained from the initial six runs. In addition, since an analysis conducted after the experiment was complete revealed no differences between the judgments of the original 36 positions and judgments of the second set of later 36 positions, the data were combined. Thus the final free-field data set consists of 12 judgments of the original 36 positions and six judgments of the second set of 36 positions. The headphone data set consists of 10 judgments of the original 36 positions and 6 judgments of the second set of 36 positions.

We computed the judgment centroid, angle of error, and κ^{-1} separately for each of the 72 source positions in both free-field and headphone conditions, for each of the eight subjects. We also computed correlations between target and centroid azimuth, and between target and centroid elevation for both the free-field and headphone conditions. Finally, we computed a three-dimensional "goodness of fit" between the set of points defined by each subject's judgment centroids and the set of points defined by the target locations. The correlations and the goodness of fit were computed according to algorithms devised by Schonemann (Lingoes and Schonemann, 1974; Schonemann and Carroll, 1970) for the purpose of fitting one matrix to another. Our use of the algorithms involves rigid rotation of the matrix of centroids to a least-squares fit with the matrix of target positions (thus we ignore constant azimuth and/or elevation biases in the judgments) and computation of a statistic S, the normalized sum of the squared residuals. The final measure, which we call correlation in one case, and goodness of fit in another, is equal to $(1-S)^{1/2}$. In the case of the azimuth or elevation correlations, this measure is nearly identical to a Pearson correlation since the data are two dimensional. The goodness of fit measure is essentially a three-dimensional Pearson correlation, which gives an overall indication of the degree of match between the targets and the judgment centroids.

Table II shows summary statistics from both the free-field and headphone conditions. The entries in this table represent averages over all 72 source positions. Table III shows similar data for six separate regions of auditory space: front, side (left and right were combined), and rear quadrants; high (+ 36 to + 54 deg), medium (0 to + 18 deg), and low (− 18 to − 36 deg) elevations. Figures 2–9 show the relation between target and response azimuth (for stimuli at all elevations) and the relation between target and response elevation (for stimuli at all azimuths) for each of the subjects in both free-field and headphone conditions.

The data reveal a number of important features of localization behavior. First, there are substantial individual differences. While these differences are less apparent in the global performance metrics such as goodness of fit and average angle of error, they emerge clearly in the correlations between target and centroid elevation, and the number of

TABLE II. Global measures of localization performance. Measures of free-field performance are in boldface type and measures of simulation performance are in parentheses.

ID	Goodness of fit	Azimuth correlation	Elevation correlation	% reversals
SDE	**0.93** (0.89)	**0.983** (0.973)	**0.68** (0.43)	**12** (20)
SDH	**0.95** (0.95)	**0.965** (0.950)	**0.92** (0.83)	**5** (13)
SDL	**0.97** (0.95)	**0.982** (0.976)	**0.89** (0.85)	**7** (14)
SDM	**0.98** (0.98)	**0.985** (0.985)	**0.94** (0.93)	**5** (9)
SDO	**0.96** (0.96)	**0.987** (0.986)	**0.94** (0.92)	**4** (11)
SDP	**0.99** (0.98)	**0.994** (0.990)	**0.96** (0.88)	**3** (6)
SED	**0.96** (0.95)	**0.972** (0.986)	**0.93** (0.82)	**4** (6)
SER	**0.96** (0.97)	**0.986** (0.990)	**0.96** (0.94)	**5** (8)

TABLE III. Regional measures of localization performance.

