The dominant role of low-frequency interaural time differences in sound localization[a)]

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 6 May 1991; revised 29 June 1991; accepted 28 October 1991)

Two experiments are described in which listeners judge the apparent directions of virtual sound sources–headphone-presented sounds that are processed in order to simulate free-field sounds. Previous results suggest that when the cues to sound direction are preserved by the simulation, the apparent directions of virtual sources are nearly the same as the apparent directions of real free-field sources. In the experiments reported here, the interaural phase relations in the processing algorithms are manipulated in order to produce stimuli in which the interaural time difference cues signal one direction and interaural intensity and pinna cues signal another direction. The apparent directions of these conflicting cue stimuli almost always follow the interaural time cue, as long as the wideband stimuli include low frequencies. With low frequencies removed from the stimuli, the dominance of interaural time difference disappears, and apparent direction is determined primarily by interaural intensity difference and pinna cues.

PACS numbers: 43.66.Qp, 43.66.Yw, 43.66.Pn (WAY)

INTRODUCTION

Human sound localization is mediated by a wide variety of cues. The primary acoustical cues are the frequency-dependent patterns of interaural time and intensity differences that result from diffraction of incoming sound waves around the head and pinnae (Kuhn, 1979; Shaw, 1974; Middlebrooks and Green, 1990; Middlebrooks *et al.*, 1989). In addition to interaural differences, monaural spectral features (e.g., the direction-dependent spectral notches or peaks provided by the filtering action of each ear's pinna) are also thought to provide cues to sound source location (Butler *et al.*, 1990; Musicant and Butler, 1985). If we add to this list the potential acoustical cues produced by source (or head) movement (Perrott and Musicant, 1981; Thurlow *et al.*, 1967), and the nonacoustical cues such as those provided by source familiarity (Coleman, 1962) and vision (Platt and Warren, 1972; Shelton and Searle, 1980) it becomes apparent that sound localization is a complex perceptual process that involves integration of information derived from multiple cues. The weighting of the various cues, the frequency range over which each of the cues is viable, and the regions of auditory space in which the cues are important are among the many issues that have stimulated research on this topic for over a century.

No completely satisfactory answers to these questions have emerged from all the research. However, consensus seems to have been reached on a few points. For example, most researchers now agree that interaural time difference (ITD) is the major cue for source azimuth (direction on the horizontal plane), and that ITD is probably encoded mostly by low-frequency auditory neurons (e.g., Middlebrooks and Green, 1990). There also seems to be agreement that source elevation (direction above or below the horizontal plane) is cued by the spectral peaks and notches, produced by pinna filtering, that occur mainly at frequencies above 5 kHz. There is less agreement on the question of whether the pinna cues are processed binaurally (i.e., via computation of interaural intensity differences) or monaurally. Moreover, there are virtually no data describing how the auditory system weights and combines the ITD and spectral cues when both are present (e.g., with a wideband stimulus), although one theory of localization (Searle *et al.*, 1976) offers a quantitative model of the process.

One obvious reason for our ignorance in the area of human sound localization is that control of the stimulus is very difficult. The complexity of the sound environment in a typical room prohibits systematic manipulation and control of the various acoustical cues, thus forcing most researchers to choose acoustically simpler settings for their work. The great number of "lateralization" studies reflects this search for simple and controllable acoustics. A lateralization experiment involves presentation of stimuli to listeners through headphones and thus allows complete control of the stimulus. However, in lateralization experiments the apparent positions of the sound images produced by the stimuli are inside the head, somewhere roughly on a line between the ears. Because of the unnaturalness of the percept, researchers have come to question the generalizability of lateralization results to localization questions. Research focused explicitly on localization issues is usually conducted either in a free sound field (e.g., an anechoic chamber) or in a sound field in which reflections can be controlled (e.g., Hartmann, 1983).

[a)] Portions of this paper were presented at the 12th midwinter meeting of the Association for Research in Otolaryngology, February 1989.

 0001-4966/92/031648-14$00.80

Even in these environments it is both difficult to measure the acoustic stimulus (at a listener's eardrums) and nearly impossible to control it in systematic, predictable ways.

Several laboratories have recently adopted stimulus generation techniques that capture some of the realism of a natural acoustical environment while retaining complete definition and control of the acoustical stimulus. Most of these techniques involve headphone presentation of stimuli that have been digitally processed in order to mimic the acoustical diffraction effects of the head and pinnae. Stimuli processed in this way typically appear to originate outside the head in contrast with the headphone-presented stimuli used in lateralization experiments. The basics of the new approach were proposed more than 30 years ago by Blauert and others (described in Blauert, 1983), but it is only now that digital technology has made it feasible.

While our approach is not fundamentally different from that described by Blauert (1983), there are two distinguishing features. First, we use an acoustical verification procedure to quantify the physical equivalence (measured at a listener's eardrums) of stimuli produced by free-field sources and stimuli produced by our headphone system (Wightman and Kistler, 1989a). Second, we offer psychophysical data as confirmation of the perceptual adequacy of stimuli presented with the headphone system (Wightman and Kistler, 1989b). These data suggest that the apparent directions of sounds presented in a simulated free field (via our headphone system) are nearly identical to the apparent directions of actual free-field sources.

This article describes two experiments that exploit the stimulus control offered by the headphone system in order to address questions about the relative salience of interaural time differences and interaural intensity differences (pinna effects included) as cues for sound localization. Both experiments ask listeners to localized sounds that are synthesized in such a way that the two cues should provide conflicting information. The first experiment concerns localization of wideband stimuli,and the second concerns localization of high-pass-filtered stimuli.

I. METHODS

The procedures used in both experiments for stimulus synthesis and for psychophysical measurement are nearly identical to those described in earlier publications (Wightman and Kistler, 1989a,b). Readers are referred to the former articles for details. Those features of the procedures that are common to both of the current experiments are briefly summarized here.

A. Subjects

Nine young adults (three males, six females) served as paid volunteers in these experiments. All had normal hearing, as verified by pure-tone audiometry, with no history of hearing problems of any kind. Four of the subjects also served in the previous study (Wightman and Kistler, 1989b). The other five had no previous experience in psychoacoustical experiments. All the subjects were naive regarding the purposes of the experiments. Eight subjects served in experiment I and six subjects served in experiment II.

B. Stimuli

As in previous experiments, the basic stimulus 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. In addition to any other modifications required by the experiments, the spectrum of the noise was "scrambled" (differently on each trial) by an algorithm that divided the spectrum into critical bands, and assigned a random intensity (uniform distribution, 20-dB range) to the noise spectrum level within each critical band. The stimulus spectrum was randomized, as in our previous study (Wightman and Kistler, 1989b), in order to prevent listeners from learning specific stimulus or transducer characteristics (in free-field conditions, no additional attempt was made to compensate for loudspeaker differences).

The noise stimuli were presented either by loudspeaker (Realistic Minimus 0.7) or by headphones (Sennheiser HD340). The vertical semicircular arc of loudspeakers was identical to that used in the previous study (Wightman and Kistler, 1989b). Stimuli could be presented at any azimuth and at one of six elevations relative to the horizontal plane passing through the subject's ears ($-35°$ to $+54°$ in 18° increments). In the various headphone conditions, stimuli were synthesized by passing a burst of "scrambled-spectrum" Gaussian noise through two digital filters. This produced two stimuli: one for the subject's left ear and the other for the subject's right ear. Each digital filter consisted of two cascaded sections. In the baseline condition, the first stage included the subject's free-field-to-eardrum transfer function (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 [Eq. (4) from Wightman and Kistler, 1989a]. The HRTF and headphone transfer functions were estimated using techniques described previously (Wightman and Kistler, 1989a). In experimental conditions, the phase characteristics of the first-stage digital filter were manipulated to produce stimuli in which the interaural time and intensity cues conflicted. The second stage of the digital filter in all conditions was a 10th-order (1024 taps) zero-phase FIR bandpass filter (0.21–14 kHz, except in experiment II) that was used to eliminate processing artifact at low and high frequencies.

Stimuli for a given subject and a given run were precomputed (using Signal Technology, Inc.'s ILS software on a DEC Vaxstation 3500) 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.

