NASA-CR-202118

Annual Progress Report

Period Covered: 5/1/94-4/30/95

NASA Cooperative Research Agreement
NCC2-542

Psychophysical Evaluation of Three-Dimensional Auditory Displays

Frederic L. Wightman, Ph. D., Principal Investigator
Waisman Center, University of Wisconsin
1500 Highland Avenue
Madison, WI 53705

To CAST

Introduction

This report describes the progress made during the first year of a three-year Cooperative Research Agreement (CRA NCC2-542). The CRA proposed a program of applied psychophysical research designed to determine the requirements and limitations of three-dimensional (3-D) auditory display systems. These displays present synthesized stimuli to a pilot or virtual workstation operator that evoke auditory images at predetermined positions in space. The images can be either stationary or moving. In previous years, we completed a number of studies that provided data on listeners' abilities to localize stationary sound sources with 3-D displays. The current focus is on the use of 3-D displays in "natural" listening conditions, which include listeners' head movements, moving sources, multiple sources and "echoic" sources. The results of our research on two of these topics, the role of head movements and the role of echoes and reflections, were reported in the most recent Semi-Annual Progress Report (Appendix A). In the period since the last Progress Report we have been studying a third topic, the localizability of moving sources. The results of this research are described below.

The fidelity of a virtual auditory display is critically dependent on precise measurement of the listener's Head-Related Transfer Functions (HRTFs), which are used to produce the virtual auditory images. We continue to explore methods for improving our HRTF measurement technique. During this reporting period we compared HRTFs measured using our standard open-canal probe tube technique and HRTFs measured with the closed-canal insert microphones from the Crystal River Engineering Snapshot system.

Detailed Progress Report

1. Localization with Moving Sources

An important requirement of a usable 3-D auditory display is synthesis of veridical auditory image movement. Sound image movement, defined as a change in the direction of a sound relative to the listener's head and ears, occurs even when the sound source itself is stationary. In a natural situation, listeners move their heads. and these movements cause a change in the position of a stationary source relative to the listener's head. The changes in relative orientation result in predictable changes in the spatial cues produced by the sound source at the listener's ears. Such changes could be important since in theory they can provide essential information to the listener about source position. We have found that listeners judge the position of both real and virtual sound sources more accurately if head movements are encouraged. Using a virtual auditory display system. we presented sound sources which appeared to be stationary to the listener by coupling the image synthesis to the listener's head position in real time. The listeners were encouraged to move their heads during the stimulus presentation. Front-back reversals often reported by some listeners when localizing virtual sources disappeared and judgments of source elevation were more accurate. The details of this experiment were presented in the Semi-Annual Progress Report (Appendix A).

The results of the first experiment on head/image movement do not address the question of whether the improvement observed in localization performance requires proprioceptive feedback from actual head movement or auditory image movement alone. Since 3-D auditory displays are likely to find application in situations in which the operator's head may not be free perception) can be obtained with source movement alone. It is possible to provide the listener with changes in the acoustical cues similar to those that accompany head movement simply by moving the source, while the listener's head remains stationary. There is very little published data on listeners judgments of apparent position of a moving source. Previous research on source movement has focussed either on listeners' ability to judge "time to contact" of a moving source or on the minimum angular movement that is detectable. We are currently conducting experiments in which listeners are asked to localize moving sources and in which listeners are allowed to move the source to aid localization.

Using the "absolute judgment" paradigm described in our publications and previous progress reports, we tested listeners in several conditions in which the stimulus was a moving source. The first condition did not provide a "naturally" moving source but simulated movement with static sources. It consisted of presenting 3 250 msec noise bursts that changed either in azimuth or elevation by 10 degrees. An example of an azimuth change would be a sequence of 3 sources at 50, 40, 30 degrees azimuth and 20 degrees elevation. An elevation change might consist of 3 sources at 160 degrees azimuth and -30, -20, -10 degrees elevation. This condition served to provide contextual information, without actually simulating a naturally moving source. Since we were primarily interested in how this condition would affect the resolution of front-back confusions, we only tested four listeners who made front-back confusions when judging the position of static virtual sources. The listener's task was to report the azimuth, elevation and distance of the last (third) source in the sequence. None of the listeners appeared to benefit from the additional cues provided by this condition. Listeners' performance in this task was remarkably similar to their performance in the static source condition. Figure 1 shows the results from a single listener in the static source (left panel), azimuth "movement" (center panel) and elevation "movement" (right panel) conditions.

In a second experiment, we presented listeners with a virtual source that moved 40 degrees in azimuth. The stimulus was a noise burst 1 sec in duration and the rate of movement was 1 degrees/25 msec. In one condition the listener reported the apparent starting position and in a second condition, the apparent ending position. We tested 7 listeners, the 4 listeners that participated in the first experiment and 3 listeners who do not make confusions. When listeners were presented moving sources, their judgments of starting (or ending) source position were no more accurate than their judgments of static sources. Front-back reversal rates in the moving source task were similar to the rates observed in the static source experiments. Data from the static and moving source conditions are presented for two subjects in Figures 2 and 3.

In the third experiment, listeners were presented a virtual source and encouraged to move the source by pressing keys on a computer keyboard. Both azimuth and elevation movement was possible. The stimulus was a dei noise that played continuously until terminated by the listener. Preliminary data suggest that when the listener is allowed to control the source movement, the apparent difficulties that some listeners experience in resolving front-back differences disappear, just as they did when head movement was encouraged. The results from a single listener in this condition are presented in Figure 4. An analysis of the source movement histories indicated that the angular movement was about 5 degrees for both azimuth and elevation for listeners who do not typically make front-back reversals and about 40 degrees for azimuth and 20 degrees for elevation for listeners who do make front-back reversals.

2. A Comparison of Open-Canal and Closed-Canal HRTF Measurements

The fidelity of a 3-D auditory display is critically dependent on accuracy with which we can measure the listener's Head-Related Transfer Functions (HRTFs) that are used to produce virtual auditory images. If the HRTF measurements are not made carefully, or if a generic set of HRTF measurements are used, the fidelity is compromised, often resulting in large increases in front-back confusions and degradations in the perception of source elevation. Currently, we measure HRTFs using an open-canal probe microphone system (Etymotic ER7-C). If the tip of the probe tube is place at the eardrum and the probe remains stable during the measurement session, this technique produces very accurate representations of both the directional and non-directional components of the HRTF. This techniques does have several disadvantages, however. First, it is sometimes difficult to place the probe tube near the eardrum because of the shape of the earcanal. Second, the probe tube microphone is relatively insensitive and noisy. Third since the canal is open, the signal level cannot exceed 75 dB to avoid contamination by the acoustic reflex. Because of the last two problems, averaging is required to obtain an acceptable signal-to-noise ratio. If HRTF measurements are made using a closed-canal insert microphone system, the microphone (a more sensitive one) is positioned at the canal entrance and the signal level can be higher, obviating the need for extensive averaging, since the earcanal is blocked. A potential disadvantage is that canal entrance measurements may not capture all of the directional characteristics of the HRTF.

Six listeners participated in an experiment designed to compare HRTF measurements made with open-canal probe microphones (Etymotic ER-7C) and closed-canal insert microphones (from the Crystal River Engineering Snapshot HRTF Measuring System). During a single session, measurements were made at 126 spatial positions using both microphone systems. The measurements were repeated several times on a different days.

In order to compare the measurements made with the two systems, we find it useful to decompose each individual HRTF into the product (in the frequency domain) or convolution (in the time domain) of two transfer functions. One represents the "average" response of the ear (at the eardrum) to sounds from all directions, and the other represents the departures from that average that are specific to each individual direction. The first we call the "diffuse-field" estimate (DFE), which formally is the response of the ear to a diffuse sound field. The second we call the "directional transfer function" or DTF. The DTFs are estimated by dividing each HRTF by the DFE. Figures 5 and 6 show the HRTF, DFE and DTF at a single source position from two listeners. the solid curves show the measurements taken at the eardrum with the probe-tube system and the dashed curves show the measurements taken at the entrance to the closed ear canal. While the two systems produce very different HRTFs and DFEs. the DTFs are very similar.

Multidimensional Scaling Analysis was used to summarize DTF differences between the two measuring systems and repeatability of each system. The levels (dB) in non-overlapping critical bands were determined for each DTF. The difference between any two sets of DTFs was represented by the Euclidean distance metric, the square root of the sum of squared dB differences. A 29 x 29 matrix was constructed, representing the differences among all 29 sets of DTFs (there were 2 or more sets of DTFs for each measurement system from each of the 6 listeners). This matrix was subjected to the scaling analysis which produced a 3-dimensional solution. accounting for 90% of the variance in the data. A 2-D projection of the 3-D scaling solution is shown in Figure 7. The letters refer to different listeners, with uppercase representing the canal entrance measurements and lowercase representing the probe measurements. The differences between the two systems appear to be no greater than differences among repeated measurements on a given listener for each system alone. For 3 of the listeners, variability among the sets of canal-entrance measurements was somewhat greater than for the probe measurements.

We also evaluated the potential utility of the closed-canal system for measuring HRTFs that can be to produce virtual auditory targets in a localization task. Two sets of virtual sound sources were synthesized, one from HRTF data obtained using the standard Etymotic probe tube system and one from data obtained with the CRE closed-canal system. In both cases the source was a single 250 ms burst of white noise presented over high-quality headphones at about 70 dB SPL. Each of the 126 virtual positions were randomly presented 5 times. Listeners judged the apparent positions of both sets of virtual sources, those made from closed-canal measurements and those made from eardrum measurements. Results from two listeners are shown in Figures 8 and 9. Data from the canal-entrance condition are shown in the left panels and data from the probe-tube system are shown in the right panels. The fact that the patterns of judgments are nearly identical for both sets of virtual sources suggests that the CRE closed-canal HRTF measuring system can be used effectively in the process of producing virtual auditory targets. Its main advantages over the conventional probe-tube system are a much higher signal/noise ratio (thus, shorter measuring time) and less discomfort for the listener.

Publications

Papers

Wightman, F. L. & Jenison, R. L. (1995). Auditory Spatial Layout. In W. Epstein & S. J. Rogers (Eds.), Handbook of Perception and Cognition. Volume 5: Perception of Space and Motion. Orlando, FL: Academic (In Press).

Wightman, F. L. & Kistler, D. J. (1995). Factors affecting the relative salience of sound localization cues. In R. Gilkey and T. Anderson (Eds.), Binaural and Spatial Hearing. Hillsdale, NJ: Erlbaum (In Press).

Macpherson, E. A. (1994). On the role of head-related transfer function spectral notches in the judgment of sound source elevation. In G. Kramer (Ed.), Proceedings of the 1994 International Conference on Auditory Displays, Santa Fe. NM. (In Press).

Zahorik, P. A., Kistler, D. J., Wightman, F. L. (1994). Sound localization in varying virtual acoustic environments. In G. Kramer (Ed.), Proceedings of the 1994 International Conference on Auditory Displays, Santa Fe, NM. (In press).

Abstracts

Jenison, R. L. (1994). Radial basis function neural network for modeling auditory space. Journal of the Acoustical Society of America. 95 (Pt. 2), 2898.

Jenison, R. L., Wightman, F. L. and Kistler, D. J. (1994). Self-organizing model of auditory maps. Abstracts of the 17th Mid-Winter Meeting. Association for Research in Otolaryngology, 70.

Wightman, F. L., Kistler, D. J., & Andersen, K. (1994). Reassessment of the role of head movements in human sound localization. Journal of the Acoustical Society of America. 95 (Pt. 2), 3003.

Wightman, F. L., & Kistler, D. J. (1994). The importance of head movements for localizing virtual auditory display objects. Proceedings of the 1994 International Conference on Auditory Displays, Santa Fe, NM. (In Press)

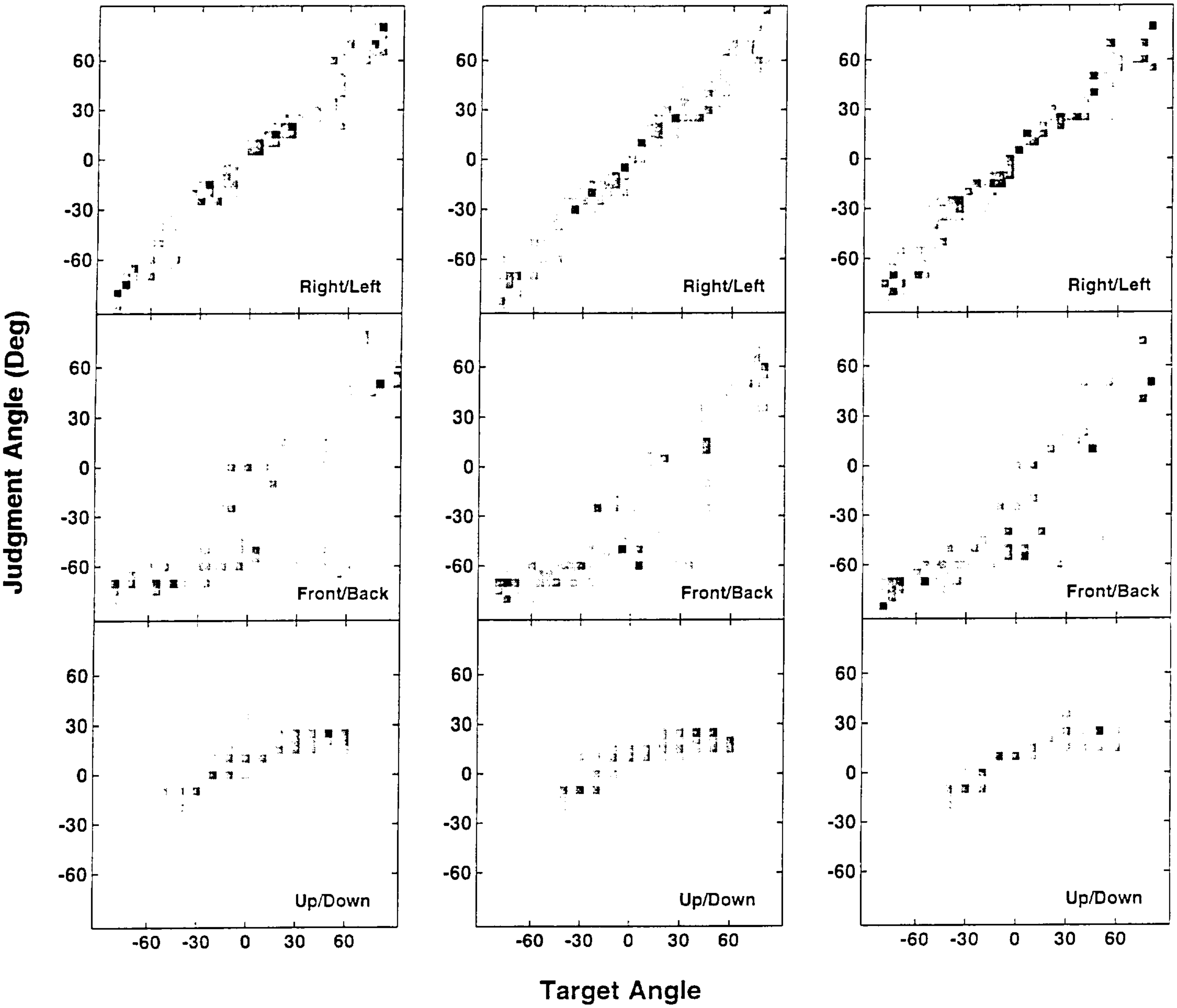


Figure 1. Judgments of apparent position of virtual sources from Listener SMQ in the static source (left panel), azimuth "movement" (center panel) and "elevation" condition (right panel).

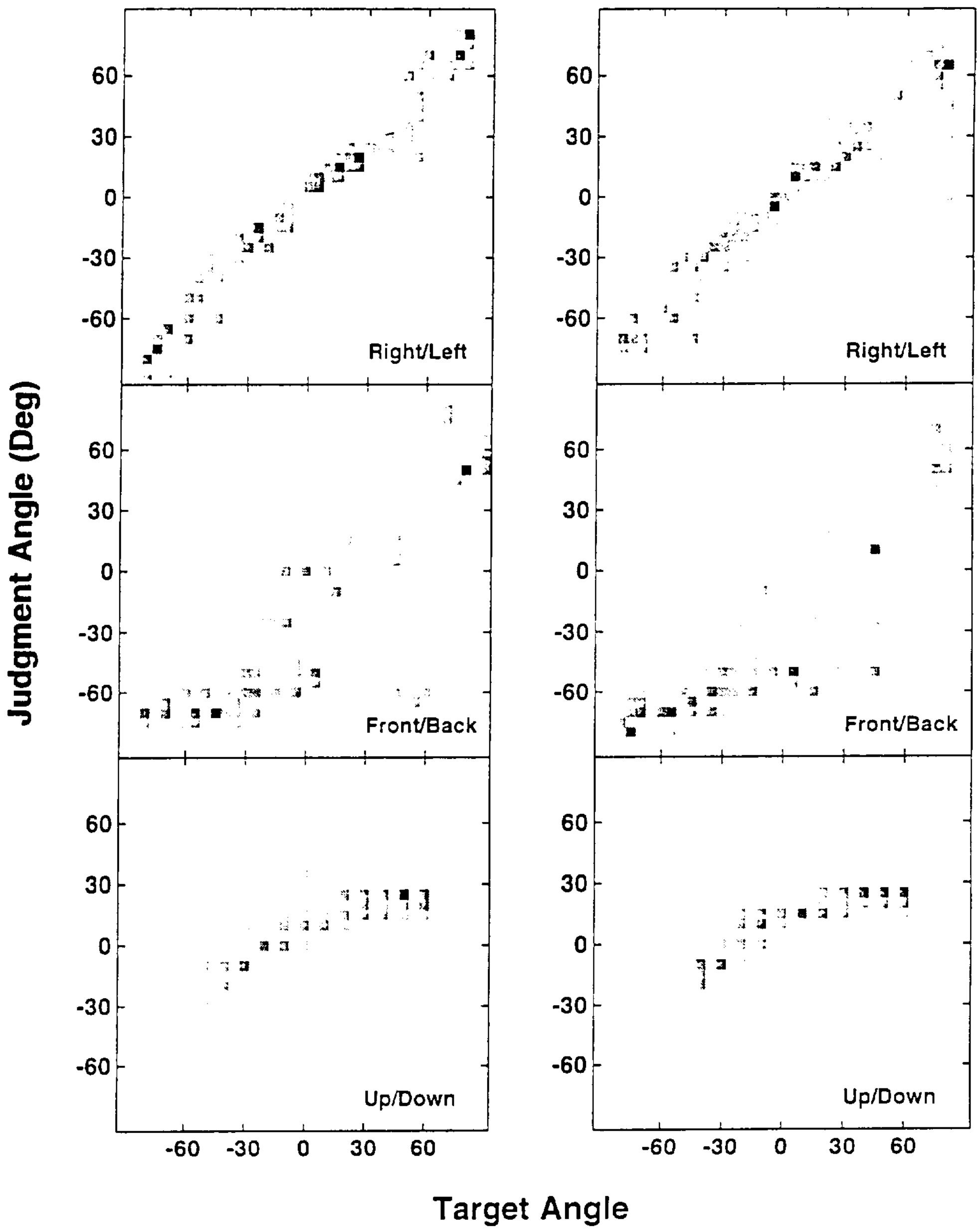


Figure 2. Judgments of apparent position of virtual sources from Listener SMQ in the static source condition (left panel) and the moving source condition (right panel).

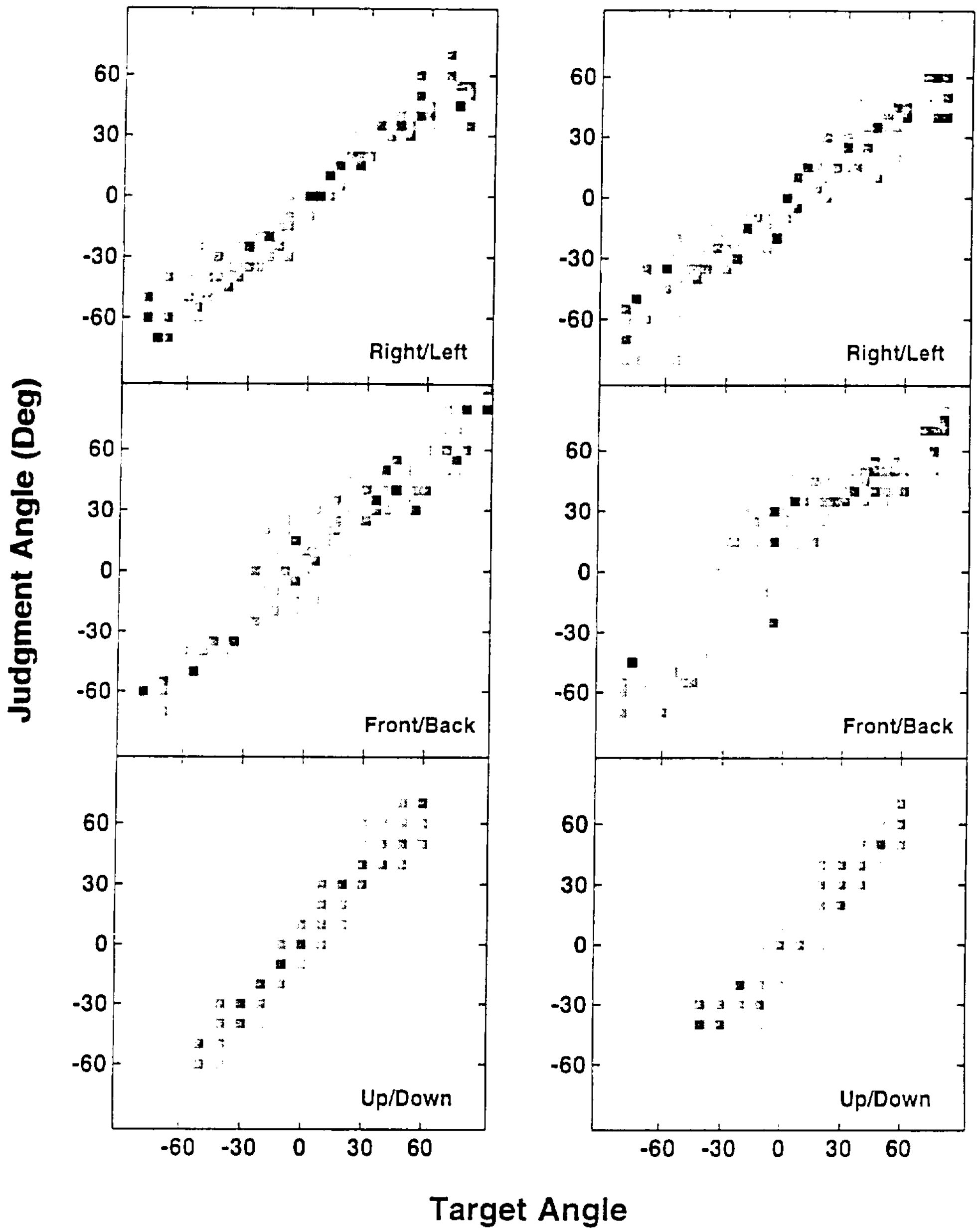


Figure 3. Judgments of apparent position for virtual sources from Listener SNJ in the static source condition (left panel) and the moving source condition (right panel).

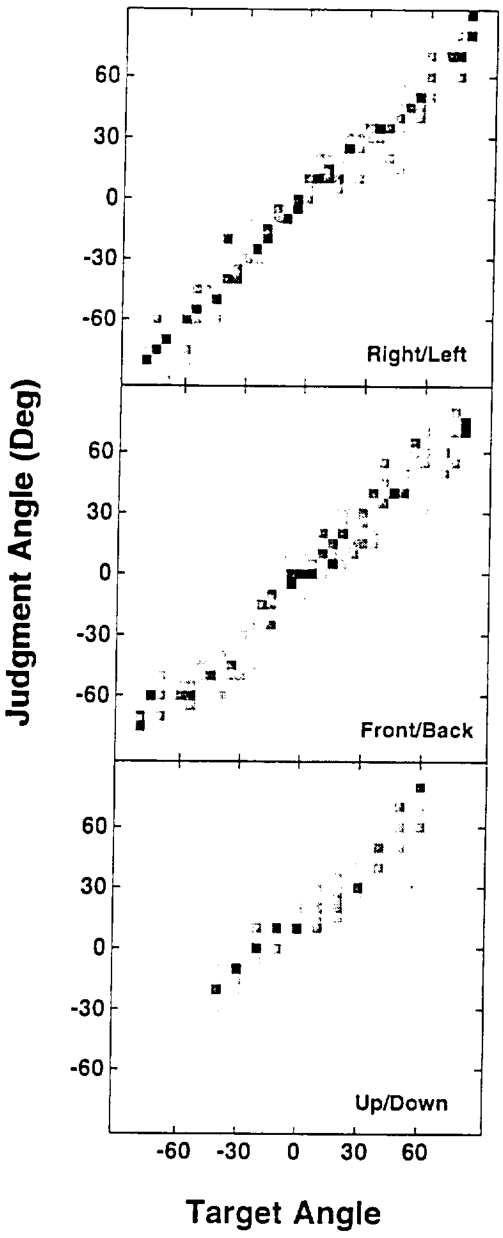


Figure 4. Judgments of apparent position of virtual sources from Listener SMQ in the condition in which the listener controls the movement of the source.

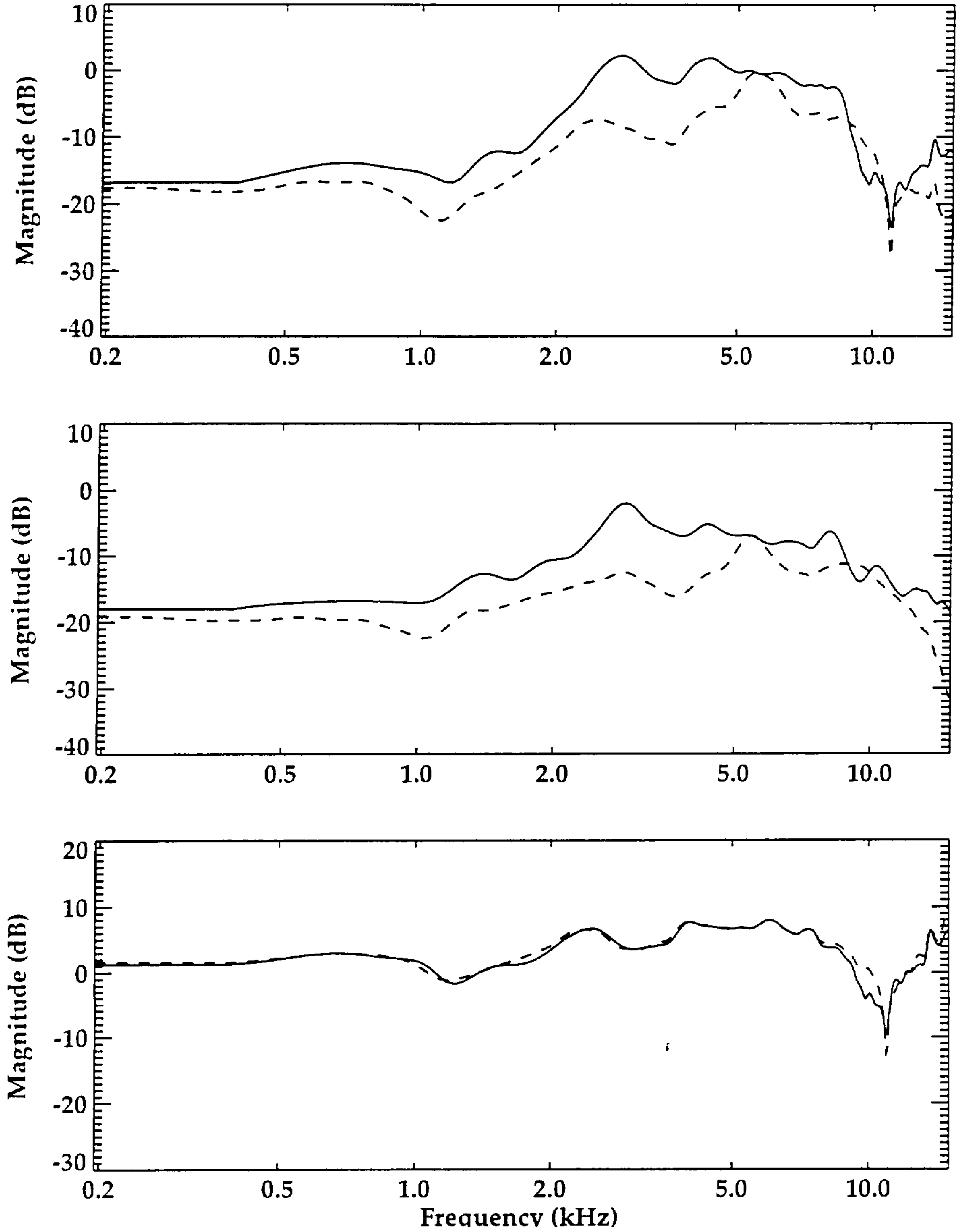


Figure 5. The top panel shows the raw HRTF magnitudes for a single source position from Listener AFW. The measurement obtained with the probe microphone is plotted with a solid line and the measurement obtained with the canal entrance microphone is plotted with a dashed line. Diffuse field estimates are plotted in the center panel and directional transfer functions are plotted in the bottom panel.

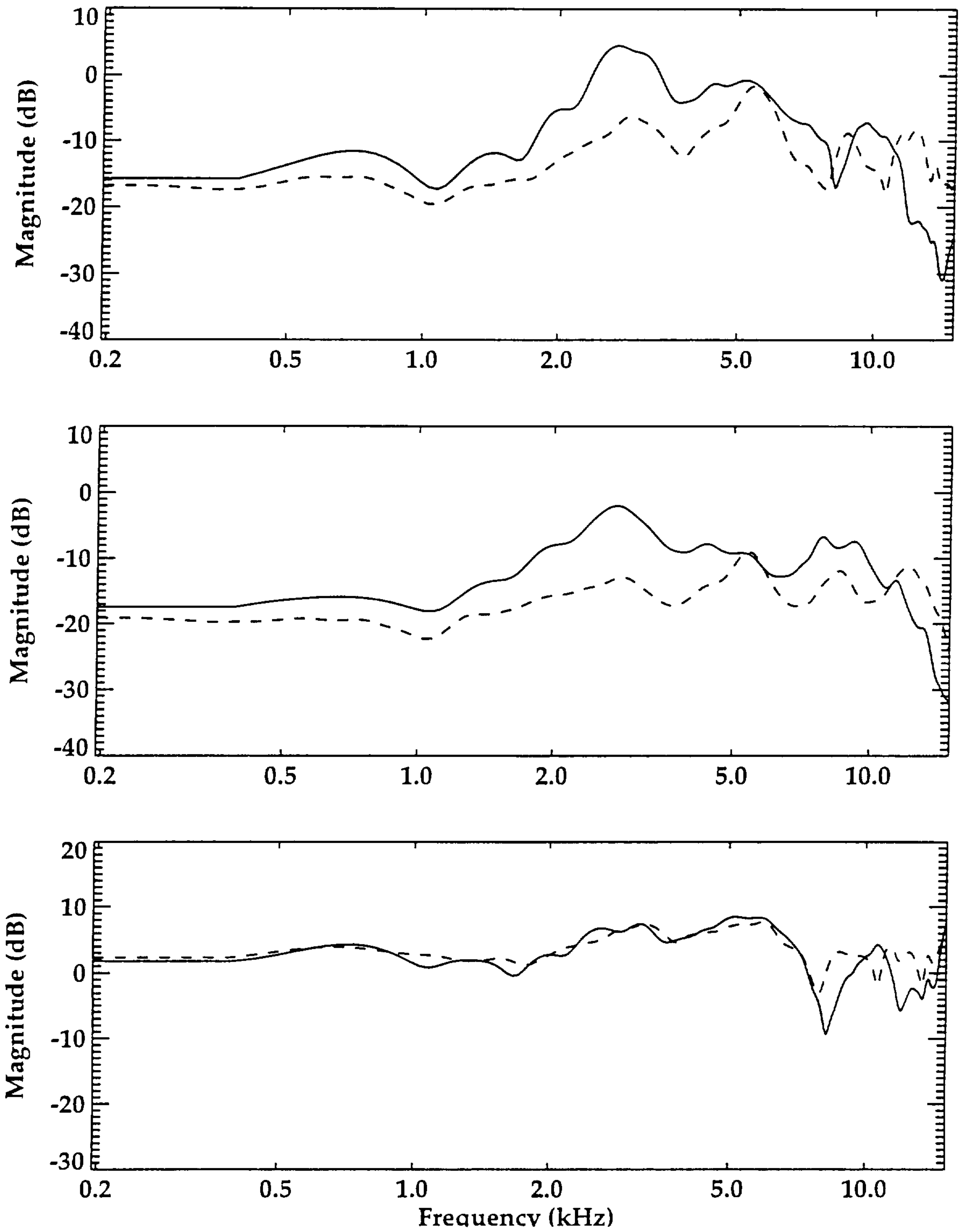


Figure 6. The top panel shows the raw HRTF magnitudes for a single source position from Listener SNF. The measurement obtained with the probe microphone is plotted with a solid line and the measurement obtained with the canal entrance microphone is plotted with a dashed line. Diffuse field estimates are plotted in the center panel and directional transfer functions are plotted in the bottom panel.

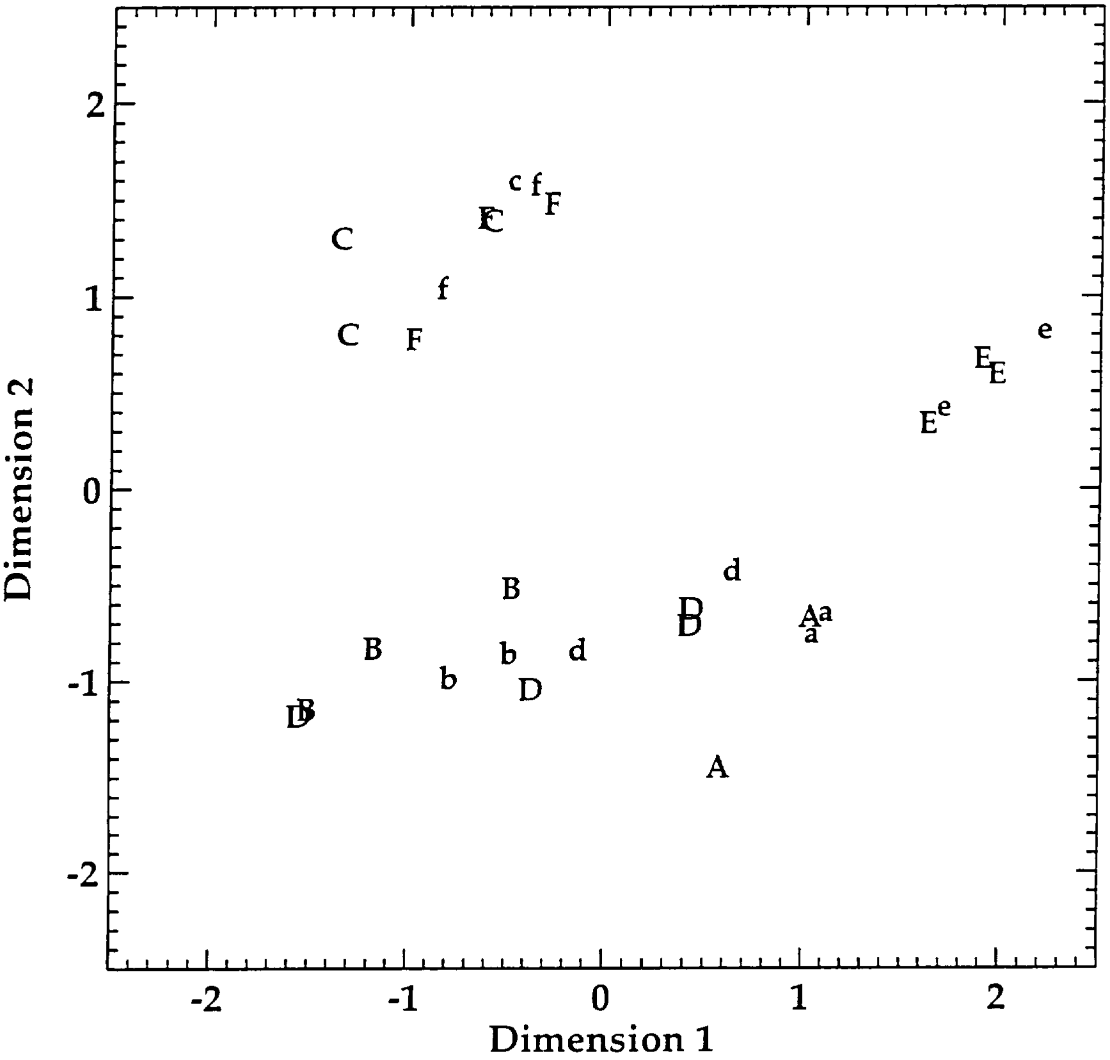


Figure 7. Two-dimensional projection of the 3-dimensional Multidimensional Scaling solution of DTFs estimated from measurements made with open-canal probe microphones (lowercase) and closed-canal insert microphones (uppercase). Each listener is represented by a different letter.

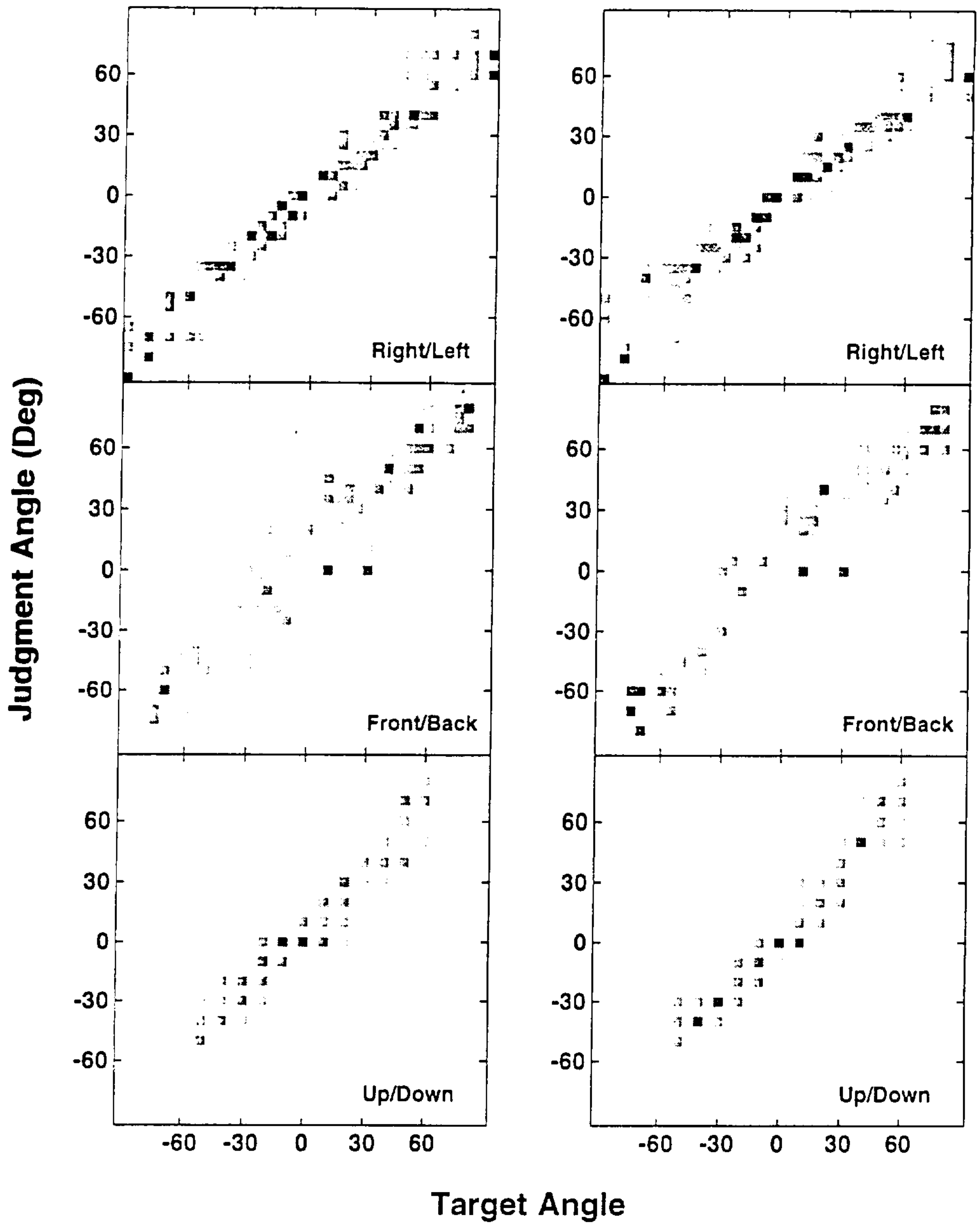


Figure 8. Judgments of apparent position of virtual sources produced from HRTF measurements with the open-canal system (left panel) and with the closed-canal system (right panel) from Listener SNJ.