		Low elevations			Middle elevations			High elevations		
Azimuth	ID	Angle of error	κ^{-1}	% reversals	Angle of error	κ^{-1}	% reversals	Angle of error	κ^{-1}	% reversals
Front	SDE	**37.4**(34.9)	**0.08**(0.05)	**21**(46)	**16.2**(19.7)	**0.04**(0.05)	**4**(21)	**28.3**(40.0)	**0.10**(0.10)	**19**(29)
	SDH	**20.0**(16.0)	**0.04**(0.03)	**4**(17)	**21.1**(37.4)	**0.03**(0.08)	**1**(8)	**19.3**(26.2)	**0.03**(0.05)	**12**(58)
	SDL	**21.9**(22.2)	**0.06**(0.03)	**3**(0)	**25.9**(29.1)	**0.07**(0.07)	**1**(11)	**27.8**(31.3)	**0.07**(0.14)	**44**(81)
	SDM	**19.4**(23.1)	**0.05**(0.05)	**0**(3)	**16.3**(15.3)	**0.02**(0.04)	**0**(5)	**24.0**(21.9)	**0.01**(0.10)	**4**(20)
	SDO	**14.2**(15.2)	**0.05**(0.02)	**1**(0)	**14.5**(21.8)	**0.04**(0.06)	**0**(0)	**25.7**(34.8)	**0.04**(0.15)	**7**(39)
	SDP	**10.9**(22.3)	**0.02**(0.06)	**0**(6)	**9.4**(13.9)	**0.01**(0.03)	**0**(9)	**17.1**(23.2)	**0.05**(0.06)	**5**(34)
	SED	**19.2**(19.0)	**0.07**(0.06)	**0**(2)	**21.2**(15.1)	**0.05**(0.03)	**0**(0)	**31.6**(29.1)	**0.08**(0.10)	**19**(30)
	SER	**19.6**(16.7)	**0.08**(0.03)	**1**(0)	**17.9**(16.7)	**0.05**(0.04)	**2**(0)	**27.4**(24.5)	**0.07**(0.13)	**11**(2)
	Mean	**20.4**(21.0)	**0.06**(0.04)	**4**(10)	**17.9**(21.5)	**0.04**(0.05)	**1**(7)	**25.2**(29.3)	**0.06**(0.10)	**16**(38)
Side	SDE	**22.9**(33.4)	**0.03**(0.06)	**6**(14)	**17.0**(18.3)	**0.04**(0.05)	**19**(20)	**24.5**(37.1)	**0.10**(0.15)	**28**(28)
	SDH	**14.7**(10.4)	**0.03**(0.02)	**17**(28)	**14.3**(13.8)	**0.03**(0.02)	**6**(3)	**17.9**(16.0)	**0.02**(0.03)	**25**(11)
	SDL	**18.4**(21.3)	**0.04**(0.04)	**0**(1)	**18.7**(14.2)	**0.04**(0.02)	**9**(6)	**24.7**(29.5)	**0.09**(0.11)	**16**(32)
	SDM	**19.4**(15.6)	**0.07**(0.04)	**11**(10)	**17.9**(19.7)	**0.07**(0.05)	**7**(12)	**19.7**(26.8)	**0.05**(0.15)	**21**(10)
	SDO	**13.2**(14.3)	**0.02**(0.03)	**7**(21)	**16.6**(17.4)	**0.05**(0.03)	**5**(12)	**21.7**(27.4)	**0.04**(0.12)	**14**(16)
	SDP	**13.8**(15.1)	**0.06**(0.04)	**4**(0)	**15.0**(10.1)	**0.02**(0.02)	**8**(6)	**17.2**(16.9)	**0.05**(0.03)	**7**(4)
	SED	**20.0**(17.1)	**0.05**(0.03)	**2**(7)	**16.5**(12.8)	**0.05**(0.03)	**2**(0)	**27.5**(37.9)	**0.14**(0.09)	**10**(17)
	SER	**20.2**(18.5)	**0.07**(0.07)	**2**(9)	**13.9**(17.3)	**0.04**(0.04)	**7**(7)	**26.7**(22.4)	**0.08**(0.06)	**22**(15)
	Mean	**17.7**(18.4)	**0.04**(0.03)	**6**(11)	**16.1**(15.1)	**0.04**(0.03)	**8**(8)	**22.6**(26.7)	**0.07**(0.09)	**18**(17)
Back	SDE	**19.6**(17.6)	**0.04**(0.04)	**4**(1)	**22.9**(24.2)	**0.08**(0.07)	**4**(2)	**44.8**(49.3)	**0.08**(0.09)	**14**(21)
	SDH	**22.4**(20.2)	**0.03**(0.02)	**0**(2)	**27.3**(25.9)	**0.04**(0.03)	**0**(0)	**24.8**(27.1)	**0.05**(0.05)	**0**(0)
	SDL	**21.9**(21.5)	**0.04**(0.03)	**0**(2)	**14.5**(17.7)	**0.04**(0.03)	**0**(0)	**24.6**(27.7)	**0.12**(0.15)	**2**(10)
	SDM	**16.0**(16.2)	**0.02**(0.06)	**0**(8)	**18.4**(20.9)	**0.05**(0.07)	**0**(0)	**18.4**(31.9)	**0.08**(0.21)	**0**(15)
	SDO	**20.8**(16.0)	**0.02**(0.02)	**0**(0)	**28.9**(22.3)	**0.04**(0.10)	**0**(2)	**35.1**(33.3)	**0.08**(0.11)	**5**(15)
	SDP	**15.7**(11.7)	**0.02**(0.01)	**0**(0)	**14.0**(11.9)	**0.02**(0.01)	**1**(0)	**21.3**(18.7)	**0.06**(0.04)	**0**(0)
	SED	**23.2**(22.5)	**0.03**(0.02)	**1**(0)	**20.5**(14.1)	**0.08**(0.04)	**0**(4)	**31.3**(32.9)	**0.16**(0.06)	**3**(5)
	SER	**13.9**(15.8)	**0.03**(0.03)	**0**(1)	**20.5**(21.3)	**0.05**(0.08)	**0**(9)	**33.0**(29.2)	**0.04**(0.10)	**3**(41)
	Mean	**19.4**(17.8)	**0.03**(0.03)	**1**(1)	**21.1**(19.7)	**0.05**(0.05)	**1**(2)	**29.8**(31.2)	**0.09**(0.10)	**4**(13)