C. Procedure

We measured apparent source direction with an "absolute judgment" procedure (Wightman and Kistler, 1989b), in which the listener indicated the apparent direction of a sound source by calling out numerical estimates of apparent azimuth and elevation, using standard spherical coordinates. For example, a sound heard directly in front would produce a response "0,0," and a sound heard on the right and slightly elevated would produce "+ 90, + 10." No feedback was given. All the subjects we tested were able to produce consistent judgments in this paradigm after several hours of practice (Wightman and Kistler, 1989b). In the experiments reported here, subjects were given 5 h of experience in the free-field listening condition before any data were recorded.

The main purpose of these experiments was to evaluate the effects of *changes* in interaural stimulus parameters on the apparent directions of sounds presented over headphones. Interpretation of the results would be facilitated if, with naturally occurring interaural difference parameters, the headphone presented sounds were perceptually equivalent to free-field presented sounds. This equivalence was the subject of a previous experiment (Wightman and Kistler, 1989b) which measured the azimuth and elevation components of the sound images. The data in that paper suggested that the free-field and headphone-presented stimuli produced similar sound images, since apparent azimuth and elevation were nearly the same for the two kinds of stimuli. Thus, for the four subjects who participated in that earlier experiment, "equivalence" (to the extent it is indicated by apparent azimuth and elevation) will be assumed. However, we felt it was important to verify the equivalence of free-field and headphone stimuli for those subjects in the current experiments who had not participated in the previous study. For this purpose, prior to undertaking the current experiments, we tested the five new subjects in the same free-field and simulated free-field (headphone) conditions as were used in the previous experiment.

The free-field condition (described in detail in Wightman and Kistler, 1989b) required subjects to estimate the apparent directions of sounds delivered from 36 different directions, selected pseudorandomly to cover a 360° range of azimuths with elevations from 36° below the horizontal plane to 54° above it. Subjects were blindfolded and seated at the center of the loudspeaker arc inside an anechoic chamber. Head movements were discouraged, but the head was not restrained. After each stimulus was presented, the subject called out azimuth and elevation estimates, and the experimenter recorded the responses. No feedback was given to the subjects. On a given run, subjects heard a stimulus from each of the 36 directions once; the order on each run was random. Each 36-trial run lasted about 15 min, and breaks of about 5 min were taken after each run.

The procedure in all the headphone conditions, including the baseline condition run only on the new subjects, was nearly identical to that used in 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. The trial sequence was the same as for the free-field condition. After each stimulus was presented, the subject called out azimuth and elevation estimates. The experimenter, who was in an adjoining room listening over an intercom, entered the responses on a PC keyboard.

Pilot testing revealed that responses in many of the conflicting cue conditions would be concentrated at a single azimuth and elevation, suggesting that the apparent directions of the stimuli were very similar. In order to compensate for the response biases that might accompany presentation of a large number of similar stimuli, the conflicting cue stimuli were intermingled with stimuli from other experiments that did not produce concentrated response patterns. The intermingling was pseudo-random, with nominally equal *a priori* probabilities of a conflicting cue stimulus or another stimulus on each trial. The constraint was that on each run exactly 36 conflicting cue stimuli and 36 other stimuli would be presented. Thus the runs in the experimental conditions required 72 direction estimates, only 36 of which were relevant to the work reported here. About six baseline or three experimental runs could typically be completed in each 90-min session.

Each of the subjects completed eight runs in the free-field condition and then, after the required acoustical HRTF measurements had been made, completed eight runs in the baseline headphone condition. In this condition, described in our earlier paper (Wightman and Kistler, 1989b), stimuli were synthesized with the aim of preserving all of the acoustical cues for sound localization that are present with free-field stimuli. The results from the free-field and baseline headphone conditions will be presented along with the results from experiment I.

D. Data analysis

The eight azimuth and elevation responses by each listener to each stimulus can be thought of as representing eight points on the surface of a sphere, with the listener's head at the center. For such spherically organized data, the usual reduction techniques (e.g., computations of means and standard deviations) are inappropriate. One obvious problem is that an azimuthal error of 30° at 0° elevation is larger in an absolute distance sense than the same error at 50° elevation. We address some of the problems by borrowing techniques from the field of spherical statistics (Fisher *et al.*, 1987). For our measure of central tendency we compute the "centroid" (a kind of average direction) of the judgments, and for our measure of dispersion we use the spherical statistic κ^{-1} (Wightman and Kistler, 1989b).

Data reduction is further complicated by the fact that in many instances the judgments are not unimodally distributed, because of front–back "confusions." Front–back confusions are common in studies of sound localization (see Makous and Middlebrooks, 1990, and Wightman and Kistler, 1989b for recent data) and are most often manifest by occasional reports of an apparent rear hemisphere direction for a stimulus usually assigned a front hemisphere direction. Thus a source at 30° azimuth may product two clusters of responses, one near 30° azimuth, and another near 150° azimuth. Measures of central tendency and dispersion computed from these data could be misleading.

Since front–back confusions are typically infrequent and occur most often for sources near the median plane, it has been our practice to "resolve" them by reflecting the azimuths of the confusion judgments about 90° (Wightman and Kistler, 1989b). In the current study, however, some stimulus conditions might be expected to produce multiple images or patterns of responses in which it would be impossible to distinguish between front–back confusions and simple error variance (e.g., concentrations of azimuth judgments near 90°). For these reasons, routine resolution of front–back confusions is inappropriate. Since high rates of front–back confusions greatly complicate interpretation of measures of central tendency and dispersion, the most conservative approach would be to abandon those statistics and display all the raw data. However, we find that the raw data displays can also be misleading because of the ambiguities sometimes encountered when visualizing large multidimensional data sets. For the purposes of this article, we feel a mixed strategy is best. Thus we routinely compute and display centroids (and, in some cases κ^{-1}) without resolving the front–back confusions, and display raw data where necessary to emphasize a point. There are a few conditions in which the front–back confusion rate is high for a small number of stimuli. Because of this, the judgment centroids for these stimuli indicate an inappropriate central tendency, and the dispersion of the judgments is greatly overestimated. These cases will be highlighted as necessary.

We normally analyze separately the azimuth and elevation components of the judgment centroids. As a result, our typical data figures show the azimuth (or elevation) judgment centroids as a function of the azimuth (or elevation) of the target and do not distinguish between responses made to stimuli at different target elevations (or azimuths). This is done for reasons of convenience and is not meant to imply that we view the two dimensions as independent. In most cases, however, there is no systematic relation between the azimuth and elevation components of a judgment, between reported azimuth and target elevation, or between reported elevation and target azimuth. Thus the reported elevations of sources at any one azimuth are generally not systematically biased up or down relative to the target elevation. Likewise, the reported azimuths of sources at particular elevations are not systematically biased toward the front or rear. There are a few important exceptions in the results described in this article, and these are mentioned where appropriate.

II. EXPERIMENT I: OPPOSITION OF INTERAURAL TIME AND INTENSITY CUES

In this experiment, we seek to assess the relative salience of interaural time and intensity difference cues by measuring the apparent directions of stimuli in which the two cues give conflicting information regarding source direction. These stimuli are, of course, "unnatural" in that no stimulus presented in a real acoustic environment would produce such a combination of cues. It is difficult to predict how this lack of realism might contaminate the data. The variance of the judgments might be inflated, or the distribution of judgments might reflect the presence of more than one sound image. We choose to ignore these potential problems until they are manifested in the data. Thus our working assumption is that the extent to which the judgments are concentrated at the location signaled by one cue, by the other cue, or by some mixture of the cues, provides evidence about how the cues are weighted and combined in the auditory system.