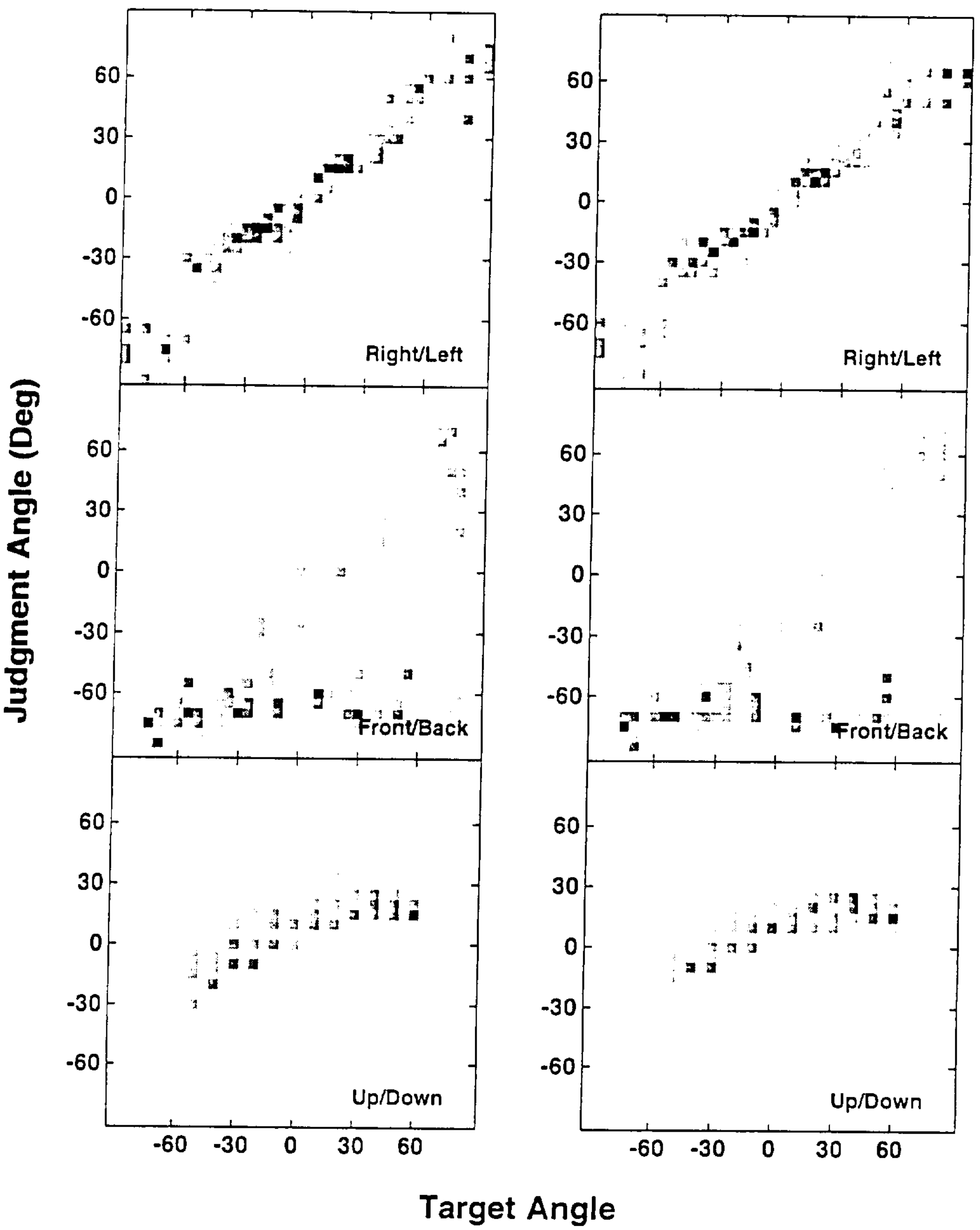


Figure 9. Judgments of apparent position of virtual sources produced from HRTF measurements with the open-canal system (left panel) and with the closed-canal system (right panel) from Listener SMQ.

APPENDIX A

Semiannual Progress Report

Period Covered: 5/1/94-11/1/94

NASA Cooperative Research Agreement
NCC2-542

Psychophysical Evaluation of Three-Dimensional
Auditory Displays

Frederic L. Wightman, Ph. D., Principal Investigator
Waisman Center, University of Wisconsin
1500 Highland Avenue
Madison, WI 53705

Progress Report

The fidelity of current virtual auditory display systems is limited primarily by the occurrence of front-back confusions and poor representation of target source elevation. Work during this reporting period attempted to achieve a better understanding of the importance of several acoustical cues that we believe are important for achieving high quality front-back and elevation perception and good externalization with virtual auditory displays. Experiments were completed on the role of dynamic cues provided by head movements and on the role of cues provided by echoes. Additionally, we continued our efforts to relate spectral features of HRTFs to perceived sound source location by formulating a model which attempts to predict elevation judgments from the frequency of the primary spectral notch in the HRTF.

1. Role of Dynamic Cues

When a listener's head moves while listening to a stationary sound source, the interaural time, interaural intensity and pinna cues change in accordance with the head movements. In an experiment described in a previous progress report, we presented 5 listeners with stationary virtual sources synthesized with the Convolvotron, which was coupled to a magnetic head tracker. The listeners were encouraged to move their heads to facilitate localization. Only one of these listeners made large numbers of front-back confusions in the baseline condition in which no dynamic cues were available. The results suggested that the cues provided by this listener's head movements could eliminate these confusions.

During the present funding period we sought to replicate this result in a second experiment with 8 new subjects, 6 of whom made front-back reversals in the baseline virtual source and in the freefield conditions. In addition to the baseline condition in which stimuli delivered to the headphones were not influenced by the movement of the listener's head ("restricted" condition), there were two movement conditions: 1) listeners were encouraged to move their heads to aid localization ("freestyle" condition); 2) listeners were told to point their noses at the sound source ("compulsory" condition). The stimuli were 2.5 s virtual sources synthesized by the Convolvotron using HRTFs measured from each listener's own ears. The position of the listener's head was tracked and the synthesis of the virtual source was modified in real time, in accordance with the head movements to simulate a stationary external source. For those listeners who made frequent front-back reversals in the baseline condition, reversal rates were near zero in the two head movement conditions. We also observed some improvement in perceived elevation, especially in the "compulsory" condition. Data from the three conditions are shown for 2 listeners in Figures 1 and 2.

Analyses of the trajectories of the listener's head movements revealed that while the tracks were idiosyncratic, they were remarkably consistent from presentation to presentation for a single listener. In general most listeners appeared to orient toward the source in the "freestyle" condition. An examination of some of the trials on which the listeners made reversals revealed that the listeners did not attempt to move their heads on the majority of these trials. The 2 listeners who did not make reversals in the baseline condition showed very little head movement in the "freestyle" condition.

Figure 3 illustrates trajectories of head movements in the "freestyle" and "compulsory" conditions for a listener who makes frequent front-back reversals in the "restricted" condition. The four panels show head movement trajectories (indicated by the dotted lines) from four trials on which the same virtual source was presented. Note the consistency in the trajectory on the four trials. Also plotted on the figures are the nominal position of the virtual source, the mean judgment made in the "restricted" condition and the judgment made on each trial in the "freestyle" condition. Figure 4 shows trajectories on two identical trials from a listener who makes few front-back confusions. Note that in the "freestyle" condition, this listener's head movements were very small.

The results strongly suggest that head movements are a natural and important component of localizing sounds and that auditory displays that incorporate head-coupled synthesis will provide a more realistic listening environment.

2. Role of Echoes

An important feature of natural listening environments is the presence of echoes and reverberation. There is anecdotal evidence that suggests that echoes may enhance the externalization of virtual sounds and that they may provide additional cues for resolving front-back ambiguities. In our first experiment, described in a previous progress report, we presented virtual sources that were synthesized to include not only the direct sound but also the first-order reflections off the four walls of an 8 x 8 x 3 m room. Reflections were attenuated by 6 dB to mimic "soft" walls. Listeners' azimuth and elevation judgments were indistinguishable from their responses to virtual sources with no reflections.

In our recent work on this topic, we tested 5 new listeners in three types of virtual stimuli: 1) "dry" virtual sources containing no echoes, 2) echoic virtual sources synthesized using the image model to predict spatial position, time delay and amount of attenuation for the first 20 reflections occurring in time after the direct source path, and 3) "perturbed" echoic sources synthesized with 20 reflections for which the time delays and attenuation factors were computed according to the predictions of the image model, but the spatial positions were chosen randomly. Listeners performed similarly in all three conditions. The details of this experiment are in a manuscript included with this report.

3. Role of Spectral Notches

There is considerable evidence to suggest that low-frequency interaural time difference is the primary determinant of perceived laterality or the "left-right" component of a sound source. It is widely believed that monaural spectral cues are important determinants of the other two dimensions of apparent source position, "front-back" and "up-down" or elevation. However, the nature of the relationship between spectral features of an HRTF measured for a particular sound source and apparent source position is not known. The most prominent features of HRTF magnitude spectra are the high-frequency notches. An examination of our HRTF data indicates that the frequency of these notches changes in a fairly systematic fashion with changes in source elevation. The pattern of change differs across azimuths and across individuals. Consequently, we sought to determine if these differences in notch frequency pattern could be used to predict elevation judgments.

A simple model was formulated which predicts that perceived elevation is determined by the frequency of the primary high-frequency notch in the HRTF of the ear closest to the source. The primary notch frequency was determined "by eye" for 132 positions spaced 30 degrees apart in azimuth and spaced 10 degrees apart in elevation (elevations ranged from -50 to +50). The model further predicts that the variability in elevation judgments is related to the notch frequency gradient such that the steeper the gradient, the lower the variability. Results from an analysis of the variability of freefield elevation judgments of 6 subjects do not support the single-notch model. We conclude that perceived elevation must depend on additional spectral features. The details of this work are provided in an attached manuscript.

Publications

Papers

Wightman, F. L. & Jenison, R. L. (1995). Auditory Spatial Layout. In W. Epstein & S. J. Rogers (Eds.), Handbook of Perception and Cognition. Volume 5: Perception of Space and Motion. Orlando, FL: Academic (In Press).

Wightman, F. L. & Kistler, D. J. (1995). Factors affecting the relative salience of sound localization cues. In R. Gilkey and T. Anderson (Eds.), Binaural and Spatial Hearing. Hillsdale, NJ: Erlbaum (In Press).

Macpherson, E. A. (1994). On the role of head-related transfer function spectral notches in the judgment of sound source elevation. In G. Kramer (Ed.), Proceedings of the 1994 International Conference on Auditory Displays, Santa Fe, NM. (In Press).

Zahorik, P. A., Kistler, D. J., Wightman, F. L. (1994). Sound localization in varying virtual acoustic environments. In G. Kramer (Ed.), Proceedings of the 1994 International Conference on Auditory Displays, Santa Fe, NM. (In press).

Abstracts

Jenison, R. L. (1994). Radial basis function neural network for modeling auditory space. Journal of the Acoustical Society of America, 95 (Pt. 2), 2898.

Jenison, R. L., Wightman, F. L. and Kistler, D. J. (1994). Self-organizing model of auditory maps. Abstracts of the 17th Mid-Winter Meeting, Association for Research in Otolaryngology, 70.

Wightman, F. L., Kistler, D. J., & Andersen, K. (1994). Reassessment of the role of head movements in human sound localization. Journal of the Acoustical Society of America, 95 (Pt. 2), 3003.

Wightman, F. L., & Kistler, D. J. (1994). The importance of head movements for localizing virtual auditory display objects. Proceedings of the 1994 International Conference on Auditory Displays, Santa Fe, NM. (In Press)

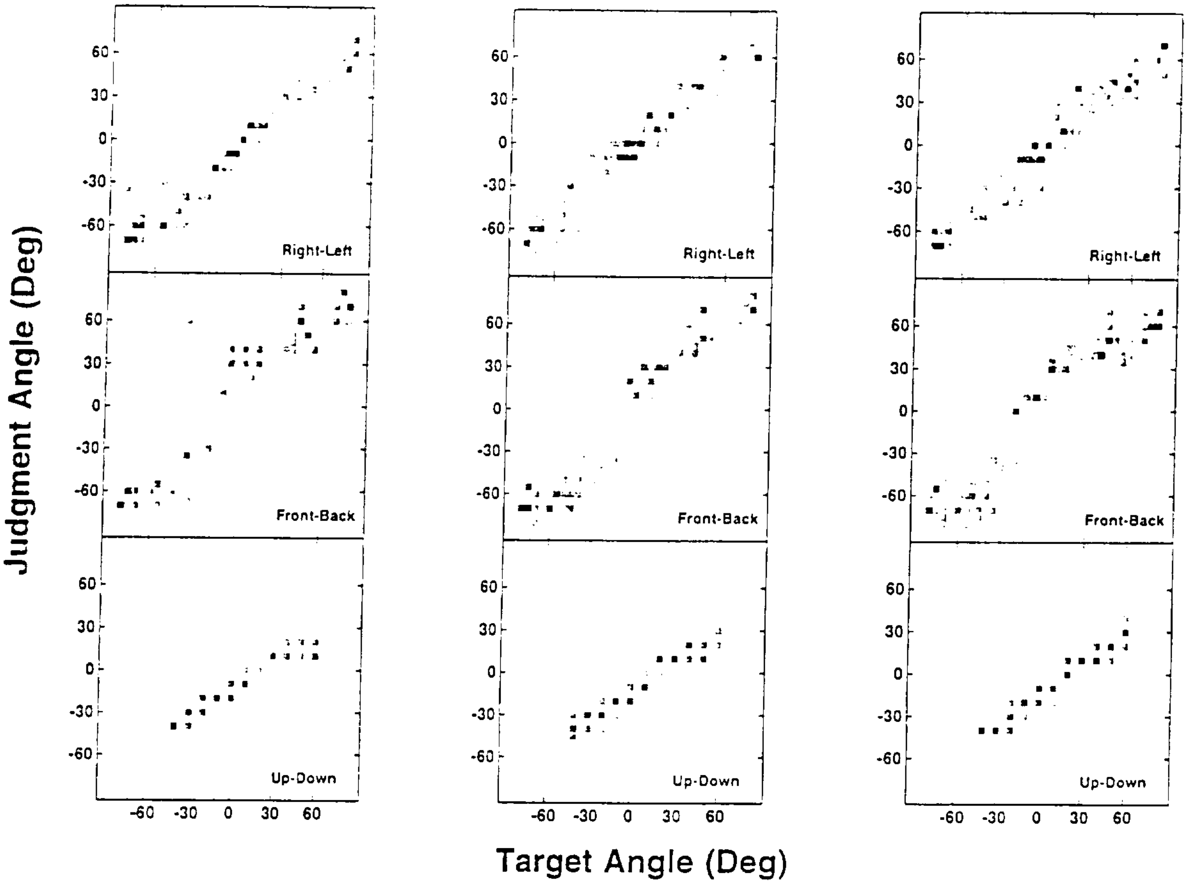


FIGURE 1. Data from Subject SNF in the three head movement conditions: "Restricted" (left panel), "Freestyle" (center panel), and "Compulsory" (right panel).

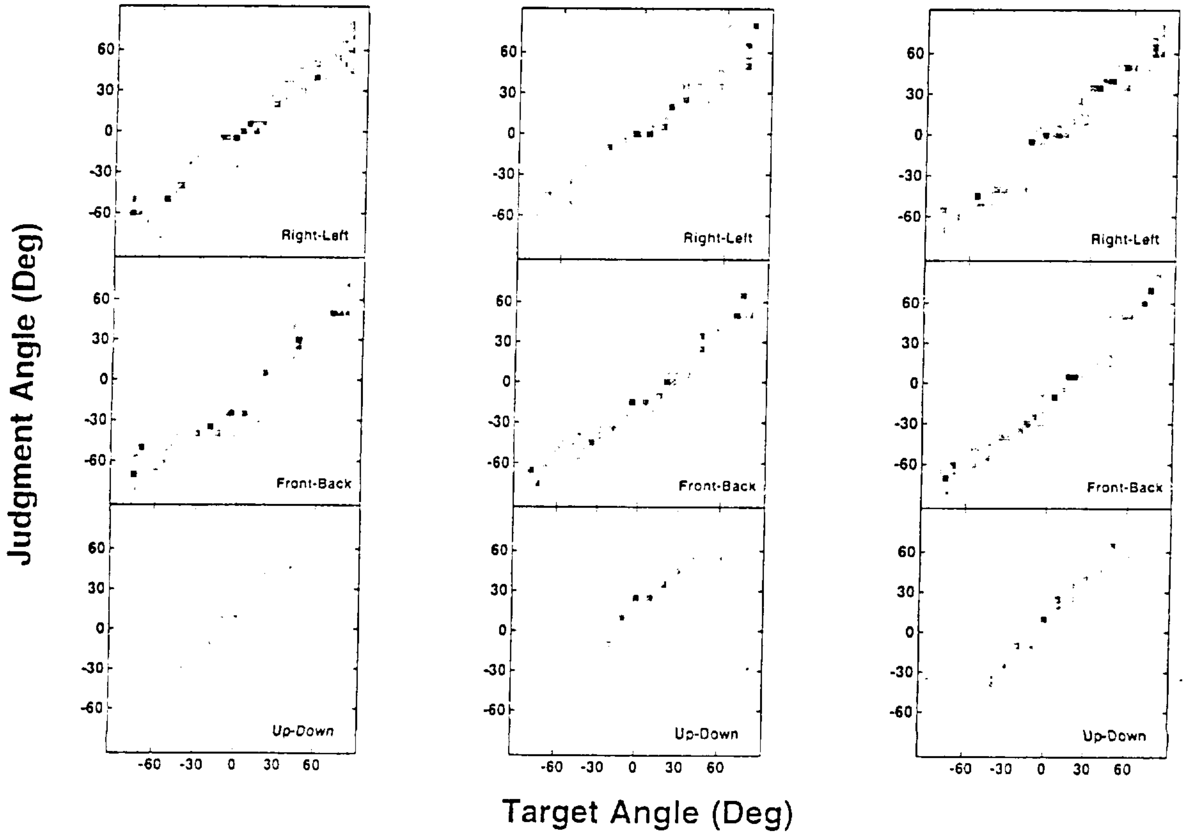


FIGURE 2. Data from Subject SNR in the three head movement conditions: "Restricted" (left panel), "Freestyle" (center panel), and "Compulsory" (right panel).

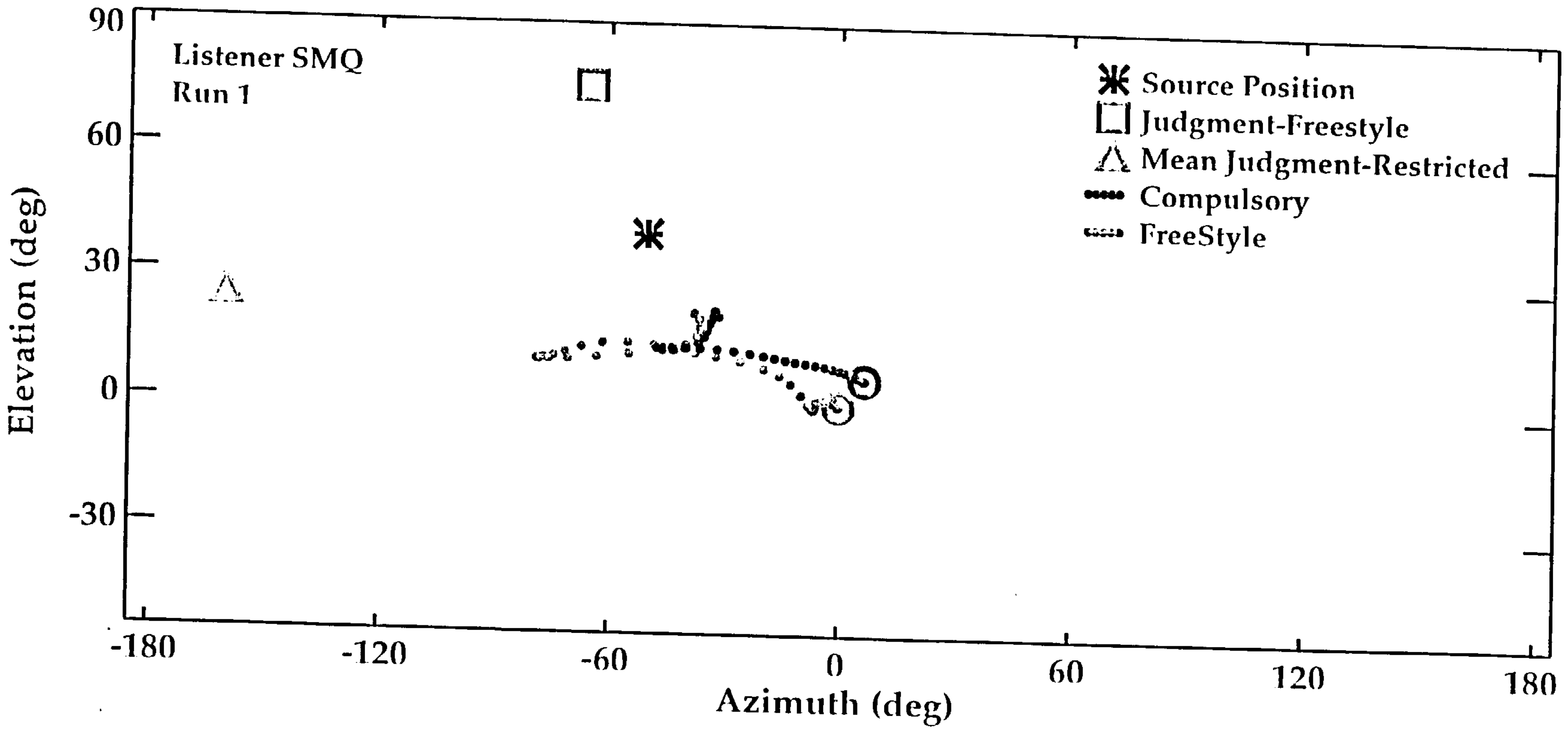

Listener SMQ
Run 1
✳ Source Position
□ Judgment-Freestyle
△ Mean Judgment-Restricted
Compulsory
FreeStyle
90
60
30
0
-30
Elevation (deg)
-180
-120
-60
0
60
120
180
Azimuth (deg)

Figure 3a

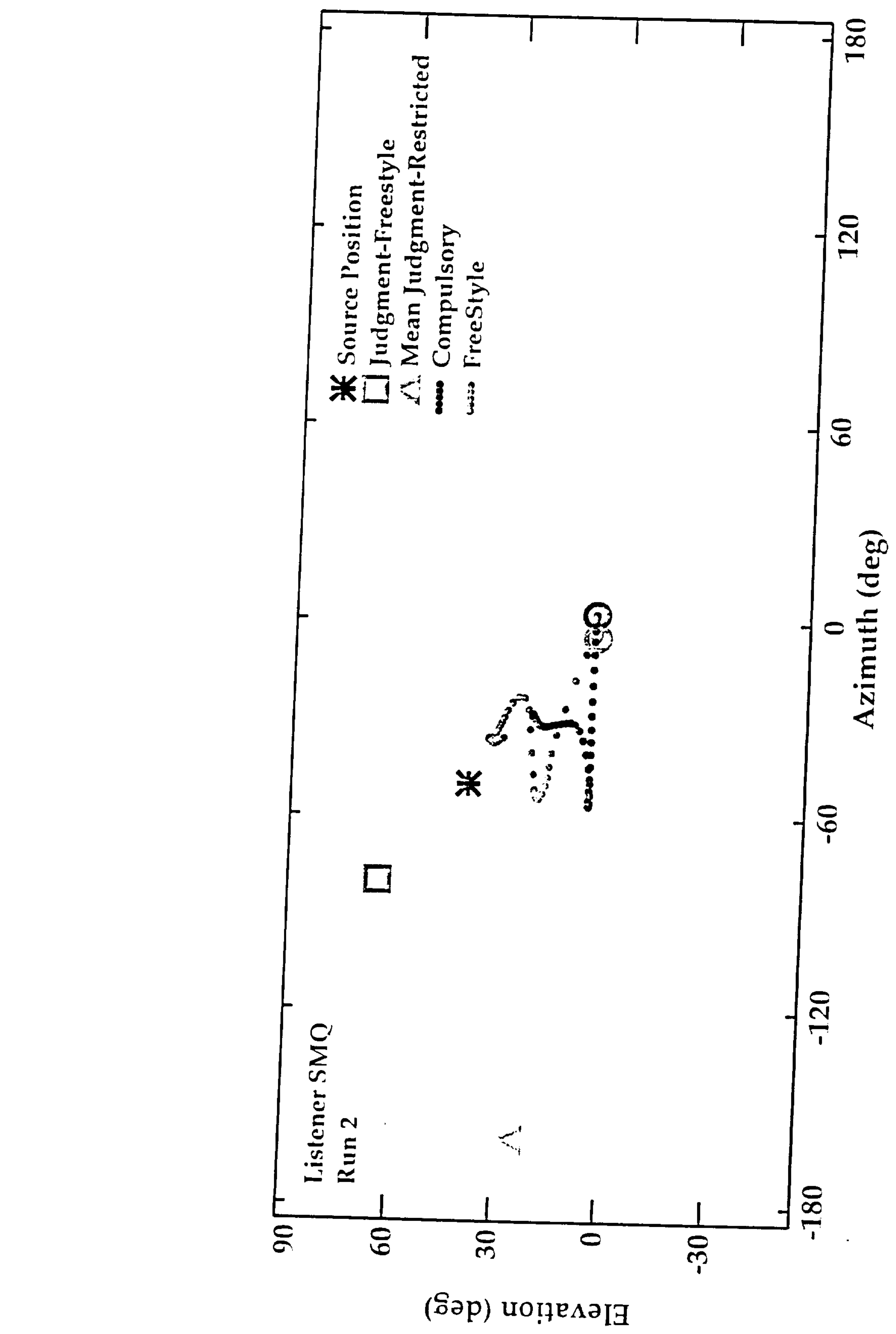

Listener SMQ
Run 2
Source Position
Judgment-Freestyle
Mean Judgment-Restricted
Compulsory
FreeStyle
Elevation (deg)
90
60
30
0
-30
Azimuth (deg)
-180
-120
-60
0
60
120
180

Figure 3b

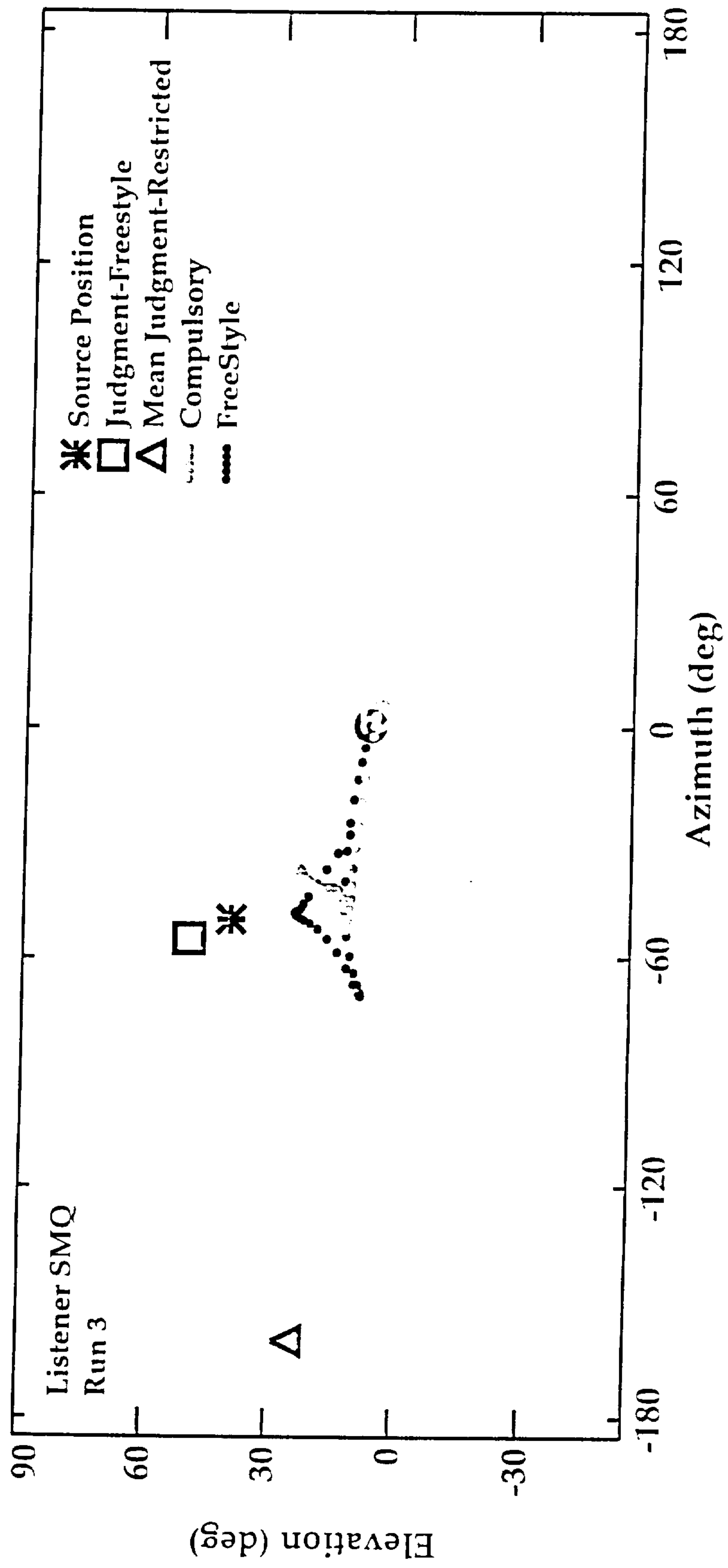

Listener SMQ
Run 3
✳ Source Position
□ Judgment-Freestyle
Δ Mean Judgment-Restricted
Compulsory
FreeStyle
Elevation (deg)
90
60
30
0
-30
Azimuth (deg)
-180
-120
-60
0
60
120
180

Figure 3c

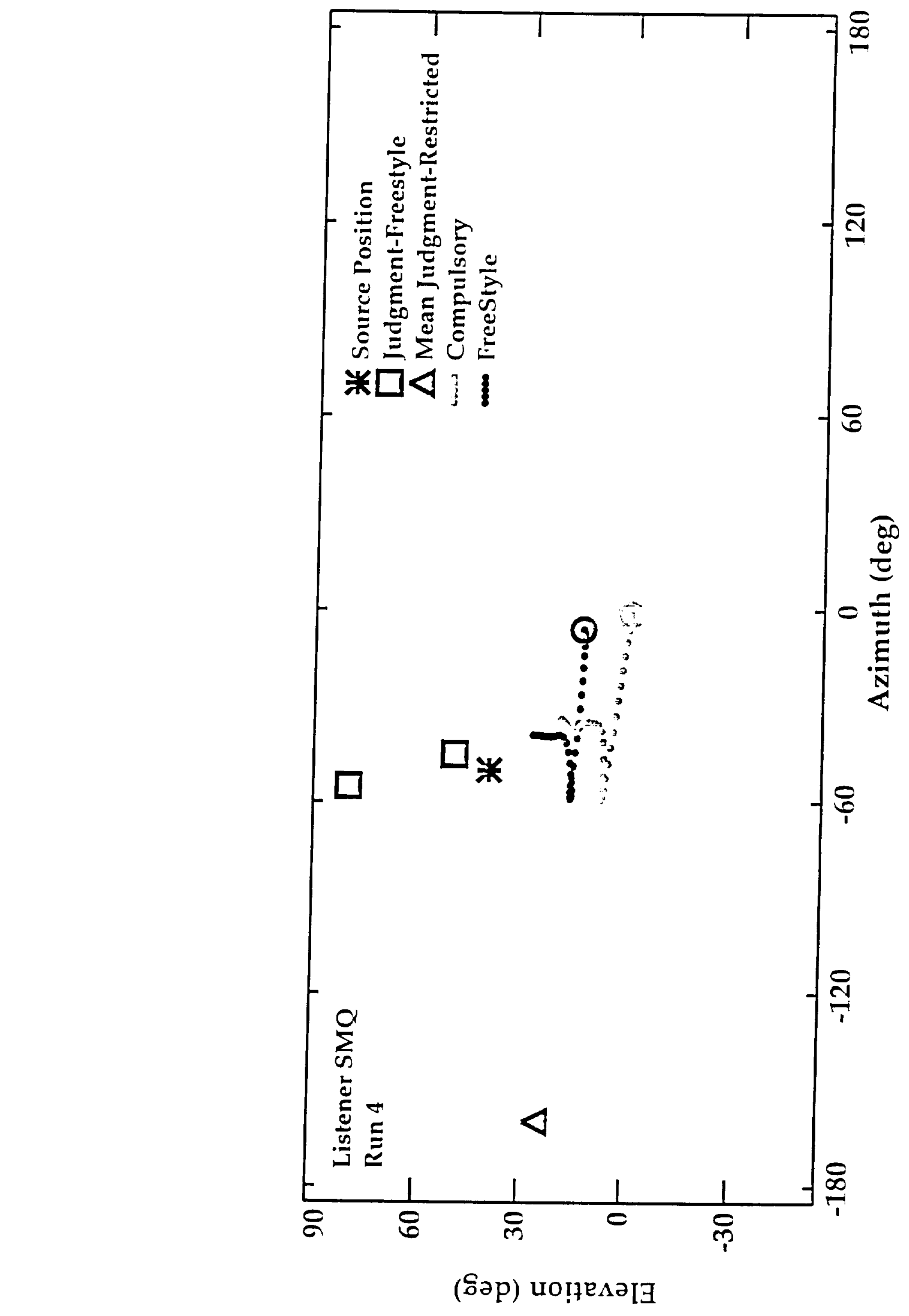

Listener SMQ
Run 4
✱ Source Position
□ Judgment-Freestyle
△ Mean Judgment-Restricted
Compulsory
FreeStyle
Elevation (deg)
90
60
30
0
-30
Azimuth (deg)
-180
-120
-60
0
60
120
180

Figure 3d

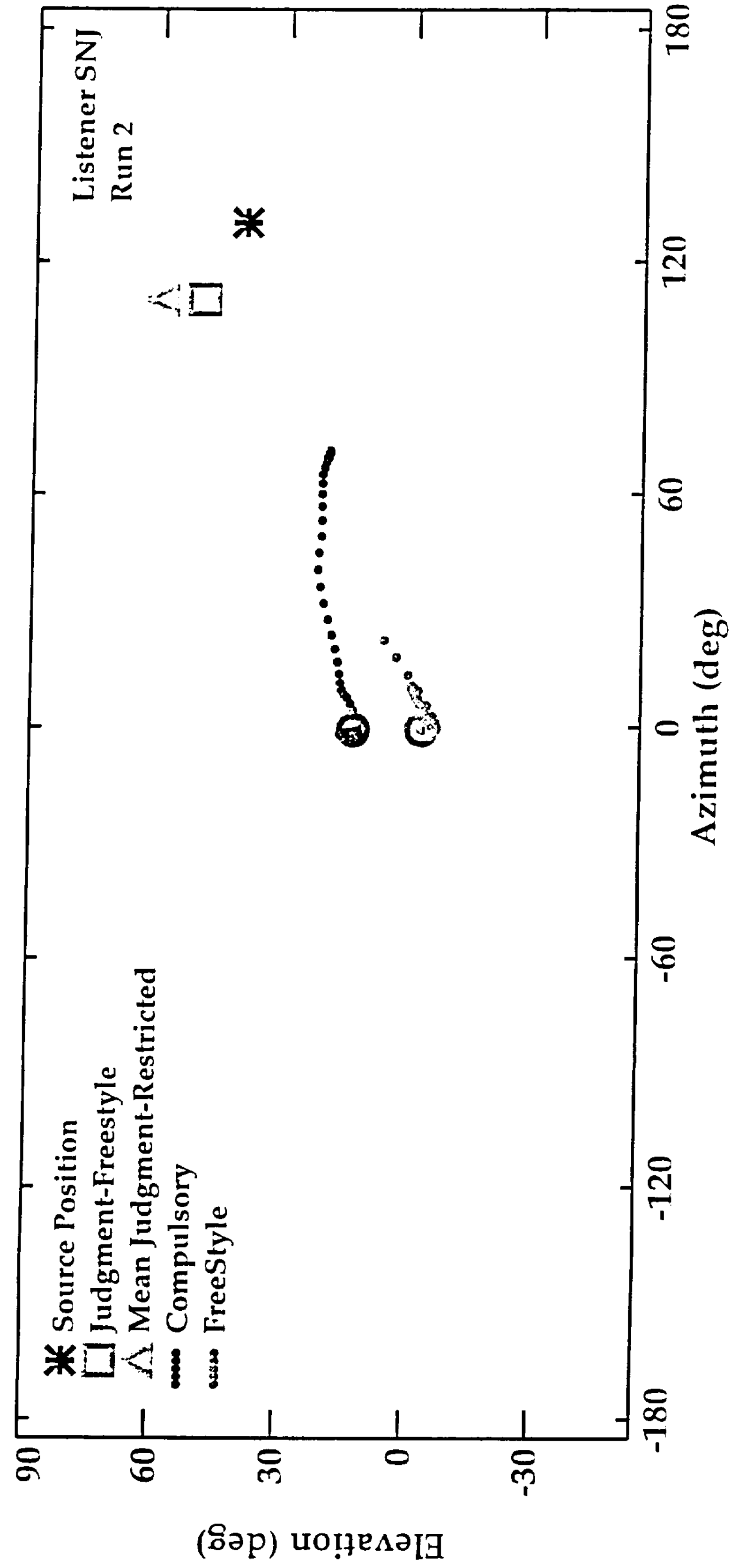

Source Position
Judgment-Freestyle
Mean Judgment-Restricted
Compulsory
FreeStyle
Listener SNJ
Run 2
Elevation (deg)
90
60
30
0
-30
Azimuth (deg)
-180
-120
-60
0
60
120
180

Figure 4b

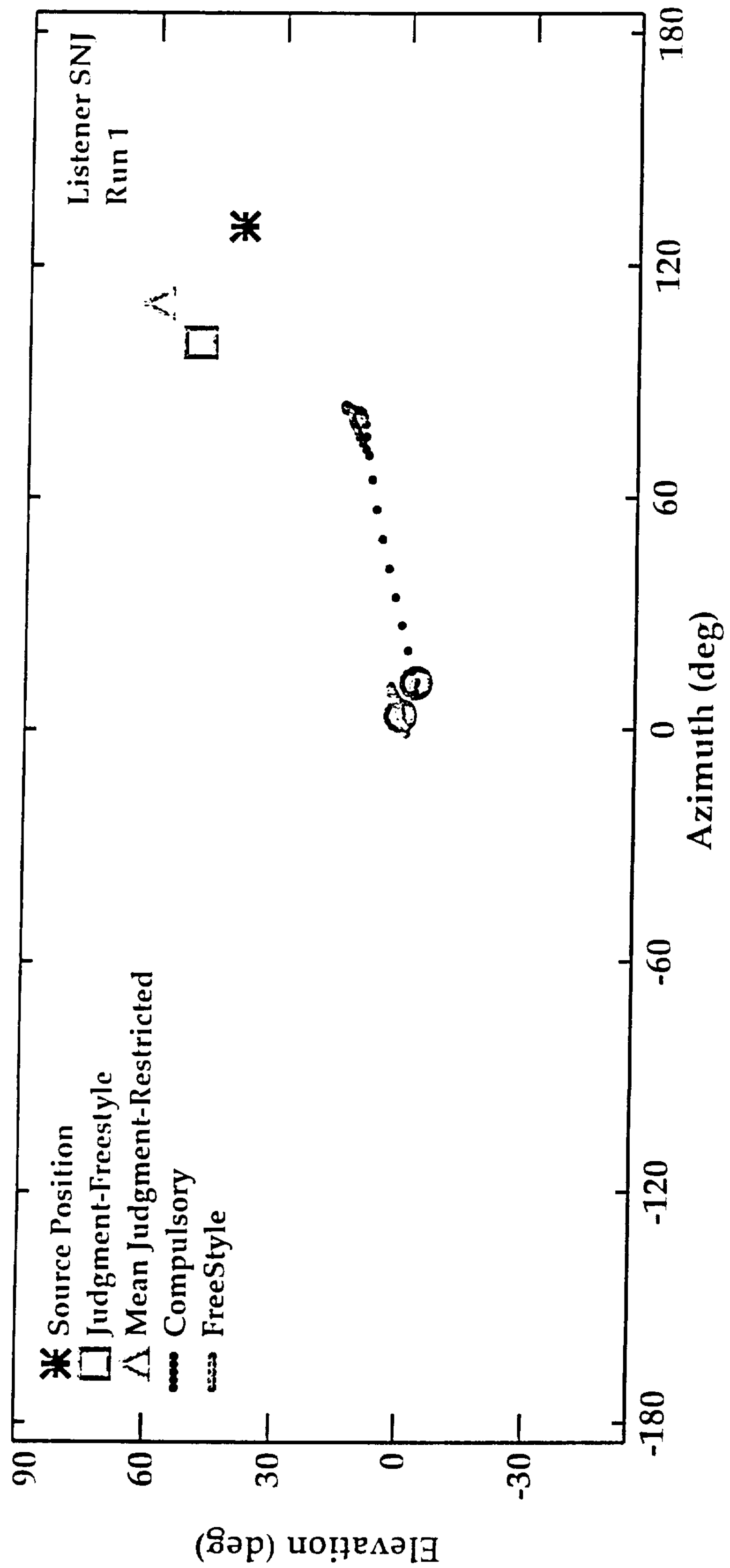

Listener SNJ
Run 1
✳ Source Position
□ Judgment-Freestyle
△ Mean Judgment-Restricted
•••• Compulsory
.... FreeStyle
Elevation (deg)
90
60
30
0
-30
Azimuth (deg)
-180
-120
-60
0
60
120
180

Figure 4a

On the role of head-related transfer function spectral notches in the judgement of sound source elevation

Ewan A. Macpherson
Waisman Center
University of Wisconsin-Madison
1500 Highland Avenue
Madison, WI 53705-2280
macpherson@waisman.wisc.edu

Abstract

Using a simple model of sound source elevation judgement, an attempt was made to predict two aspects of listeners' localization behavior from measurements of the positions of the primary high-frequency notch in their head-related transfer functions. These characteristics were: 1) the scatter in elevation judgements, and 2) possible biases in perceived elevation introduced by front-back and back-front reversals. Although significant differences were found among the notch-frequency patterns for individual subjects, the model was not capable of predicting differences in judgement behavior. This suggests that a simple model of elevation perception based on a single spectral notch frequency is inadequate.