front–back confusions. It is clear, for example, that the performance of subjects SDO and SER is superior to that of subjects SDE and SDH. It is curious that the differences among subjects emerges most clearly in the judgment of source elevation. Another obvious feature of the data (Table III) is that localization performance is not the same in all regions of auditory space. In general, it appears that precision is greatest on the side, slightly poorer in front, and poorest at high elevations in the rear. Since there have been few studies of the ability of subjects to localize sources (i.e., source position *identification*) outside the median plane or the horizontal plane, there are few data in the literature with which our results can be directly compared. The classic data on source position *discrimination* ability are at least indirectly relevant to such a comparison. Our data contrast with the results of some of the discrimination studies. For example,

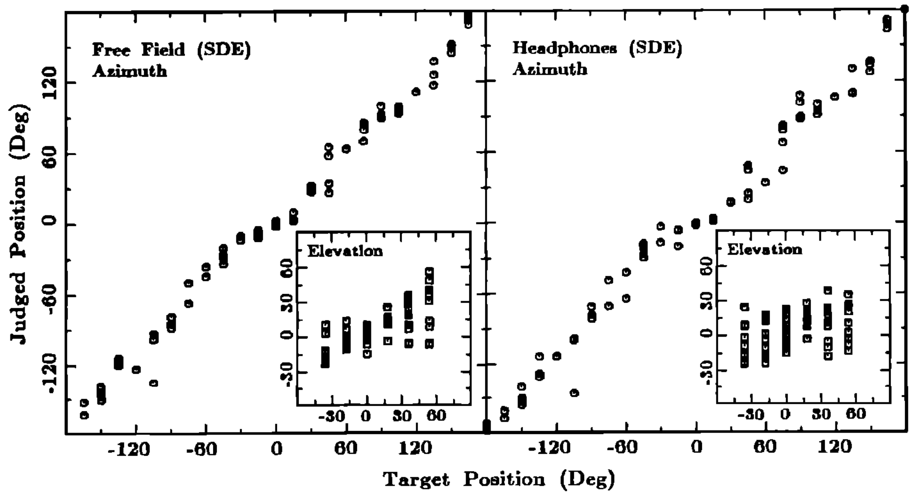


FIG. 2. Scatterplots of actual source azimuth (and, in the insets, elevation) versus judged source azimuth for subject SDE in both the free-field and headphone conditions. Each data point represents the centroid of at least six judgments. All 72 source positions are represented in each panel. Thus data from six different source elevations are combined in the azimuth panels, and data from 24 different azimuths are combined in the elevation panels. Note that the scale is the same for azimuth and elevation plots.

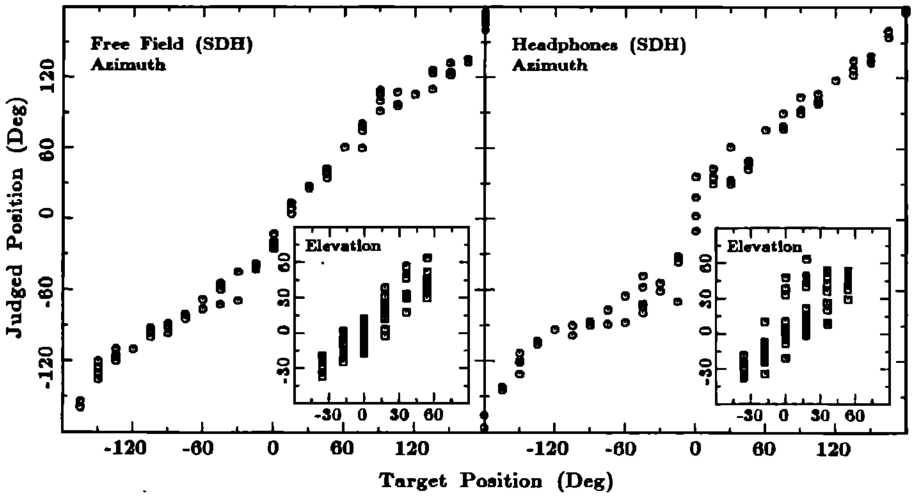


FIG. 3. Same as Fig. 2 but for subject SDH.