A. Stimuli

The stimuli for this experiment were trains of eight 250-ms bursts of scrambled-spectrum Gaussian noise, as described above. Prior to their presentation over headphones the stimuli were digitally filtered, as in our previous work (Wightman and Kistler, 1989a), to simulate 36 different free-field source positions. In the current experiment, however, the interaural time cues naturally present in the stimuli were altered by manipulating the phase components of the digital filters. The other cues, such as the interaural intensity differences at each frequency and the spectral shape cues at each ear, were unaffected by this manipulation. In one condition, "ITD:0," the phase components of the left and right digital filters for each of the 36 source positions were replaced with the left–right average phase at each position. This created a set of stimuli in which the ITDs at each of the 36 target positions were zero but the interaural intensity and spectral shape cues were normal. One could say that for these stimuli the interaural time cues "point" consistently to a median plane source direction (azimuth: 0° or 180°, elevation: any value), while the interaural intensity and spectral shape cues point to the nominal direction of the target on each trial. In a second condition, "ITD: − 45," the phase functions for the left and right filters at each source position were replaced with the left- and right-phase functions extracted from the filters used for the single position − 45,0 (− 45° azimuth, or 45° left of straight ahead, 0° elevation, or on the horizontal plane). The interaural time difference at − 45,0 is roughly the same as that produced by all sources on the circle centered at − 90,0 that intersects − 45,0 and lies parallel to the median plane (Middlebrooks and Green, 1990). Thus we might say that in this condition the interaural time cues point to any of the directions indicated by this circle. As in the ITD:0 condition, the interaural intensity and spectral shape cues for these stimuli were "normal," or appropriate to each of the 36 target positions. In the third condition, "ITD:90," the single left–right pair of phase functions used for all 36 filter pairs was extracted from the filters used for position 90,0 (90° right from straight ahead, and on the horizontal plane).

To verify that the phase manipulations had the desired effect on the ITD cues, we analyzed representative stimuli from each of the three conditions. For this analysis, we passed both left-ear and right-ear stimuli through a bank of zero-phase, critical-bandwidth digital filters and computed left–right cross-correlation functions from the envelopes of the waveforms in each critical band. The time delay at the maximum in the cross-correlation function was taken as the "effective" ITD cue at the critical band center frequency (Middlebrooks and Green, 1990). Ten scrambled-spectrum noise bursts from each of six source positions, each of the three phase conditions, and all eight subjects were analyzed in this way.

Figures 1 and 2 show the results of our analysis of stimuli synthesized for two representative subjects and the three experimental conditions of this experiment. Note that in the

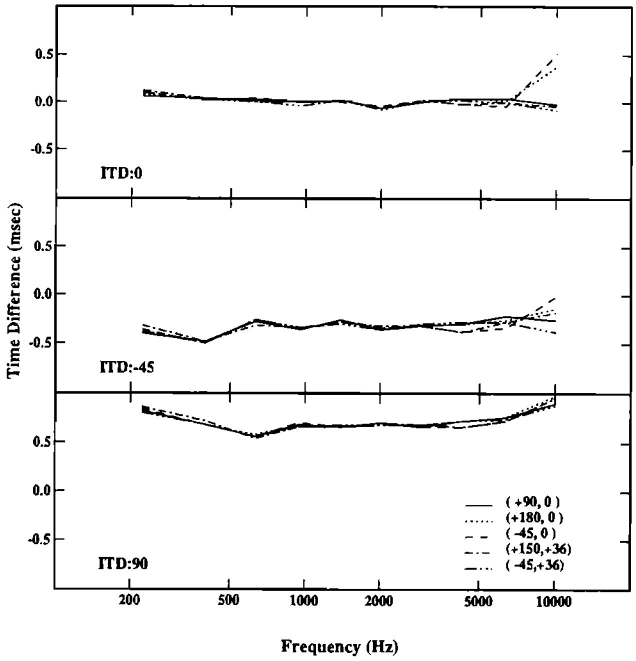

FIG. 1. Interaural time differences in sample stimuli synthesized for subject SDO in the three conflicting cue conditions of experiment I. Each curve shows the average (across 10 stimuli) interaural time difference within a critical band as a function of critical band center frequency. Interaural time difference was estimated by crosscorrelating the envelopes of critical band filtered left-ear and right-ear stimuli. The different line types indicate the five different source positions, listed on the bottom panel, which the stimuli were synthesized to simulate.

ITD:0 condition (top panels of Figs. 1 and 2), the effective ITD cue is close to zero. The departures from zero at low and high frequencies are most likely a result of measurement error. At low frequencies, the cross-correlation functions have very broad peaks, making determination of the maximum somewhat uncertain. At high frequencies, the multiplicity of peaks creates a similar uncertainty. In the ITD:90 condition (bottom panels) the effective cue is around 700 μs at each frequency. In the ITD: $-$ 45 condition, the effective cue is negative, appropriate to the $-$ 45° azimuth, and intermediate between the ITD:0 and ITD:90 cues. Moreover, the cue is larger at low frequencies in the ITD: $-$ 45 condition, reflecting the dependence of ITD on frequency, which is greatest at this intermediate azimuth (Kuhn, 1979). The results of this analysis suggest that the phase modifications had exactly the effect we anticipated, namely fixing the ITD cue.

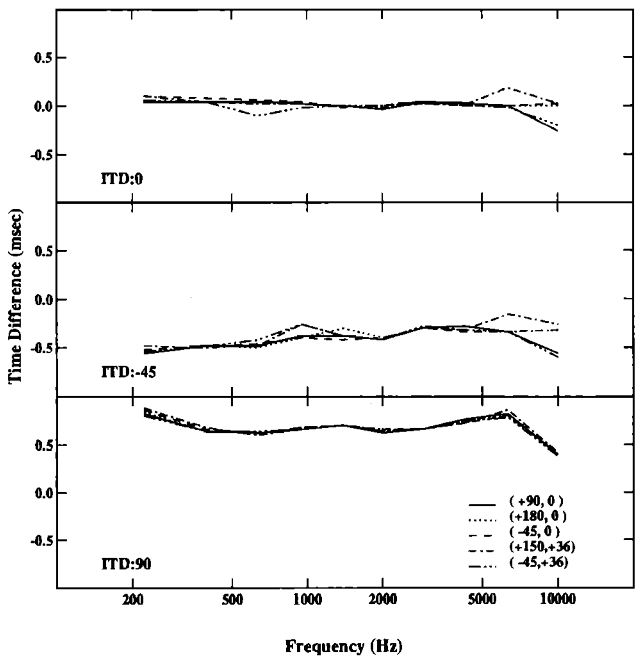

FIG. 2. Same as Fig. 1, but for subject SGG.

B. Results

We might expect that in conditions involving conflicting cues listeners would produce more variable judgments than in conditions in which the cues were normal. The conflicting cues might be expected to produce multiple images, to which listeners might respond differently on different trials. Or, the conflicting cues might produce an auditory image that is more diffuse and thus more difficult to localize consistently. There is little indication in the data that either occurred in this experiment.

Our measure of the dispersion of the apparent direction judgments is κ^{-1} an index that is analogous to standard deviation. An informal analysis of the values of κ^{-1} in this experiment suggests that the variability is equal in experimental and baseline conditions. For two of the eight subjects, the values of κ^{-1} at 9 of the 36 directions are higher in one of the conflicting cue conditions (a different condition for each of the two subjects) than in the baseline (all cues normal) condition. However, the judgments with higher κ^{-1} are not confined to target directions where the two cues are most discrepant. For the other six subjects, there is no difference in κ^{-1} between baseline and experimental conditions. This suggests that listeners responded to the conflicting cue stimuli in a generally consistent way, and that if multiple or diffuse images were produced, these effects were not reflected in the judgments. We will return to the issue of variability in connection with discussion of specific results.

The results from the five new subjects in the free-field and baseline conditions are indistinguishable from the results of the four subjects who served in our earlier study (Wightman and Kistler, 1989b). To quantify the similarity of the two groups, we computed the same summary statistics as before (Wightman and Kistler, 1989b: Table II, p. 872). For this comparison only, front–back confusions were resolved. The "goodness of fit" between the matrix of target directions and the matrix of judgment centroids (Lingoes and Schonemann, 1974), the correlations between target and centroid azimuth and between target and centroid elevation, and the percentages of front–back confusions were comparable in the two groups in both the free-field and baseline conditions. This suggests, as before, that the apparent directions of stimuli in the free-field and baseline headphone conditions are quite similar.