1 Introduction

The role of spectral cues in auditory localization is known to be significant but is as yet poorly understood. While it has been established that low-frequency interaural time difference is the primary determinant of the left-right component of perceived source position [1], no simple and reliable cue for the elevation or front-back components has been found.[1] There does exist a regular dependence of spectral notch frequency on position for the head-related transfer functions of the cat [2,3] and somewhat similar feature motion for humans [4], and some researchers have proposed that this may be the principal elevation cue [5]. Although notch frequency clearly depends on position, it may be the case that the pattern is not as regular for humans as it is for the cat, and no causal relationship between this aspect of the physical acoustics and listener behavior has been confirmed.

The aims of the present study were to examine the differences among the notch frequency patterns of a number of individuals and to attempt to predict patterns in their elevation judgements on the basis of these differences. Predictions were made using the following simple model of elevation perception, which will be referred to as the single-notch model:

[1] The position of a source in space can be defined in a three-pole coordinate system with dimensions of left-right, back-front and elevation (up-down). The left-right dimension corresponds to the angle between the source and the vertical median plane. Sources with equal left-rightness lie on a "cone of confusion", so-called because the interaural time-difference cue is approximately constant and hence ambiguous.

> Given that a source is localized to a particular cone of confusion (determined by interaural time difference) and to either the front or rear hemisphere (determined by some unknown spectral cue), then perceived elevation is determined by the frequency of the primary high-frequency notch in the head-related transfer function of the ear nearest the source.

Whatever plausibility of this model possesses rests on the observation that contours of equal notch frequency tend to intersect each cone of confusion only twice - once in the frontal hemisphere and once in the rear. This is generally true for moderate positive and negative elevations. Musicant and Butler [6] established that spectral features due to the filtering by the near ear are the dominant cues for resolving source position on the cone of confusion. Observations made in our laboratory and by Morimoto and Aokata [7] confirm that listeners are accurate in determining on which cone of confusion a source lies and that errors are primarily made in resolving position on the cone.

Using this model and measured notch patterns, two predictions pertaining to listeners' localization judgements were made. The first concerned the variance of the elevation responses and the second response bias under conditions of front-to-back or back-to-front confusion. The predictions were evaluated using free-field localization response data.

2 Methods

2.1 Subjects

Data were collected from six members of the Hearing Development Research Laboratory subject pool. There were three female and three male subjects ranging in age from 20 to 24. All reported normal hearing. For each subject head-related transfer functions were measured and free-field localization judgement data were collected as described below.

2.2 HRTF notch measurements

The procedure used to measure head-related transfer functions is described in detail by Wightman and Kistler [8]. Using probe-tube microphones positioned as close to the eardrum as possible, source-to-eardrum impulse responses were measured for positions spaced by 10° in both azimuth and elevation.

The location of the primary high-frequency spectral notch in each transfer function was located "by eye" on a computer screen plot of the spectrum and was marked using a mouse input device. Some judgement was required to select the desired notch; care was taken to follow particular features to higher elevations where they tended to peter out. The primary notch is visible in Figure 1, which shows directional transfer functions (HRTFs normalized by the diffuse-field response) as a function of elevation at 0° azimuth for subject SNF. Note the motion of the high-frequency notches as elevation increases. Since the extraction was a time-consuming task, the analysis was limited to positions spaced by 30° in azimuth and to elevations lying between -50° and +50°. This was done for both left and right ears and resulted in 264 data points for each subject.

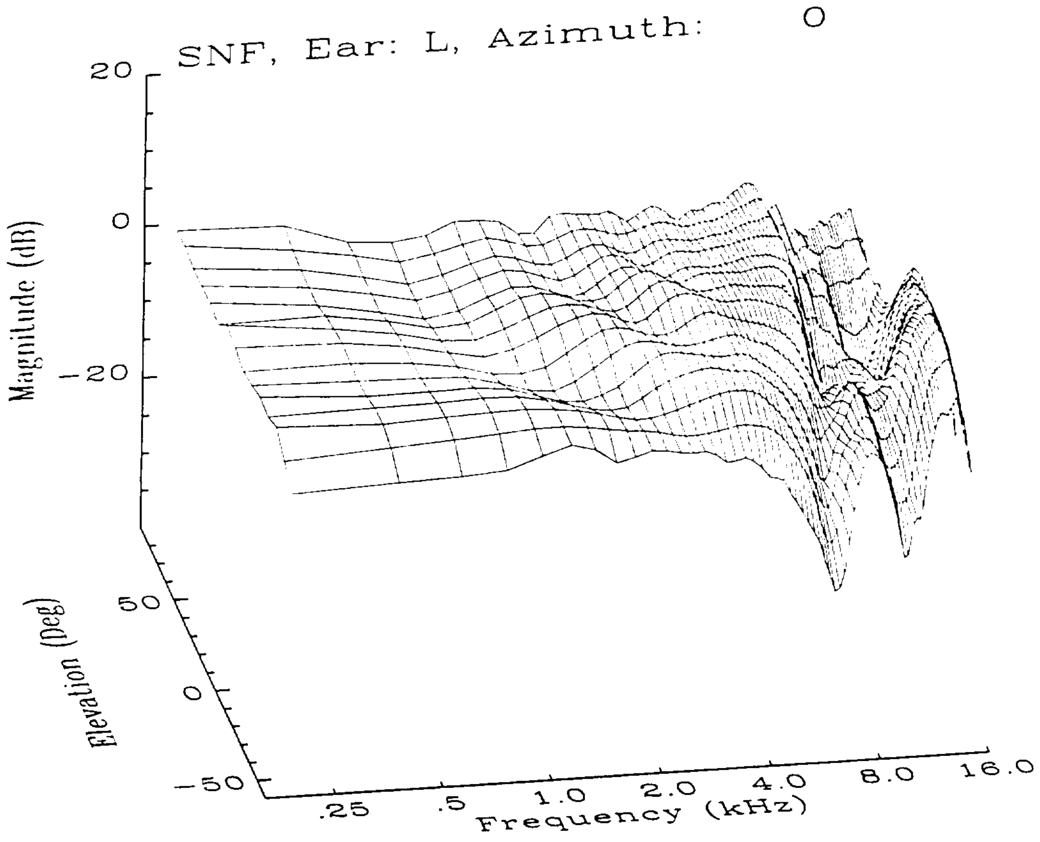


Figure 1: Directional transfer functions as a function of elevation at 0° azimuth for subject SNF.

2.3 Free-field judgements

Free-field localization judgements were collected with participants seated blindfolded in an anechoic chamber. Broadband (200-14000 Hz) noise bursts of 250 ms duration were played from loudspeakers mounted on a moveable arc. Subjects responded verbally with the azimuth and elevation of the perceived source location.

3 Individual notch frequency patterns

Contour plots of left-ear primary notch frequency as a function of direction are plotted in Figures 2-5 for four representative subjects. The dotted curves in these plots show the cones of confusion. Subjects SNF and SNX show approximately horizontal orientation of the notch contours on the ipsilateral side (negative azimuths). The contours for SNF are generally more closely spaced than those for SNX, revealing that notch frequency varies more slowly with position for the latter. Subjects SNT and SNY show upwards tilting of the contours towards the front. This is extreme in the case of SNY.

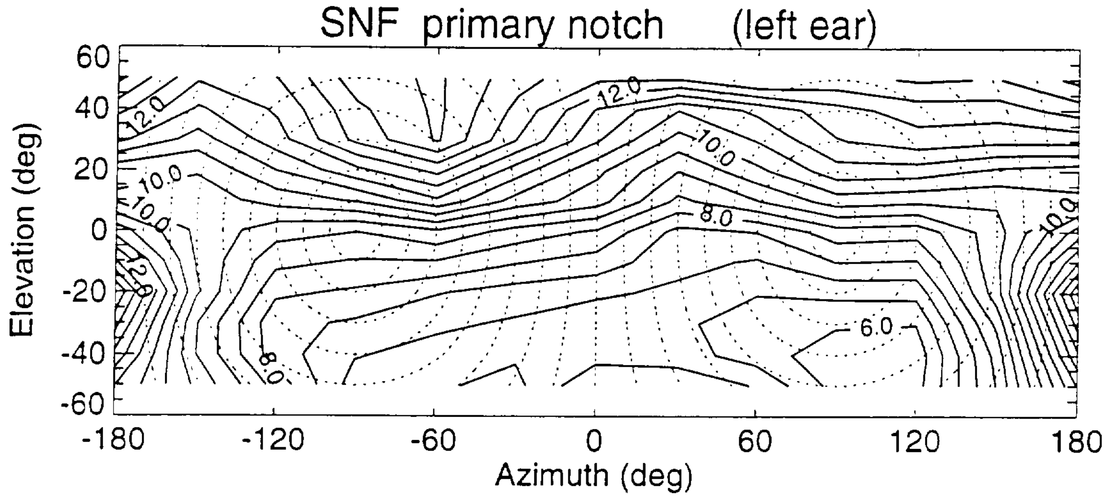


Figure 2: Contours of equal notch frequency (in kHz) for subject SNF. Dotted curves indicate cones of confusion.

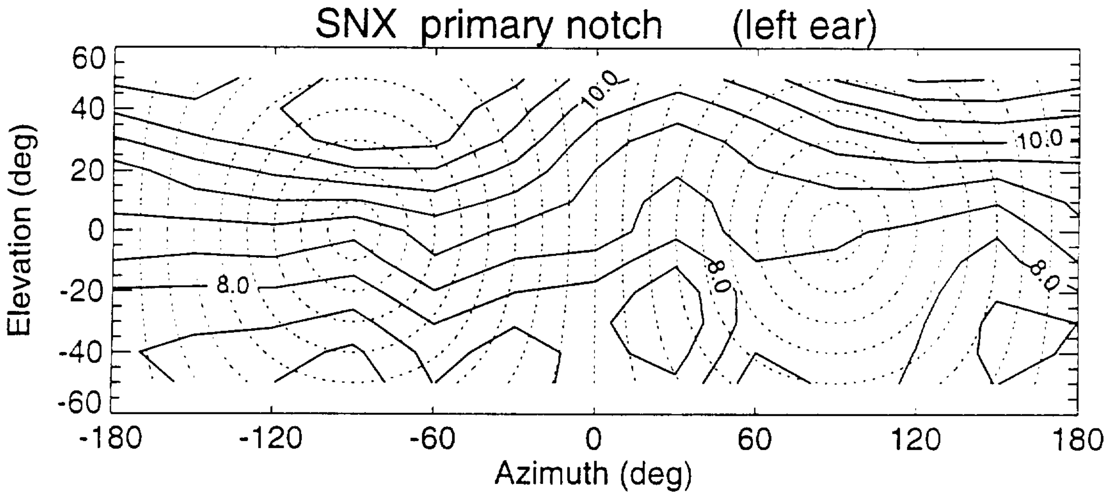


Figure 3: Contours of equal notch frequency for subject SNX.

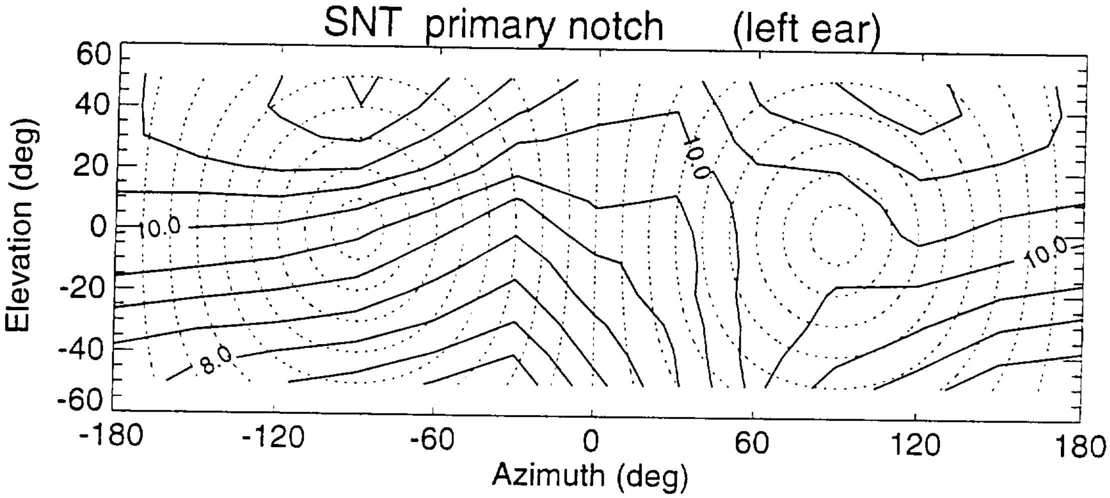


Figure 4: Contours of equal notch frequency for subject SNT.

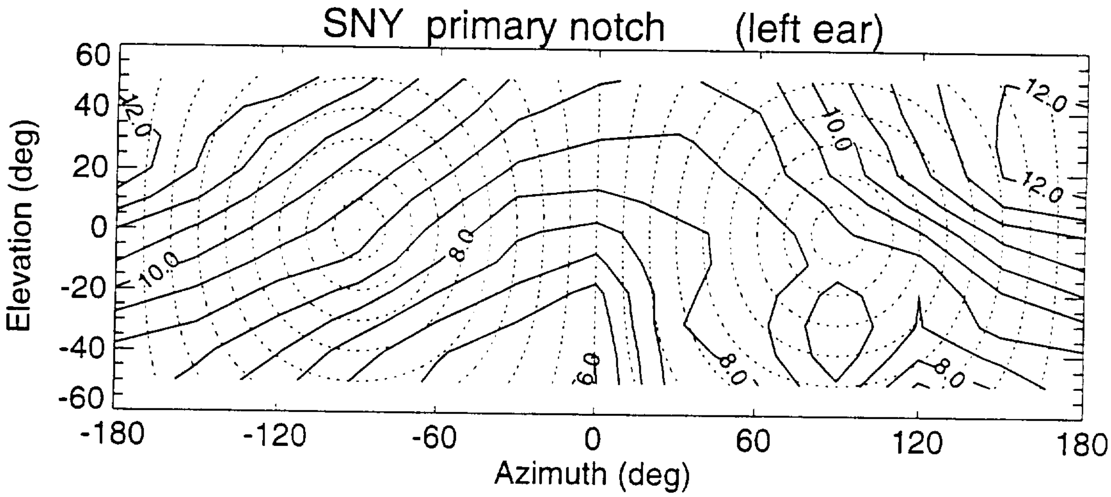


Figure 5: Contours of equal notch frequency for subject SNY.

4 Scatter of elevation judgements

4.1 Single-notch model prediction

The first prediction generated using the single-notch model concerns the relationship between the rate at which notches change frequency with position and the variance of the elevation component of subject's responses. If notch frequency determines elevation, then subjects for whom notches move more rapidly should show less spatial scatter in their responses to individual real source locations. Uncertainty about notch frequency should correspond to relatively greater uncertainty about elevation for subjects with "slow" notches.

4.2 Data analysis

To evaluate this prediction, some measure of the spatial dependance of notch frequency for each subject was required. The magnitude of the near-ear notch frequency gradient averaged over the region of the sphere under consideration was chosen as a suitable metric of overall notch "speed". This value (∇) was calculated over all available positions and also for positions within the region of the upper hemisphere lying between -30° and +30° (the high-front case). The latter region was considered to be of particular interest since, unlike in the coronal plane, elevation judgement is likely to be almost entirely spectrally-based near the median plane due to the near-zero values of interaural difference cues.

To characterize the degree of scatter in subjects' judgements, the standard deviation of the elevation responses elicited by each physical source position was calculated and then averaged over the region of interest, yielding the value σ. Only responses classified as unconfused were analyzed; those deemed to be examples of front-back, back-front, or up-down reversal were excluded. The number of responses remaining at each location ranged from 4 to 7. The criteria for these classifications are discussed in Section 5.2.

4.3 Results

The results of these analyses are presented in Table 1. Linear regression revealed the correlations between ∇ and σ to be -0.21 in the overall case and -0.61 in the high-front case.

5 Bias in Front-Back Confused Elevation Judgements

5.1 Single-notch model prediction

The second prediction made by the single-notch model concerns the effects of front-to-back and back-to-front confusions on elevation error. If, as in the cases of subjects SNT and SNY, the contours of constant notch frequency are tilted significantly away from the horizontal, and if the single-notch model is correct, then a front-back or back-front reversal should have a significant and consistent effect on elevation judgement errors. For example, it might be expected that SNY would experience over-elevation in cases of back-to-front reversal since the notch contours sweep upwards towards the front. Similarly, front-to-back reversals should be under-elevated.

Table 1. Averaged Standard Deviation of Elevation Judgements

Subject	All Positions		High-Front	
	∇ (Hz/deg)	σ (deg)	∇ (Hz/deg)	σ (deg)
SNF	67.1	13.0	91.6	11.3
SNJ	61.9	15.5	59.6	16.0
SNR	43.5	16.5	42.5	21.4
SNT	47.0	16.6	37.2	13.2
SNX	40.1	11.8	36.3	17.4
SNY	39.1	16.3	22.9	17.8
	correlation = -0.21		correlation = -0.61	

5.2 Data analysis

The available responses for each physical source location in the left hemisphere were classified as one of: correct front (F), correct back (B), front-to-back reversed (FB), back-to-front reversed (BF), or up-down reversed (UD). If a judged position lay closer to the real location when reflected in the coronal plane, it was deemed to be a BF or FB confusion. Errors of elevation of greater than 90° were classed as up-down confusions and were excluded from the analysis. The mean difference between the reported and actual elevations was calculated at each position, and then these mean differences were averaged over the region of interest. This procedure was carried out for subjects SNF, SNX, SNT, and SNY, who all had orderly notch patterns and made significant numbers of front-back reversals.

5.3 Results

The results are presented in Table 2, in which arrows in cells indicate the predicted direction of the bias. Both of the subjects with more horizontal contours tend to over-elevate, although their bias patterns differ. Subject SNT does show significant over-elevation of back-to-front reversed judgements, but also over-elevates sources correctly localized in the front. SNY, for whom the notches were even more strongly tilted, shows no significant bias in any condition. The striking result is that for both of these listeners there are no differences between confused and unconfused judgements.

Table 2. Elevation Judgement Bias (bias in degrees).

Response Type	Subjects with Horizontal Contours		Subjects with Tilted Contours	
	SNF	SNX	SNT	SNY
F	12.1	10.0	26.7	0.7
BF	17.1	1.1	↑ 24.7	↑ 2.7
B	8.4	6.1	10.3	-4.3
FB	-1.2	5.0	↓ 11.2	↓ 2.1

6 Discussion and conclusions

Although significant differences exist among the notch patterns for different subjects, attempting to predict localization behavior on the basis of these differences using the single-notch model cannot be termed a success. There appears to be no strong relationship between the average magnitude of the notch frequency gradient and response scatter either for the all-positions or the high-front case. Although the correlation coefficient of -0.61 is suggestive it is not a convincing result, and its magnitude is due mainly to one outlying point (subject SNF). There appears to be little evidence of a relationship when the quantities are averaged over all positions. It is not surprising that the observed correlations, while low, were in the appropriate direction since the rate of notch movement with position must be positively correlated with the rate of change of overall spectral shape with position.

In the case of front-back and back-front reversals, the predictions of the simple notch model were not observed. The two subjects (SNT and SNY) with tilted notch contours had very different error bias patterns and, more importantly, showed no effect of front-back reversals on elevation judgements.

The results of these analyses clearly do not support the single-notch model of elevation perception. The observed individual differences in notch-frequency variation do not yield strong predictive power for localization behavior when coupled with this model. Therefore, elevation judgements must depend on additional spectral cues which have yet to be identified and verified.

Acknowledgements

The author is grateful to Dr. Frederic Wightman and Dr. Doris Kistler for their advice and assistance. The tedious extraction of notch-frequency from the HRTF plots was done by Kristin Andersen. This research was supported in part by grants from NIH, NASA, and ONR.

References

[1] Wightman, F.L. and D.J. Kistler. "The Dominant Role of Low-Frequency Interaural Time Differences in Sound Localization." *Journal of the Acoustical Society of America* **91** (1992): 1648-1661.

[2] Rice, J.J., B.J. May, G.A. Spirou, and E.D. Young. "Pinna-Based Spectral Cues for Sound Localization in Cat." *Hearing Research* **58** (1992): 132-152.

[3] Neti, C., E.D. Young, and M.H. Schneider. "Neural Network Models of Sound Localization Based on Directional Filtering by the Pinna." *Journal of the Acoustical Society of America* **92** (1992): 3140-3156.

[4] Kuhn, G.F. "Physical Acoustics and Measurements Pertaining to Directional Hearing." In *Directional Hearing*, edited by W.A. Yost and G. Gourevitch, 3-25. New York: Springer-Verlag, 1987.

[5] Butler, R.A. and K. Belendiuk. "Spectral Cues Utilized in the Localization of Sound in the Median Sagittal Plane." *Journal of the Acoustical Society of America* **61** (1977): 1264-1269.

[6] Musicant, A.D. and R.A. Butler. "The Influence of Pinnae-Based Spectral Cues on Sound Localization." *Journal of the Acoustical Society of America* **75** (1984): 1195-1200.

[7] Morimoto, M. and H. Aokata. "Localization Cues of Sound Sources in the Upper Hemisphere." *Journal of the Acoustical Society of Japan* **5** (1984): 165-173.

[8] Wightman, F.L. and D.J. Kistler. "Headphone Simulation of Free-Field Listening I: Stimulus Synthesis." *Journal of the Acoustical Society of America* **85** (1989): 858-867.

Sound localization in varying virtual acoustic environments

Pavel Zahorik
Doris J. Kistler
Frederic L. Wightman
Waisman Center
University of Wisconsin - Madison
1500 Highland Avenue
Madison, WI 53705
zahorik@waisman.wisc.edu

Abstract

Localization performance was examined in three types of headphone presented virtual acoustic environments: an anechoic virtual environment, an echoic virtual environment, and an echoic virtual environment for which the directional information conveyed by the reflections was randomized. Virtual acoustic environments were generated utilizing individualized head-related transfer functions and a three-dimensional image model of rectangular room acoustics - a medium sized rectangular room (8m x 8m x 3m) with moderately reflective boundaries (absorption coefficient, $\alpha = 0.75$) being modeled. Five listeners reported the apparent position of a wide spatial range of virtual sound sources. Judgments of apparent source position were unaffected by acoustic environment manipulation even though sound sources presented in each of the three environments were informally discriminable. These findings question the necessity of spatialized room reflection information for high localization performance in virtual auditory displays, as well as provide further evidence for the robustness of precedence phenomena.

1 Introduction

In standard instantiations of headphone delivered three-dimensional auditory displays, errors in sound source localization may be roughly assigned to one of three categories:

1. Small judgment variation, or "blur", of apparent sound source position about target virtual source position.
2. Reversal of position judgment about the coronal plane - so called "front to back" or "back to front" reversals.
3. Judgment errors in degree of cranial externalization.

At present, precise explanation for the existence of these localization error types in three-dimensional auditory displays is unavailable. However, it seems clear that such auditory displays are in a number of senses not faithful to the reproduction of auditory stimulation occurring in natural, everyday situations. It therefore appears conceivable that localization errors in 3-D auditory displays are in some fashion a result of non-natural simulation.

One way in which standard 3-D auditory displays may be regarded as non-natural is the lack of reflection and reverberation simulation. Several studies have shown the inclusion of reflection information representative of indoor room environments affects localization errors. Specifically, Begault reports that for listeners localizing sounds in such virtual echoic environments constructed with nonindividualized head-related transfer functions (HRTFs), egocentric distance (or externalization) judgments increased by approximately a factor of three relative to localizing in virtual anechoic environments [1]. Durlach and his colleges [2] concur with Begault's findings, further claiming that it is most likely a decrease in direct-to-reverberant energy ratio, thought to be an important cue for the perception of auditory distance [3], that accounts for the increase in cranial externalization of auditory images presented with synthetic reflections. Interestingly, these benefits in externalization as a result of reflection simulation have been reported to be at the expense of increases in reversal errors [1]. It should also be noted that these synthetic reflection findings appear to challenge the classical notions of "precedence" as a purely echo suppressive mechanism [4].

Virtual simulation of echoic space involves three principle parameters in addition to those utilized by standard headphone 3-D auditory displays: reflection time delay, reflection attenuation (potentially frequency dependent), and reflection spatial position. Correct simulation of reflection spatial position is perhaps the most computationally demanding parameter. Hence, the greatest gains in implementational simplicity of virtual echoic space simulations would be realized by constraining this parameter in some sense. As a result of informal listening tests with virtual echoic environments constructed from nonindividualized HRTFs, Begault reports no difference in apparent source position between simulations where reflection spatial information is properly represented and simulations where reflection spatial information is chosen randomly [5]. Such results suggest that it in fact may not be necessary to properly simulate reflection spatial information in virtual echoic displays.

It is the goal of this study to further examine localization performance in virtual echoic environments with two principal additions. First, displays will be constructed with individualized HRTFs. Second, the echoic environment will be manipulated by varying the spatial information contained in the reflections. The latter addition will seek to formally determine the necessity of spatially correct reflection information for successful localization performance.

2 Method

2.1 Listeners

Three male and two female paid volunteers served as listeners. All had audiometrically verified normal hearing, as well as previous experience with localization judgment tasks.

2.2 Stimuli

Three classes of stimuli were used: virtual anechoic stimuli, virtual echoic stimuli, and virtual perturbed-echoic stimuli. The virtual anechoic stimuli were produced by filtering 250ms gaussian noise-bursts (chosen at random from a sample of 50 pre-computed noise bursts, then bandpass filtered from 200-14000 kHz and windowed with a 10ms ramp up/down raised cosine function) with left/right pairs of HRTFs corresponding to an array of 144 source positions. HRTFs were derived from individual listener probe-tube microphone measurements taken from 450 source positions in anechoic space (Wightman and Kistler provide a detailed description of this HRTF measurement procedure [6]).

Virtual echoic stimuli were constructed using a three dimensional image-source room acoustics model [7]. Such a model provides information as to the spatial position of each reflection (i.e. the incident angle of the reflection on the listener), as well as time delay and attenuation information. In this study, an 8m x 8m x 3m rectangularly shaped room with a centrally located listener was modeled. Each of the six reflecting surfaces were defined to have uniform frequency 0.75 absorption coefficients, α, which were independent of incidence angle. Loss of intensity due to distance of sound travel obeys the inverse square law in the acoustic free-field and was computed as such in this setting. Therefore, total attenuation of each reflection is a function of the number of reflector contacts and the total distance of sound wave travel. It should be noted the a variety of other room acoustic models exist. The image-model was chosen in this rectangular room setting for its simplicity and computational efficiency [8].

Binaural room impulse responses (BRIRs) were constructed from the information provided by the image-model. Specifically, right/left pairs of HRTFs (the time-domain equivalents thereof) corresponding to the appropriate spatial positions of the direct sound source path and each of its reflections were individually scaled and time-shifted the appropriate amounts, and then summed together. An interpolation algorithm was implemented when the spatial positions of reflections were disparate from measured HRTF positions. The resulting BRIRs were then convolved with the same type of noise-burst as described previously. In this study, the BRIRs were limited to include only the first 20 reflections occurring in time after the direct source path.

The third type of stimuli, the virtual perturbed-echoic stimuli, were constructed in a fashion analogous to the construction of the virtual echoic stimuli, but with one crucial difference. In this case the spatial positions of the reflections were chosen at random, rather than as prescribed by the image-model. All other stimulus parameters (including attenuations values, and time delay) remained the same as for the virtual echoic stimuli.

All stimuli were pre-computed and stored for subsequent experimental presentation over headphones.

2.3 Procedure

Three virtual acoustic conditions were presented: A baseline condition with the virtual anechoic stimuli described above, a "correct" reflection condition with the virtual echoic stimuli, and a random reflection condition with the virtual perturbed-echoic stimuli. Listeners verbally reported apparent sound source position in terms of azimuth, elevation and distance via a polar coordinate system. The three virtual acoustic conditions where presented in successive blocks of the same 144 virtual source positions. Order of presentation within a block was randomized. The 144 source positions were chosen at random from the possible 450 positions at which HRTF measurements were performed. Four replications in each of the virtual acoustic conditions yielded 576 judgments per condition for each listener.

3 Results

The three virtual acoustic environments examined here were found to have little effect on localization performance. Figures 1-3 display localization data from three representative listeners. Virtual source position is plotted as a function of apparent source position (for each of the experimental conditions) in three different transformed coordinated systems: right/left, front/back, up/down. The right/left dimension is determined by collapsing sources and judgments across both the front/back and up/down dimensions, such that a -90° angle is directly to the listener's left, a 0° angle directly in front of the listener, and a 90° angle to the listener's right. Front/back and up/down dimensions are determined analogously, by collapsing across the remaining two dimensions.

Listener	Baseline	Spatially Correct Reflections	Spatially Random Reflections
SMQ	0.2083	0.1424	0.1441
SNF	0.1892	0.1563	0.1319
SNJ	0.0677	0.0434	0.0522
SNX	0.1267	0.1094	0.0922
SNY	0.0838	0.0991	0.1270

Table 1: Reversal proportions

Listener	Baseline	Spatially Correct Reflections	Spatially Random Reflections
SMQ	5.20	5.03	5.03
SNF	3.58	3.73	3.72
SNJ	3.06	3.00	3.00
SNX	6.08	6.08	6.03
SNY	2.65	2.90	2.90

Table 2: Distance judgments (ft.)

Symbol shading is proportional to the number of judgments at a given position. Visual examination of Figures 1-3 suggests the existence of little within-subject difference across experimental condition. Front-Back and Back-Front reversals may be seen on Figures 1-3 as judgments lying on or near the negative diagonal (i.e. y = -x) of the Front-Back dimension panel.

Table 1 displays combined Front-Back and Back-Front reversal rates for each listener in each acoustic condition. A within-subjects ANOVA on the arcsine transformed reversal rates (a recommenced transformation for small proportional scores [9]) revealed no significant differences across experimental conditions, $F(2,4) = 1.95$, $p = .204$.

Table 2 shows listener's distance judgments for each condition. Results of a within-subject ANOVA suggest that distance judgments were also unaffected by experimental condition, $F(2,4) = 0.17$, $p = .844$.

4 Conclusion

These null results are perhaps somewhat surprising, given both the findings of Begault and others, and the fact that the stimuli presented in these three virtual acoustic conditions, upon subjective evaluation, were markedly different. It is not inconceivable to attribute these differences to, at least in part, the use of individualized HRTFs, since it has been shown that the use of nonindividualized HRTFs (such as [1] and [5]) suffers from both a degradation in externalization and an increase in reversal error rates [10]. Therefore, it is quite possible that the lack of increase in distance judgments, as well as reversal error rates, for echoic conditions as compared to anechoic conditions is a result of the use of individualized HRTFs. It should be noted that the constancy of reversal error rates across experimental conditions is in fact an encouraging result when compared to the results of previous studies.

Regardless of cause, a clear difference in results between this study and previous studies exists. Localization performance, in terms of apparent sound source position, has been shown to be quite robust with respect to the varied virtual acoustic environments examined. Therefore, if particular applications of 3-D auditory displays are concerned only with localization performance, and individualized HRTFs are available, two conclusions exit: Reflection spatial information need not necessarily be realistic, and further, such reflection information is perhaps wholly unnecessary.

Acknowledgments

Financial support provided in part by grants from NIH, NASA and ONR.

References

[1] Begault, D. R. "Perceptual Effects of Synthetic Reverberation on Three-Dimensional Audio Systems." *Journal of the Audio Engineering Society* 40 (1992): 895-904.

[2] Durlach, N. I., W. S. Woods, A. Kulkarni, H. S. Colburn, and E. M. Wenzel. "On the Externalization of Auditory Images." *Presence* **1** (1992): 251-257.

[3] Mershon, D. H. and L. E. King. "Intensity and Reverberation as Factors in the Auditory Perception of Egocentric Distance." *Perception and Psychophysics* **18** (1975): 409-415.

[4] Blauert, J. *Spatial Hearing.* Cambridge, MA: MIT Press, 1983.

[5] Begault, D. R. "Binaural Auralization and Perceptual Veridicality." *Journal of the Audio Engineering Society* (1992): Preprint 3421 (M-3).

[6] Wightman, F. L. and D. J. Kistler. "Headphone Simulation of Free-Field Listening. I: Stimulus Synthesis." *Journal of the Acoustical Society of America* **85** (1989a): 858-867.

[7] Allen, J. B. and D. A. Berkley. "Image Method for Efficiently Simulating Small-Room Acoustics." *Journal of the Acoustical Society of America* **65** (1979): 943-950.

[8] Vorlander, M. "Simulation of the Transient and Steady-State Sound Propagation in Rooms Using a New Combined Ray-Tracing/Image-Source Algorithm." *Journal of the Acoustical Society of America* **86** (1989): 172-178.

[9] Kirk, R. E. *Experimental Design*, Second edition. Monterey, CA: Brooks/Cole, 1982.

[10] Wenzel, E. M., M. Arruda, D. J. Kistler, and F. L. Wightman. "Localization Using Nonindividualized Head-Related Transfer Functions". *Journal of the Acoustical Society of America* **94** (1993): 111-123.