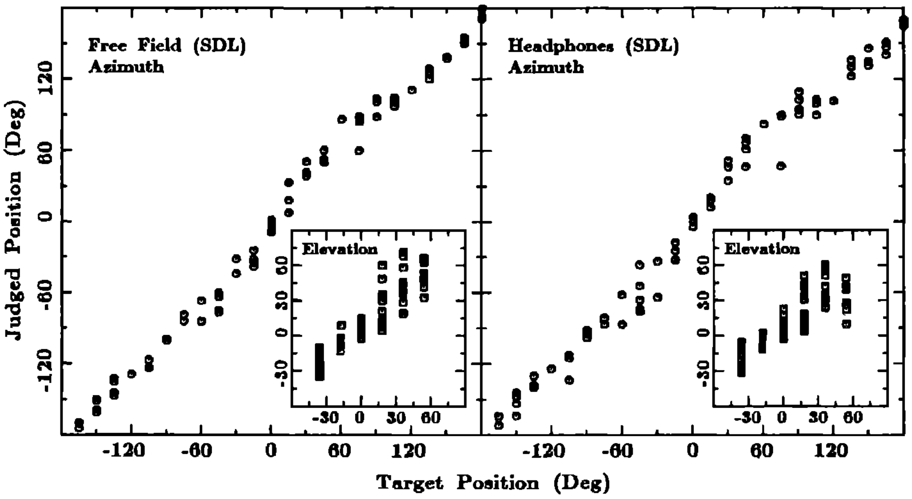


FIG. 4. Same as Fig. 2 but for subject SDL.

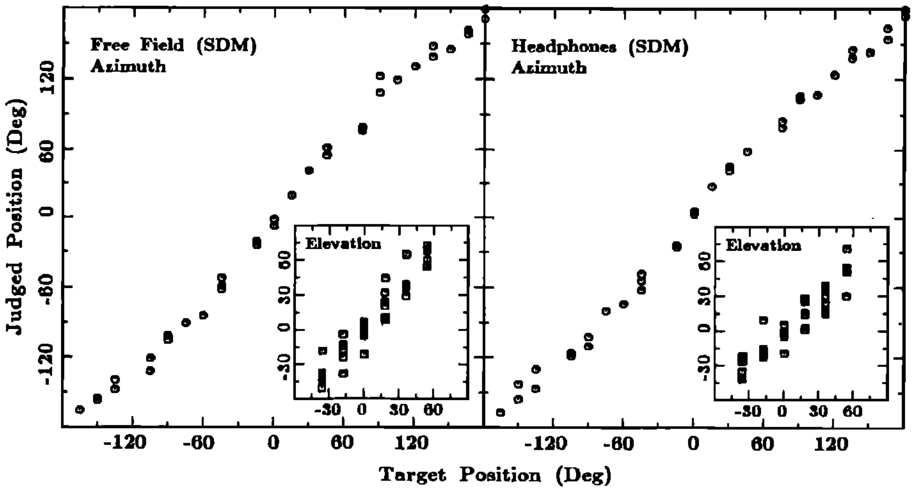


FIG. 5. Same as Fig. 2 but for subject SDM. This subject judged apparent positions of only 36 source locations in both free-field and headphone conditions.