Figure 3 shows data from the baseline headphone condition and from the three conflicting cue conditions (ITD:0,

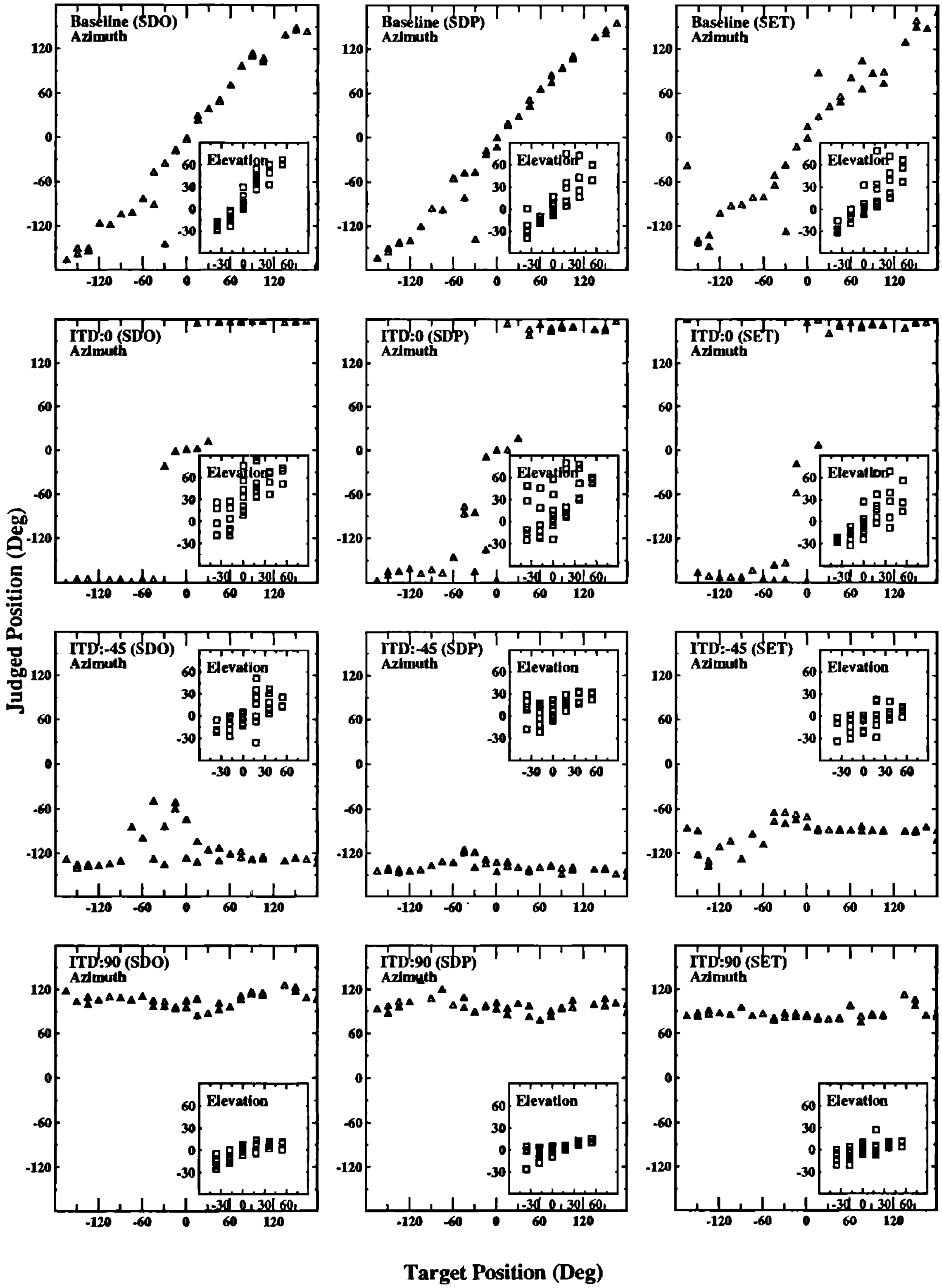

FIG. 3. Scatterplots of target direction versus judged direction for the baseline headphone condition and the three experimental conditions of experiment I. Each data point represents either the azimuth (main part of each panel) or the elevation (insets) component of the centroid of eight judgments. Thirty-six centroids, corresponding to the 36 stimulus directions, are plotted in each panel. Data from three representative subjects are shown.

ITD: − 45, and ITD:90). Results produced by three representative subjects, two from our previous study and one of the new subjects, are shown in this figure. The results from the other five subjects who participated in this experiment are quite similar to those shown in Fig. 3 and, in order to conserve space, are not shown here. Each data point represents the azimuth and elevation component of the centroid (average direction) of eight judgments. Note that the azimuth and elevation components of the centroids are plotted separately. In each case, the abscissa represents the nominal direction of the target, and the ordinate represents the centroid of the judgments. The several points plotted at each target azimuth (or elevation) reflect the fact that in the set of 36 stimuli there were stimuli presented at several elevations (or azimuths) for each azimuth (or elevation) represented.

As indicated above, the data from the baseline condition (top panels of Fig. 3) are nearly identical to the data from the free-field condition, a result consistent with our previous finding (Wightman and Kistler, 1989b). Fixing the ITD cue at zero (ITD:0 data) has a dramatic effect on the azimuth components of the judgments. Note that all but a few of the azimuth judgments are clustered either near 0° (directly ahead) or near 180° (directly behind), regardless of the actual target azimuth. Note also that the elevation components of the judgments are relatively unaffected by the phase manipulation. In other words, while the apparent azimuth of all stimuli is shifted to 0° or 180°, the apparent elevation of most of the stimuli is near normal. This pattern of results is consistent with the hypothesis that apparent direction is cued primarily by interaural time difference. A zero ITD cue signals a source anywhere on the median plane: at azimuths of 0° or 180° and at any elevation.

Fixing the ITD cue at a large value (ITD:90) also has a dramatic effect on the judgments. Recall that in this condition the interaural time cues in all 36 stimuli are set to those interaural time cues in the stimulus at position + 90,0. The intensity parameters of the stimuli (interaural differences and spectral shapes) are unaffected by this manipulation and thus are appropriate to each of the 36 source positions. The azimuth components of the judgments from the ITD:90 condition are consistent with those from the ITD:0 condition; apparent azimuth seems to be completely determined by interaural time difference. In this condition judgments of the apparent azimuths of all 36 sources are clustered around a single azimuth, which for some subjects is almost exactly 90° and for others is closer to 120°. In other words, even when the nominal stimulus direction is on the opposite side of the head (an azimuth of − 90° for example), and large interaural intensity differences are produced that signal the opposite side, the apparent azimuth is still determined by interaural time difference. The data are unambiguous on this point; in all of the eight runs from eight subjects in this condition there was not a single negative azimuth judgment indicating a perceived source on the opposite side of the head.

The pattern of elevation judgments is quite different in the ITD:90 condition than in the baseline or ITD:0 conditions. Note that for all subjects, the range of elevation judgments in the ITD:90 condition is much smaller than in the baseline or ITD:0 conditions. In particular, stimuli with nominally high elevation are judged to have much lower elevation. This result is wholly consistent with the hypothesis that apparent direction is determined by interaural time difference. In the ITD:90 condition, the only source elevation that is consistent with the large interaural time cue is 0. At an azimuth of 90°, sources at elevations higher or lower than 0° would produce smaller interaural time differences. In the ITD:0 condition there is no compression of elevation judgments since all source elevations are consistent with the zero interaural time difference cue.

The results from the ITD: − 45 condition are compatible with the results from the other two experimental conditions. The azimuth judgments are concentrated around one or two values (close to − 45 and − 135) and, regardless of the interaural intensity and spectral shape cues, there are no azimuth judgments indicating a source on the opposite side of the head. For most subjects the pattern of elevation judgments is intermediate between those obtained in the ITD:0 and ITD:90 conditions. While the range of elevation judgments is compressed in this condition compared to that in the baseline and ITD:0 conditions, it is not compressed nearly as much as in the ITD:90 condition. This is expected, since the intermediate ITD cue in the ITD: − 45 condition could signal a source anywhere on the "circle of confusion" (contour of constant ITD) that intersects the horizontal plane at − 45,0. The results from the ITD: − 45 condition thus add support to our previous suggestions about the dominance of the ITD cue with wideband sources.