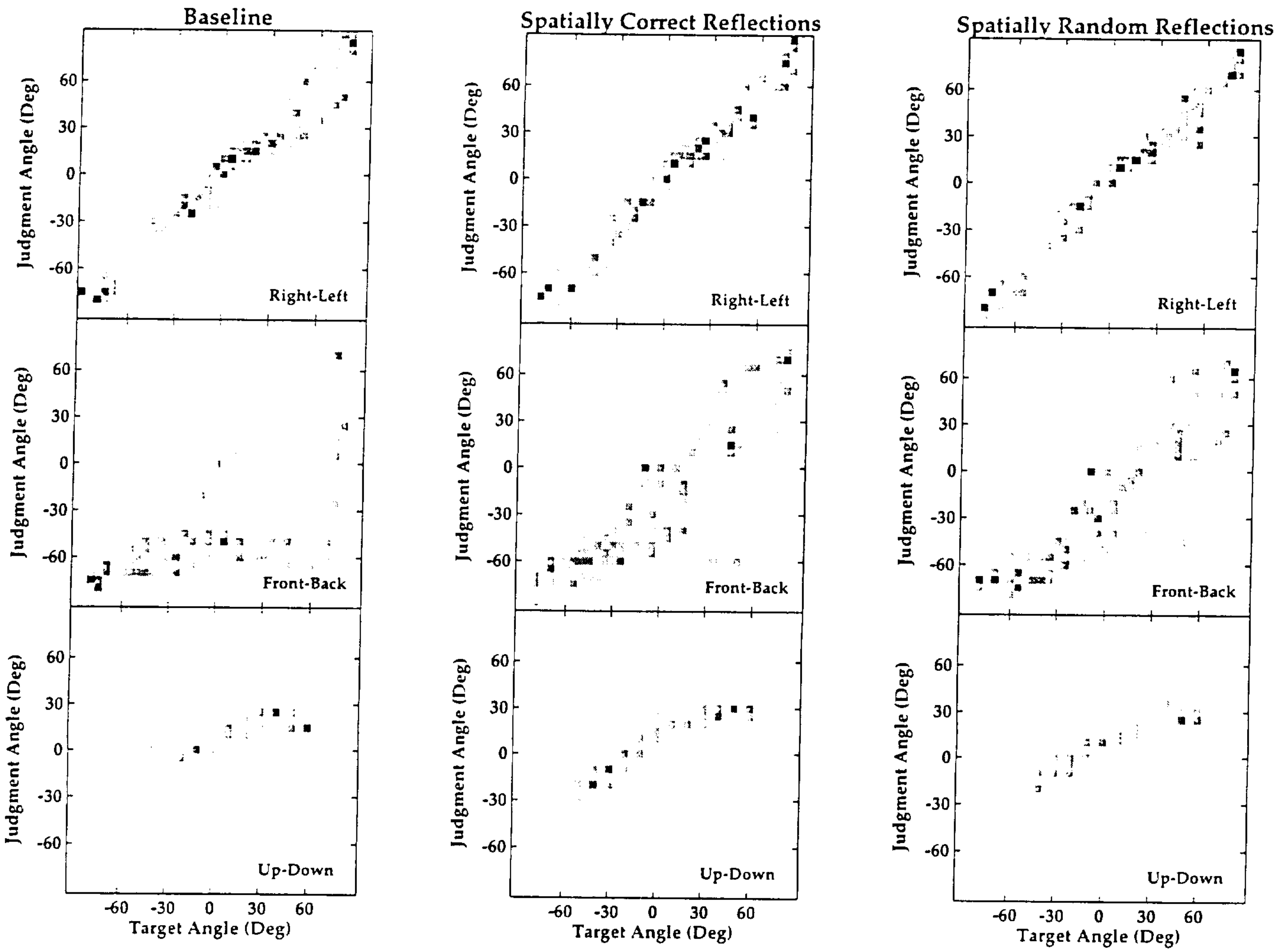


Figure 1: Localization data for listener SMQ.

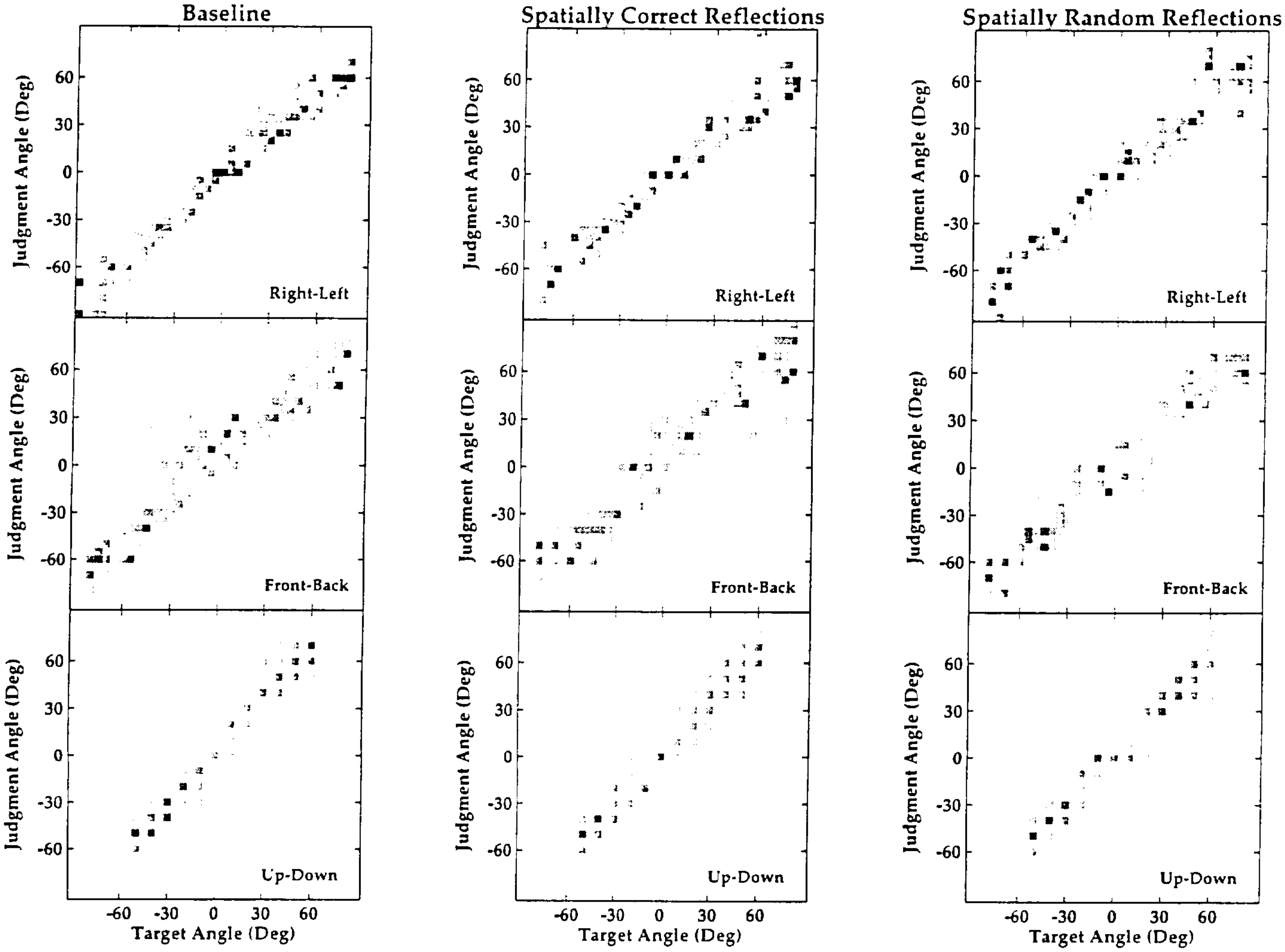


Figure 2: Localization data for listener SNJ.

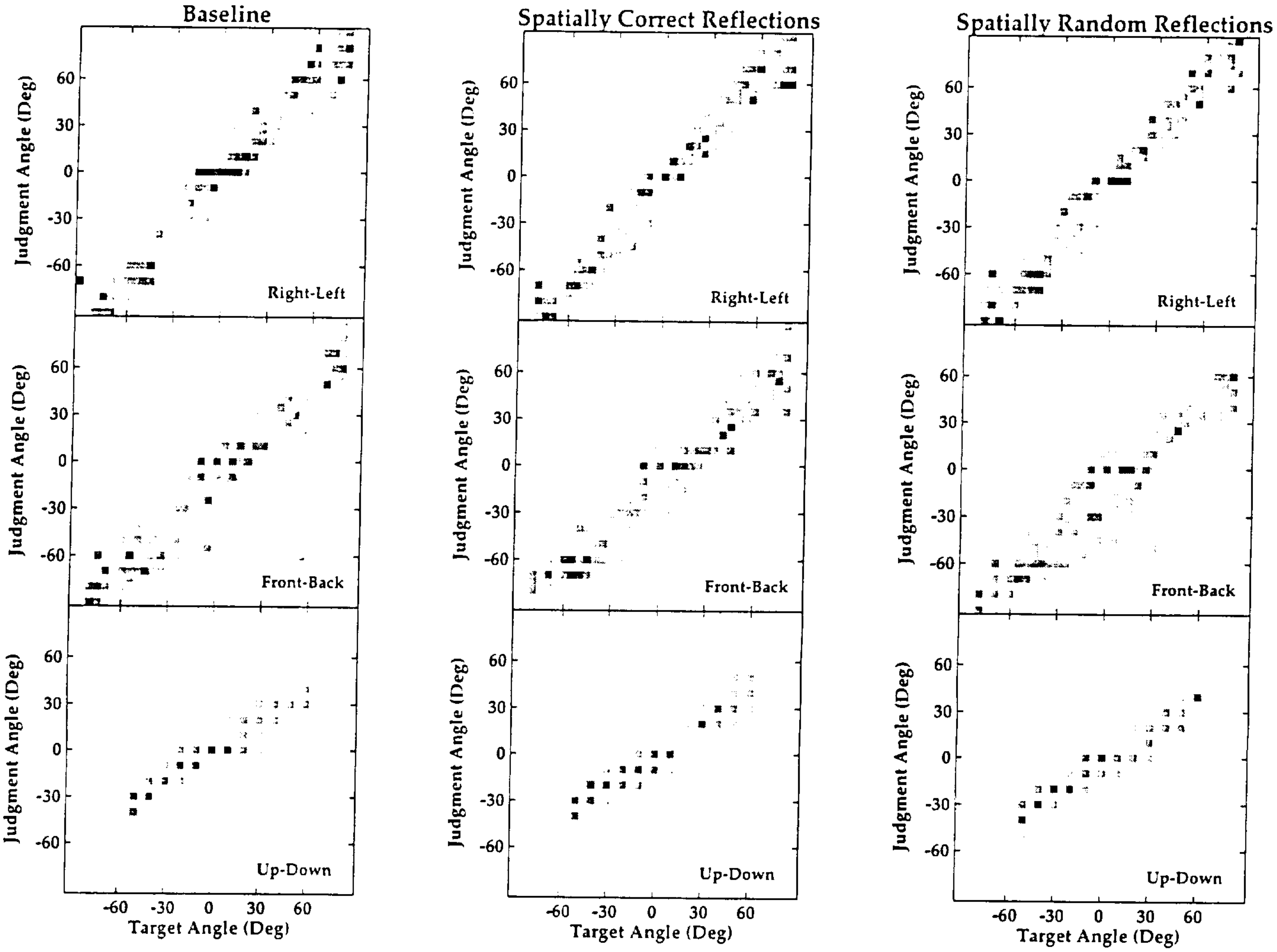


Figure 3: Localization data for listener SNY.

FACTORS AFFECTING THE RELATIVE SALIENCE OF SOUND LOCALIZATION CUES

Frederic L. Wightman
Waisman Center and Department of Psychology
University of Wisconsin, Madison
Madison, WI 53706

and

Doris J. Kistler
Waisman Center
University of Wisconsin, Madison
Madison, WI 53706

Note: This is a draft of a chapter that will appear in Binaural and Spatial Hearing, edited by R. Gilkey and T. Anderson, and published by Lawrence Erlbaum, Hillsdale, NJ. Please do not cite or quote.

I. Introduction

This chapter is about the relative salience of the acoustical cues to apparent sound source position. It consists of a rather loose collection of hypotheses and data. Most of the data come from our own experiments but a few are from the work of others; some of the data are shown here for the first time, but many have been presented elsewhere. Our discussion focusses on the factors that influence the apparent position of a sound source. Very little attention is given here to the discriminability of sounds from different spatial positions or to the accuracy with which listeners can identify the true spatial origin of a sound source.

We begin with a brief review of the potential acoustical determinants of apparent position and follow with some educated guesses about which of these might be more or less salient in various listening conditions. We conclude by discussing the results of several experiments in which listeners indicated the apparent positions of sounds that had been modified to isolate the contributions of one or more of the potential cues.

II. Acoustical Determinants of Apparent Position

Given the extensive treatment of this topic elsewhere (e.g., Middlebrooks and Green, 1991; Wightman and Kistler, 1993; Shaw and Duda chapters in this volume), most readers will be quite familiar with the acoustical determinants of apparent sound position, which we will call localization cues. Thus, there is little need to review them here. However, at the risk of being repetitious, we will discuss the cues from a slightly different perspective in order to emphasize a few simple points.

In our view, a potential acoustical localization cue is any physical aspect of the acoustical waveforms reaching a listener's ears that is altered by changes in the position of the sound source relative to that of the listener. For our purpose here we will limit our discussion to the direction component of relative position and ignore the distance component. Given this limitation, a taxonomy of potential cues can be described as in Table 1. The temporal-spectral distinction represented in Table 1 is artificial, given the isomorphism between a waveform and its spectrum. However, since the auditory mechanisms thought to subserve temporal and spectral processing are different, we find it useful to consider the two kinds of cues separately. The monaural-binaural distinction is included to emphasize the fact that changes in sound source position produce changes in the waveform at each ear individually (monaural), as well as changes in the relation between the waveforms at the two ears (binaural).

Consider the monaural cues first. The monaural temporal cue is the position-dependent change in the waveform at one ear caused by the change in the impulse response of the acoustical system consisting mostly of the head and pinna. The transfer function of this system is usually called the head-related transfer function or HRTF for short. Fig. 1 shows the impulse-response of the HRTF from a listener's left ear for two source positions. Note that there are substantial differences in the temporal fine structure of the two impulse responses. Some investigators suggest that this temporal fine structure, in particular the time differences among the major peaks, provides important information about sound source position that is extracted directly from the stimulus waveform by the auditory system (e.g., Batteau, 1967).

There are at least two reasons why such monaural temporal cues are not likely to be relevant for human sound localization. First, since the HRTF impulse responses are short, on the order of about 2 ms, the limited temporal resolving power of the auditory system, also about 2 ms, probably renders the temporal fine structure of the impulse responses undetectable (Green, 1971). Second, the results of a psychophysical experiment (Kistler and Wightman, 1992) suggest that changes in the temporal fine structure of the HRTF impulse responses do not produce subsequent changes in the apparent positions of sound sources. In this study, listeners judged the apparent positions of virtual sound sources presented via headphones (Wightman and Kistler, 1989a, 1989b). The virtual sources were synthesized using HRTFs that had been measured on the same listeners. In one condition of the experiment the HRTFs used to produce virtual sources were modelled as minimum-phase systems, thus producing the same amplitude spectrum as the measured HRTFs but different phase spectra and hence different impulses responses. The apparent positions of sources synthesized using minimum-phase HRTFs were indistinguishable from the positions of sources synthesized from measured HRTFs. While it was not reported in that paper, an additional condition tested the effect of using linear-phase HRTFs. The impulse responses of linear-phase HRTFs were quite different from either the minimum-phase or measured impulse responses, yet apparent position judgments were unaffected. We conclude that, to a first approximation, monaural temporal cues are unimportant.

The monaural spectral cues are the well known direction-dependent changes in the pattern of spectral peaks and valleys superimposed on an incoming stimulus by the filtering action of the pinna. In other words, they are the direction-dependent changes in the amplitude spectrum of the HRTF. These changes are large and systematic, as can be seen in Fig. 2, which shows HRTF magnitude functions recorded from two listeners at a single source azimuth and several elevations. The prominent spectral notch between 5 kHz and 10 kHz, which moves in a regular way as source elevation changes, is thought by some to be an important cue for source elevation (Rice, May, Spirou, and Young, 1992; Musicant and Butler, 1984). While there is little doubt that spectral peaks and notches such as those shown in Fig. 2 are detectable (Moore, Oldfield, and Dooley, 1989), their role in sound localization is not yet clear.

For monaural spectral cues to be generally useful, a listener must have some knowledge not only of the relevant HRTF features and how they vary with source position, but also of the spectral characteristics of the sound source itself. It might be reasonable to assume that listeners commit to memory the important features of their own HRTFs. However, since the spectrum of the signal received at each ear is the product of the HRTF and the source spectrum, in order for a listener to recover the HRTF and compare it to a remembered template, the source spectrum must be known in advance. The requirement for a priori knowledge about the source spectrum can be mitigated by assuming that most real-world sounds have wideband spectra that are locally smooth (Zakarauskas and Cynader, 1993). However, the proportion of real-world sound spectra that meets the locally smooth criterion has yet to be determined. Fig. 3 shows amplitude spectra of six real-world sounds and illustrates our conviction that the wide variability among such sounds precludes many simplifying assumptions about their spectral characteristics.

There are two additional characteristics of the monaural spectral cues that might bear on their utility. First, they are highly idiosyncratic. Fig. 4 illustrates this point by showing the directional features of the HRTFs from 10 listeners for one ear and a single source position. These "directional transfer functions" or DTFs are computed by dividing each HRTF by the RMS average of the HRTFs from all directions measured. Note that in certain frequency regions the differences in the DTFs from one listener to another are as great as 20 dB. This suggests that the specific strategies used to obtain source position information from the spectral shape of HRTFs may vary from one listener to another. Second, the monaural spectral cues exist only at high frequencies, as might be expected given the dimensions of the pinnae. A principal components analysis of the DTFs from 10 listeners and a large number of spatial positions produces basis functions that are essentially flat up to 5 kHz (Kistler and Wightman, 1992). Since each DTF can be represented as a weighted sum of these basis functions, we can conclude that the directional components of the HRTFs themselves are essentially flat up to 5 kHz. Thus, the utility of monaural spectral cues will depend both on adequate high-frequency content in the sounds to be localized and adequate high-frequency sensitivity on the part of the listener.

The binaural cues are presumed to be derived by some kind of differencing operation on the information retrieved from each ear. How this might be accomplished in the nervous system is not our concern here, so for the purposes of simplicity we will assume the binaural cues are derived from a ratio of the HRTFs at the two ears. Because the spectrum of the sound source appears in both numerator and denominator of this ratio, it cancels. Thus the utility of the binaural cues does not depend critically on the characteristics of the source or on the listener's a priori knowledge of them.

Interaural time difference (ITD) is related to the phase of the HRTF ratio, and is generally thought to be one of the most important localization cues. To a first approximation, the ITD is the same at all frequencies. While the ITD in measured HRTFs is higher at low frequencies (below 1.5 kHz) than at high frequencies

(Wightman and Kistler, 1989a), the observed low-frequency increase in the ITD is not as large as the 50% increase expected on theoretical grounds (Kuhn, 1977). Our view is that the larger ITD at low frequencies is perceptually irrelevant. Psychophysical evidence of this can be found in the results of the experiment reported by Kistler and Wightman (1992), in which listener's judged the apparent positions of sources in which the ITD was either natural or constant across frequency. The patterns of judgments in the two conditions were indistinguishable.

Fig. 5 shows the ITD cue for two listeners. For these plots the ITD was estimated by computing the time delay at the maximum in the cross-correlation between left and right HRTF impulse responses at each spatial position. Note that the change in the ITD with changes in source position is smooth and roughly the same for the two listeners. Note also that the contours of constant ITD are roughly circular, in agreement with theoretical predictions made by assuming the head is a rigid sphere. The consequence of constant ITD contours is that a given ITD indicates not just one but a whole locus of potential source positions. We will return to both of these details later.

Interaural level difference (ILD), derived from the amplitude of the HRTF ratio, is a complicated function of frequency since for any given source position the peaks and valleys in the HRTF occur at different frequencies in the two ears. Moreover, The ILD is small at low frequencies, regardless of source position, because the dimensions of the head and pinna are small compared to the wavelengths of sound at frequencies below about 1500 Hz. For these reasons, we suggest that The ILDs in individual frequency bands are much more likely to be useful localization cues than overall ILD. Fig. 6 shows The ILDs in various frequency regions derived from the HRTF measurements obtained from a typical listener. Note that the ILDs in the low-frequency band are small, regardless of source position. Note also the complexity of the pattern of ILDs in the high-frequency bands. While the overall pattern of ILDs in each of the bands is similar, there is sufficient detail in each one so that extraction of useful localization cues would require that listeners remember the details of the pattern. Otherwise, The ILD can provide only coarse information about source position, and even that is likely to be ambiguous, since like the ITD a given ILD indicates a whole locus of potential source positions.

III. Factors that influence the salience of the cues

In this section we present the results of experiments that reveal the stimulus or listener factors that appear to determine the relative importance or salience of the various cues. Four factors will be considered: 1) the reliability or consistency of the cue across stimulus conditions, listeners, and frequency; 2) a priori knowledge of stimulus characteristics; 3) the frequency content of the stimulus; and 4) the plausibility or realism of the cue.

A. Methodological Considerations

Most of the experiments described in this section were conducted in our laboratory, so a brief review of our psychophysical procedures may be useful here. The essential elements of the methods by which we generate and present stimuli and ask listeners to indicate the apparent spatial position of the sound source have been published elsewhere (Wightman and Kistler, 1989a, 1989b, 1992; Kistler and Wightman, 1992), so only an outline will be given here.

1. Listeners: With few exceptions, the listeners in our research are University of Wisconsin undergraduate students who serve 4-6 hours per week over long periods

of time and are paid an hourly rate for their services. They are always blindfolded before being led into the testing room, which is either an anechoic chamber or a small soundproof room. The blindfolds are kept in place the entire time the listeners are in the testing room. The listeners receive minimal training (2 hours at most) before data collection. The only purpose of the training is to familiarize the listeners with the response procedures.

2. Stimuli: The standard stimulus in our research is a 250-ms burst of Gaussian noise with a nominally flat spectrum between 200 Hz and 14 kHz. In some conditions the spectrum of the stimulus is "scrambled" by assigning the spectrum level within each critical band randomly, drawing from a uniform distribution with either a 20-dB or 40-dB range. This manipulation assures a very different stimulus spectrum on each trial, thus reducing the possibility that listeners will learn stimulus characteristics. In any one experiment stimuli are presented from a large number of real or virtual spatial positions all around the listener. The set of potential positions includes 24-36 azimuths (from -180° to +170°) and 6-10 elevations (from -50° to +60°). The stimuli are delivered either through small loudspeakers (Realistic Minimus 0.35) or headphones. The virtual source stimuli are synthesized using the standard FIR digital filtering techniques described in previous publications (e.g., Wightman and Kistler, 1989a).

3. Responses: Listeners report the apparent position of each stimulus verbally. Apparent azimuth and elevation are given in degrees, in accordance with standard single-pole world coordinates (the "North" and "South" poles are above and below the listener and the "equator" defines the horizontal plane that passes through the ears). Apparent distance is reported in feet. No feedback of any kind is given, except that

when listeners appear to make a large sign error, for example, reporting a negative azimuth (left side) for a positive azimuth (right side) source, they are asked if they are sure they made the intended response. In any one condition, listeners make between 600 and 1000 responses at the rate of about 2 per minute.

4. Data handling: Because of the difficulties in dealing with front-back confusions we make no attempt here to generate summary statistics or measures of central tendency from our data. Thus, the figures show raw data; each and every response is represented on the figures. For ease of interpretation we represent the data in a 3-pole coordinate system (Kistler and Wightman, 1992). The result is that each response (azimuth, ϕ, and elevation, θ) appears on three different plots. The azimuth component (ϕ) of each response is decomposed into a left-right component (λ) and a front-back component (ψ) according to the following equations:

$$\lambda = \arcsin(\cos\theta\sin\phi)$$

$$\psi = \arcsin(\cos\theta\cos\phi)$$

The elevation component of each response (θ) becomes the up-down component without transformation.

B. Cue reliability or consistency:

There are several dimensions on which one might rate the "reliability" of a localization cue. Among them are the extent to which the cue depends on source characteristics, provides the same information in all bands across the frequency spectrum, is roughly the same from listener to listener, and is unambiguous. Our view is that a reliable cue will contribute more to the determination of apparent source position than a less reliable cue, and in situations in which cues conflict a reliable cue will be dominant.

Given the set of cues described earlier in this chapter and our criteria for reliability, the ITD cue would seem to score the highest. The ITD does not depend on source characteristics, provides roughly the same information in each frequency band, and the relationship between the ITD and source position is not highly idiosyncratic. However, as mentioned above, the cue is ambiguous since a given ITD indicates a range of potential source positions. This is an issue to which we will return shortly.

A published experiment in which the ITD cue conflicted with the other localization cues revealed the dominance of the ITD cue (Wightman and Kistler, 1992). Listeners judged the apparent positions of virtual sources in which the ITD signalled one position and all other cues signalled another position. As long as the wideband noise stimulus contained low-frequency energy the listeners' judgments were completely determined by the ITD cue. In other words, listeners judgments always indicated the position signalled by the ITD cue, even when, for example, all other cues pointed to a position on the opposite side of the head. When low frequencies were removed from the stimulus, by highpass filtering above about 1500 Hz, the dominance of the ITD cue was eliminated, and listeners' judgments seemed to be determined by the other cues, ILDs and the monaural spectral cues.

In the experiments on ITD dominance, as well as in several other experiments involving localization of both real and virtual sources, some listeners made frequent front-back confusions (Wightman and Kistler, 1989b, 1992; Kistler and Wightman, 1992). We believe that these front-back confusions reflect not only the ambiguity of the ITD cue but also the dominance of that cue. While the ILD cues are also ambiguous, the contours of constant ILD and hence the confused positions are different in each frequency band. Thus, it seems unlikely that the source of front-back confusions is ILD ambiguity. In fact, one might argue that since the pattern of The ILDs across frequency is not ambiguous it could actually serve as a cue for resolving front-back confusions.

The pattern of ILDs across frequency is also a reliable localization cue in that it does not depend critically on stimulus characteristics. However, the facts that The ILDs are prominent only at high frequencies and are highly idiosyncratic (Fig. 6) may detract from their utility as localization cues. There is some evidence that The ILDs may be used primarily to resolve front-back confusions, as suggested above. In an unpublished conflicting cue experiment similar to the one described above (Wightman and Kistler, 1992), listeners localized virtual sources in which the pattern of ILDs was "zeroed," by using the leading ear's HRTF magnitude to synthesize both left and right ear stimuli. In addition, the spectrum of the noise stimulus was scrambled in this condition to prevent listeners from using monaural spectral cues. Since the ILD manipulation affected only the magnitude of the filters used to synthesize the virtual sources, the ITD cues were undisturbed. Fig. 7 shows typical results from this condition along with baseline results from a condition in which all the cues were intact. Note that the consequences of setting the ILD cue to zero in all bands were to increase front-back confusions and to decrease the range of elevation judgments. The latter effect is observed in the data from only about half of the listeners. There is no hint of an overall bias of the judgments toward the median plane (0° on the left-right plot) as would be expected if the ILD cue were contributing significantly to the apparent position judgments.

On our scale of cue reliability the monaural spectral cues are clearly the least reliable. They are highly idiosyncratic and their utility depends critically on a listener's a priori knowledge of source characteristics. The impact of a listener's knowledge or expectations about source characteristics is the topic of the next section.

C. The role of a priori knowledge of source characteristics

To the extent that apparent source position depends on the binaural cues (ITD and ILD), the characteristics of the source should be irrelevant. The source spectrum cancels in the HRTF ratio from which the ITD and ILD are derived. The evidence presented above suggests that indeed, the binaural cues are the most salient. Nevertheless, there is no doubt that the monaural spectral cues, which are influenced by source characteristics, contribute in important ways to the determination of apparent source position.

One experiment that reveals the importance of monaural spectral cues involves a comparison between the apparent positions of sources with scrambled spectra and the apparent positions of comparable sources with flat spectra. Fig. 8 shows typical results from a single listener presented with flat-spectrum stimuli in free field (top left) and virtual free field (top right), and with scrambled-spectrum stimuli in free field (bottom left) and virtual free field (bottom right). In the scrambled-spectrum conditions, the free field sources were scrambled over a 40-dB range and the virtual free field sources were scrambled over a 20-dB range. Note that the effects of scrambling the source spectrum on each trial are an increase in front-back confusions and distortions of elevation perception. If only binaural cues were important, there should be no effect of scrambling the source spectrum. Other data from a variety of scrambled-spectrum conditions indicate that, as shown in Fig. 8, it seems to require more scrambling in free field than in virtual free field to reveal comparable effects. A possible reason for this is the absence of cues provided by normal head movements in the virtual source conditions. We will return to this issue later in the chapter.

In monaural listening conditions, in which one ear is plugged and covered with a muff (for free-field presentations) or in which the signal to one earphone is turned off (for virtual free-field presentations), the binaural cues to apparent source position are distorted. It might be expected that in such conditions listeners asked to localize sound sources would rely more completely on the monaural spectral cues. It is not surprising, then, that scrambling the source spectrum has much more dramatic effects on apparent source position judgments in monaural listening conditions. Fig. 9 illustrates this point. Note that while some traces of source localizability remain in the flat-spectrum condition (judgments clustered around major diagonal), all evidence is gone in the scrambled-spectrum condition. The fact that all of the judgments in the monaural scrambled-spectrum condition are within 25° of the horizontal plane is curious and is a result we cannot readily explain.

D. Source frequency content

Accurate sound localization is possible only with wideband sound sources. For a source consisting of a sinusoid or a narrow band of noise, the apparent position and actual position are rarely coincident and often very far removed from one another. There are many reasons for our inability to localize narrowband sources. Narrowband stimuli provide an impoverished and typically ambiguous set of cues, since neither the pattern of ILDs across frequency nor the monaural spectral cues are available. This issue has recently received considerable attention elsewhere (see chapters by Butler and by Middlebrooks in this volume; Middlebrooks and Green, 1991;

Middlebrooks, 1992; Wightman and Kistler, 1993), so we will not deal with it here. Rather, we will consider the importance of specific frequency regions.

The experiment on ITD dominance discussed earlier (Wightman and Kistler, 1992) suggests that the salience of the ITD cues diminishes at high frequencies. On the other hand, the spectral cues (the ILDs in the various frequency bands and monaural spectral cues) might be expected to be more salient at high frequencies since it is there that these cues are acoustically more robust. An experiment in which listeners judged the apparent positions of filtered sound sources suggests that one way the high-frequency information is used is to resolve front-back confusions. Fig. 10 shows apparent position judgments from a typical listener presented with wideband virtual sources (left) and sources with the frequencies from 5 kHz to 10 kHz removed with a bandstop filter. The most significant effect of the filtering seems to be an increase in front-back confusions. While not shown here, the effect of lowpass filtering at 5 kHz is quite similar.

D. The role of cue realism or plausibility

The extent to which the constellation of localization cues presented to listeners matches their experience and expectation has significant effects on the apparent positions of sounds and on the relative weight assigned to the various cues. The results of several experiments we have conducted using virtual sound sources suggest that those cues that are unnatural or unusual are generally weighted less in the determination of apparent source position.

Some evidence on this point comes from experiments in which listeners hear sounds as if "through someone else's ears" (Wenzel, Arruda, Kistler, and Wightman, 1993). The virtual sources in these experiments are synthesized using HRTFs from a different listener than the one judging the apparent positions of those virtual sources. In such conditions one might expect that the ITD cues in the stimuli would match closely the ITD cues normally experienced by the listener (assuming comparable head sizes), but that the ILD and spectral cues would be very different. The most obvious consequence of listening "through someone else's ears" is a dramatic increase in front-back confusions (Wenzel et al., 1993). We feel that this result reflects the fact that the spectral cues normally used to resolve front-back confusions are given less weight because they are unusual or unnatural.

In everyday listening, sound sources produce localization cues that are "consistent" across the frequency spectrum. In other words, because the sounds originate from a real source, the position indicated by the ITD, the ILD, and the monaural spectral cues is the same (with the natural ambiguities, of course) regardless of the frequency band considered. ITD, for example, is roughly the same at 500 Hz as it is at 5000 Hz. With real sources a situation could not occur in which The ILD in one frequency region indicated a source on one side of the head and The ILD in another frequency region indicated a source on the other side. Such sources can be easily synthesized, however, and a listener's judgments of their apparent positions can be revealing.

In our research on the cue realism issue, we studied the apparent positions of virtual stimuli in which cues in one frequency region conflict with cues in another frequency region. In one condition, for example, the ILD and spectral cues were the same throughout the frequency range (200 Hz -14000 Hz), and indicated one of five possible directions on the horizontal plane. The ITD cue in each of four bands of equal width on a log scale (roughly 1.5 octaves wide) indicated a different direction. Thus, the ITD cue was "inconsistent" across the frequency range and the ILD and spectral cues were "consistent". In other conditions the ITD cue was consistent and the other cues inconsistent.

The results were the same for all 5 listeners tested and were unambiguous. The apparent position judgments always followed the consistent cue. Even if the ITD cue was inconsistent only in a single high-frequency band (above 5 kHz), listeners appeared to ignore the ITD altogether and put maximum weight on the ILD and spectral cues, which were consistent across the spectrum. This is an important result. It suggests not only that "realistic" cues are given greater weight than “unrealistic” cues but also that high-frequency ITD cues can be just as important as low-frequency ITD cues. In this condition, the fact that the high-frequency ITD cue was different from the low-frequency ITD cue was recognized and apparently led the listener to ignore both ITD cues.

IV. Additional cues - resolution of front-back confusions

Many of the experimental manipulations we have described in this chapter have produced an increase in the frequency with which listeners make front-back confusions. Scrambling the source spectrum, removing the high-frequency energy from the source, and listening to unfamiliar spectral cues all increased the front-back confusion rate in our listeners. The obvious conclusion from these results is that the cues provided by source familiarity and high-frequency content are normally used by listeners to resolve confusions. However, there remains the problem that even in our free-field listening conditions, when the whole suite of cues is available, including normal ITDs, ILDs, and spectral cues, some listeners still make large numbers of front-back confusions. Fig. 11 shows one example. Since there is no evidence that these individuals are handicapped by their localization errors in real life, we conclude that source familiarity and high-frequency content are not the only stimulus parameters that facilitate resolution of confusions and that in everyday listening additional cues must be used.

There are several differences between our free-field testing environment and everyday listening situations. The most obvious difference is that our environment lacks the echoes and reverberation present in nearly all everyday listening settings. We tested the influence of normal echoes by adding the first 20 reflections from a simple rectangular room to our normal virtual source stimuli. There was no change in front-back confusion rate.

The primary acoustical difference between sources in the front and sources in the rear appears in the frequency range between 3 kHz and 7 kHz. Fig. 12 illustrates this difference by showing averaged HRTF magnitude functions for front and rear sources. We reasoned that emphasizing the acoustical difference between front and rear sources might allow better front-rear distinction and lower confusion rate. To emphasize front-rear differences we squared the magnitude of the HRTFs used to synthesize virtual sources. Listeners' judgments of the apparent positions of the spectrally emphasized sources did not show any decrease in front-back confusion rate.

In all our previous work, involving both free-field and virtual-source conditions, listeners are asked not to move their heads. Thus, the usual changes in the localization cues that accompany head movements were not available. Since there are good reasons to believe that information from the changes in localization cues could be used to resolve confusions (e.g., Wallach, 1940), we have begun an experiment to assess the role of head movements. In this experiment listeners localize virtual sources (2.5s wideband noise bursts) in two conditions. In one, the virtual stimuli are presented over headphones, and the listeners are asked not to move their heads during the test. This condition is identical to our usual virtual-source condition except that the stimulus is longer. In the second condition, using the same stimuli, listeners are encouraged to move their heads during stimulus presentation if they feel it would facilitate localization. A magnetic head tracker is used to sense head position and the virtual synthesis algorithm is modified according to the head tracker's reports in real time, using a Convolvotron (Foster, Wenzel, and Taylor, 1991), in order to simulate a stationary external source. Apparent position judgments are made verbally after each stimulus presentation. Preliminary results from a single listener are shown in Fig. 13. Note that in the head-stationary condition this listener makes frequent front-back confusions, as evidenced by the off-diagonal responses in the "front-back" panel. These data are from the same listener whose free-field judgments are shown in Fig. 11. In the head-movement condition, however, the front-back confusions are nearly eliminated.

The preliminary results of this experiment strongly suggest that among the additional cues we have considered, those provided by head movements can be important. It appears that head movements should be viewed as a natural and important component of the sound localization process. Future research designed to assess the salience of the other cues, ITD, ILD, and spectral cues will need to acknowledge the importance of the dynamic information provided by head movements and to appreciate the situations in which this information might be important.

V. Conclusions

The main point we have tried to emphasize here is that the apparent position of a sound source is determined by much more than just the low-frequency ITDs and high-frequency ILDs highlighted in Lord Rayleigh's original Duplex Theory (Strutt, 1907). Many other cues are involved, such as monaural spectral cues, and the relative contributions of the cues seem to be determined by a variety of stimulus and listener factors, including stimulus dynamics, source familiarity, listener expectations, and cue plausibility. While the general outline of a comprehensive theory of sound localization is beginning to emerge, many important questions remain unanswered and many details are missing. Modern technology has only recently given us the tools needed to address those questions and to fill in the details through systematic, controlled research. We can expect rapid progress in the years ahead.

TABLE 1 - POTENTIAL ACOUSTICAL LOCALIZATION CUES

	TEMPORAL	SPECTRAL
MONAURAL	Monaural Phase (*Batteau*)	1) Overall Level 2) Monaural Spectral Cues
BINAURAL	Interaural Time Difference (ITD)	1) Interaural Level Difference (ILD) 2) Binaural Spectral Differences

Figure Legends

Fig. 1: Examples of HRTF impulse responses recorded from a listener's left ear for two source positions on the listener's left side.

Fig. 2: DTFs (HRTFs divided by the RMS of HRTFs from all directions) recorded from two listeners and sources at 90° azimuth.

Fig. 3: Magnitude spectra of six "everyday" sounds.

Fig. 4: DTFs recorded for a source located at 90° azimuth and 0° elevation from the right ear of 10 listeners.

Fig. 5: ITD measured from two listeners plotted as contours of constant ITD (in μs) on a globe. Listeners are faced toward a "longitude" of 0°, and the "equator" or 0° latitude describes the plane passing through the ears.

Fig. 6: ILDs in three different frequency bands derived from the HRTFs measured from a single listener. The "floor" of each panel shows the contours of constant ILD.

Fig. 7: Apparent position judgments from an experiment in which the ILD and ITD cues were set in conflict. The left panels show data from the condition in which cues were normal. The right panels show the results of setting the ILD to 0 dB at all frequencies. All responses are shown in each panel. The darkness of the data point indicates the proportion of possible judgments in that area. Front-to-back confusions are revealed in the "front-back" panels by negative judgments at positive target angles.

Fig. 8: Apparent position judgments with flat-spectrum stimuli (top panels) and scrambled-spectrum stimuli (bottom panels). The stimuli were presented either in free field (left) or virtual free field (right).

Fig. 9: Apparent position judgments with monaural free-field presentation. The stimuli had either flat spectra (left) or scrambled spectra (right). In the case of scrambled spectra, the range of scrambling was 40 dB.

Fig. 10: Apparent position judgments with bandstop stimuli. The left panels show data from a baseline condition in which the wideband stimulus had a scrambled (on average flat) spectrum. The right panels show data from the condition in which the scrambled spectrum stimuli had energy between 5 kHz and 10 kHz removed by sharp bandstop filtering.

Fig. 11: Apparent position judgments from a single listener presented with flat-spectrum stimuli in free field. Note the large number of front-back confusions in the "front-back" panel.

Fig. 12: Averaged HRTF magnitude functions (12 listeners) for sources in the front (solid line, sources between -30° and 30° azimuth and -40° and 40° elevation), and for sources in the rear (dashed line, sources between -150° and 150° azimuth and -40° and 40° elevation).

Fig. 13: Apparent position judgments with wideband flat-spectrum virtual sources in two conditions. The data in the panels on the left are from the stationary-head condition and the data in the panels on the right are from the head-movement condition.

References

Batteau, D. W. (**1967**). "The role of the pinna in human localization," Proc. R. Soc. London, Ser. B, **168**, 158-180.

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

Green, D. M. (**1971**). "Temporal auditory acuity," Psych. Rev. **78**, 540-551.

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

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

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

Middlebrooks, J. C., and Green, D. M. (**1991**). "Sound localization by human listeners," Annu. Rev. Psychol. **42**, 135-159.