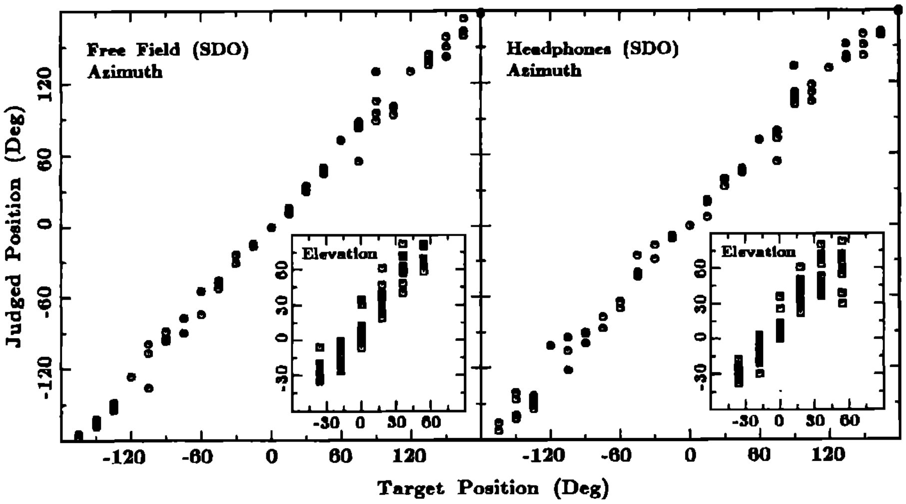


FIG. 6. Same as Fig. 2 but for subject SDO.

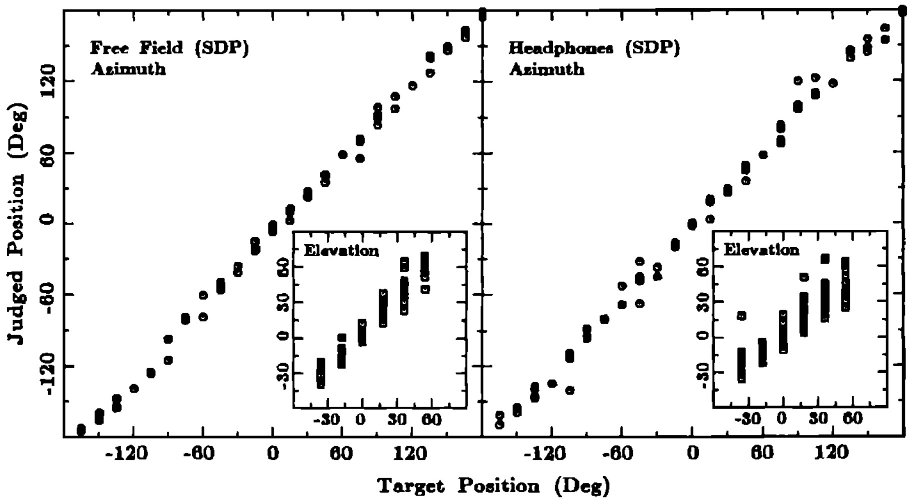


FIG. 7. Same as Fig. 2 but for subject SDP.

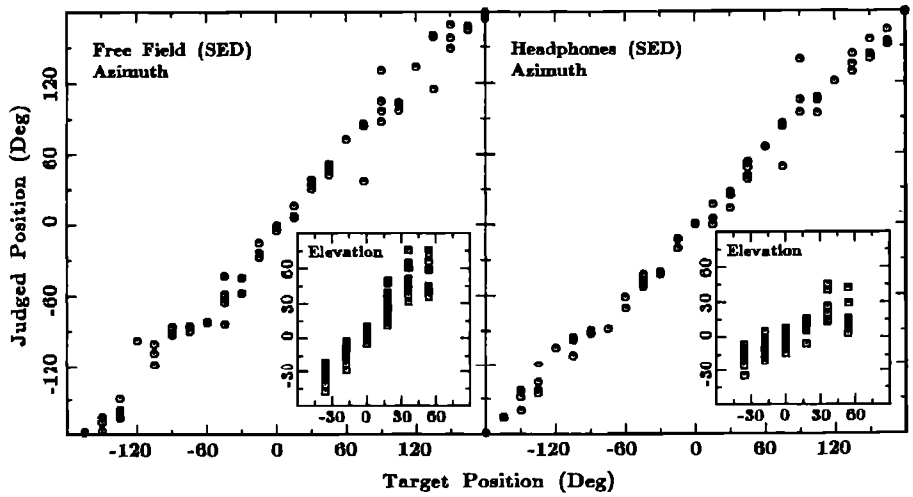


FIG. 8. Same as Fig. 2 but for subject SED.