The major results of this experiment are summarized in Figs. 4 and 5. Figure 4 shows the distributions of all the azimuth judgments from the eight subjects in this experiment. Note that in most cases the azimuth judgments are concentrated around a single value, roughly that expected given the ITD cue. Moreover, there are no azimuth judgments on the side opposite to that signaled by the ITD cue in the ITD:45 and ITD:90 conditions. Figure 5 shows the mean slopes of straight lines fit to the raw elevation data (not centroids) in each condition. This figure documents our claim that the range of elevation judgments is compressed (i.e., lower slope) in the ITD: − 45 and ITD:90 conditions, as would be expected if interaural time difference were the dominant cue.

While not pertinent to the main thrust of this work, there are two additional features of the data from experiment I that are noteworthy. First, there is the tendency of nearly all subjects in the ITD:0 and ITD: − 45 conditions (Fig. 4) to report apparent directions in the rear (i.e., near 180° of − 180° in the ITD:0 condition and near − 135° in the ITD: − 45 condition). Second, some subjects (e.g., SDO and SDP), particularly in the ITD:0 condition (Fig. 3), consistently report elevations much higher than that of the target for a few stimuli. Although not indicated in the figure, the "over-elevated" stimuli are those with target azimuths in the frontal quadrant. Both of these features are representative of a kind of response bias that is exhibited by many of our listeners when they are presented stimuli that (by design or by accident) incorporate less than optimal localization cues. The reported apparent directions of these stimuli are often displaced toward regions where there is normally greater

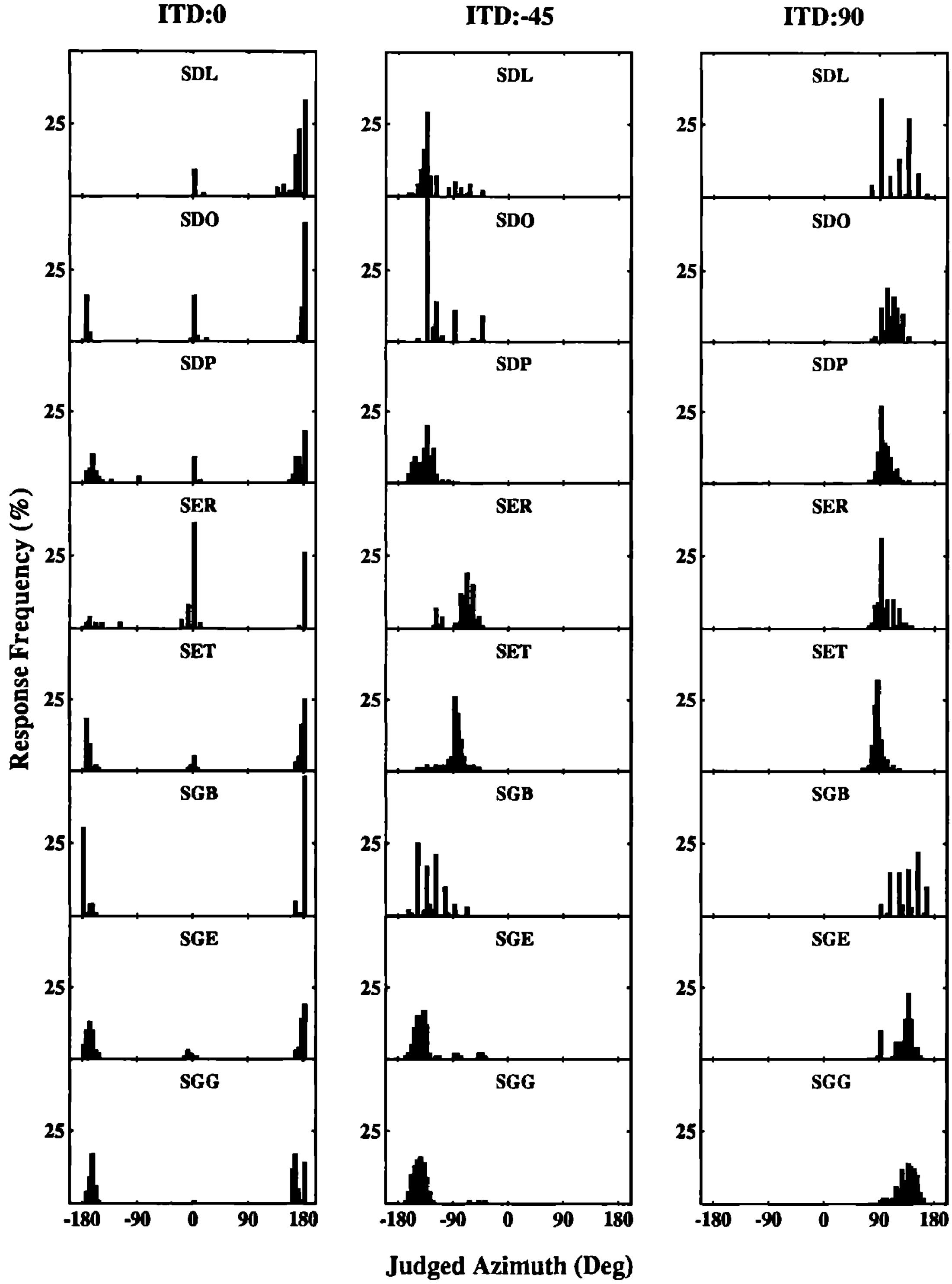

FIG. 4. Distributions of the azimuth components of apparent direction judgments (raw data, not centroids) from the eight subjects who participated in experiment I. The experimental condition is indicated at the top of each column of histograms.

uncertainty regarding source direction (i.e., above and behind).

III. EXPERIMENT II: CONFLICTING CUES IN BANDLIMITED STIMULI

The duplex theory of sound localization (Strutt, 1907) argues that interaural time difference is an important cue only at low frequencies. One rationale for this argument derives from the fact that with sinusoidal stimuli, the interaural time (phase) difference cue becomes ambiguous above about 1500 Hz, where the wavelengths are such that time leads and lags cannot be distinguished. However, many studies have shown that interaural time differences in stimulus *envelopes*, which do not suffer the ambiguity problem, can be

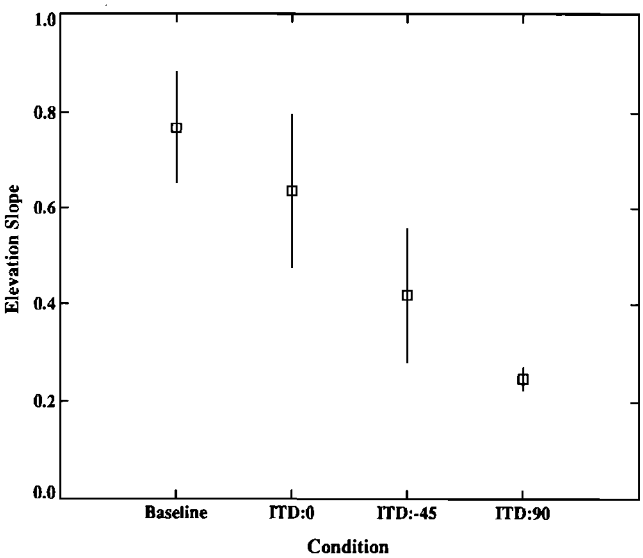

FIG. 5. Mean slopes (error bars indicate plus/minus two standard errors) of straight lines fit to the elevation scatterplots (e.g., those shown in Fig. 3) from experiment I. Elevation slopes from all eight subjects are represented in this figure.

detected at much higher frequencies (e.g., Henning, 1974; McFadden and Pasanen, 1976; McFadden and Moffitt, 1977). Therefore, it is of considerable interest to determine the extent to which the dominance of the ITD cue that we observed in experiment I is contingent on contributions from low or high frequencies. Experiment II examines this issue by asking subjects to localize conflicting cue stimuli that are identical to those used in experiment I, except that in experiment II they are high-pass filtered. All aspects of the psychophysical procedures are the same as in experiment I, and five of the eight subjects from experiment I participated in experiment II. A sixth subject was recruited only for this experiment.