Moore, B. C. J., Oldfield, S. R., and Dooley, G. J. (**1989**). "Detection and discrimination of spectral peaks and notches at 1 and 8 kHz," J. Acoust. Soc. Am. **85**, 820-835.

Musicant, A. D., and Butler, R. A. (**1984**). "The influence of pinnae-based spectral cues on sound localization," J. Acoust. Soc. Am. **75**, 1195-1200.

Rice, J. J., May, B. J., Spirou, G. A., and Young, E. D. (**1992**). "Pinna-based spectral cues for sound localization in cat," Hear. Res. **58**, 132-152.

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

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

Wenzel, E. M., Arruda, M., Kistler, D. J., and Wightman, F. L. (**1993**). "Localization using nonindividualized head-related transfer functions," J. Acoust. Soc. Am. **94**, 111-123.

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

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

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

Wightman, F. L., and Kistler, D. J. (**1993**). "Sound Localization," in *Springer Series in Auditory Research: Human Psychophysics*, edited by W. A. Yost, A. N. Popper, and R. R. Fay (Springer-Verlag, New York), Chap. 5, pp. 155-192.

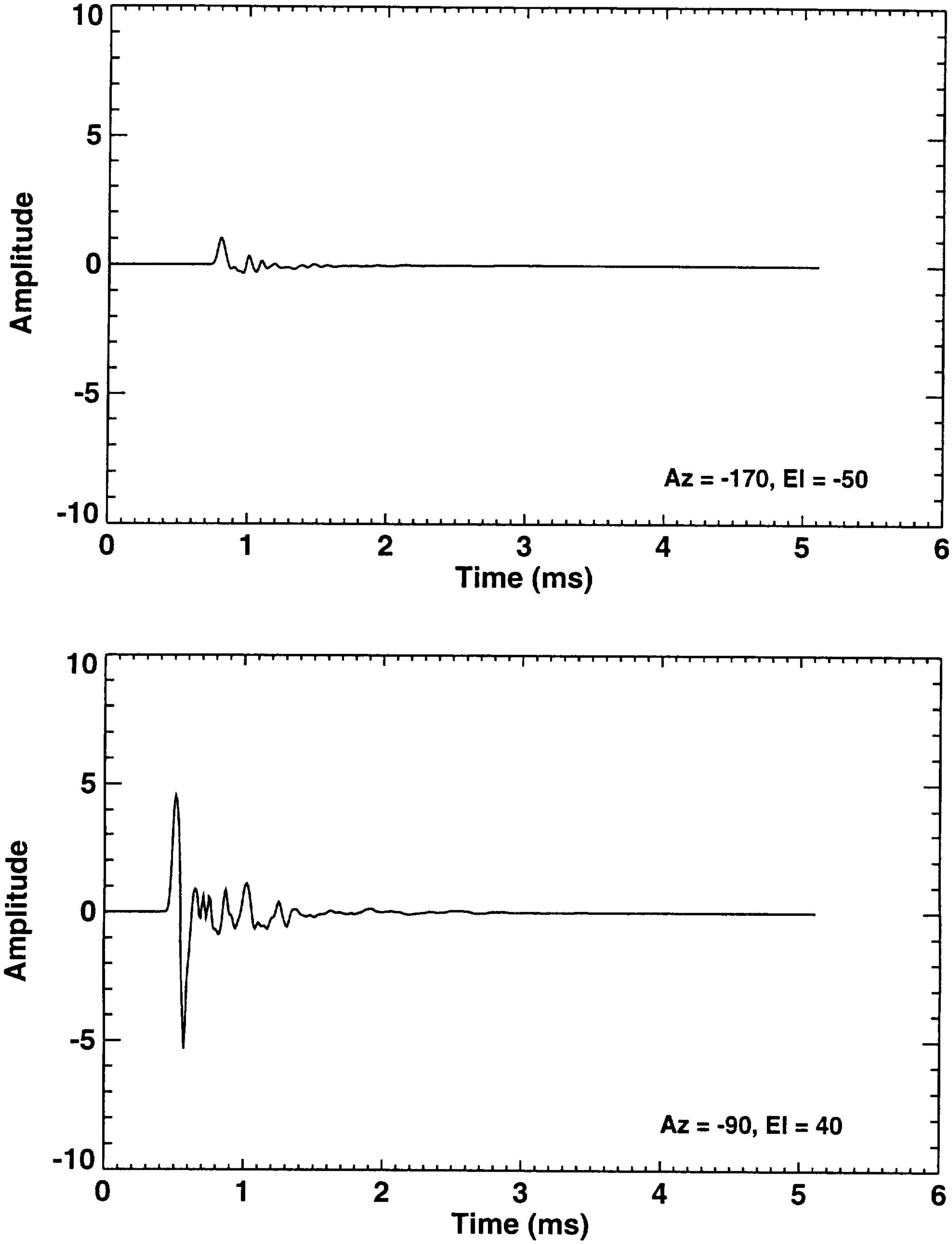
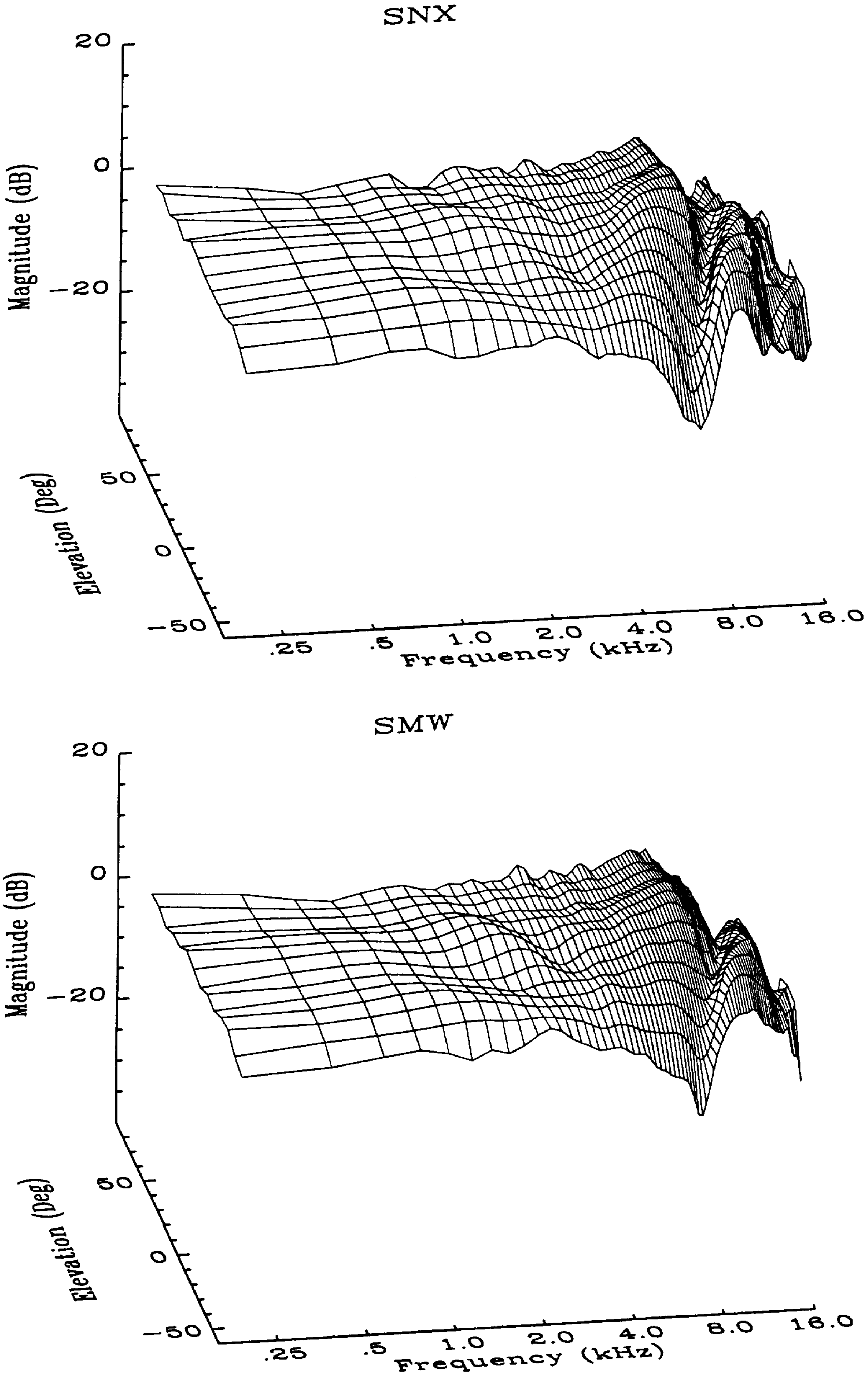

10
5
0
-5
-10
Amplitude
0
1
2
3
4
5
6
Time (ms)
Az = -170, El = -50
10
5
0
-5
-10
Amplitude
0
1
2
3
4
5
6
Time (ms)
Az = -90, El = 40

Fig 1



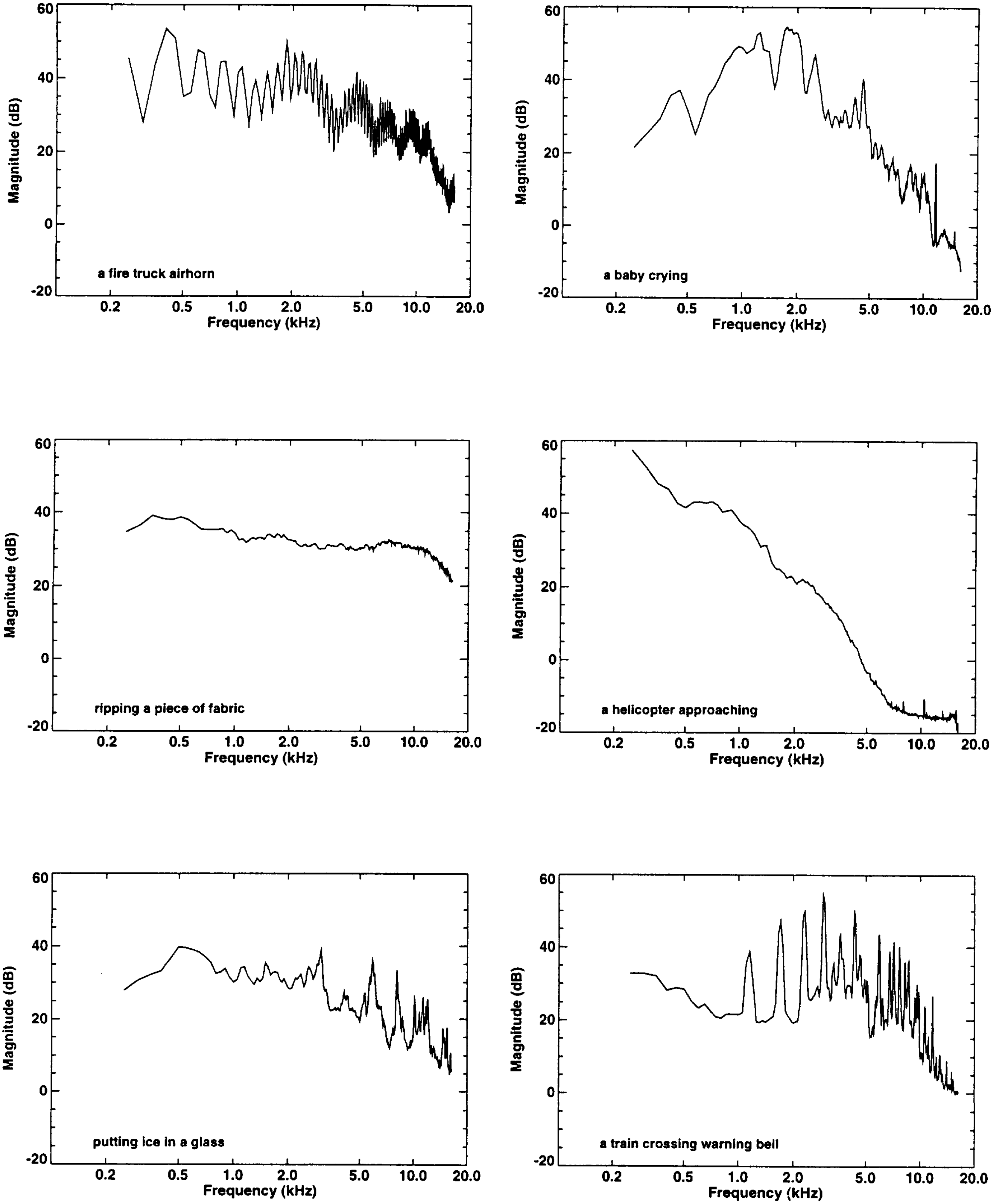
SNX

20
0
−20
Magnitude (dB)
50
0
−50
Elevation (Deg)
.25
.5
1.0
2.0
4.0
8.0
16.0
Frequency (kHz)

SMW

20
0
−20
Magnitude (dB)
50
0
−50
Elevation (Deg
.25
.5
1.0
2.0
4.0
8.0
16.0
Frequency (kHz)

Fig 2



60
40
20
0
-20
Magnitude (dB)
0.2
0.5
1.0
2.0
5.0
10.0
20.0
Frequency (kHz)
a fire truck airhorn
a baby crying
ripping a piece of fabric
a helicopter approaching
putting ice in a glass
a train crossing warning bell

Fig 3

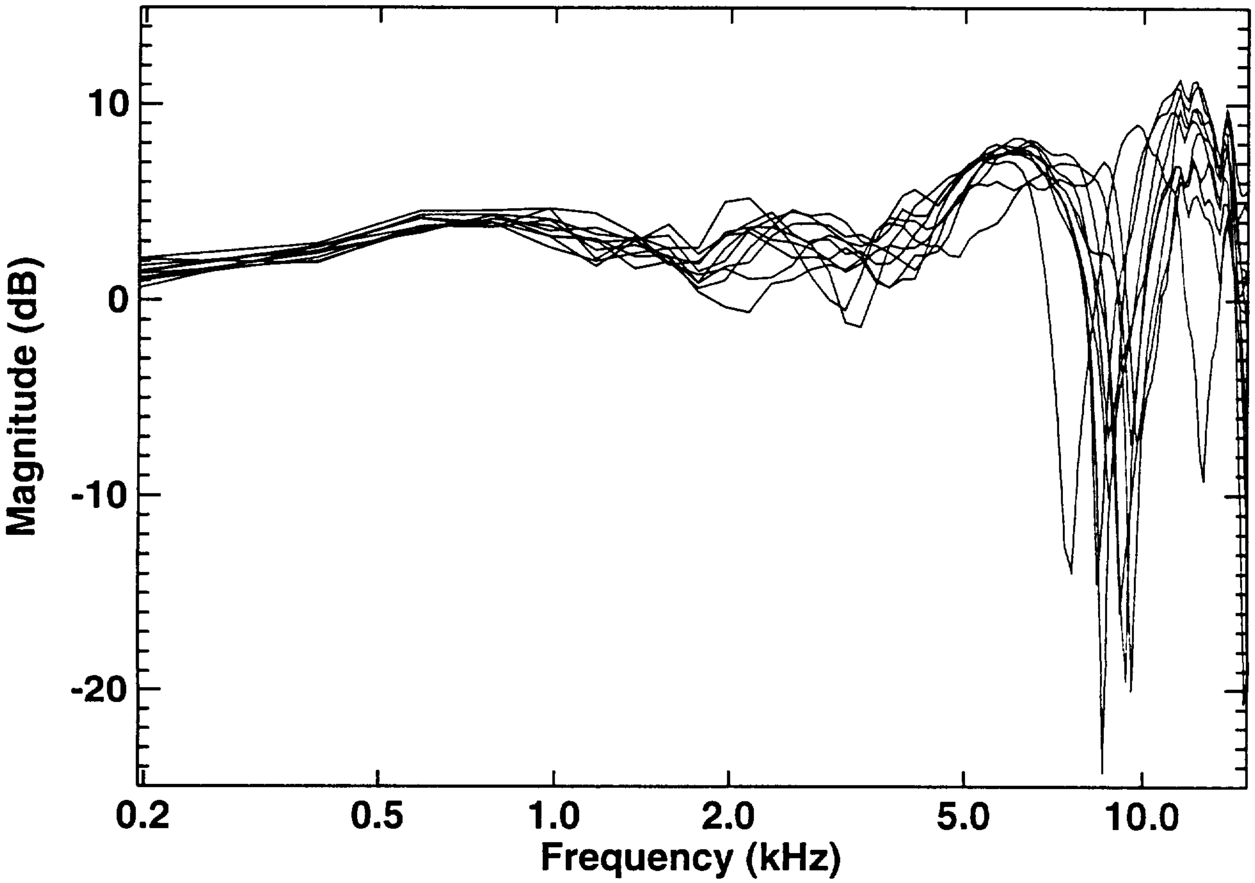


Fig 4

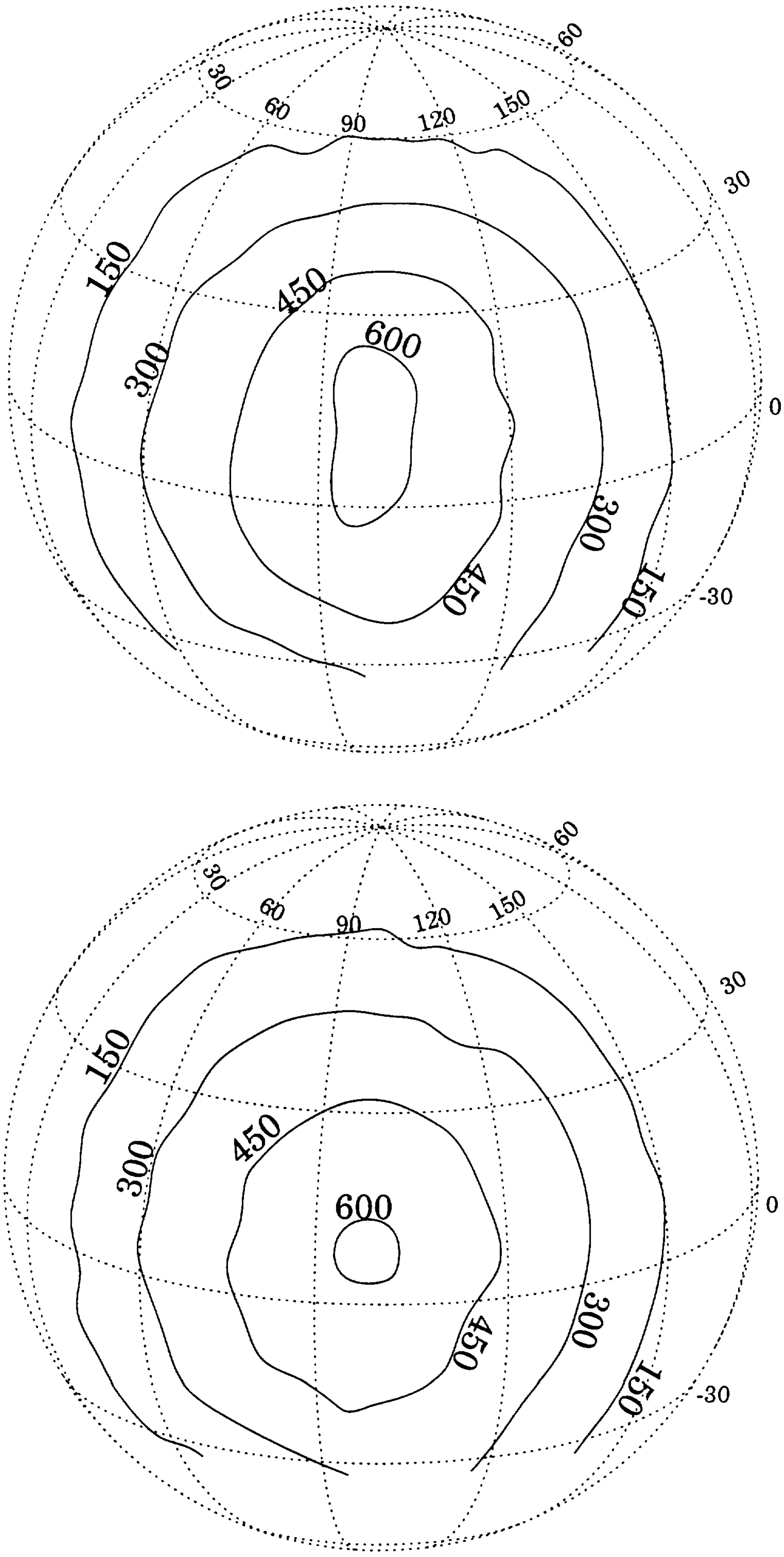

60
30
60
90
120
150
30
0
-30
150
450
300
600
300
450
150
60
30
60
90
120
150
30
0
-30
150
450
300
600
300
450
150

Fig 5

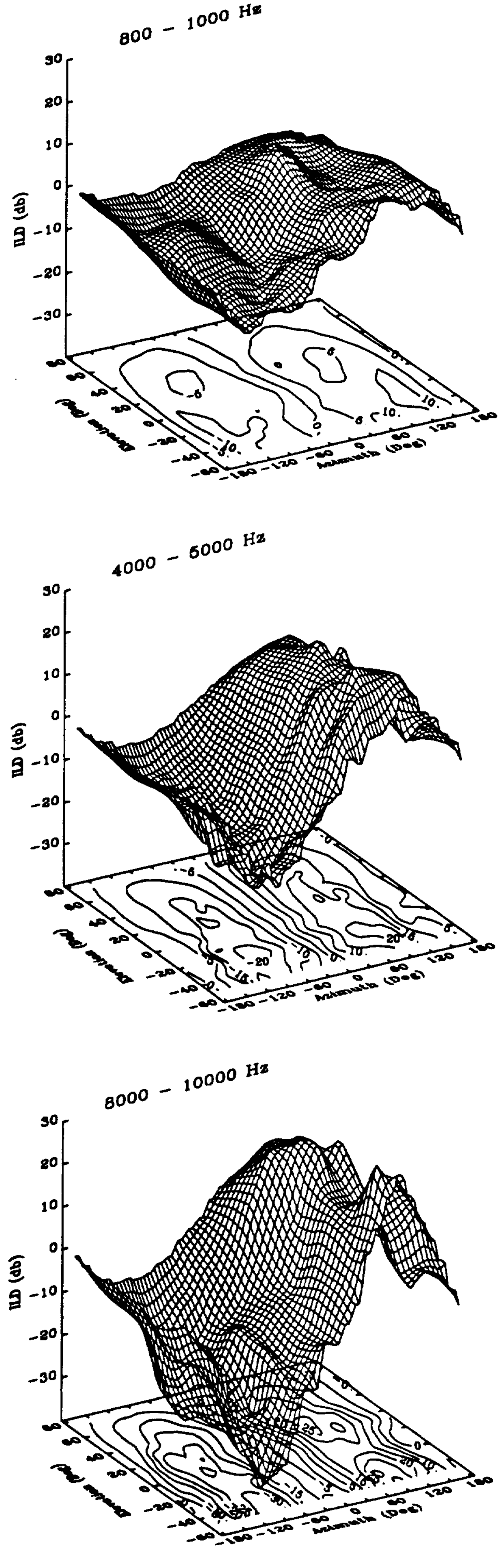
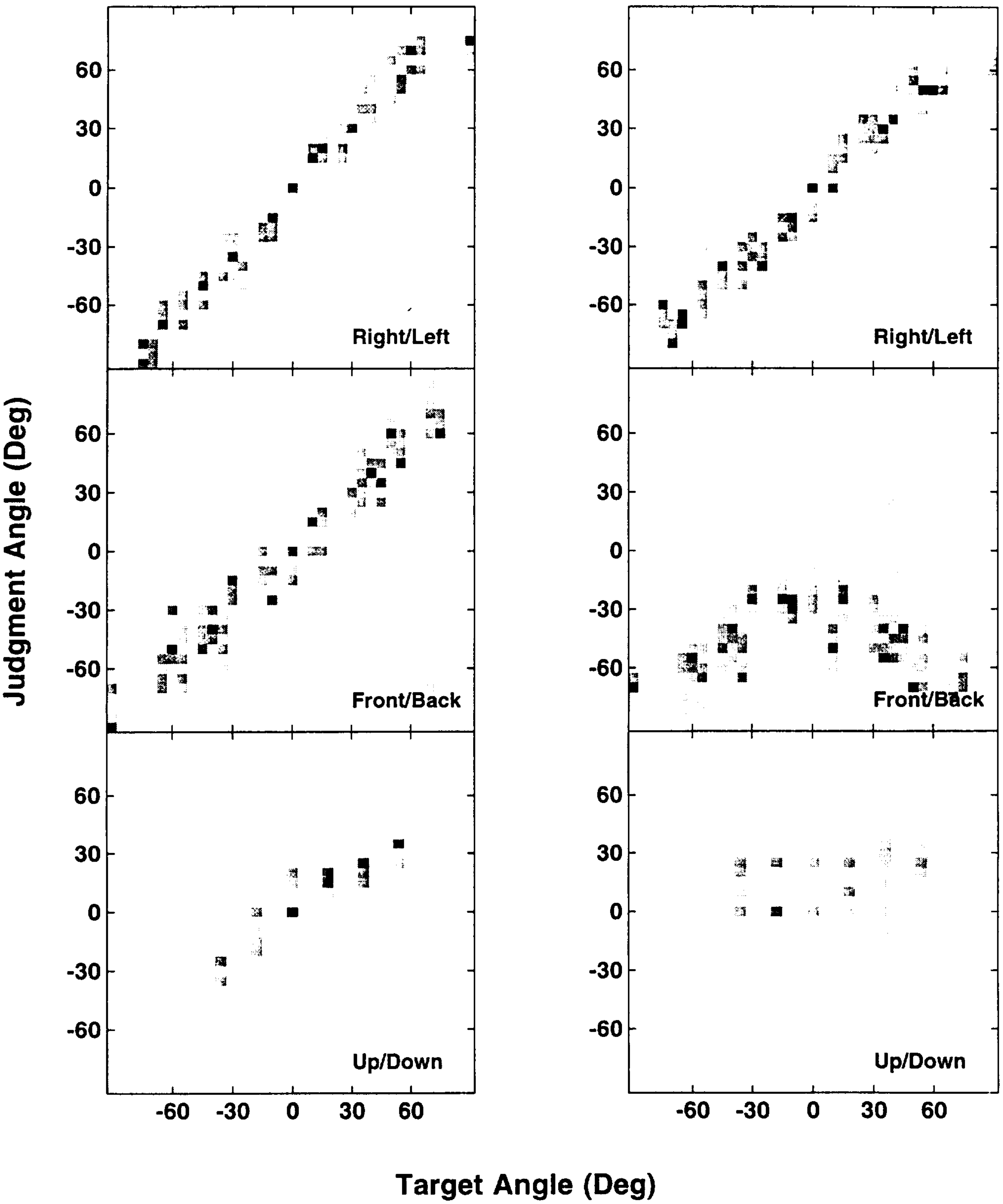

800 - 1000 Hz
4000 - 5000 Hz
8000 - 10000 Hz
ILD (db)
30
20
10
0
-10
-20
-30
Elevation (Deg)
60
40
20
0
-20
-40
-60
Azimuth (Deg)
-180
-120
-60
0
60
120
180

Fig 6



Judgment Angle (Deg)
60
30
0
-30
-60
Right/Left
Front/Back
Up/Down
-60
-30
0
30
60
Target Angle (Deg)

Fig 7

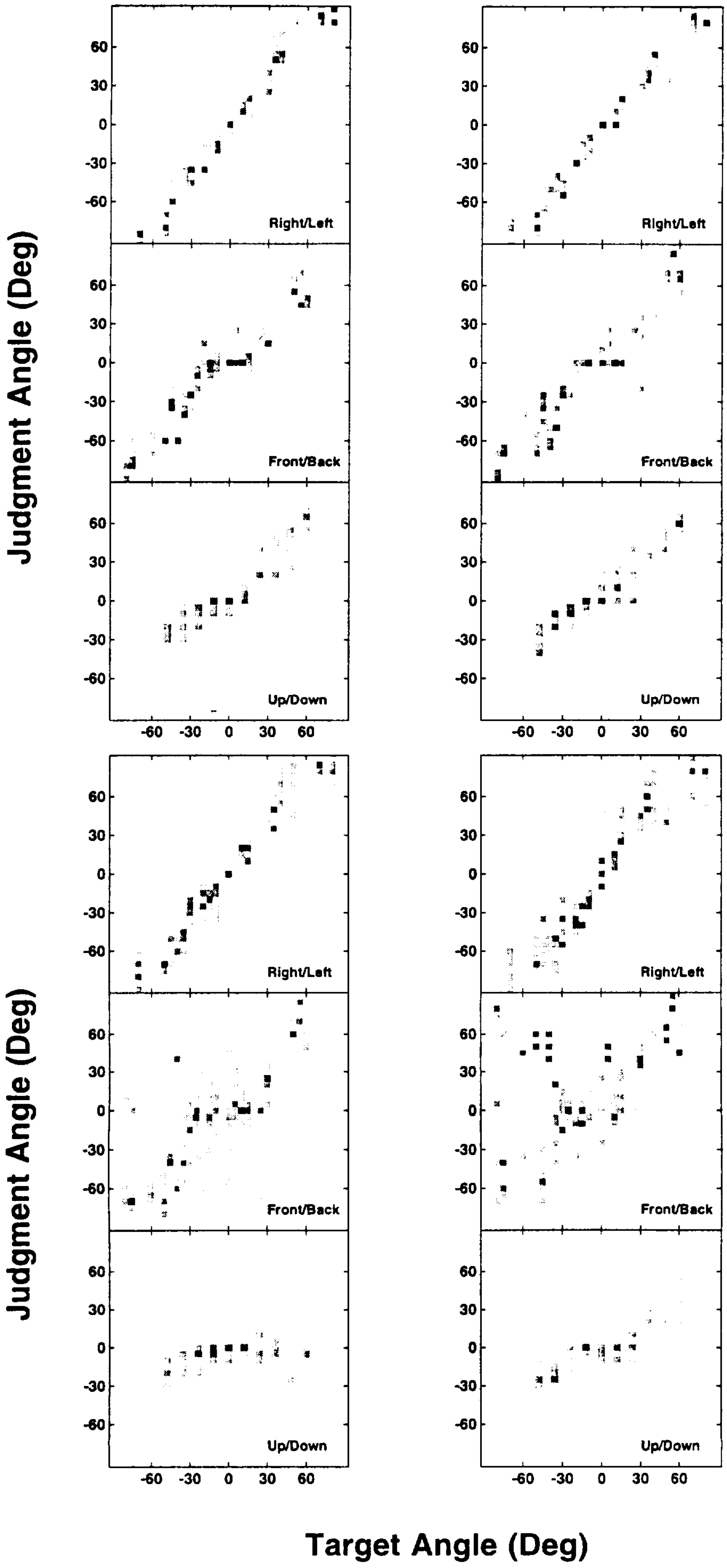
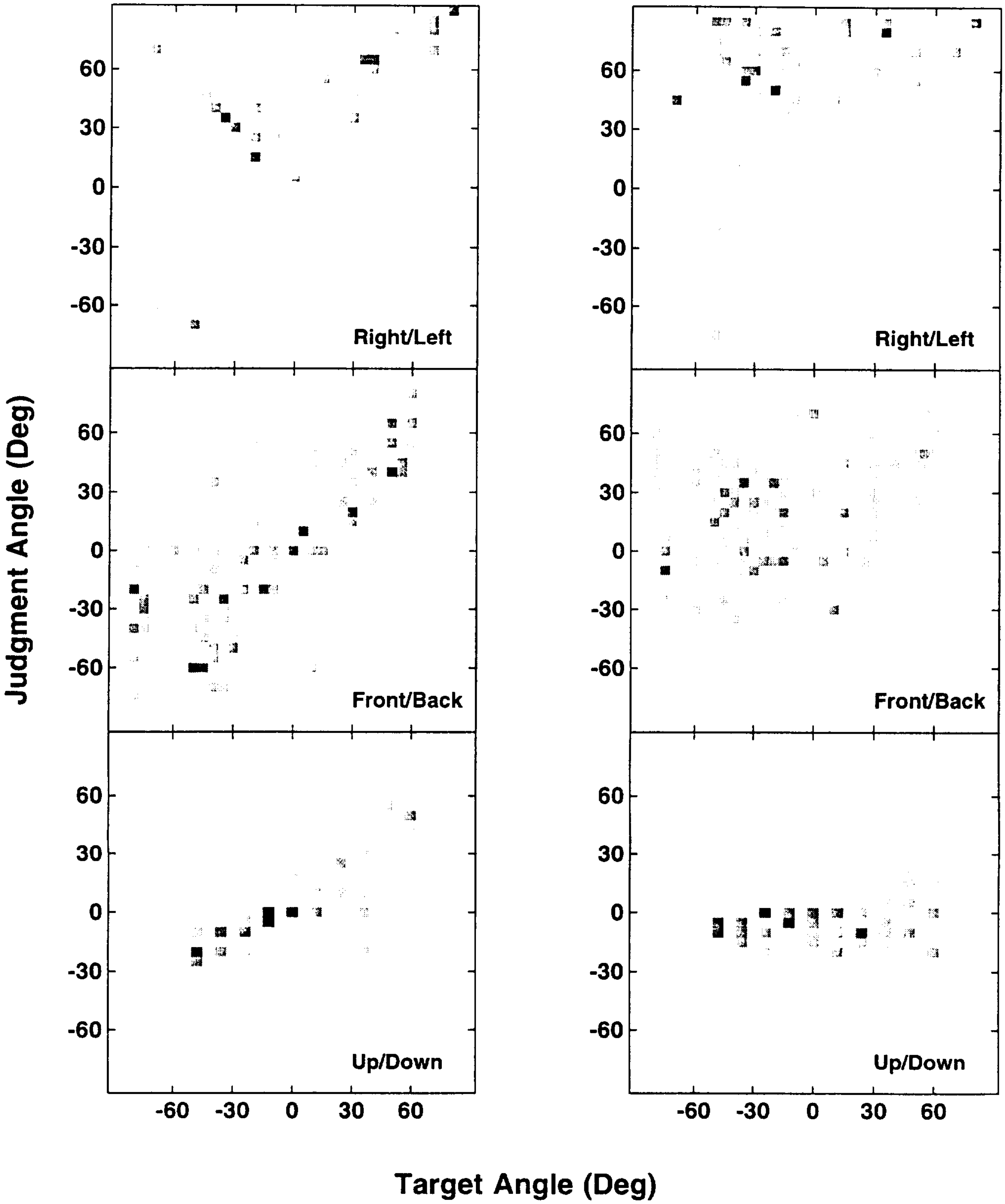
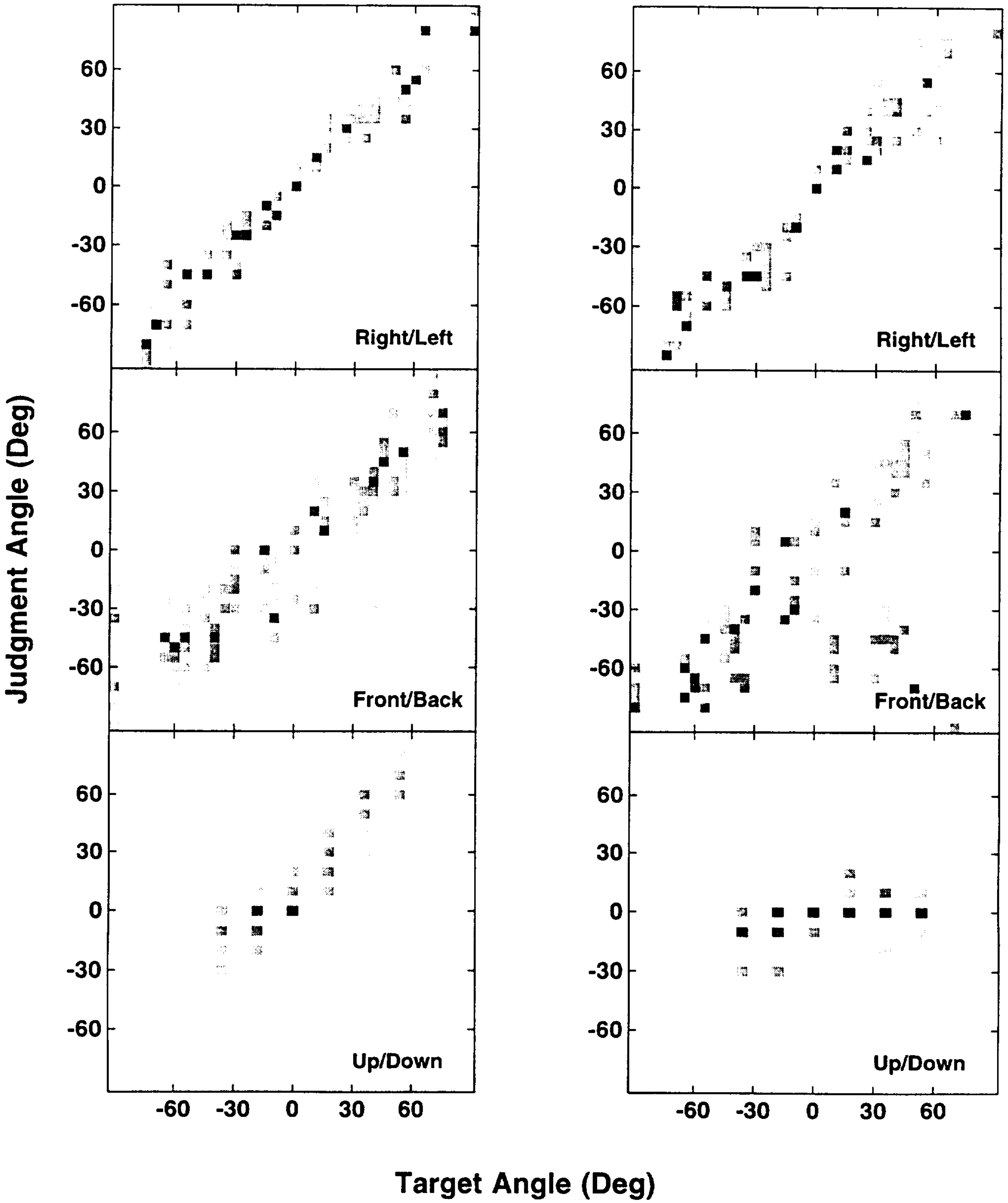

Judgment Angle (Deg)
Judgment Angle (Deg)
Right/Left
Front/Back
Up/Down
Right/Left
Front/Back
Up/Down
Right/Left
Front/Back
Up/Down
Right/Left
Front/Back
Up/Down
60
30
0
-30
-60
-60
-30
0
30
60
Target Angle (Deg)

Fig 8



Judgment Angle (Deg)
60
30
0
-30
-60
Right/Left
Front/Back
Up/Down
-60
-30
0
30
60
Target Angle (Deg)

Fig 9



Judgment Angle (Deg)
60
30
0
-30
-60
Right/Left
Front/Back
Up/Down
-60
-30
0
30
60
Target Angle (Deg)