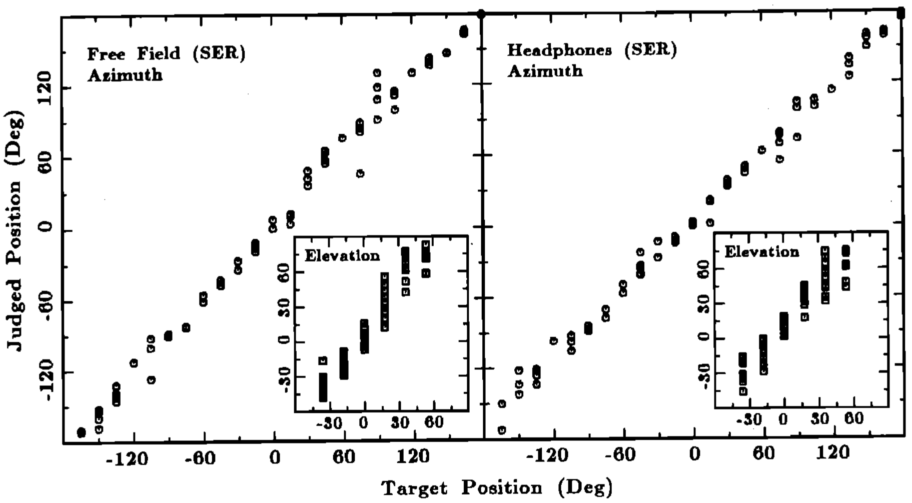


FIG. 9. Same as Fig. 2 but for subject SER.

Mills (1958) reports that discrimination is considerably poorer on the side than in front. Other discrimination data, however, argue for little effect of azimuth on source position discrimination performance (Harris, 1972).

One recent study that tested localization performance at a large number of source positions is that reported by Oldfield and Parker (1984a). In spite of substantial differences in procedure and data analysis techniques, our free-field data are at least qualitatively similar to those reported by these investigators. While Oldfield and Parker did not report individual subject data, their group results are basically consistent with ours, both in terms of the magnitude of the average error, and in terms of the regions of auditory space in which localization is best and poorest.

The data in Tables II and III from the headphone condition are nearly identical to the data from the free-field condition. Moreover, those differences that do emerge seem to be confined to the elevation components of the judgments. For each subject, the correlation between target elevation and response elevation is slightly lower in the headphone condition. However, as can be seen in Fig. 2 (subject SDE), the difference between free-field and headphone elevation judgments does not appear to be as great as the correlation statistic indicates. This subject is relatively poor at judging source elevation in both conditions. Finally, there is a clear increase (approximately double) in the frequency of front–back confusions in the headphone condition.

III. DISCUSSION

The data indicate that our attempt to simulate free-field listening with headphone stimulus delivery was largely successful. For each subject in each of the six regions of auditory space there was a very close correspondence between judgments of the apparent positions of real sources in free field, and judgments of the apparent positions of digitally synthesized virtual sources presented under headphones. Comparisons of the goodness of fit values, which indicate the overall correlation between target and response positions, and comparisons of the target-response correlation plots (Figs. 2–9) verify the close correspondence. There are also no obvious differences in any region of auditory space between free-field and headphone conditions in the average angle of error or in the stability of the judgments as indicated by κ^{-1}.

The data also indicate that there are some relatively subtle aspects of free-field listening, which are not captured in the headphone simulation. For example, there is a substantial increase in the number of front–back confusions in the headphone condition. We are aware that others who have attempted to simulate free-field listening under headphones have also observed this (e.g., Blauert, 1983), and we have no explanation.

We feel it is potentially important that the differences among the subjects and the differences between the free-field data and the headphone data appear almost entirely in the elevation components of the responses. For each subject in both free-field and headphone conditions, the correlation between target and response azimuth is nearly perfect (0.950 is the lowest correlation). However, the correlation between target and response elevation is always a bit lower, and for all subjects it is lower in the headphone condition than in free field. One reason for the generally lower correlations in the elevation components of the responses may be that the range of elevations studied is much smaller than the range of azimuths. It is well known that other things being equal, restricting the range of one or both variables will lower the correlation.

One subject, SDE, was especially poor at judging sound source elevation. This prompted us to conduct a preliminary study of the potential acoustical bases of elevation coding, as revealed in the HRTFs measured on each subject's ears. After removing from each measurement the characteristics of the pseudorandom noise signal (used for the measurement) and the loudspeaker used to present the signal, we computed "interaural elevation dependency" functions for each subject. These functions were computed by first dividing all leading ear HRTFs by the corresponding trailing ear HRTF, to produce "interaural difference" functions. Thus, for sources on the right, the right ear HRTFs were divided by the left ear HRTFs, and vice versa for sources on the left. Next, the interaural difference functions were normalized to zero elevation. This was done by dividing all the interaural