A. Stimuli

The stimuli for this experiment were produced in virtually the same way as those used in the ITD:90 condition of experiment I. They were wideband scrambled-spectrum noise bursts which contained conflicting cues to sound location. The interaural intensity and spectral shape cues signaled the normal target direction for all 36 sources in a stimulus set, while the ITD cue always signaled a source at 90,0. The only difference in experiment II was that the high-pass cutoff of the final filter, previously a 0.2- to 14-kHz zero-phase bandpass filter, was raised, thus producing high-pass-filtered versions of the stimuli. Four conditions were studied, corresponding to four high-pass cutoff frequencies: 5, 2.5, 1.0, and 0.5 kHz.

B. Results

Figures 6 and 7 show the results from all six subjects in the four conditions of experiment II. Individual differences in performance were much greater in this experiment than in experiment I, so all the data are displayed. The format of the data display in Figs. 6 and 7 is the same as in Fig. 3; the azimuth and elevation components of the centroid direction judgments are plotted separately in each panel, and the data from different subjects and conditions appear in different panels.

The most revealing result from this experiment is that the data from the 5-kHz high-pass conflicting cue condition (top panels) are similar to the data from the baseline conditions of experiment I (Fig. 3). There are two notable exceptions. The azimuth centroids for frontal sources from two of the subjects (SGG and SGE, in Fig. 7) are shifted rearward by a large number of front–back confusions. In all other cases, apparent source direction corresponds closely to the direction signaled by the interaural intensity and spectral shape cues. Thus, with low frequencies removed from the stimuli, fixing the ITD cue had little effect. It appears that with stimuli containing only high frequencies, both azimuth and elevation are faithfully cued by interaural intensity and spectral shape cues, and the interaural time differences present in the envelope of the stimuli at high frequencies are ineffective cues. This result contrasts sharply with the result from the ITD:90 condition of experiment I in which wideband stimuli were presented. In that condition the ITD cue clearly dominated both azimuth and elevation perception.

The data from the remaining three conditions of experiment II (2.5, 1.0, and 0.5-kHz high-pass cutoff frequencies) show the expected trend. As lower and lower frequencies are added to the stimuli, the dominance of the interaural time cue is revealed. For one subject (SDL), the dominance is nearly complete in the 2.5-kHz high-pass condition in that the pattern of judgments is similar to that observed in the ITD:90 condition of experiment I. This suggests that for SDL salient interaural time information comes from the 2.5- to 5-kHz region. For another subject (SDP), however, it does not appear that the interaural time cue is completely dominant even in the 0.5-kHz high-pass condition. For that subject, frequencies as low as 0.2–0.5 kHz appear to carry important interaural time cues.

Figure 8 shows the mean slopes of straight lines fitted to the elevation data from this experiment. The data plotted at the 200-Hz cutoff frequency are those from the ITD:90 condition of experiment I. Note that as the cutoff frequency is raised the slope of the elevation data approaches 0.8, the slope in the baseline condition. We argue that slopes lower than 0.8 reflect the constraints imposed on apparent elevation by the large, dominant low-frequency interaural time cue.

IV. DISCUSSION

We feel that the data shown here constitute strong evidence in favor of the hypothesis that low-frequency interaural time differences are dominant cues for localization of broadband sound sources. In all conditions in which low-frequency interaural time cues conflict with other cues, apparent sound source direction is determined largely by the interaural time cue. The clearest result is that when a large ITD in the low frequencies signals a source far to the right and close to the horizontal plane (ITD:90 condition), all judgments of apparent source direction cluster around these values. This occurs in spite of the presence in many stimuli of

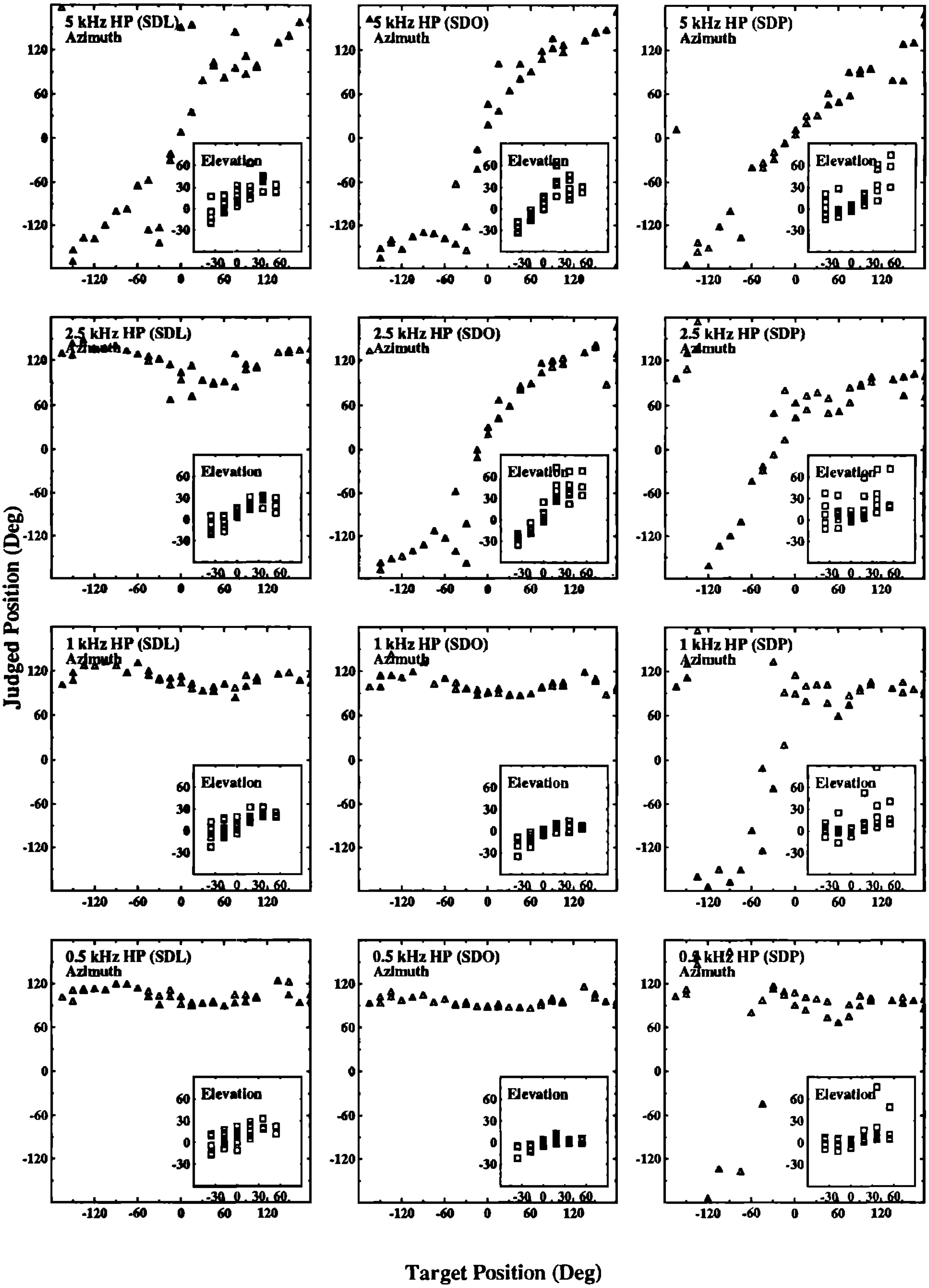

FIG. 6. Scatterplots showing the data obtained from three of the six subjects in the four conditions of experiment II. The format of each panel and the meaning of each data point are the same as in Fig. 3.