Fig 10

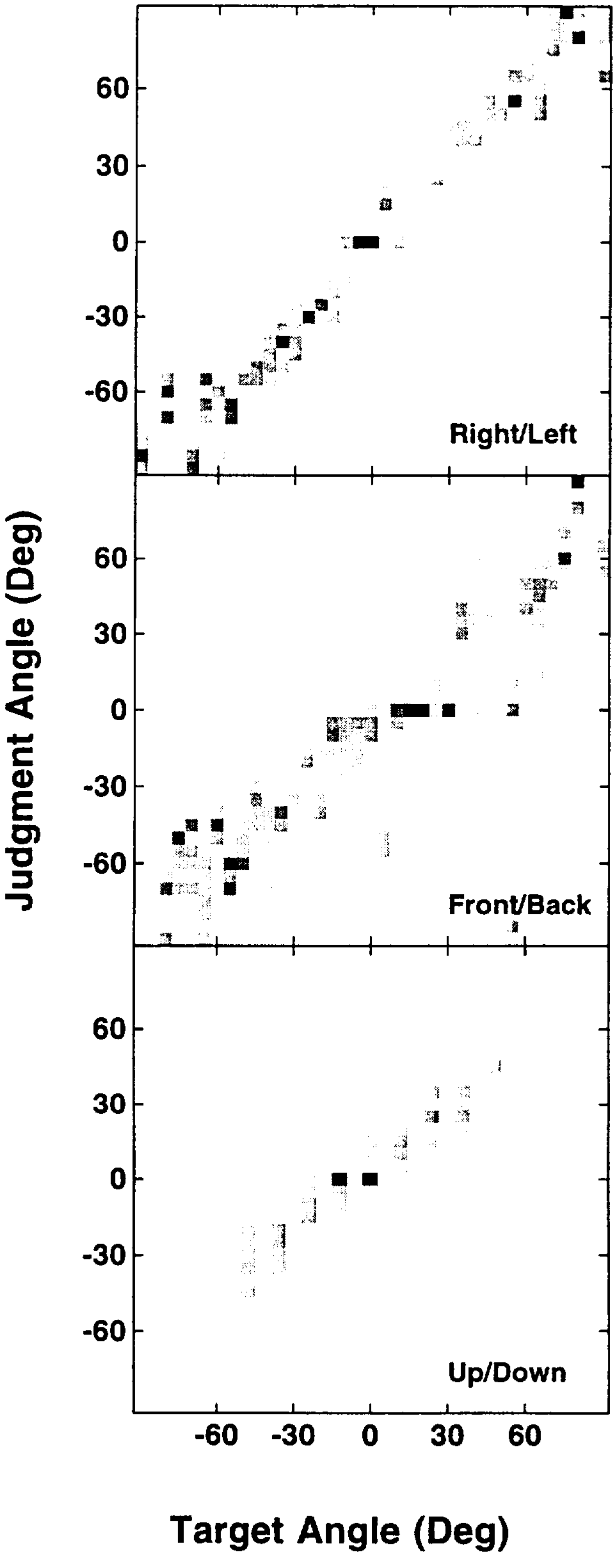

Judgment Angle (Deg)
60
30
0
-30
-60
Right/Left
60
30
0
-30
-60
Front/Back
60
30
0
-30
-60
Up/Down
-60
-30
0
30
60
Target Angle (Deg)

Fig 11

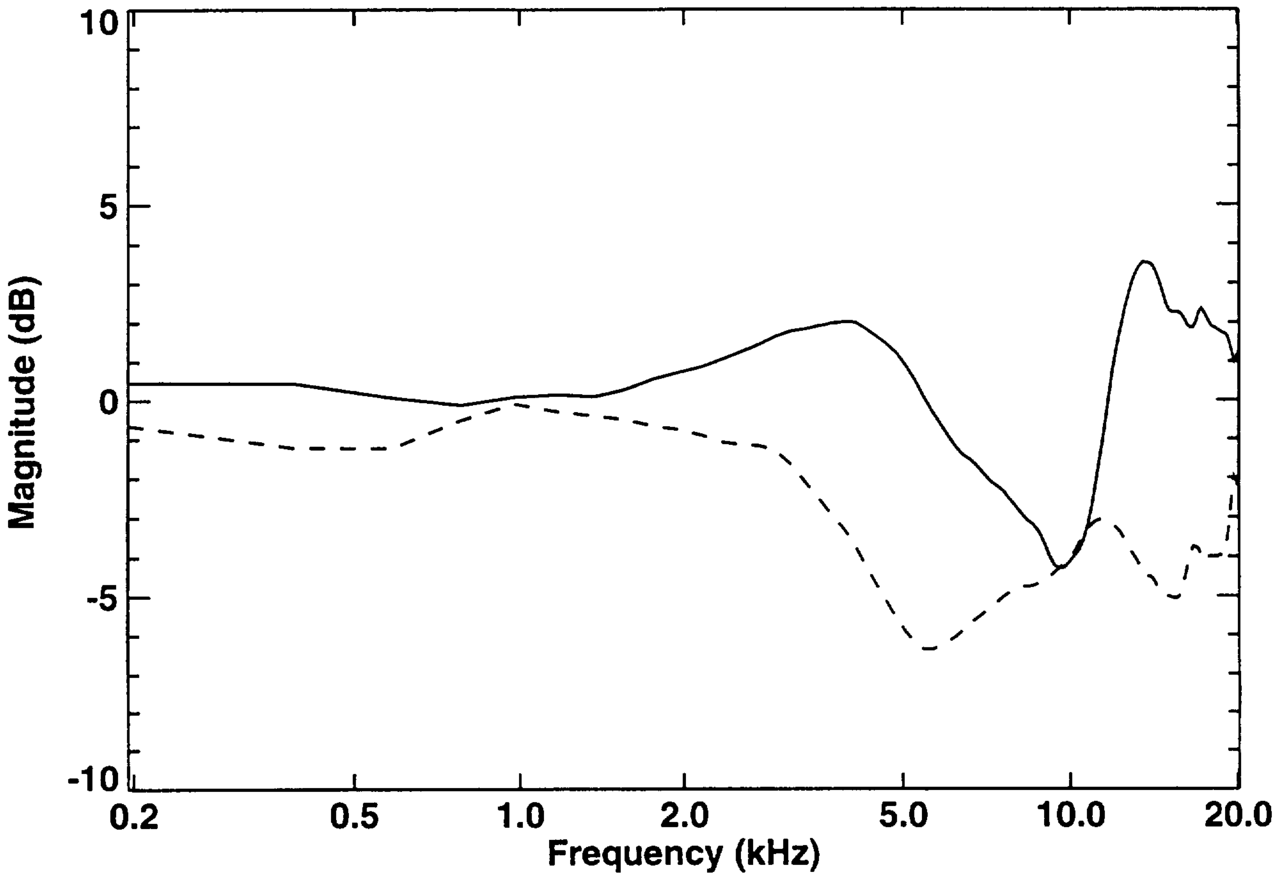
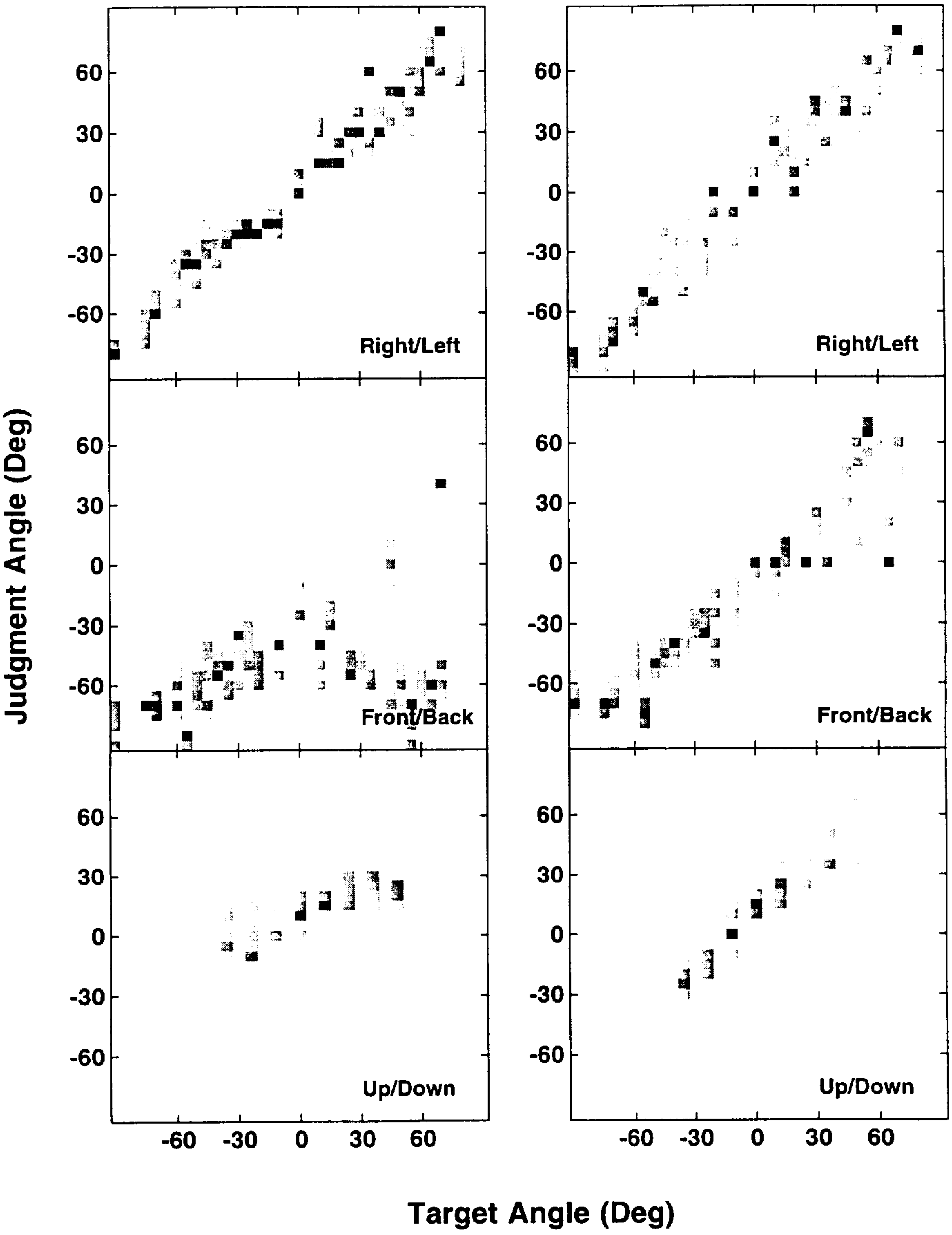


Fig 12



Judgment Angle (Deg)
60
30
0
-30
-60
Right/Left
Front/Back
Up/Down
-60
-30
0
30
60
Target Angle (Deg)

Fig 13

AUDITORY SPATIAL LAYOUT

Frederic L. Wightman

Waisman Center and Department of Psychology

University of Wisconsin, Madison

Madison, WI 53706

and

Rick Jenison

Department of Psychology

University of Wisconsin, Madison

Madison, WI 53706

Note: This is a draft of a chapter that will appear in Volume 5 of the Handbook of Perception and Cognition: Perception of Space and Motion, edited by W. Epstein and S Rogers, and published by Academic Press, Orlando, FL. Please do not cite or quote.

I. Introduction

Everyday sights and sounds are typically described with reference to the environmental object that produced them and not to the physical pattern of stimulation at the sensory receptor. Thus, we say that we see a house rather than an array of points and edges and that we hear a bell rather than a complex of inharmonic partials. This object-oriented view of perception has come to be known as "object perception". In the case of vision the physical features of environmental objects map directly to patterns of stimulation on the retina. Quite naturally, then, the study of visual object perception concentrates on revealing the details of further processing of the peripheral representation, on such issues as size and shape invariance under various transformations of the retinal image. In contrast, hearing offers no direct peripheral representation of environmental objects. All auditory sensory information is packaged in a pair of acoustical pressure waveforms, one at each ear. While there is obvious structure in these waveforms, that structure (temporal and spectral patterns) bears no simple relationship to the structure of the environmental objects that produced them. The properties of auditory objects and their layout in space must be derived completely from higher-level processing of the peripheral input. Thus many of the issues central to the study of auditory object perception are different from those involved in visual object perception.

The definition of what constitutes an auditory object is an issue of some controversy and considerable importance. Many acoustical waveforms evoke a mental reference to the source of the waveform. These are clearly auditory objects. We hear a church bell, for example, or ice tinkling in a glass. We hear the objects themselves and are generally unaware of the spectral and temporal structure of those waveforms. However, reference to an identifiable physical object may not be a necessary condition for auditory "objectness". As we will mention later, waveforms made up of sequences of pure tones can also contain what most would agree are primitive auditory objects, even though no known physical object could have produced the sounds.

That the study of auditory object perception is immature is reflected in the fact that there are few empirical data on the important issues. Thus, while we can be precise here in our descriptions of the physical features of auditory stimuli and somewhat certain about the details of the peripheral encoding of those features, discussion of the higher level processing that subserves auditory object formation and segregation must be speculative. In the context of our discussion of the spatial layout of auditory objects, for example, we can and will review the substantial body of evidence on the factors that determine the apparent spatial positions of single, static sound sources. However, since there are relatively few data on the perception of moving sources and virtually no data on perception of the spatial relations among auditory objects, our treatment of these important issues will be limited to an analysis of the potential sources of information and will not attempt to address in detail the questions related to how those sources of information may be utilized.

The chapter begins with a discussion of the peculiarities of acoustical stimuli and how they are received by the human auditory system. A distinction is made, following Gibson (1966), between the ambient sound field and the effective stimulus in order to differentiate the perceptual distinctions among various simple classes of sound sources (ambient field) from the known perceptual consequences of the linear transformations of the sound wave from source to receiver (effective stimulus). Next we deal briefly with the definition of an auditory object, specifically the question of how the various components of a sound stream become segregated into distinct auditory objects. The remainder of the chapter will focus on issues related to the spatial layout of auditory objects. Stationary objects will be considered first. Since much of the material relevant to this subject has been recently reviewed elsewhere (e.g., Middlebrooks and Green, 1991, Wightman and Kistler, 1993), the section will concentrate on topics not covered in those previous reports. The sources of information related to the apparent distance of an auditory object is one such topic. The spatial layout of moving auditory objects is discussed next, and in this context we offer a detailed treatment of the acoustics of moving sound sources. A distinction between source movement and observer movement is made in order to draw attention to the possible role of proprioceptive feedback in the perception of auditory spatial layout. The chapter concludes with a brief treatment of experimental evidence on the importance of input from other senses (vision, primarily) in establishing auditory spatial layout.

II. Acoustical Information - The ambient sound field and the effective stimulus

As we use the term here, "information" is an abstract construct that serves as the bridge between an organism and its environment. It has a structure that is not related to the characteristics of either the transmitting medium or the receptor surface. For example, the "squareness" of a visual object is specified by information (e.g., relationships among visual patterns) that is not defined in terms of the physics of light or the anatomy and physiology of the retina. In the case of auditory objects, the mechanical events that produce them have lawful acoustical consequences in the sound patterns that are represented to the peripheral auditory system. If those patterns map in a one-to-one or many-to-one fashion onto the object properties, then they constitute information that potentially specifies those properties. In principle, then, for any physical property of an environmental object to be recoverable by an organism there must be information available to the perceiver that specifies that property.

The specific property of auditory objects that is of interest here is spatial layout. The information about auditory spatial layout is conveyed acoustically, and thus the stimulus that must be decoded by the perceiver in order to determine spatial layout is a sound wave. There is information about spatial layout contributed both by the specific type of sound wave that is generated and by the transformations that sound waves undergo in their passage from the source to our ears. This section of the chapter provides an overview of the broad classes of simple sound sources and the characteristics of the waves they produce (the ambient field), and then discusses in detail the source-to-receiver transformations that convey information about the spatial layout of the sound sources (the effective stimulus).

The ambient sound field:

Waves in general are important means by which information about a physical event is conveyed to a perceiver. Discussion of wave generation and propagation is beyond the scope of this chapter since both are extraordinarily complex topics, especially in the case of naturally occurring physical events and natural environments. Simplifying assumptions are not only useful, but mandatory for our purposes here. In the case of sound-producing events a convenient assumption is that the sound is produced by a so-called "point" source, or acoustic monopole, and that the propagation equations are linear. Any small object vibrating in a mass of fluid (air) has all the attributes of an acoustic monopole provided the dimensions of the object are small relative to the sound wavelengths produced and the sound field of interest is several object lengths away. The sound field produced by a monopole is omnidirectional, i.e. the same in any direction equidistant from the source.

The sound fields produced by two or more simultaneously active monopoles can be assumed to combine linearly. Thus, an acoustic "dipole", a very common type of sound source in nature, can be described as the superposition of two spatially separated monopole sources that are 180° out of phase. In contrast with monopole sources, which are omnidirectional, dipole sources have both magnitude and orientation. The structure of the dipole field can best be understood by considering the dipole in terms of its canceling monopoles. The field has an angular dependence with no sound at all produced at 90° to the dipole axis where the sound fields of the constituent monopoles exactly cancel.

The intensity of a sound wave (proportional to pressure squared per unit area) diminishes as the wave travels away from the source. Several factors are responsible for this. One which applies to all sound waves, including those proposed by monopoles and dipoles, is atmospheric absorption. Absorption is the result of nonadiabatic propagation caused by temperature differentials between compressions and rarefactions in the propagating wave and in air depends on temperature, humidity, wavelength. The attenuation coefficient in air at 20° C with 50 % humidity is approximately 1 x 10^{-10} f^2/meter, where f is frequency in Hz. For a monopole source, intensity also decreases with the inverse square of the distance from the source because the total acoustical power is spread out over the surface area of a sphere, the radius of which is the distance from the source. Considering both geometrical spreading and absorption, the intensity of a monopolar source as a function of distance can be written

$$I(r)=\frac{P}{4\pi r^2}e^{-\alpha r}$$

where r is the distance from the sound source, P is the total power produced by the source, and α is the attenuation coefficient. Sometimes the term "attenuation length", $1/\alpha$, is used to describe the distance over which the intensity decreases to 1/e. At short distances the decrease in intensity with distance is dominated by spherical spreading, while at distances well beyond the attenuation length, absorption is dominant.

The intensity of the sound field produced by a dipole decreases somewhat differently with distance. For a dipole field it is simplest to discuss the decrease in pressure (proportional to the square root of intensity). The equation governing the pressure decrease is complicated, but its essential elements are a magnitude and a direction component. The magnitude part has two terms, one decreasing with the inverse square of distance and the other linearly. The inverse square dependence dominates the field near the source and the linear component dominates at large distances.

The characteristics of sound radiation, whether modeled as a monopole or as a dipole, may contribute significant information to aid source identification as well as to determine spatial layout. As described above, monopoles radiate sound evenly in all directions, but dipoles have a figure eight directivity pattern. While the compression and rarefaction components cancel in a plane perpendicular to the dipole axis, a pressure gradient does exists in the field near the source that may be useful for tracking a sound source. An example of a dipole source that we are particularly interested in tracking is a flying insect near our ear. There are also more complex sources in nature which can be modelled as the sum of several constituent dipoles.

The effective stimulus:

For our purposes here the effective stimulus will be defined in terms of the acoustical pressure waveforms produced by an ambient sound field as they exist just prior to transduction at the listener's eardrums. For simplicity we will assume that the ambient field is produced by one or more acoustical monopoles. The relationship between the ambient field and the effective stimulus is defined by a series of linear transformations of the acoustical waveform which incorporate a number of potential sources of information about the spatial layout of sound sources in the environment. In this section of the chapter we will identify the relevant transformations and to describe the spatial information that each incorporates. A later section will examine in detail the evidence on whether or not the information is perceptually relevant.

The acoustics of the local environment which includes the source and the listener contribute several potentially important sources of information about spatial layout. For example, because of the long wavelengths and slow propagation velocity of sound, the reflections and diffractions of an emitted sound wave off the walls, floor, ceiling, and contents of a typical room enrich the ambient sound field considerably. There is information about the size of the room in the timing of the reflections, information about the wall coverings and contents in the pattern of reverberation, and information about the distance between source and listener in the ratio of direct to reflected sound. If long distances are involved, such as in large rooms or in open spaces, the high-frequency content of the effective stimulus is reduced by atmospheric absorption. There is ample evidence that all these effects are detectable by a normally-hearing listener.

The listener's shoulders, head, and outer ear structures (especially the pinnae) are significant components of the local acoustical environment and as such contribute additional information relevant to auditory spatial layout. The pattern of reflections and diffractions of an incident sound wave off these structures is heavily dependent on the direction from which the sound arrives, and thus, the information contributed by these effects relates primarily to the direction of auditory objects. The pinnae, in particular, are highly directional, modifying incident sound waves in ways that are specific to each different angle of incidence. As in the case of room effects, there is ample evidence of the detectability of pinna effects.

The fact that we have two ears separated by an acoustically opaque head suggests that information about auditory spatial layout may come from three sources: the effective stimulus at the left ear, the effective stimulus at the right ear, and the difference. These are clearly not independent sources of information. However, there are reasons to believe that all are important. Information from the difference signal, for example, is uniquely independent of the characteristics of the source, and because of the insensitivity of the auditory system to the absolute timing of events is the only source of information on the direction-dependent difference in the time-of-arrival of an acoustic waveform. Because of the approximate lateral symmetry of the head, interaural difference information is ambiguous. Interaural time difference, for example, is the same for sources in the front and sources in comparable positions (on the same side of the head, and at the same angles relative to the interaural axis) in the rear. Information from each of the individual ears can potentially resolve these ambiguities.

The information relevant to auditory spatial layout that is contained in the effective stimuli at the two ears can be described as either temporal or spectral patterns. At a formal mathematical level the two descriptions are isomorphic so one might think the choice is arbitrary. However, when higher-level processing of the information is considered the distinction becomes important because temporal and spectral processing mechanisms in the auditory system are thought to be so different. For this reason we will discuss temporal and spectral separately. Because of the auditory system's relative insensitivity to monaural phase (the phase spectrum of a stimulus at one ear), our discussion of temporal information will concentrate on interaural time differences and the temporal patterns of room reflections. Interaural phase, defined as the difference between the phase spectra of the left and right ear stimuli, is relevant only when considering single frequency components of a stimulus. Our discussion of the spectral information in effective auditory stimuli will focus on the direction-dependent changes in the magnitude components of the complex source-to-eardrum transformation.

III. Auditory Objects

It seems obvious that before any discussion of the rules that govern the spatial layout of auditory objects we should know what an auditory object is. Unfortunately there is little consensus on what might constitute a satisfactory definition of an auditory object, nor on what alternative terms might better serve. One alternative that has been proposed is "sound event" (Blauert, 1983), but this term seems to refer more directly to a disturbance of the ambient sound field than to any aspect of the perception of that disturbance. Another alternative is "sound stream" (Bregman, 1990), but this term does not convey the obviously close association between everyday auditory stimuli and the environmental objects that produced them. The term "auditory object" is borrowed from the field of visual perception where the features of environmental objects map directly to features of the effective stimulus, a pattern of light on the retina. Its use in auditory perception is less satisfying, since there is no straightforward mapping of object features to stimulus features. Nevertheless, the fact that auditory percepts in daily life are so naturally and immediately associated with the objects that produced the sounds is undeniable and gives currency if not clarity to the term "auditory object".

The effective stimulus at each ear consists of a one-dimensional acoustical pressure-waveform. This waveform contains the superposition of the acoustic outputs from all the objects in the listener's environment. A complete understanding of what constitutes an auditory object would therefore include specification of the rules whereby the various components of the single pressure waveform are segregated into discrete auditory objects. These rules are the object of considerable current interest in the auditory research community (e.g., Bregman, 1990, and Handel, 1989), and it is not our purpose to summarize them here. Rather we will focus on the contributions to this segregation process offered by spatial separation. For the purposes of our discussion, it may be helpful to distinguish between two kinds of auditory objects, "concrete" and "abstract". Concrete auditory objects are formed by sounds emitted by real objects in the environment. Although experimental data are scarce, segregation of concrete objects seems to be determined primarily by spatial and temporal rules. Abstract auditory objects do not often correspond to real environmental objects. They consist typically of more primitive sound elements and are formed by simpler frequency and temporal relations. There has been considerable research on the rules governing the formation of abstract auditory objects (e.g., Bregman, 1990). We concentrate here exclusively on concrete auditory objects.

IV. Spatial layout of stationary auditory objects

Much of the experimental literature on auditory spatial layout concerns the accuracy with which the spatial position of a sound-producing object is indicated to a listener, that is, the degree of correspondence between the actual position of the object and its apparent position. It is our view that experiments which focus on accuracy can fail to consider other important features of the auditory percept. For example, consider experiments on monaural listening. The results generally show that the apparent positions of auditory objects are strongly biased toward the interaural axis and the side of the functioning ear. However, those same results are often reported as indicating that monaural localization accuracy is near normal on the side of the functioning ear and progressively poorer off the interaural axis on that side. The emphasis on accuracy obscures the fact that in monaural listening all the sounds appear to emanate from one place. For reasons such as this, we prefer to ignore the accuracy component of spatial layout altogether, and discuss only the factors that govern the apparent spatial positions of auditory objects.

The apparent spatial position of an auditory object is defined by its apparent direction and its apparent distance relative to the listener. The potential sources of information for apparent direction and the stimulus features that appear to govern apparent direction have been extensively and recently discussed elsewhere (Wightman and Kistler, 1993; Middlebrooks and Green, 1991). Therefore, the material on apparent direction will only be summarized here. Much less attention has been paid to apparent distance, and while data are scarce, they will be covered in some detail in this chapter.

Acoustical sources of information about static spatial layout:

The spatial position of each sound-producing object in a listener's environment is specified by several acoustical sources of information which for brevity we will call "cues". Many of the cues are a result of the interactions of the sound waves with the listener's head and pinnae. These interactions are conveniently summarized by a linear transformation, the so-called "head-related transfer function", or HRTF, which represents the changes in the amplitude and phase of the sound wave from the sounding object's position to the listener's eardrum. Mathematically, HRTFs are usually specified in terms of the sound wave's spectrum. Thus, if X(jω) is the source spectrum (j is the complex operator and ω is angular frequency) and Y(jω) is the spectrum of the waveform at the eardrum, the HRTF, H(jω), is given by:

$$H(jw)=\frac{Y(jw)}{X(jw)}$$

More generally, since the HRTF varies with source direction and distance and thus is different at each ear, we must write two equations for H(jω), one for the left ear and one for the right ear. Each depends on source azimuth (θ), elevation (ϕ), and distance (d) relative to the listener:

$$H_l(\theta,\phi,d,j\omega)=\frac{Y_l(\theta,\phi,d,j\omega)}{X(j\omega)}$$

$$H_r(\theta,\phi,d,j\omega)=\frac{Y_r(\theta,\phi,d,j\omega)}{X(j\omega)}$$

All the information about sound source position are represented in the pair of HRTFs shown above. These HRTFs vary in complicated ways with changes in source position, so simplifying assumptions must be made in order to appreciate the essential elements. Two convenient assumptions are that the acoustical space enclosing the source and listener is anechoic, and that the listener's head is spherical with pinna-less ears at opposite ends of a diameter of the sphere. The anechoic assumption allows the main effect of distance to be modelled as a simple attenuation of 6 dB for every doubling of distance from the source. The spherical head assumption leads to a greatly simplified account of the effects of diffraction of the sound wave around the head. Figure 1 illustrates the latter point. Ignoring the details for the moment (the spherical model is described in detail in Kuhn, 1977) we see that at each ear individually, variations in source azimuth (or elevation, not shown in the figure) can be expected to produce mainly variations in effective stimulus intensity, a result of the "head shadow" effect when the source is on the opposite side of the head from the ear under consideration. The head shadow effect can be expected to be much larger at high frequencies than at low frequencies. This is because at low frequencies sound wavelengths would be long with respect to the dimensions of the head, and thus the sound waves would travel around the head without attenuation. The covariation of stimulus intensity with azimuth (and elevation) which occurs at each ear individually can be viewed as a potential "monaural cue" to sound source position. Figure 1 also illustrates the potential "binaural cues" to sound source position that are offered by interaural differences (defined by the ratio of the two HRTFs). Note that for all source azimuths other than 0° and 180° the acoustical path from source to ear has a different length for the two ears. This path-length difference produces a small difference in the time of arrival of the sound wave at the two ears. The interaural-time-difference (ITD) varies systematically with source azimuth and is largest for azimuths of +90° and -90°. In addition, because of the head shadow effect mentioned earlier, there will be an interaural level difference (ILD) that varies with azimuth in roughly the same way as ITD and which is large at high frequencies and small at low frequencies.

The utility of monaural cues is compromised by the fact that some or all features of the sound source waveform must be known in order for the cue to be unambiguous. In the simple spherical head case described above, while stimulus intensity at a given ear varies systematically with source azimuth, a listener with access only to the effective stimulus at that ear would have no way of knowing whether a weak stimulus was produced by a source on the opposite side of the head or by a weak source. In more general terms, note that (from Equation 3) the effective stimulus at one ear, say the right ear, is defined by the product of the source spectrum and the HRTF:

$$Y_r(\theta,\phi,d,j\omega)=X(j\omega)H_r(\theta,\phi,d,j\omega)$$

Thus, even if a listener had perfect memory for the HRTF at each and every possible source position, a given effective stimulus could unambiguously indicate a specific source position only if the source spectrum were known.

Binaural cues to source position are derived from the ratio of the transduced representations of the two effective stimuli. Thus, the utility of these cues does not require knowledge of the source spectrum, since that term appears in both numerator and denominator and hence cancels. Nevertheless, to the extent that the spherical head model is accurate, binaural cues are also ambiguous. Note, as shown in Figure 1, that the difference in acoustical path length from the source to the two ears, which gives rise to the ITD, is the same for sources in front and in the rear. A source at an azimuth of 30°, for example, would produce the same ITD as a source at 150° azimuth. The same could be said for ILDs and for sources at complementary positions above and below the horizontal plane. In fact, the spherical head model predicts conical surfaces projecting outward from the ears along which ITD and ILD are constant, and thus along which cues based on ITD and ILD would be ambiguous. These are the so-called "cones of confusion". We should mention here that cone-of-confusion ambiguities could be resolved by head movements, as Wallach (1940) pointed out in his now-classic treatise on the issue. If a listener knew both the direction of movement of the head and the direction of change of the ITD or ILD cue, the direction of the sound source could be derived without ambiguity.

Detailed measurements of human HRTFs (Shaw, 1974; Wightman and Kistler, 1989a; Middlebrooks, Makous, and Green, 1989; Middlebrooks and Green, 1990; Pralong and Carlile, 1994) provide a complete catalog of the potential acoustical cues to apparent sound position and highlight the limitations of the spherical head model. The most prominent features of HRTFs not anticipated by the spherical head model are the directional filtering characteristics of the pinnae and the large listener-to-listener differences in HRTFs. The multiple ridges and cavities of the pinna produce resonant peaks and antiresonant notches in the magnitude response of the HRTF. The frequencies at which these peaks and notches appear are dependent on sound source direction, and thus could serve as potential spatial position cues, provided some a-priori information about the source was available. Figure 2 shows an example of how the frequency of a given notch in the HRTF changes with sound source elevation. HRTFs from two listeners are shown in this Figure to illustrate individual differences. Note that while the general characteristics of the notches are the same from listener to listener, the frequencies at which the notches appear are highly listener dependent.

The spherical head model provides a reasonably accurate prediction of the ITDs derived from actual HRTF measurements. Figure 3 shows ITDs from the horizontal plane HRTFs of a representative listener, estimated by Wightman and Kistler (1989a). Also plotted in the figure are the ITDs predicted by:

$$ITD=\frac{d}{2c}(\theta+\sin\theta)$$

where θ is the azimuth angle as in Figure 1, c is the velocity of the sound wave (cm/sec), and d is the interaural distance (cm), chosen for this example to fit the HRTF data shown. While this equation is usually cited as representing the predictions of the spherical head model (e.g., Woodworth, 1938; Green, 1976), it is really just a first-order approximation (Kuhn, 1977). Nevertheless, as Figure 3 shows, it provides an accurate representation of horizontal plane ITDs. Figure 4 (from Wightman and Kistler, 1993) shows a more complete set of ITD data from the same listener. This figure also shows the contours of constant ITD, which for the spherical head model would be circular. Clearly the spherical head model provides a good first-order approximation to measured ITDs. Just as clearly, ITD is an ambiguous cue to sound source direction since any given ITD signals not one but a whole locus of potential directions.

Interaural level differences derived from HRTF measurements are complicated functions of frequency at each and every source direction, a situation caused at least in part by pinna filtering effects. Figure 5 shows ILD functions derived from a single listener's HRTF measurements at a source elevation of 0 and azimuths of 0° and 90°. Note that even for a source on the median plane (0° azimuth), where ILDs would result only from interaural asymmetries, ILDs are large enough (greater than 0.5 dB, the ILD threshold) to be considered potential sources of information about source position. For a source at 90° ILDs are generally much larger, especially at high frequencies as would be expected because of head shadowing.

The elaborate frequency dependence of ILDs complicates our discussion of them as potential cues to sound source position. We can discuss the interaural level cue either as an "interaural spectral difference", referring to the entire pattern of ILDs across frequency, or as ILD averaged across one or more frequency bands. Figure 6 illustrates the latter approach. In the upper panel we show one extreme, ILD averaged across the entire frequency spectrum. The bottom panels illustrate the other extreme, ILDs in two high frequency "critical bands". Note that the general pattern of ILD as a function of sound source direction is the same regardless of the bandwidth over which ILD is considered or the center frequency of the band. Note also that the general pattern of ILDs is the same as the pattern of ITDs, showing a similar kind of "cone-of-confusion" ambiguity. Thus, unless a listener could analyze the idiosyncratic details of ILD patterns in narrow bands, ILD information could not be used to disambiguate errors resulting from dependence on ITDs, and vice-versa. As mentioned above, information provided by head movements can, in theory, offer such disambiguation.

The acoustical sources of information about the distance of a sound producing object are not well understood. Nor have they been well documented by systematic measurements. In an anechoic environment, the two most obvious stimulus features that depend on distance are overall level and spectral content. Overall level decreases by 6 dB for every doubling of the distance between the source and the listener (the inverse square law), and atmospheric absorption gradually attenuates the high frequency components of a sound as the distance between source and listener is increased (about 2 dB per hundred feet at 6 kHz, and 4 dB per hundred feet at 10 kHz). The utility of both of these monaural cues, of course, depends on knowledge of source characteristics. However, the requirement for a-priori knowledge about the source can be eliminated if the perceiver is allowed two or more "looks" at the stimulus from different vantage points. For example, Lambert (1974) pointed out that just two "looks" at stimulus intensity, as might be obtained if the perceiver's head is rotated, would provide sufficient information for a determination of source distance, without the need for knowledge of source characteristics.

There are two potential binaural distance cues, ITD and ILD; both vary slightly with the distance between source and listener (Coleman, 1963). In the case of ITD, for a source at 90° azimuth, there can be as much as a 150 microsecond difference in the ITD produced by a near source and a far source. A near source produces a larger ITD than a far source. This change in ITD with distance occurs because with a source close to the head the extra distance around the head is greater than if the source were far from the head. Distance affects ILDs in a comparable way, although in this case the effect is highly frequency dependent. At low frequencies the distance effect is greatest. For a 300 Hz tone at 90° azimuth, for example, the ILD for a source far from the head (several wavelengths) is about 0.5 dB but for a source at 44 cm it is over 10 dB. The effects at higher frequencies and at source azimuths off the interaural axis are considerably smaller.

In a non-anechoic environment, which of course includes nearly all everyday listening situations, there is an additional distance cue provided by the mix of the direct sound wave from source to listener with the reflections of that sound wave off the surfaces of the listening room. When the sound source is close to the head the direct sound dominates, since because of the extra distance traveled and absorption at the surfaces the level of the reflected sound is always lower. However, as the source to listener distance increases, the direct sound level decreases, and the ratio of direct to reflected sound level decreases. Given a specific enclosure, then, this ratio is perfectly correlated with source to listener distance. Moreover, even though it is a monaural cue, its validity does not depend on a-priori knowledge of stimulus characteristics.

Acoustical determinants of apparent spatial position:

Our purpose in this section is to review what is currently known about how the acoustical information about the spatial position of stationary sources is actually used. Most of the experiments in this area have considered apparent source direction and apparent distance separately, and for convenience we maintain this separation here. Several comprehensive reviews of this area have appeared recently (Wightman and Kistler, 1993; Middlebrooks and Green, 1991), so the material will only be summarized here.

In the vast majority of experiments on the apparent spatial position of stationary auditory objects only apparent direction (azimuth and elevation) has been considered. Until recently the dominant theoretical position, epitomized by the Duplex Theory (Strutt, 1907), was that ITD provided the dominant source of information about apparent direction at low frequencies and that ILD was dominant at high frequencies. The duplex theory derived from the facts that the auditory system was much less sensitive to ITDs at high frequencies than at low frequencies (Yin and Chan, 1988; Joris and Yin, 1992) and from the fact that ILDs are much larger at high frequencies than at low frequencies (see Figure 5). Information provided by pinna filtering was not considered in the Duplex Theory.

Few empirical data on apparent source direction contradict the Duplex Theory. However, there are many natural circumstances which reveal the limitations of the theory and which argue for a situation dependent weighting of the various sources of information about apparent sound direction. Localization of narrowband sounds is one such circumstance. Most narrowband sounds offer conflicting cues to apparent direction, so it is not surprising that they are not often localized accurately. The extreme case of a narrowband sound is a sinusoid. Sinusoids offer doubly ambiguous ITD cues. A 1000 Hz sinusoid, for example, could provide a 400 μs ITD leading to the right ear while at the same time indicating a 600 μs ITD leading to the left ear. As Figure 4 shows, each ITD signals a whole range of potential source directions. It should not be surprising that unless a sinusoid has a broadband transient associated with onset or offset its apparent position is unclear (Hartmann, 1983). Other narrowband sounds are somewhat less ambiguous but still inaccurately localized. The apparent azimuth of a high-frequency noise band is given by ILD, as suggested by the Duplex Theory (Middlebrooks, 1992). However, the apparent elevation seems to be determined by a learned association between spatial position and the spectral peaks and valleys produced by pinna filtering (Middlebrooks, 1992). The resultant apparent direction is often far removed from the actual source direction and well off the contour of directions indicated by ILD alone. In this case and others (e.g., monaural localization, as described by Butler, Humanski, and Musicant, 1990) the learned association between spatial position and pinna filtering details appears to be a favored source of information about apparent sound direction. In general the data suggest that in the absence of unambiguous (i.e., wideband) ITD the information provided by pinna filtering appears to dominate.

If a wideband source contains both low and high frequencies apparent direction seems to be governed primarily by ITD (Wightman and Kistler, 1992). In the Wightman and Kistler experiments (1992) free-field noise sources were synthesized using algorithms based on listeners' own HRTFs. The "virtual sources" were then presented via headphones, affording complete control over the acoustical stimulus. When the ITD information was manipulated to signal one direction and all other cues were left to signal another direction, the listeners' judgments of apparent direction always followed the ITD cue. Thus, even in the presence of opposing ILDs of as much as 20 dB, ITD was dominant. The dominance of ITD occurred for all listeners so long as the stimuli contained energy below about 1500 Hz. When the low frequencies were filtered out ITD was effectively ignored and judgments of apparent position followed the ILDs and pinna filtering cues.