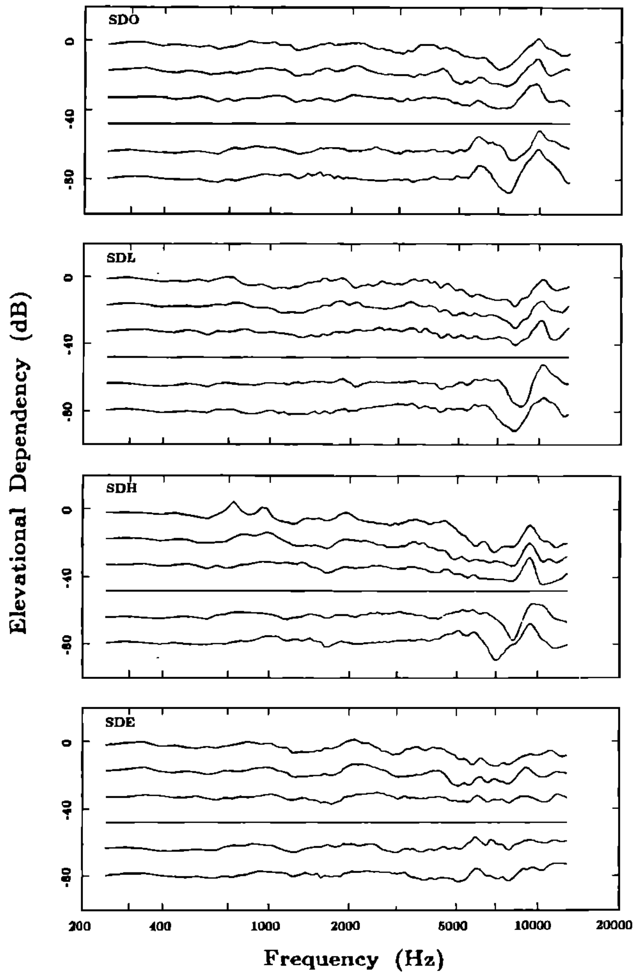


FIG. 10. Interaural elevation-dependency functions for four subjects. In each panel, the lower five functions are displaced downward by successive increments of 15 dB for display clarity. Thus each function has a low-frequency asymptote of 0 dB. Each function represents the change in interaural amplitude caused by a source elevation shift from 0 deg to the elevation represented by each function. From top to bottom, the functions in each panel represent elevations of 54, 36, 18, 0 (the reference), − 18, and − 36 deg. The elevation-dependency functions for the four subjects not shown (SDM, SDP, SED, and SER) are similar to those of subjects SDL, SDH, and SDO.

difference functions at a given azimuth by the interaural difference function at that same azimuth and zero elevation. Finally, these normalized elevation-dependency functions were averaged over all azimuths. The result for each subject is a set of six functions that show the change in interaural difference (we examined only the interaural intensity difference), which results from moving a source from zero elevation to the five other elevations. Figure 10 shows the elevation dependency results from four subjects. As can be seen, in spite of large intersubject differences in the measured HRTFs (see the companion article, Wightman and Kistler, 1989), the elevation dependency functions from three of the subjects are remarkably similar. The elevation dependency functions from the four subjects that are not shown are virtually indistinguishable from these three. Subject SDE's elevation-dependency functions, however, are radically different, and, in fact, show very little elevation dependency compared to the other subjects. From this preliminary analysis, it appears that subject SDE's poor performance in judging the elevation of sound sources both in free-field and under headphones may have an acoustical basis.

IV. CONCLUSIONS

Within the limited range of conditions studied here, appropriately synthesized stimuli presented over headphones are judged to have the same spatial positions as stimuli presented in free field. Results from eight subjects in a psychophysical experiment that directly compared free-field and headphone listening confirmed the adequacy of the simulation procedures. We feel that the real importance of the simulation techniques described here is the potential they offer as a means for studying aspects of human sound localization that were heretofore inaccessible. Headphone stimulus presentation allows independent and precise manipulation of all aspects of the stimuli presented to the two ears. This experimental advantage has been exploited for years in research on what has been called "lateralization" or "in-head localization." Lateralization studies have contributed much to our understanding of coding of interaural difference cues, but the generalizability of this work to questions about sound localization has always been limited, since stimuli that are lateralized are always heard inside the head, a very unnatural percept. Now, the same advantage of headphone stimulus delivery can be brought to bear on issues more directly related to localization of sounds in auditory space.