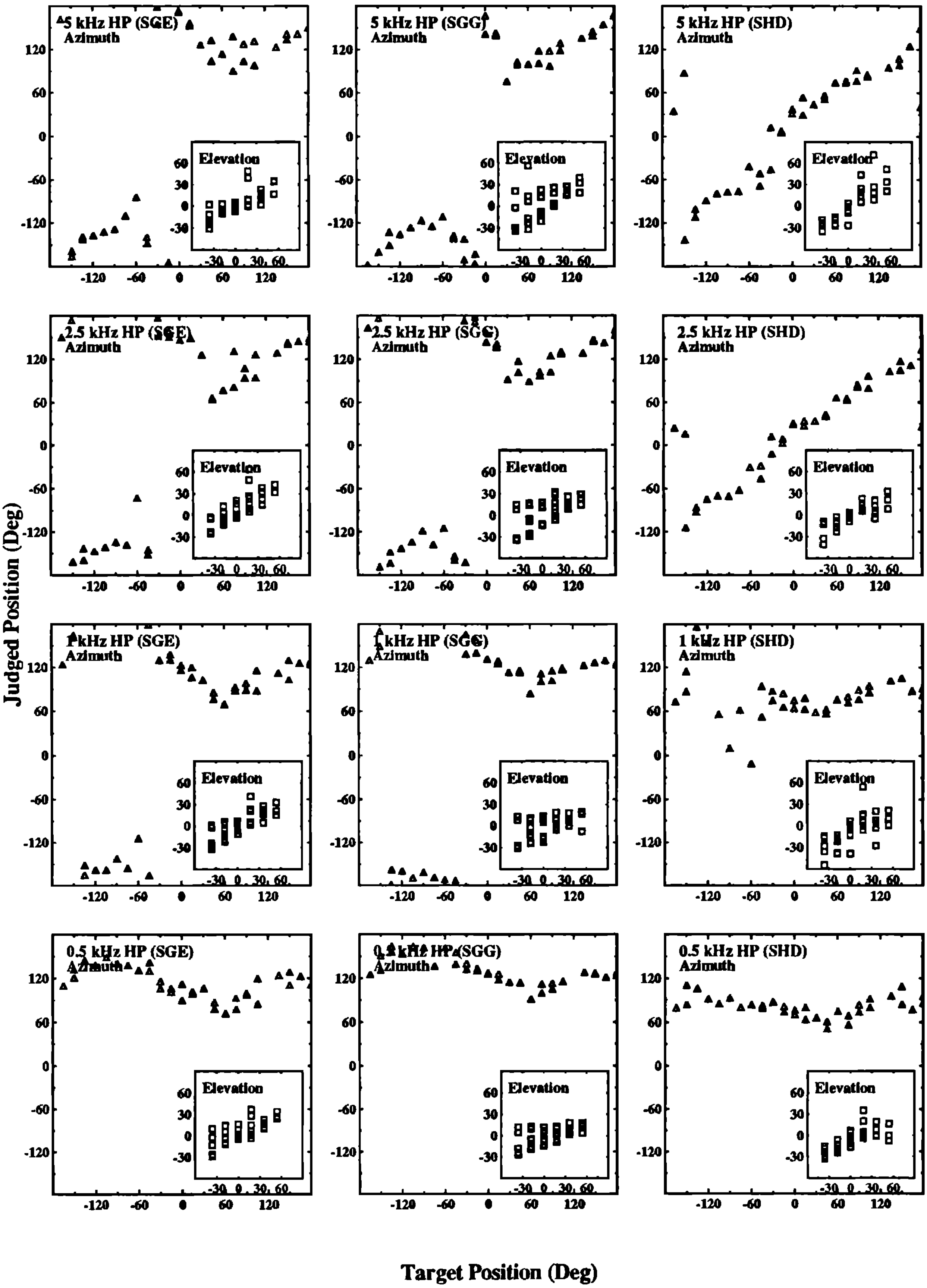

FIG. 7. Same as Fig. 5, except this figure shows the data from the remaining three subjects in experiment II.

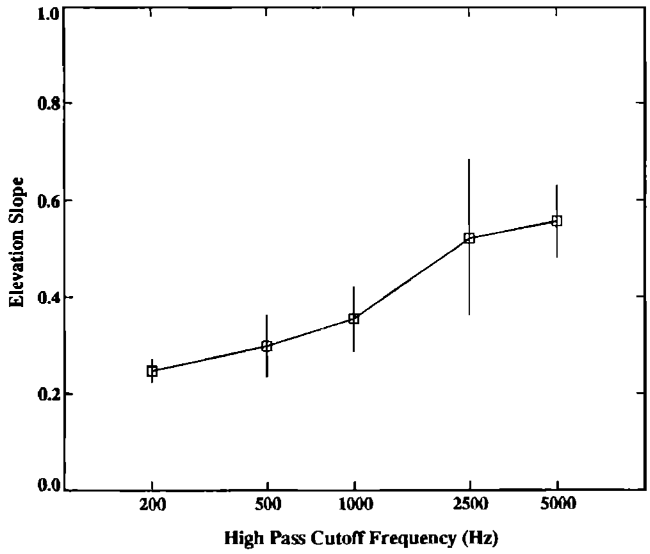

FIG. 8. Mean slopes (error bars indicate plus/minus two standard errors) of straight lines fit to the elevation scatterplots obtained in experiment II (those in Figs. 6 and 7). The data at the 200-Hz cutoff frequency are those from the ITD:90 condition of experiment I, also shown in Fig. 5.

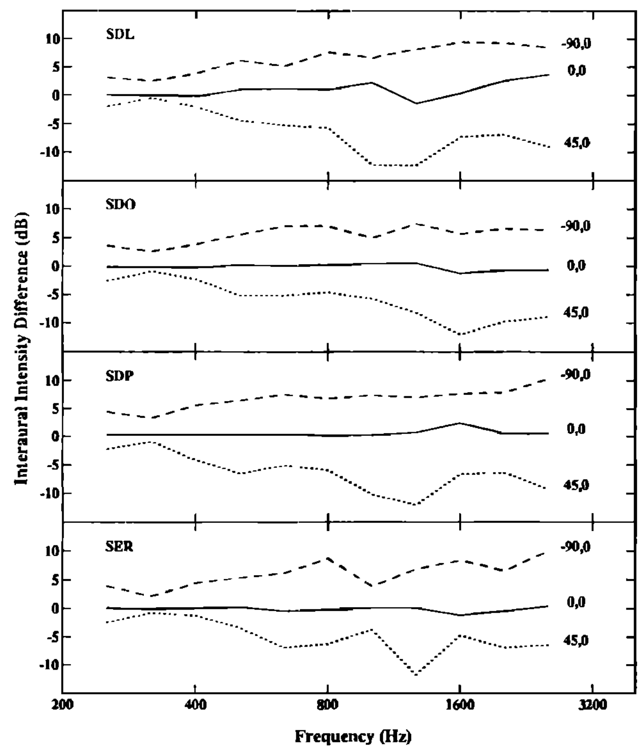

FIG. 9. Measured interaural intensity differences in 1/3-octave bands for stimuli at three different azimuths on the horizontal plane. Data from four of the eight subjects in this experiment are shown here. The data from the remaining four subjects are comparable. The source positions are indicated at the side of each line. Note that only data from the low-frequency bands (center frequencies less than 2.5 kHz) are shown.

large interaural intensity and spectral shape cues that indicate source directions on the opposite side and at higher or lower elevations. Similarly, when the low-frequency interaural time difference is zero (ITD:0 conditions), signaling a source on the median plane, all judgments of apparent source direction are close to the median plane even though in many stimuli the interaural intensity and spectral shape cues signal directions far to the right or left. When low frequencies are removed from the stimuli (e.g., in the 5-kHz highpass condition) the ITD cue is less salient, and apparent direction is determined by the interaural intensity and spectral shape cues.