The importance of the ITD cue is further emphasized by the fact that listeners' make frequent front-back confusions in certain conditions (Stevens and Newman, 1936; Oldfield and Parker, 1984a,b; Wightman and Kistler, 1989b; Wenzel, Arruda, Kistler, and Wightman, 1993). Recall that if apparent direction were governed by ITD, front-back confusions would be expected given the spherical symmetry of the head (Figure 4). While the rate of front-back confusions in everyday life is unknown, with laboratory stimuli and especially virtual source stimuli, front-back confusion rates can be as great as 25% (Oldfield and Parker, 1984a,b; Wightman and Kistler, 1989b). Contours of constant ITD from actual measurements are smooth and regular, as predicted by the symmetry argument, though slightly different for different listeners (Wightman and Kistler, 1993). Contours of constant ILD, on the other hand, are quite irregular and variable from one frequency band to another (Figure 6). We suggest that the fact that listeners make consistent and frequent front-back confusions argues at least indirectly for the dominance of ITD cues and the lesser importance of ILD and pinna filtering cues.

The relative salience of the various acoustical cues to the spatial layout of auditory objects also depends on the "realism" of the cues. In experiments with virtual sources similar to those described above in which ITD was in conflict with other cues (Wightman and Kistler, 1992), we have produced stimuli in which cues in one frequency region conflict with cues in another frequency region. In one condition, for example, the ILD and spectral cues were the same throughout the frequency range (200 Hz -14000 Hz), and signalled, or "pointed to" one of five possible directions on the horizontal plane. The ITD cue in each of four bands (roughly 1.5 octaves wide) pointed to a different direction. Thus, the ITD cue could be said to be "inconsistent" across the frequency range and the other cues "consistent". In other conditions the ITD cue was consistent and the other cues inconsistent, and in still other conditions, the frequency range was divided somewhat differently. The results were unambiguous. Listeners' judgments always followed the consistent cue. Even if the ITD cue was inconsistent only in a single high-frequency band (above 5 kHz), listeners appeared to ignore ITD and put maximum weight on the ILD and spectral cues which were consistent across the spectrum. Not only does this result suggest that high-frequency ITD cues are encoded as well as low-frequency ITD cues, but it also suggests that cues which are "realistic" are given greater weight than unrealistic cues. With real sources and real listening environments it is highly unlikely that either ITD or the other cues could be inconsistent across the frequency spectrum.

The fidelity of the ITD, ILD and spectral cues to spatial position is compromised in most natural listening situations by the presence of echoes. These echoes, which to a first approximation are filtered copies of the sound wave, are produced when a sound wave bounces off objects or surfaces in the environment and because of the extra distance they have to travel they reach the listener slightly later than the original, or direct sound wave. Typically, the intensities of the echoes are considerably weaker than the intensity of the direct sound, both because of the additional path length and because most objects and surfaces absorb some of the sound energy, particularly at high frequencies. Nevertheless, when the echoes combine with the direct sound the acoustical cues that signal the spatial position of the sound source are disrupted. With echoes the effective stimulus at each ear consists of the superposition of sounds from a number of different directions. Thus both the monaural and binaural cues are distorted.

It might be expected that the presence of echoes would seriously impair a listener's ability to determine the spatial layout of sound sources. In fact, in all but the most extreme cases the echoes are hardly noticed, and localization performance is not impaired (Hartmann, 1983; Begault 1992). The substantial body of empirical data on this phenomenon can be summarized in the hypothesis that listeners attend only to the first few milliseconds of a stimulus, the time before echoes arrive, in order to determine the spatial position of a source. The spatial information arriving later, which would be corrupted by echoes, is somehow suppressed. This is the well-known "precedence effect" (Wallach, Newman, and Rosenzweig, 1949; Zurek, 1980, Clifton and Freyman, 1989). While many of the characteristics of the phenomenon and most of the underlying mechanisms are not well understood, it is clear that the precedence effect is of central importance to the determination of auditory spatial layout in natural listening situations

Compared with our well-developed understanding of how various sources of acoustical information are combined to determine the apparent direction of auditory objects, relatively little is known about how listeners might form a judgment of apparent distance. Available evidence suggests that perception of auditory distance is not well developed in humans. Apparent distance is typically very different than real distance (e.g., Gardner, 1968; Mershon and King, 1975), and only relative distance can be determined with any accuracy (Cochran, Throop, and Simpson, 1968; Holt and Thurlow, 1969). While there are suggestions in the literature that the distances of familiar sounds are judged more accurately (Coleman, 1962; McGregor, Horn, and Todd, 1985), the classic demonstration by Gardner (1968) shows that in an anechoic room with levels equalized even the apparent distance of speech is not accurately reported. The most reliable finding seems to be that sounds presented with reverberation are judged to be more distance than the same sounds presented without reverberation (e.g., Mershon and King, 1975).

From several different perspectives inaccuracies in judging the distance of an auditory object are not surprising. First, the primary acoustical correlates of distance, level and spectrum, are unambiguous only if the characteristics of the source are known. Second, in everyday life the absolute distance of an auditory object carries little significance. Direction is clearly much more important; it serves to orient our gaze. Of course, if an auditory object is moving, and especially if that movement is toward the listener, distance carries considerable significance. Experiments on estimation of distance of a moving auditory object typically ask listeners to judge the time at which the object will reach to listener's position, called "time to contact". The available data on listeners' judgements of auditory time to contact will be reviewed in a later section of this chapter.

V. Spatial Layout of Dynamic Auditory Objects

In everyday life an individual's auditory world is constantly in motion. The orientations of sound-producing objects with respect to a listener's head and ears are ever changing, either because the objects themselves are moving or because the listener's head is moving. In either case the result is a constantly changing pattern of directional cues at the ears and, if conditions are right, the introduction of additional cues to movement such as doppler shift. This section of the chapter will describe those additional movement cues in some detail and then will discuss the available psychophysical data on listeners' processing of dynamic spatial information.

Additional acoustic information from moving sounds. Moving sounds can be described using the mathematics of kinematics (Jenison and Lutfi, 1992). *Kinematics* is the branch of mechanics that describes pure motion, employing the variables of displacement, time, velocity, and acceleration. Doppler shifts, changes in ITD (described earlier) and intensity can be shown to have dependencies based on kinematics. In addition to ITD, Doppler shift, and time-varying intensity, the first differentials of these observed variables may be sensed directly as well. Figure 7 shows the geometry of the sound source moving relative to an observer. φ_t is the angle of the incident wavefront at any time *t* and is dependent on the distance D_t to a point p on the median plane. θ_0 is the angle at the anticipated closest point of approach (CPA) and β is the angle of the source trajectory relative to the median plane. Angle β is equivalent in magnitude to $\theta_0 + \pi/2$. R_t is the distance from the sound source to the observer.

Movement of either the sound source or the observer changes the relative wavelength of the sound waves. This change is known as the Doppler shift. The well known lawful dependence of the Doppler shift on velocity of the sound source relative to an observer is

$$\omega = \frac{\omega_0}{(1 - M\cos\varphi_t)}$$

where ω_0 is the intrinsic frequency, ω is the shifted frequency, M is the Mach number defined as velocity divided by the speed of sound and φ_t is the angle of trajectory relative to the observer (see Figure 7). The frequency shift depends only on the velocity component directed toward the observer. This result holds true regardless of the time history of the trajectory. The Doppler-shifted frequency at a given time and position are affected only by the source's velocity and frequency at the instant the wave is generated. Furthermore, the source need not be traveling at a constant velocity or in a straight line for it to apply. When the sound source is far from the observer and approaching (φ_t is small, thus $\cos(\varphi_t)$ is near 1), the angle φ_t changes very little, hence little change in the frequency shift. However, the magnitude of the shift will be at its maximum. Since the sound source is approaching the observer, the shift is toward a higher frequency. As the sound source approaches the observer, φ_t increases rapidly resulting in a rapid decrease in frequency (see Figure 9). As the sound source passes and recedes, there is a corresponding decrease in frequency relative to the intrinsic frequency of the sound source. This of course is the experience we've all had listening to a passing train whistle that decreases in pitch as it passes by and recedes into the distance.

These observed variables, ITD, time-varying intensity, and Doppler, along with their first-order differentials with respect to time, all have characteristic spectrotemporal patterns. Zakarauskas and Cynader (1991) analyzed intensity patterns for actual moving sound sources along various trajectories and derived mathematical expressions for the observed variables that are related to the inverse-square distance relationship. Jenison (1994) extended these analyses to include Doppler and ITD patterns. The simplest trajectory is that of the rectilinear approach with constant velocity as shown in Figure 8. For illustration, the starting point for the moving sound source in these examples is located some distance R_s directly on the median line as shown in the Figure 8.

The characteristic patterns for the three sound source trajectory angles (β) of 90°, 120°, and 150° are shown in Figure 9. For the purpose of this example we have assumed a source of moderate intensity, a velocity of 5 m/s and a starting distance from the observer of 5 m. Note that all of the ITD functions begin at 0 delay due to the midline starting point. The intensity functions will also start at the same intensity for a given distance from the observer. In the case of the Doppler shift, the shift is toward a higher frequency when the sound is approaching the observer and toward a lower frequency when receding. So for β_1 equal to 90°, the frequency shift will start at unity and decline. For the cases of β_2 and β_3, where the source is initially approaching, passes through a closest point of approach and then recedes, the frequency shift will initially be greater than unity and then decline.

Jenison (1994) has shown that acoustical kinematics sufficiently convey velocity (trajectory and speed) information regarding the moving sound source directly from the observed Doppler shift together with time-varying ITD. While the theoretical analyses show that sufficient information is available to the observer regarding higher order variables such as the velocity and time-to-contact of the moving sound source, it remains to be known whether the human observer has sufficient sensory mechanisms to detect this information, particularly under conditions of uncertainty.

Most of the empirical research on perception of moving sound sources has focussed, either directly or indirectly, on the question of whether or not dynamic spatial changes are processed with some kind of specialized "movement detectors". There is considerable neurophysiological evidence that differential information lawfully related to motion is directly detected by the visual system (Maunsell and VanEssen, 1983). Recent evidence suggests that there are also direction-sensitive neurons spatially segregated in auditory cortex (Stumpf, Toronchuk and Cynader, 1992). Other findings suggest that neural processing of auditory motion involves mechanisms distinct from those involved in processing stationary sound location (Spitzer and Semple, 1991; Spitzer and Semple, 1993; Toronchuk, Stumpf and Cynader, 1992). Thus, while converging physiological evidence supports the existence of motion sensitive neurons, the psychophysical evidence for specialized motion detectors is inconclusive. The two lines of research that have addressed this question involve measurements of the "minimum audible movement angle", or MAMA, and measurements of auditory motion aftereffects.

The MAMA experiments are variations of the classical "minimum audible angle", or MAA experiments conducted with stationary sources. They are both detection or discrimination experiments that measure the threshold for discriminating small changes in spatial parameters. In the case of MAAs, what is measured is the smallest spatial separation of two static sources that can be reliably detected. The MAMA represents the smallest amount of spatial displacement or movement of a single source that can be reliably detected. While both experiments can inform us about the processing capabilities of the auditory system, it is important to note that since they involve discrimination or detection paradigms the extent to which the results can be generalized to questions about apparent spatial position may be quite limited. In other words, that listeners can discriminate between two sources at slightly different spatial positions does not necessarily imply that the apparent positions of the sources were different. Similarly, discrimination between a moving source and a static source does not necessarily imply that movement itself was perceived.

While the investigators involved in the MAMA research may quibble over details, most would probably agree that the results do not support the existence of specialized motion detectors in the auditory system. Measured MAMAs, when expressed in terms of the total angle traversed at threshold, are roughly the same as or slightly larger than the MAAs measured with stationary sources, or about 2° (Grantham, 1986; Perrott and Musicant, 1977; Harris and Sergeant, 1971; Perrott and Tucker, 1988). A simple explanation of the basic MAMA results is that the listener takes an acoustic "snapshot" of the position of the source at the beginning and end of its trajectory (Grantham, 1986) and discriminates on the basis of static positional changes. Not all the available data support this view, but the exceptions are relatively minor (Perrott and Marlborough, 1989).

Gibson took issue with the notion of a series of perceptual snapshots, which requires fusion or composition to account for the perception of a single moving object (1966). By redefining information for motion perception, Gibson eliminated the need for a concept such as fusion. Since motion information is available to the observer, even through discrete "looks", the additional step of reconstruction to a continuous event is simply not necessary. To Gibson, the mechanics of the mediating sensory system were not germane to the perception of motion. To have "dynamic event perception", in contrast to the less elegant "motion perception plus inference", it must be shown that even though dynamic properties, such as mass and inertia, are not present in the optic (or acoustic) array, they are specified by the kinematics. That is, the information regarding the physical motion of an object is conveyed through the kinematics, whether discrete or continuous.

Research on motion aftereffects provides indirect evidence on the question of the existence of specialized motion detectors. The idea is that exposure to an adapting stimulus that is moving in one direction fatigues the neural elements that respond to movement in that direction. The aftereffect, a perception of movement in the opposite direction, is presumed to reflect the spontaneous activity of the neural elements sensitive to movement in the opposite direction. Movement aftereffects are common in vision, one variation of which is called the "waterfall illusion" (Sekular and Pantle, 1967).

Grantham (1989, 1992) has reported reliable though weak evidence for motion aftereffects in audition. After prolonged exposure to a free-field adapting stimulus that was moving in the horizontal plane, listeners' judgments of the direction of movement of a subsequently presented probe stimulus were slightly biased in a direction opposite to that of the adapting stimulus. While the effects were disappointingly small, the results were nevertheless suggestive.

Some of the research on perception of moving sound sources has been less concerned with the existence of specialized motion detectors and more broadly focussed. For example, several studies have attempted to quantify the relative salience of the various sources of acoustical information that signal source movement. These experiments ask listeners to indicate the time at which a moving source is closest to them (time to interception) or the time at which they would make contact with the source (acoustic "tau"). In a theoretical study, Shaw, McGowan and Turvey (1991) analyzed the acoustic intensity field produced by collinear relative movement between a sound source and an observer and showed the acoustic-tau to be related to the inverse of the relative change in average intensity. Jenison (1994) extended the analysis to the more general case, including "time-to-interception", showing that time-averaged intensity and time-varying ITD and their corresponding first-order derivatives are sufficient for conveying both collision and interception information.

Empirical studies of auditory time-to-contact or time-to-interception include that reported by Rosenblum, Carello, and Pastore, 1987, in which listeners heard sound sources over headphones. Three stimulus variables were manipulated, interaural time difference, overall level, and Doppler shift. Each was presented both in isolation and in competition such that each indicated a different point of closest approach, or interception. The results suggested that while any of the three stimulus parameters could accurately indicate point of closest approach, overall level was the dominant cue. The authors argue that overall level should be dominant since it is the only cue of the three that is, in all environmental circumstances, unequivocal. Todd (1981) investigated how well subjects could discriminate time-to-contact for visual stimuli by simulating two simultaneously approaching objects on a computer display. Subjects were asked to judge which object would arrive first. We have recently launched analogous experiments that examine subjects' ability to discriminate the arrival of two sound sources. Sounds were synthesized according to the simple kinematics of a moving sound composed of three harmonics using ITD, average intensity, and Doppler shift. A sound arriving to the left of the listener was mixed with a sound arriving differentially in time to the right of the observer. Subjects were asked to choose which sound would arrive sooner. Figure 10 shows preliminary results from 24 subjects. In Todd's experiment relative time-to-contact was 75 % correctly discriminated when the difference in time-to-contact was about 50 ms. In contrast the relative auditory time-to-contact in our preliminary studies was 75 % correctly discriminated when the difference was about 300 ms. Schiff and Oldak (1990) examined observers' accuracy in using visual and acoustical estimates of time-to-arrival from film and sound-recorded approaching vehicles. Their data indicate that sighted subjects were significantly more accurate in estimating time-to-arrival with sight than sound, however blind subjects performed as well or better than sighted with only the acoustic channel. While the evidence is only suggestive at this point, human observers have the capacity to efficiently estimate relative time-to-contact regardless of how the information is conveyed as long as the temporal window for estimation is within several seconds. This restricted window should not be surprising given the pattern of the observables described above. Significant changes in ITD, intensity, and Doppler occur only in a spatial region (hence the temporal region as well) about the closest-point-of-approach. This relationship holds for subtended angle in the visual domain as well.

Head movements provide a somewhat different kind of dynamic auditory stimulus from movement of the sound source. Because head movements typically involve changes only in the direction of the sound source with respect to the head there is very little doppler shift and very little change in overall level. However, interaural parameters change more rapidly with head movements than with typical source movement. In addition, head movements provide additional information to the perceiver via proprioceptive feedback from the neck musculature. While there has been speculation about the role of head movements for decades, there have been few empirical studies of their role (Thurlow and Runge,1967; Pollack and Rose, 1967; Simpson and Stanton, 1973). Only recently has empirical research begun to provide firm evidence of the importance of head movements for perception of the spatial layout of auditory objects.

Given a stationary auditory object in the environment there is a change in the angular relation of the object and a listener's head that accompanies normal head movement. This change in relative orientation produces a systematic and predictable change in the pattern of spatial cues (ITD, ILD, spectral cues) produced by the object at the listener's ears. If these normal changes in the spatial cues are disrupted the apparent position of the auditory object is often disturbed. Young (1931) reported one of the first demonstrations of this phenomenon. In this experiment sounds were routed to the ears through rubber tubes attached to fixed ear trumpets. With this arrangement the normal coupling between a listener's head movements and changes in the acoustical stimulus at the ears was eliminated. Listeners reported all sounds as originating behind the head, outside of the listeners' visual fields, regardless of the actual position of the sound source. Similar front-back confusions are reported in the modern studies of virtual sound sources that are synthesized and presented to listeners via headphones (Wightman and Kistler, 1989b).

As mentioned above, front-back confusions are not entirely unexpected given the rough spherical symmetry of the head and the salience of ITD cues. The idea that in everyday life a listener's head movements might provide the information needed to avoid them is usually attributed to Wallach (1940). Wallach showed that if a listener could monitor the direction of change in ITD that accompanied a head movement, the front-back ambiguity could be avoided. For example, suppose a sound is presented at an azimuth of 45° and an elevation of 0° (on the horizontal plane, roughly 45° to the right of the median plane). A front-back confusion would be represented by an apparent azimuth report of roughly 135°. If the listener's head moved to the right, the ITD produced by the source initially at 45° would decrease because the angle of the source relative to the head would approach 0°, the point of minimum ITD. However, if the source were actually at 135° azimuth, the ITD would have increased. Thus, the direction of change in ITD unambiguously indicates whether the source was in the front or the rear.

In spite of the simplicity and face validity of Wallach's arguments, conclusive evidence that head movements are used to resolve front-back confusions has not appeared. One obvious reason for this is that experiments which control both head movements and the associated auditory stimulus dynamics have been technically too demanding until recently. Advanced technology now allows synthesis of virtual sources in such a way that the effects of head movements can be studied directly. Using magnetic head trackers and real-time convolution devices such as the Convolvotron (Foster, Wenzel, and Taylor, 1991), one can monitor a listener's head position continually during an experiment and adjust the synthesis algorithms dynamically (20-40 times per second) to simulate a stationary source. As the listener's head moves, the device compensates for changes in the relative positions of the stationary virtual source and the head by using different left-right pairs of HRTF-based filters for each updated head position. The movement compensation is smooth and the resultant percept of an external sound source in a stationary position is compellingly realistic (Wenzel, 1992).

We have recently begun some research on the role of head movements that takes advantage of the new technology and attempts to clarify some of the issues raised by the earlier work (Wightman, Kistler, and Andersen, 1994). The essential elements of the paradigm were as described in earlier work (Wightman and Kistler, 1989b). Listeners localized virtual sources (2.5s wideband noise bursts) in two conditions. In one, the virtual stimuli were presented over headphones with no head-tracking, and the listeners were asked not to move their heads during the test. In the other, a magnetic head tracker was used to sense head position and the virtual synthesis algorithm were modified in real time according to the head tracker's reports. In the second condition, listeners were encouraged to move their heads during stimulus presentation if they felt it would facilitate localization. Apparent position judgments were made verbally after each stimulus presentation. Preliminary results from a single listener are shown in Figure 11. Note that in the head stationary condition this listener made frequent front-back confusions, as evidenced by the off-diagonal responses in the "front-back" panel. In the head-movement condition, however, the front-back confusions were nearly eliminated. The listeners' gave no indication of other differences between the two conditions, either in their apparent position judgments or their subjective reports. Thus, in contrast with suggestions in the literature, apparent source distance was the same with and without head movements (cf: Simpson and Stanton, 1973), and the images were equally well externalized in the two conditions (cf: Durlach, et al. 1992). We conclude on the basis of these results that the primary role of head movements is resolution of confusions about the spatial layout of auditory objects.

VI. The role of auditory-visual interactions in the spatial layout of auditory objects

The sensory environment of most individuals includes both visual and auditory objects, and in many cases sound-producing objects can be seen as well as heard. Thus, while it is useful and informative to consider audition alone when discussing the spatial layout of auditory objects.

it is important to be mindful of the potential role played by vision. Indeed, some auditory-visual interactions are quite powerful and their consequences well documented.

The so-called "ventriloquism effect" is perhaps the best known of the auditory-visual interactions (e.g., Pick, Warren, and Hay, 1969). The typical manifestation of the effect is a strong biasing of the apparent position of an auditory object in the direction of a simultaneously present visual object. Evidence of the potency of this effect is familiar to anyone who has watched the image of someone speaking at the movies or on television. While the sound of the voice clearly seems to originate at the mouth of the person speaking, the actual source of the sound, a loudspeaker, is usually displaced far to one side. Clearly one's perception of the spatial layout of auditory objects will be heavily influenced by whether or not the source of the sound is visible.

Additional evidence for auditory-visual interactions comes from research on visual facilitation (e.g., Warren, 1970). Visual facilitation refers to the fact that the variance of localization judgments is lower when listeners hear the test stimulus in a lighted room than when they hear it in the dark. The source of sound is invisible in either case, and whether or not the listener makes the response in the light or the dark is irrelevant to the outcome. It is as if the listener is able to establish a frame of reference within which to place the auditory objects, and the presence of the frame of reference facilitates localization. Some investigators argue that eye movements, even in the absence of visual input, are the basis of the facilitation effect (Jones and Kabanoff, 1975), but the issue is far from being resolved. What is especially interesting about the visual facilitation effect is that it occurs only in adults. Children as old as 12 years do not show the effect (Warren, 1970).

VII. Conclusions

The study of auditory object perception in general and the spatial layout of auditory objects in particular is in its infancy. In the case of the spatial layout of single stationary sound sources in anechoic space much is known about the sources of information and how that information is processed. The salience of ITD cues, the importance of monaural spectral cues derived from pinna filtering, the role of head movements, etc., have been thoroughly documented in studies of single stationary sources. Relatively few investigators have ventured beyond the relative security of this constraint, so that experiments involving non-anechoic listening conditions and moving sources are scarce, and studies of multiple sources are virtually non-existent. The potential sources of information are reasonably well understood, but how that information might be used in the auditory system is completely unknown.

The state of affairs in hearing contrasts sharply with the relative maturity of the study of visual spatial layout, in which research on such complex topics as optic flow has been in progress for decades. One reason for the slower progress on the hearing side may be that the experiments are technically more demanding. For example, it is easier to present an arbitrary visual pattern to a retina than an arbitrary sound waveform to an eardrum. Technology is changing this situation rapidly, so we can expect significant advances in our understanding of auditory object perception in the near future.

References

Bregman, A. (1990). Auditory Scene Analysis. The MIT Press, Cambridge, Ma.

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

Begault, D. (1992) Perceptual Effects of Synthetic Reverberation on Three-Dimensional Audio Systems. Journal of the Audio Engineering Society, 40, 895-904.

Clifton, R. & Freyman, R. (1989) Effect of click rate and delay on breakdown of the precedence effect. Perception & Psychophysics, 46, 139-145.

Cochran, P., Throop, J., & Simpson, W. E. (1968). Estimation of distance of a source of sound. American Journal of Psychology, 81, 198-207.

Coleman, P. (1963). An analysis of cues to auditory depth perception in free space. Psychological Bulletin, 60, 302-315.

Coleman, P. D. (1962). Failure to localize the source distance of an unfamiliar sound. Journal of the Acoustical Society of America, 34, 345-346.

Durlach, N. I., Rigopulos, A., Pang, X. D., Woods, W. S., Kulkarni, A., Colburn, H. S., & Wenzel, E. M. (1992). On the Externalization of Auditory Images. Presence, 1 (2), 251-257.

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

Gardner, M. B. (1968). Proximity image effect in sound localization. Journal of the Acoustical Society of America, 43, 163.

Gibson, J. J. (1966). The Senses Considered as Perceptual Systems. (Houghton Mifflin, Boston, MA).

Grantham, D. W. (1986). Detection and discrimination of simulated motion of auditory targets in the horizontal plane. Journal of the Acoustical Society of America, 79, 1939- 1949.

Grantham, D. W. (1989). Motion aftereffects with horizontally moving sound sources in the free field. Perception & Psychophysics, 45 (2), 129-136.

Grantham, D. W. (1992). Adaptation to auditory motion in the horizontal plane: Effect of prior exposure to motion on motion detectability. Perception & Psychophysics, 52 (2), 144-150.

Green, D. M. (1976). An Introduction to Hearing. John Wiley & Sons, Inc., New York.

Harris, J. D., & Sergeant, R. L. (1971). Monaural/binaural minimum audible angle for a moving sound source. Journal of Speech and Hearing Research, 14, 618-629.

Hartmann, W. M. (1983). Localization of sound in rooms. Journal of the Acoustical Society of America, 74, 1380-1391.

Holt, R. E., & Thurlow, W. R. (1969). Subject orientation and judgment of distance of a sound source. Journal of the Acoustical Society of America, 6 (2), 1584.

Jenison, R. L. (1994). On acoustic information for auditory motion, Perception, submitted

Jenison, R. L. and Lutfi, R. A. (1992). Kinematic synthesis of auditory motion, Journal of the Acoustical Society of America, 92, 2458.

Jones, B., & Kabanoff, B. (1975). Eye movements in auditory space perception. Perception & Psychophysics, 17, 241-245.

Kuhn, G. F. (1977). Model for the interaural time differences in the azimuthal plane. Journal of the Acoustical Society of America, 62, 157-167.

Lambert, R. (1974). Dynamic theory of sound-source localization. Journal of the Acoustical Society of America, 56, 165-171.

Maunsell, J. H. R. and VanEssen, D. C. (1983). Functional properties of neurons in middle temporal visual area (MT) of macaque monkey: I. Binocular interactions and the sensitivity to binocular disparity. Journal of Neurophysiology, 49, 1148-1167.

McGregor, P., Horn, A. G., and Todd, M. A. (1985). Are familiar sounds ranged more accurately?. Perceptual and Motor Skills, 61, 1082.

Mershon, D. H., & King, L. E. (1975). Intensity and reverberation as factors in the auditory perception of egocentric distance. Perception & Psychophysics, 18 (6), 409-415.

Middlebrooks, J. C. (1992). Narrow-band sound localization related to external ear acoustics. Journal of the Acoustical Society of America, 92(5), 2607-2624.

Middlebrooks, J. C., & Green, D. M. (1990). Directional dependence of interaural envelope delays. Journal of the Acoustical Society of America, 87(50), 2149-2162.

Middlebrooks, J. C., & Green, D. M. (1991). Sound localization by human listeners. In Annual Review of Psychology (pp. 135-159)., Annual Reviews Inc.

Middlebrooks, J. C., Makous, J. C., & Green, D. M. (1989). Directional sensitivity of sound-pressure levels in the human ear canal. Journal of the Acoustical Society of America, 86(1), 89-108.

Oldfield, S.R., and Parker, S.P.A. (1984a) Acuity of sound localization: A topography of auditory space. I. Normal hearing conditions. Perception, 13, 581-600.

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

Perrott, D. R., & Marlborough, K. (1989). Minimum audible movement angle: Marking the end points of the path traveled by a moving sound source. Journal of the Acoustical Society of America, 85, 1773-1775.

Perrott, D. R., & Musicant, A. D. (1977). Minimum auditory movement angle: Binaural localization of moving sound sources. Journal of the Acoustical Society of America, 62, Spitzer, M. W. and Semple, M. N. (1993). Responses of inferior colliculus neurons to time-varying interaural phase disparity: effects of shifting the locus of virtual motion. Journal of Neurophysiology, 69, 1245-1263.

Spitzer, M. W. and Semple, M. N. (1991). Interaural phase coding in auditory midbrain: influence of dynamic stimulus features. Science, 254, 721-724.

Stevens, S.S., and Newman, E.B. (1936) The localization of actual sources of sound. American Journal of Psychology, 48:297-306.

Strutt, J. W. (1907). On our perception of sound direction. Philosophical Magazine, 13, 214-232.

Stumpf, E., Toronchuk, J. M. and Cynader, M. S. (1992). Neurons in cat primary auditory cortex sensitive to correlates of auditory motion in three-dimensional space. Experimental Brain Research, 88, 158-168.

Thurlow, W.R., and Runge, P.S. (1967) Effect of induced head movements on localization of direction of sounds. Journal of the Acoustical Society of America, 42 (2):480-488.

Todd, J. (1981). Visual information about moving objects. Journal of Experimental Psychology: Human Perception and Performance, 7, 795-810.

Wallach, H. (1940). The role of head movements and vestibular and visual cues in sound localization. Journal of Experimental Psychology, 27 (4), 339-368.

Wallach, H., Newman, E., & Rosenzweig, M. (1949). The precedence effect in sound localization. The American Journal of Psychology, 62, 315-336.

Warren, D. H. (1970). Intermodality Interactions in Spatial Localization. In W. Reitman (Ed.), Cognitive Psychology (pp. 114-133). New York, Academic Press.

Wenzel, E. M. (1992). Localization in Virtual Acoustic Displays. Presence, 1(1), 80-107.

Wenzel, E. M., Arruda, M., Kistler, D.J. & Wightman, F. L. (1993) Localization using nonindividualized head-related transfer functions. Journal of the Acoustical Society of

1463- 1466.

Perrott, D. R., & Tucker, J. (1988). Minimum audible movement angle as a function of signal frequency and the velocity of the source. Journal of the Acoustical Society of America, 83, 1522-1527.

Pick, H. L., Warren, D. H., & Hay, J. C. (1969). Sensory conflict in judgments of spatial direction. Perception & Psychophysics, 6, 203-205.

Pollack, I., and Rose, M. (1967) Effects of head movements on the localization of sounds in the equatorial plane. Perception & Psychophysics, 2, 591- 596.

Pralong, D., & Carlile, S. (1994). Measuring the human head-related transfer functions: A novel method for the construction and calibration of a miniature in-ear recording system. Journal of the Acoustical Society of America, 95, 3435-3444.

Rosenblum, L. D., Carello, C., & Pastore, R. E. (1987). Relative effectiveness of three stimulus variables for locating a moving sound source. Perception, 16, 175-186.

Schiff, W. and Oldak, R. (1990). Accuracy of judging time to arrival: effects of modality, trajectory, and gender. Journal of Experimental Psychology: Human Perception and Performance, 16, 303-316.

Sekular, R., & Pantle, A. (1967). A model for after-effects of seen movement. Vision Research, 7, 427-439.

Shaw, B. K., McGowan, R. S. and Turvey, M. T. (1991). An acoustic variable specifying time-to-contact. Ecological Psychology, 3, 253-261.

Shaw, E. A. G. (1974). Transformation of sound pressure level from the free field to the eardrum in the horizontal plane. Journal of the Acoustical Society of America, 56 (6), 1848-1861.

Simpson, W., & Stanton, L. (1973). Head movement does not facilitate perception of the distance of a source of sound. American Journal of Psychology, 86, 151-160.

America, 94, 111-123.

Wightman, F. L., & Kistler, D. J. (1989a). Headphone simulation of free-field listening I: stimulus synthesis. Journal of the Acoustical Society of America, 85, 858-867.

Wightman, F. L., & Kistler, D. J. (1989b). Headphone simulation of free-field listening II: psychophysical validation. Journal of the Acoustical Society of America, 85, 868-878.

Wightman, F. L., & Kistler, D. J. (1992). The dominant role of low-frequency interaural time differences in sound localization. Journal of the Acoustical Society of America, 91 (3), 1648-1661.

Wightman, F. L., & Kistler, D. J. (1993). Sound Localization. In R. Fay, A. Popper, & W. Yost (Eds.), Springer Series in Auditory Research: Human Psychophysics (pp. 155-192). New York, Springer-Verlag.

Wightman, F. L., Kistler, D. J., & Andersen, K. J. (1994). Reassessment of the role of head movements in human sound localization. Journal of the Acoustical Society of America, 95, 3003.

Woodworth, R. S. (1938). Experimental psychology. Holt, New York.

Young, P. T. (1931). The role of head movements in auditory localization. Journal of Experimental Psychology, XIV (2), 95-124.

Zakarauskas, P. and Cynader, M. S. (1991). Aural intensity for a moving source. Hearing Research, 52, 233-244.

Zurek, P. (1980). The precedence effect and its possible role in the avoidance of interaural ambiguities. Journal of the Acoustical Society of America, 67, 952-964.

Figure Legends

Figure 1: Schematic top-down representation of a listener and a sound source. The source is assumed to be sufficiently far from the listener that the acoustical wavefronts are planar, and the listener is assumed to have a spherical head with ears at opposite ends of a diameter.

Figure 2: Directional transfer functions from two listeners produced by a source at 90° azimuth. Directional transfer functions (DTFs) are HRTFs divided by the RMS average of the HRTFs from all spatial positions measured. Thus, the DTFs represent the deviation in dB from the average response of the ear. (Adapted with permission from Wightman and Kistler, 1993.)

Figure 3: Interaural time differences (ITDs), produced by a source at 0° elevation, predicted by the spherical head model (solid line) and ITDs measured from a typical listener using a wideband correlation technique. (Reproduced with permission from Wightman and Kistler, 1993.)

Figure 4: Interaural time differences from HRTF measurements from a typical listener plotted as a function of the azimuth and elevation of the sound source. Note the contours of constant ITD below the surface plot. (Adapted with permission from Wightman and Kistler, 1993.)

Figure 5: Interaural level difference (ILD) aa a function of frequency from a typical listener produced by a source at 0° elevation and 0° azimuth (dashed line) or 90° azimuth (solid line).

Figure 6: Interaural level difference from a typical listener in different frequency regions. Figure 6a shows ILDs across the entire frequency spectrum, and Figures 6b and 6c show

ILD in two high frequency critical bands. (Adapted with permission from Wightman and Kistler, 1993.)

Figure 7: Schematic diagram showing angular relations between a listener and a sound source that is moving along a straight path (represented by the arrow).

Figure 8: Schematic diagram showing three example trajectories for a moving sound source.

Figure 9: Results of kinematic analysis of the ITD (panel a), intensity (panel b), and doppler shift (panel c) cues produced by a moving sound source. The rates of change of those cues are shown in panels b, d, and f.

Figure 10: Average psychometric function from 24 listeners in the time-to-contact experiment. Percent correct discrimination between two sounds arriving at different times is plotted as a function of the arrival time difference.

Figure 11: Apparent source position judgments from a single listener in an experiment in which the listener heard virtual sources presented over headphones. In one condition (left panels) was required to hold his/her head still, and in the other condition (right panels) head movements were encouraged and the virtual stimuli were modified in real time according to the listener's head position to simulate a stationary external source. Each judgment of apparent azimuth and elevation is represented in 3 panels that reflect the extent (expressed as an angle from -90° to +90°) to which the judged position is on the right or left (top), in the front or back (middle), and above or below the horizontal plane (bottom). The darkness of each symbol represents the number of judgments that fell in the local area of the symbol.

Acknowledgments

The authors are indebted to Dr. Doris Kistler and many others in the Hearing Development Research Laboratory for their assistance in preparing this chapter and in conducting the research that led to it. Preparation of the manuscript was supported in part by research grants from the NIH (NIDCD), NASA, and ONR.