ACKNOWLEDGMENTS

This work was supported by grants from the National Institutes of Health (NINCDS) and the Air Force Office of Scientific Research, and by a Joint Research Interchange with NASA. The authors gratefully acknowledge the assistance of Marianne Arruda, Douglas Swiggum, and Linda Pettersson for their contributions to the technical phases of the work, and the helpful comments of Dr. Robert Lutfi on several earlier drafts of this article.

APPENDIX: MODELING THE DISTRIBUTION OF SPHERICAL DATA

The most commonly used probability distribution for modeling spherical data is the von Mises–Fisher distribution (Fisher *et al.*, 1987). A polar plot of this distribution shows that in any plane through its axis of symmetry it has the form of a normal or Gaussian distribution wrapped around a circle. In fact, the von Mises distribution, which is used to model distributions of data points on a circle, and from which the von Mises–Fisher distribution is derived, is very similar to the wrapped normal distribution. The von Mises, and its spherical extension, the von Mises–Fisher, are generally preferred over the circular and spherical wrapped normals because the parameters are easier to estimate. The von Mises–Fisher distribution, in probability density form is given by

$$f(\theta,\phi) = C_f \exp\{\kappa[\sin\theta \sin\alpha \cos(\phi-\beta) + \cos\theta\cos\alpha]\}\sin\theta,$$

where

$$C_f = \kappa/(4\pi \sinh\kappa)$$

and where θ represents elevation $(0<\theta<\pi)$ with 0 in this context meaning directly overhead (note that in our coordinate system 0 elevation means straight ahead, on the horizontal plane) and ϕ represents azimuth $(0<\phi<2\pi)$. The parameters α and β are location elevation and azimuth parameters; the distribution has rotational symmetry about the direction (α,β). Thus α and β define a kind of "mean direction" for the distribution. Here, κ is called the "concentration parameter"; the larger the value of κ, the more the distribution is concentrated about the direction (α, β).

If the distribution is expressed in standardized form, with a vertical axis of symmetry, in the direction $(\alpha = 0, \beta = 0)$, the expression for the distribution is greatly simplified, since the azimuth variable drops out:

$$f(\theta,\phi) = C_f \exp(\kappa \cos \theta) \sin \theta,$$

with C_f given as above.

Blauert, J. (1969). "Sound localization in the median plane," Acustica **22**, 205–213.

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.

Butler, R. A. (1975). "The influence of the external and middle ear on auditory discriminations," in *Handbook of Sensory Physiology*, edited by W. D. Keidel and W. D. Neff (Springer, Berlin).

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.

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

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

Harris, J. D. (1972). "A florilegium of experiments on directional hearing," Acta Oto-Laryngol. Suppl. 298.

Hebrank, J., and Wright, D. (1974a). "Are two ears necessary for localization of sound sources on the median plane?," J. Acoust. Soc. Am. **56**, 935–938.

Hebrank, J., and Wright, D. (1974b). "Spectral cues used in localization of sound sources on the median plane," J. Acoust. Soc. Am. **56**, 1829–1834.

Lingoes, J. C., and Schonemann, P. H. (1974). "Alternative measures of fit for the Schonemann–Carroll matrix fitting algorithm," Psychometrika **39**, 423–427.

Mills, A. (1958). "On the minimum audible angle," J. Acoust. Soc. Am. **32**, 132–134.

Morimoto, M., and Ando, Y. (1982). "On the simulation of sound localization," in *Localization of Sound: Theory and Applications*, edited by R. W. Gatehouse (Amphora, Groton).

Oldfield, S., and Parker, S. (1984a). "Acuity of sound localisation: a topography of auditory space. I. Normal hearing conditions," Perception **13**, 581–600.

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

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

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

Schonemann, P. H., and Carroll, R. M. (1970). "Fitting one matrix to another under choice of a central dilation and rigid motion," Psychometrika **35**, 245–255.

Searle, C. L., Braida, L. D., Cuddy, D. R., and Davis, M. F. (1975). "Binaural pinna disparity: Another auditory localization cue," J. Acoust. Soc. Am. **57**, 448–455.

Stockham, T. G. (1966). "High speed convolution and correlation," Proc. AFIPS Spring Joint Comp. Conf. **28**, 229–233.

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

Watson, G. (1983). *Statistics on Spheres* (Wiley Interscience, New York).

Wightman, F. L., and Kistler, D. J. (1980). "A new 'look' at auditory space perception," in *Psychophysical, Physiological, and Behavioral Studies in Hearing*, edited by G. van den Brink and F. Bilsen (Delft U.P., Delft, The Netherlands).

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