It is important to note that our results argue for the dominance of low-frequency *time cues* not just the dominance of low frequencies. While the interaural intensity differences in free-field (and simulated free-field) stimuli are generally smaller at low frequencies than at high frequencies, they are not zero, and they vary in systematic ways with the direction of a stimulus. Thus both interaural time and interaural intensity differences could serve as potential localization cues at low frequencies. In our ITD:90 condition, for example, nearly half of the stimuli contain interaural intensity differences *at both low and high frequencies* that signal directions on the opposite side. Figure 9 shows the interaural intensity differences in low-frequency 1/3-octave bands measured at three source directions from four subjects in the current experiments. These interaural intensity values agree with those measured by Shaw (1974) and others. Note that even at 500 Hz, a source at $-90,0$ produces a 5-dB interaural intensity difference that signals a direction directly opposite that signaled by the interaural time difference in the ITD:90 condition. With a pure-tone stimulus, a 5-dB interaural intensity difference at 500 Hz is sufficient to move the apparent position of a lateralized source about half way to the side (Yost and Hafter, 1987; Fig. 3.2). Thus, at least in some conditions, a 5-dB interaural intensity difference at 500 Hz must be considered a salient cue. A source 45° to the side produces interaural intensity differences as great as 8 dB (Fig. 9) in the low frequencies. Moreover, it must be remembered that interaural intensity differences close to 0 dB, which occur in all low-frequency bands for stimuli presented on the median plane (Fig. 9), are no less legitimate as direction cues than interaural intensity differences of 5 or 10 dB. A zero interaural intensity difference signals a source on the median plane. The fact that low-frequency interaural time differences appear to override all of the low-frequency (and high-frequency) interaural intensity and spectral shape cues must be taken as evidence supporting our conclusion that low-frequency time cues are dominant.

In the experiments reported here interaural time and intensity cues were independently manipulated in stimulus conditions that otherwise closely simulated free-field localization conditions (e.g., pinna cues were present, and the naturally occurring frequency dependence of all cues was maintained). Thus these experiments are quite different than most lateralization experiments, in which less attention is given to mimicking realistic localization situations. For this reason, the generalizability of most lateralization data to the conditions that we studied may be somewhat limited. That class of lateralization experiments in which listeners judged the extent of laterality of wideband stimuli with independently controlled interaural time and intensity parameters seems most comparable to our own work (e.g., Blauert, 1982, 1983; Trahiotis and Bernstein, 1986). Our results are compatible with some of the data from these experiments and incompatible with others.

In both the Blauert (1982, 1983) and Trahiotis and Bernstein (1986) studies, stimuli were presented with a fixed interaural time difference and frequency-independent interaural intensity difference of 0 dB. Thus, as in our work, the stimuli contained conflicting cues, with interaural intensity signaling "middle" and interaural time signaling "side". Blauert (1982) reports that the apparent image position of a high-pass noise stimulus with a 600 μs interaural time difference was determined by the time difference (i.e., was far to one side) as long as frequencies below about 2 kHz were present. Otherwise, the apparent image position was at the center—the position signaled by the interaural intensity cue. Blauert reported comparable results with critical-band-filtered noises (1983, p. 319), as long as the center frequency of the noise band was below about 2 kHz. Trahiotis and Bernstein (1986), using bandpass noise stimuli with interaural time differences of 250 or 500 μs, report that with most bandwidths and most listeners, the apparent image position was governed by the time difference at low frequencies and the intensity difference at high frequencies. This result, as well as Blauert's results, are qualitatively consistent with our own. However, Trahiotis and Bernstein (1986) report that for two of their four subjects, interaural time difference determines the apparent image position even at high frequencies (3.8 kHz). While between-subjects differences are present in our data as well, we have no evidence that interaural time differences at frequencies as high as 3.8 kHz can override the interaural intensity and spectral shape cues at those frequencies. In fact, for five of our six subjects, it appeared that only the interaural time differences below 2.5 kHz were dominant.

V. CONCLUSION

The importance of interaural time differences for sound localization is one of the fundamental tenets of the classic duplex theory. The theory also argues that interaural time cues are effective only at low frequencies. The results of the experiments reported here clearly support both claims of the duplex theory and suggest further that when low-frequency interaural time cues are present, they override the interaural intensity and spectral shape cues present in both low and high frequencies, and thus provide the dominant cue. With wideband stimuli, interaural intensity and spectral shape cues appear to play a secondary role in cueing apparent source azimuth and elevation. While considerably more data are needed before any firm conclusions can be drawn, the results of the current experiments suggest that the interaural time cue may be used primarily to establish the locus of possible source directions. Interaural intensity and spectral cues are they analyzed to resolve confusions. Given the geometry of the head and ears, a specific interaural time difference by itself signals a range of potential source directions (the "cone of confusion"), and thus cannot unambiguously cue any one azimuth or elevation (Middlebrooks and Green, 1990). Integration of all the available cues leads to accurate localization. This concept has been explicitly embodied, in a model of localization (Searle *et al.*, 1976), but the model does not disentangle the relative importance of interaural time and intensity differences. Our work makes it quite clear that in many circumstances (e.g., with wideband stimuli) interaural time differences make a much more significant contribution to apparent source direction than interaural intensity differences.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the assistance of Ms. Marianne Arruda and Mr. Douglas Swiggum in the technical phases of the work, and the many helpful comments from Dr. John Middlebrooks, Dr. David Green, and Dr. Ervin Hafter on earlier versions of this paper. Financial support was provided by NASA (Cooperative Agreement NCC2-542) and the NIH-NIDCD (DC00116).

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

Blauert, J. (1983). "Binaural localization: Multiple images and applications in room and electroacoustics," in *Localization of Sound: Theory and Applications*, edited by W. Gatehouse (Amphora, Groton, CT), Chap. 5, pp. 65–84.

Butler, R. A., Humanski, R. A., and Musicant, A. D. (1990). "Binaural and monaural localization of sound in two-dimensional space," Perception **19**, 241–256.

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

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

Hartmann, W. M. (1983). "Localization of sound in rooms," J. Acoust. Soc. Am. **74**, 1380–1391.

Henning, G. B. (1974). "Detectability of interaural delay in high-frequency complex waveforms," J. Acoust. Soc. Am. **55**, 84–90.

Kuhn, G. F. (1979). "The pressure transformation from a diffuse sound field to the external ear and to the body and head surface," J. Acoust. Soc. Am. **65**, 991–1000.

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

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

McFadden, D., and Moffitt, C. M. (1977). "Acoustic integration for lateralization at high frequencies," J. Acoust. Soc. Am. **61**, 1604–1608.

McFadden, D., and Pasanen, E. (1976). "Lateralization at high frequencies based on interaural time differences," J. Acoust. Soc. Am. **59**, 634–639.

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

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.

Musicant, A. D., and Butler, R. A. (1985). "Influence of monaural spectral cues on binaural localization," J. Acoust. Soc. Am. **77**, 202–208.

Perrott, D. R., and Musicant, A. D. (1981). "Dynamic minimum audible angle: Binaural spatial acuity with moving sound sources," J. Aud. Res. **21**, 287–295.

Platt, B. B., and Warren, D. H. (1972). "Auditory localization: The importance of eye movements and a textured visual environment," Percept. Psychophys. **12**, 245–248.

Searle, C. L., Braida, L. D., Davis, M. F., and Colburn, H. S. (1976). "Model for auditory localization," J. Acoust. Soc. Am. **60**, 1164–1175.

Shaw, E. A. G. (1974). "Transformation of sound pressure level from the free field to the eardrum in the horizontal plane," J. Acoust. Soc. Am. **56**, 1848–1861.

Shelton, B. R., and Searle, C. L. (1980). "The influence of vision on the absolute identification of sound-source position," Percept. Psychophys. **28**, 589–596.

Strutt, J. W. (1907). "On our perception of sound direction," Philos. Mag. **13**, 214–232.

Thurlow, W. R., Mangels, J. W., and Runge, P. S. (1967). "Head movements during sound localization," J. Acoust. Soc. Am. **42**, 489–493.

Trahiotis, C., and Bernstein, L. (1986). "Lateralization of bands of noise: Effects of spectral locus and bandwidth," J. Acoust. Soc. Am. **79**, 1950–1957.

Yost, W. A., and Hafter, E. R. (1987). "Lateralization," in *Directional Hearing*, edited by W. Yost and G. Gourevitch (Springer-Verlag, New York), Chap. 3, pp. 49–84.

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.