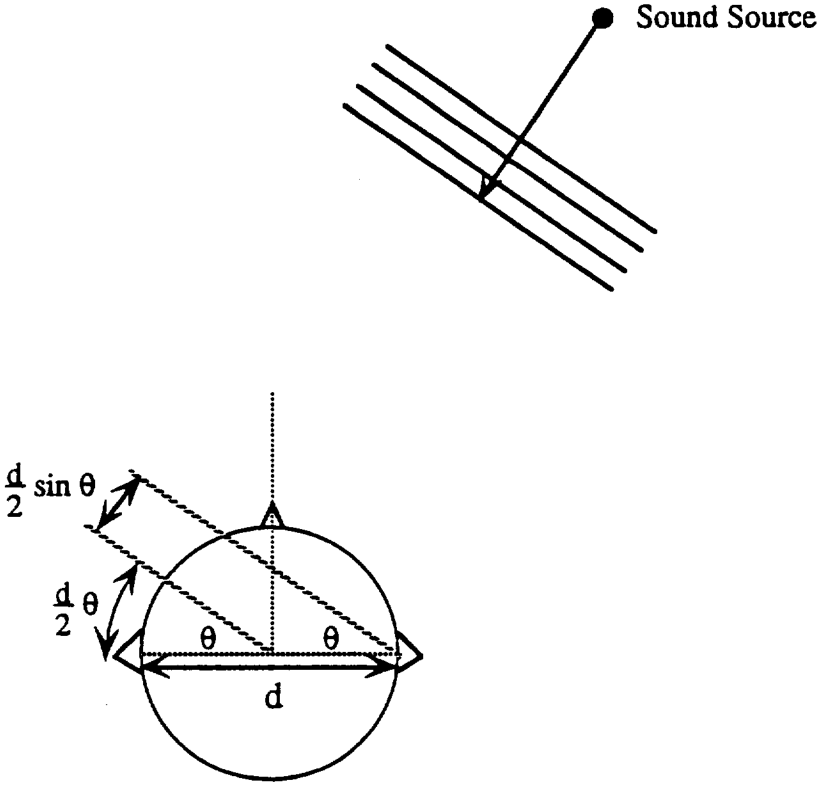


Figure 1

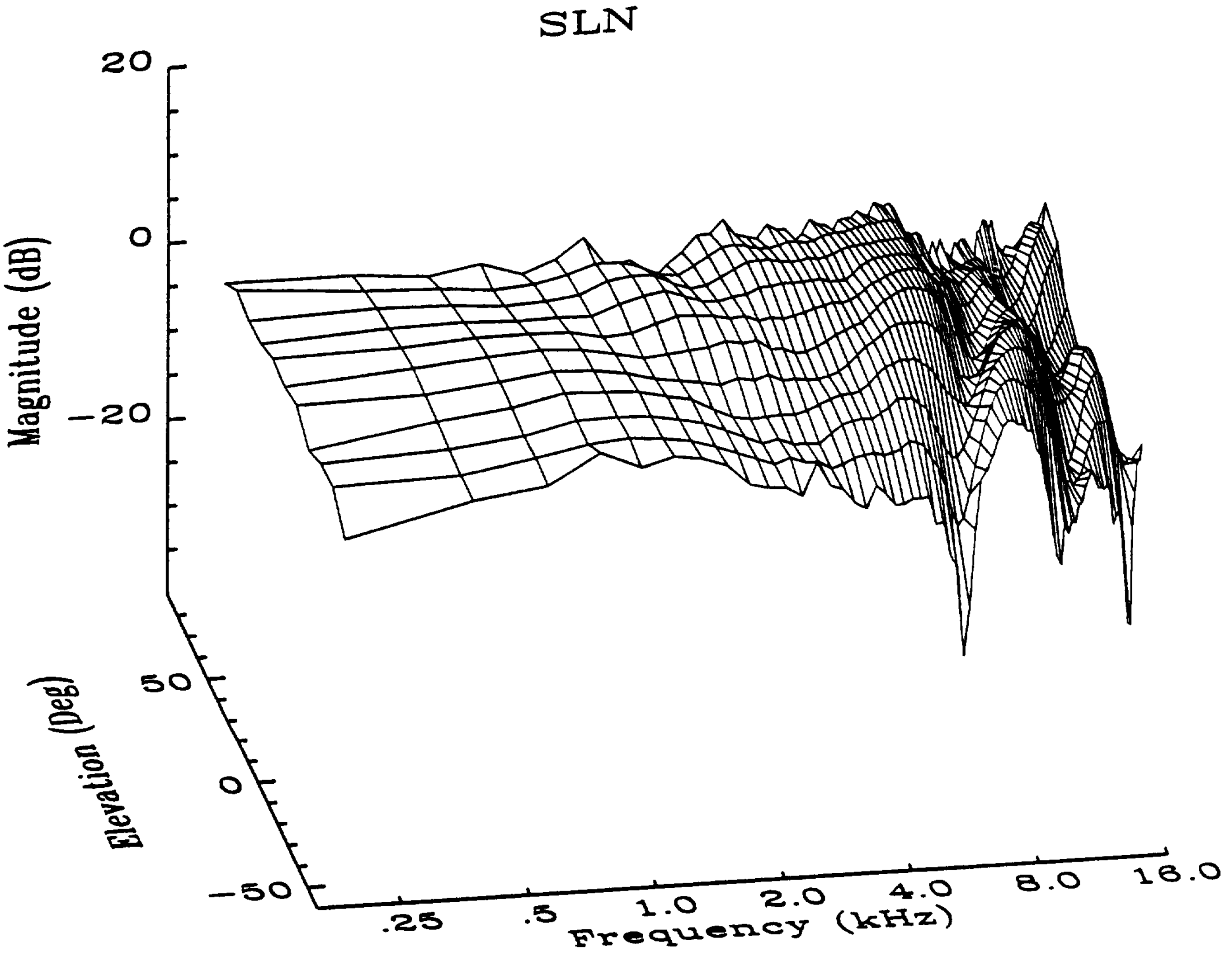

SLN
Magnitude (dB)
20
0
-20
Elevation (Deg)
50
0
-50
.25
.5
1.0
2.0
4.0
8.0
16.0
Frequency (kHz)

Figure 2a

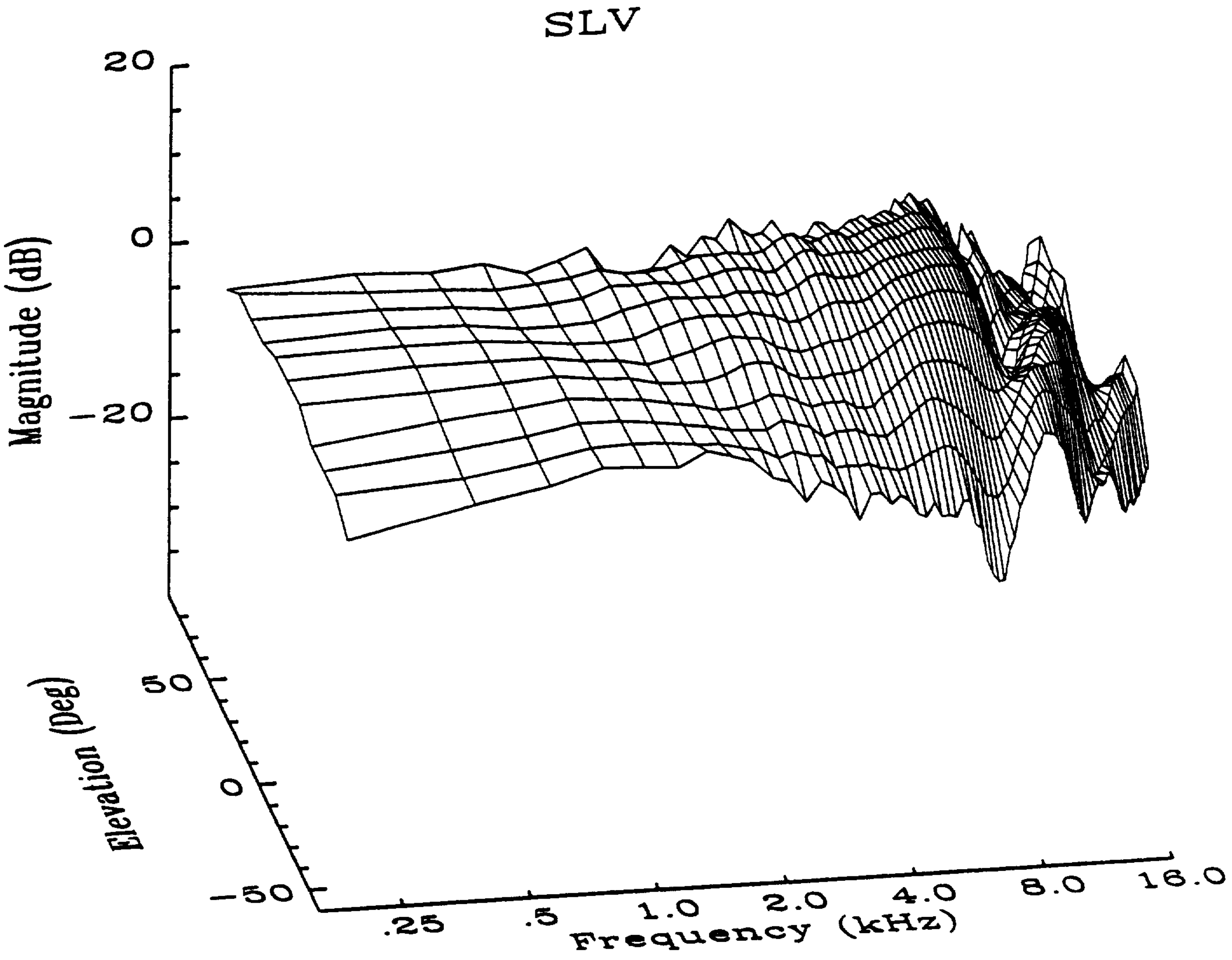

SLV
20
0
−20
Magnitude (dB)
50
0
−50
Elevation (Deg)
.25
.5
1.0
2.0
4.0
8.0
16.0
Frequency (kHz)

Figure 2b

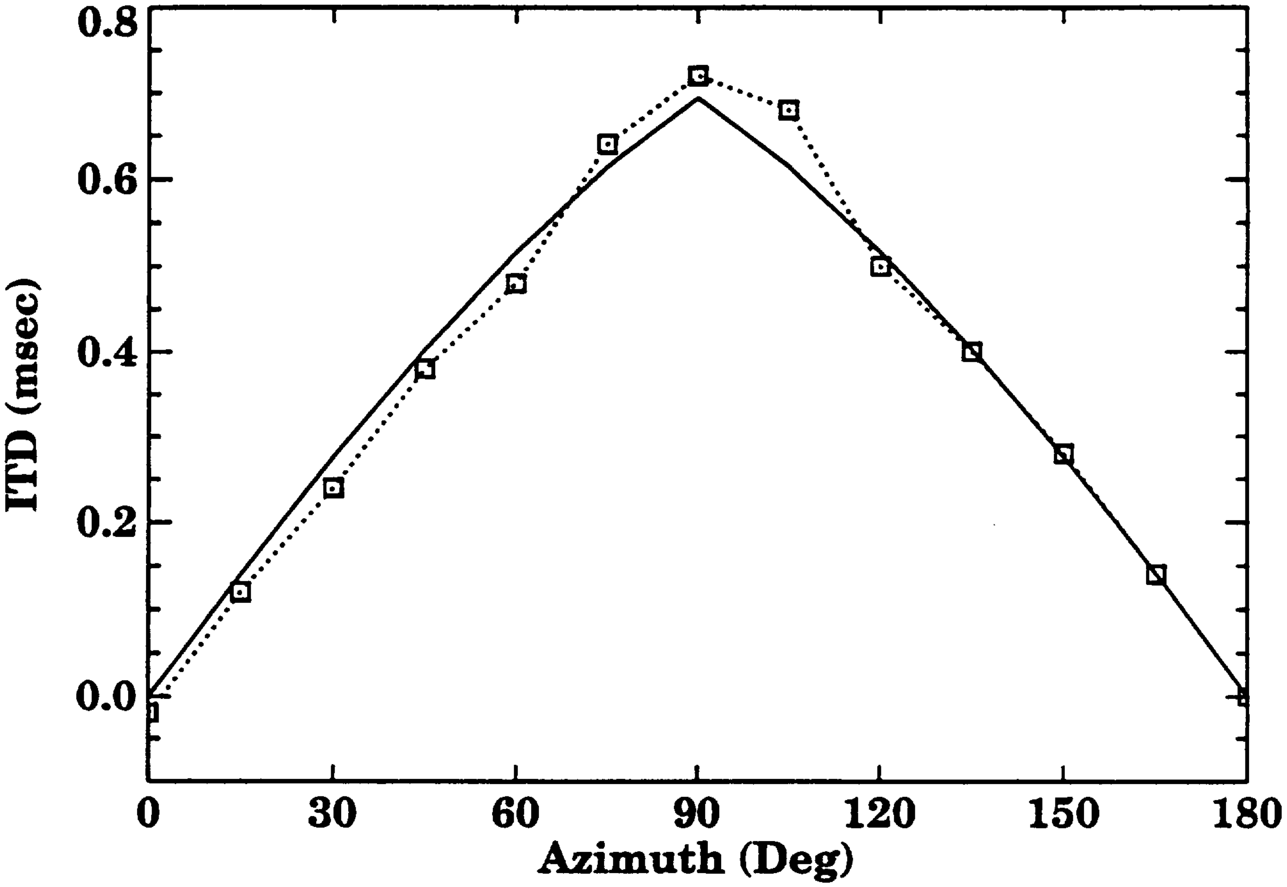


Figure 3

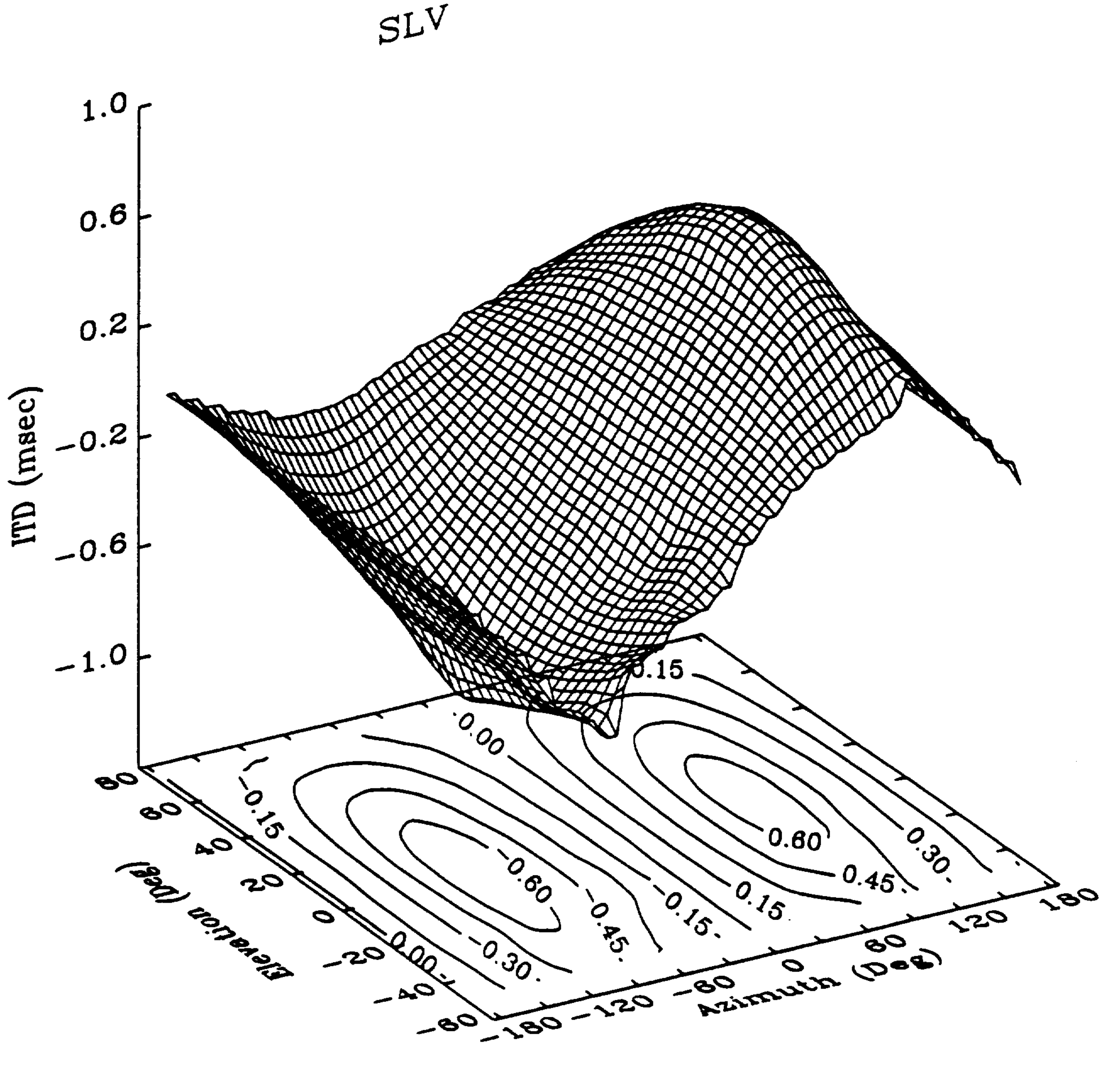

SLV
1.0
0.6
0.2
-0.2
-0.6
-1.0
ITD (msec)
80
60
40
20
0
-20
-40
-60
Elevation (Deg)
-180
-120
-60
0
60
120
180
Azimuth (Deg)
0.15
0.00
-0.15
-0.60
-0.30
-0.45
0.00
-0.15
0.15
0.60
0.45
0.30

Figure 4

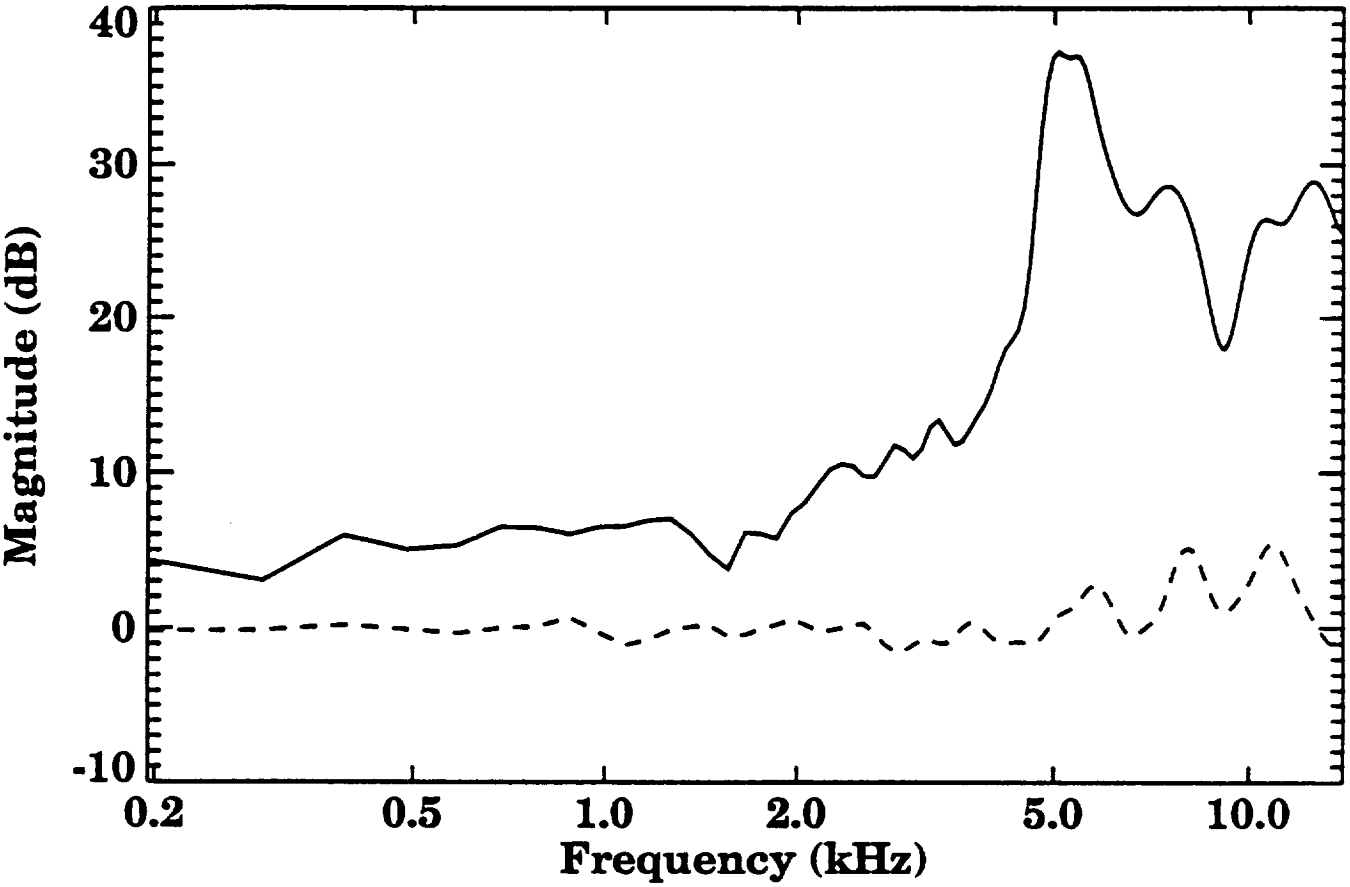


Figure 5

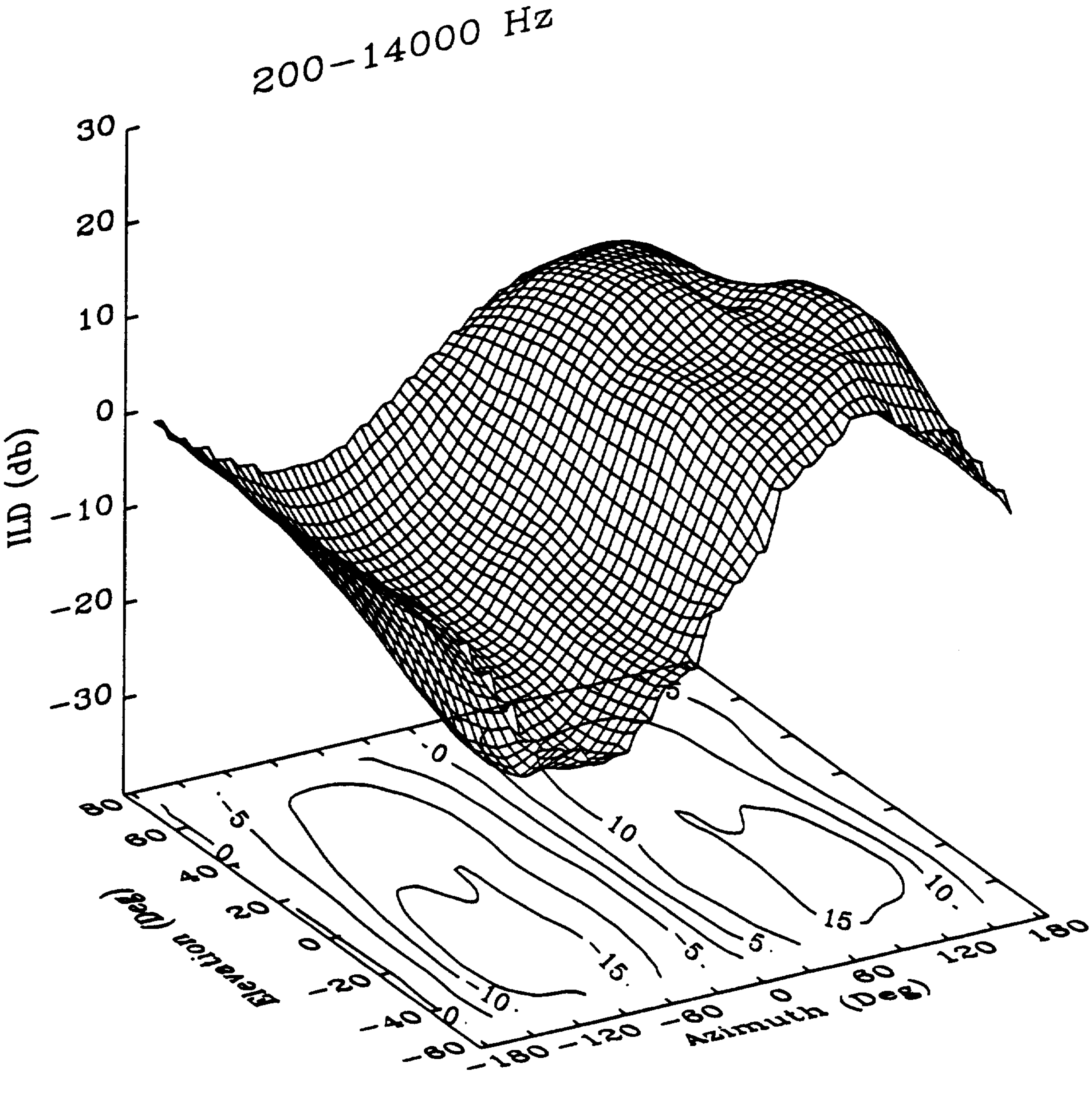

200-14000 Hz
30
20
10
0
-10
-20
-30
ILD (db)
80
60
40
20
0
-20
-40
-60
Elevation (Deg)
-180
-120
-60
0
60
120
180
Azimuth (Deg)
-5
-0
0
5
10
15
-15
-10
-5
5
10
15

Figure 6a

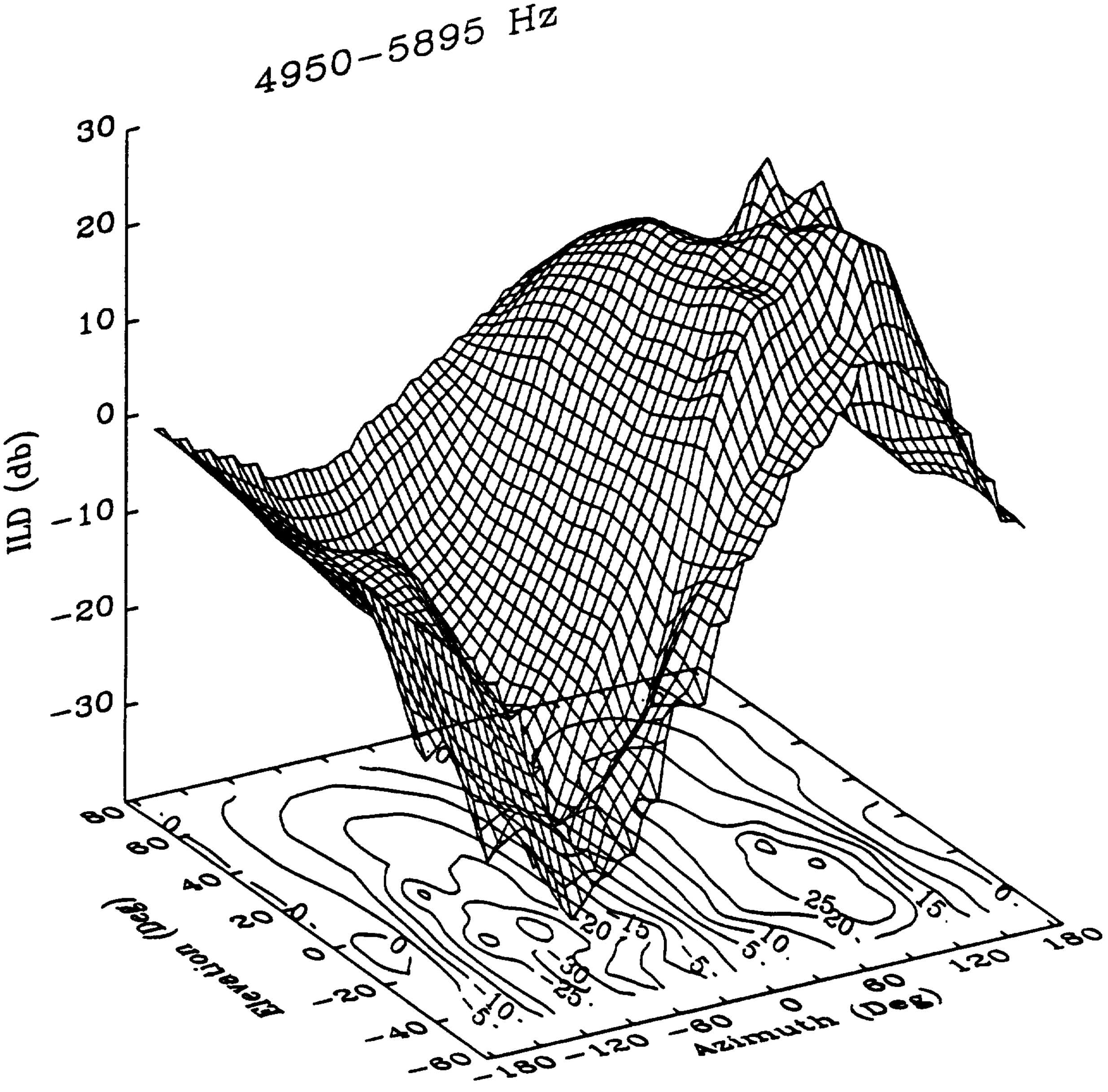

4950-5895 Hz
ILD (db)
30
20
10
0
-10
-20
-30
Elevation (Deg)
80
60
40
20
0
-20
-40
-60
Azimuth (Deg)
-180
-120
-60
0
60
120
180
0
-5
-10
-15
-20
-25
-30
5
10
15
20
25

Figure 6b

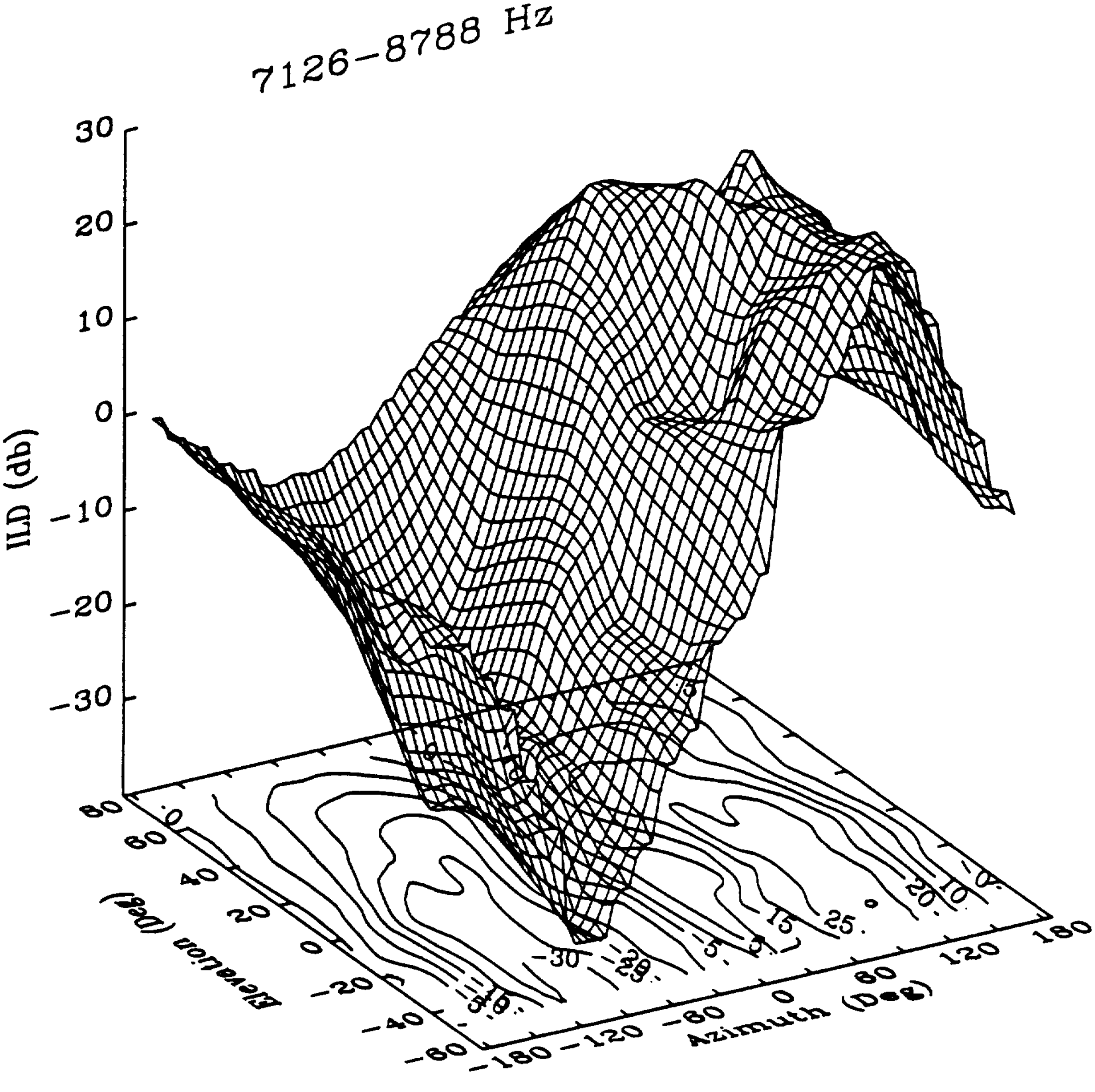

7126-8788 Hz
ILD (db)
30
20
10
0
-10
-20
-30
Elevation (Deg)
80
60
40
20
0
-20
-40
-60
Azimuth (Deg)
-180
-120
-60
0
60
120
180
-30
-25
-20
-5
5
15
25
20
10
0
-10

Figure 6c

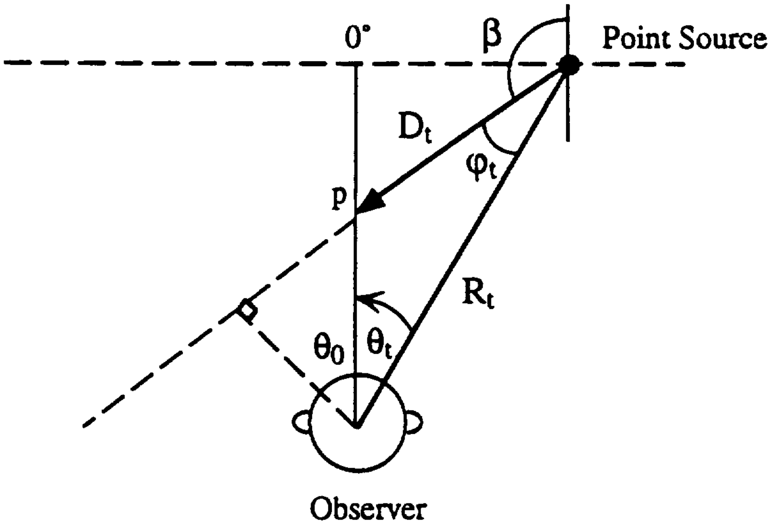


Figure 7

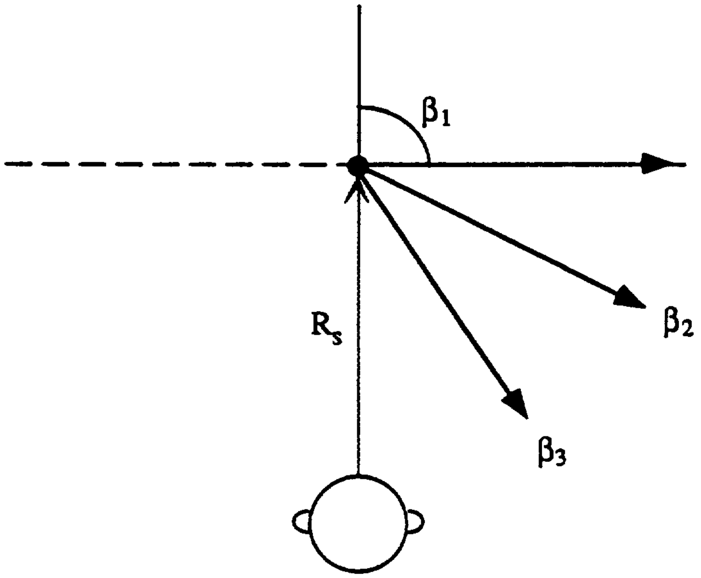


Figure 8

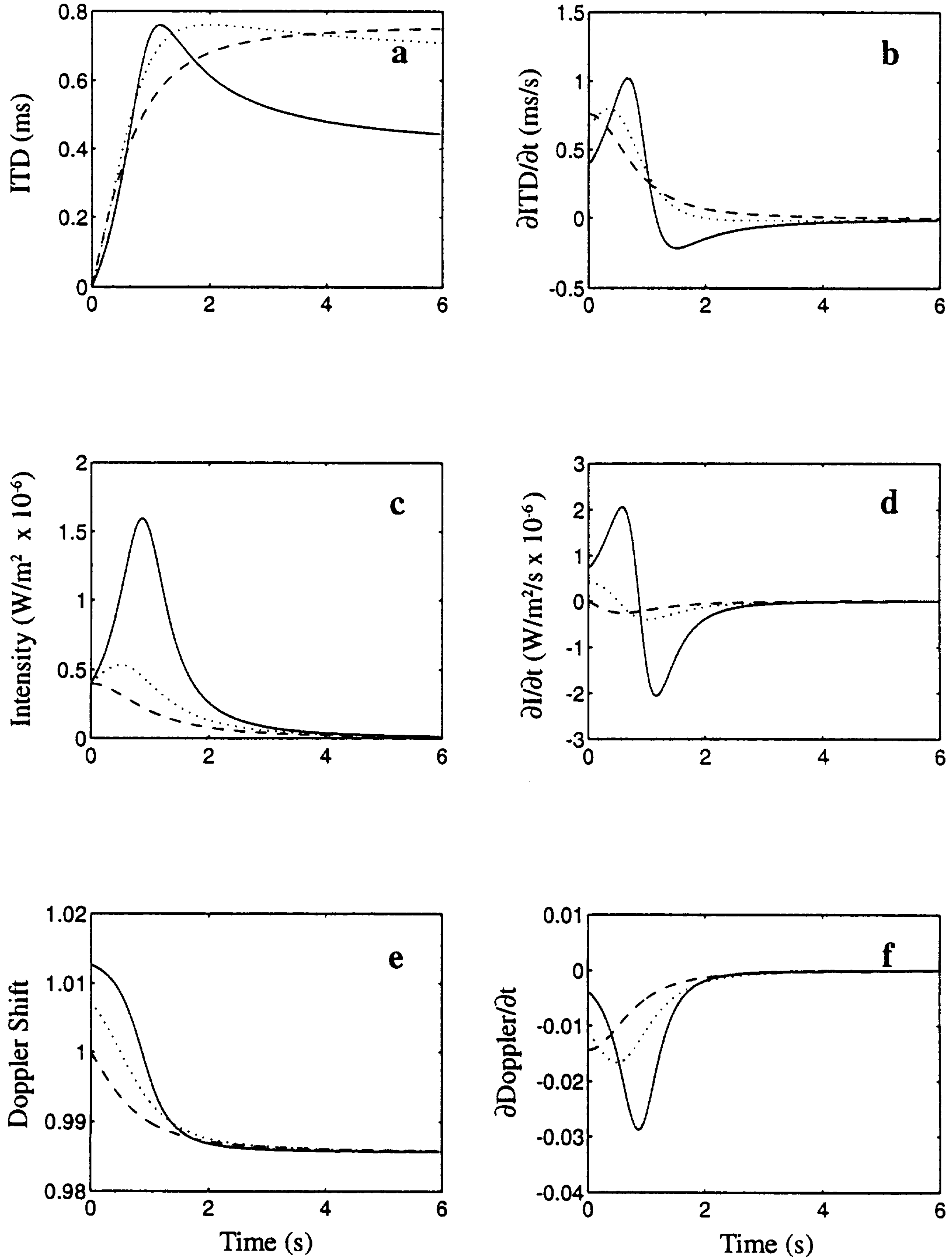

a
ITD (ms)
0.8
0.6
0.4
0.2
0
0
2
4
6
b
∂ITD/∂t (ms/s)
1.5
1
0.5
0
-0.5
0
2
4
6
c
Intensity (W/m² x 10⁻⁶)
2
1.5
1
0.5
0
0
2
4
6
d
∂I/∂t (W/m²/s x 10⁻⁶)
3
2
1
0
-1
-2
-3
0
2
4
6
e
Doppler Shift
1.02
1.01
1
0.99
0.98
0
2
4
6
Time (s)
f
∂Doppler/∂t
0.01
0
-0.01
-0.02
-0.03
-0.04
0
2
4
6
Time (s)

Figure 9

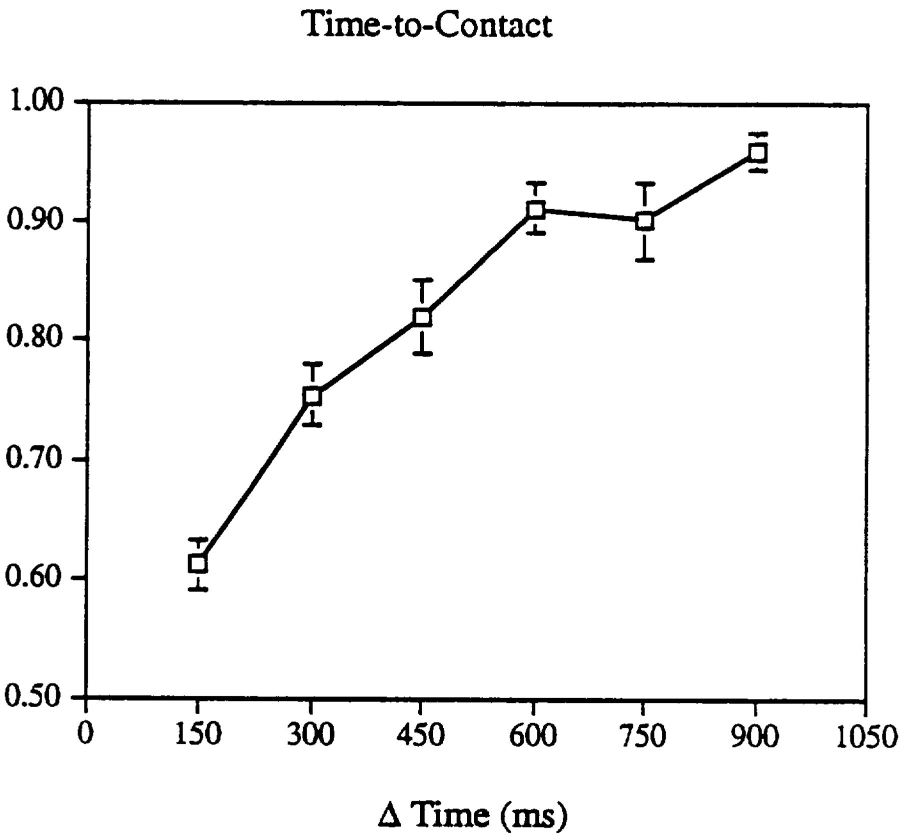

Time-to-Contact
1.00
0.90
0.80
0.70
0.60
0.50
0
150
300
450
600
750
900
1050
Δ Time (ms)

Figure 10

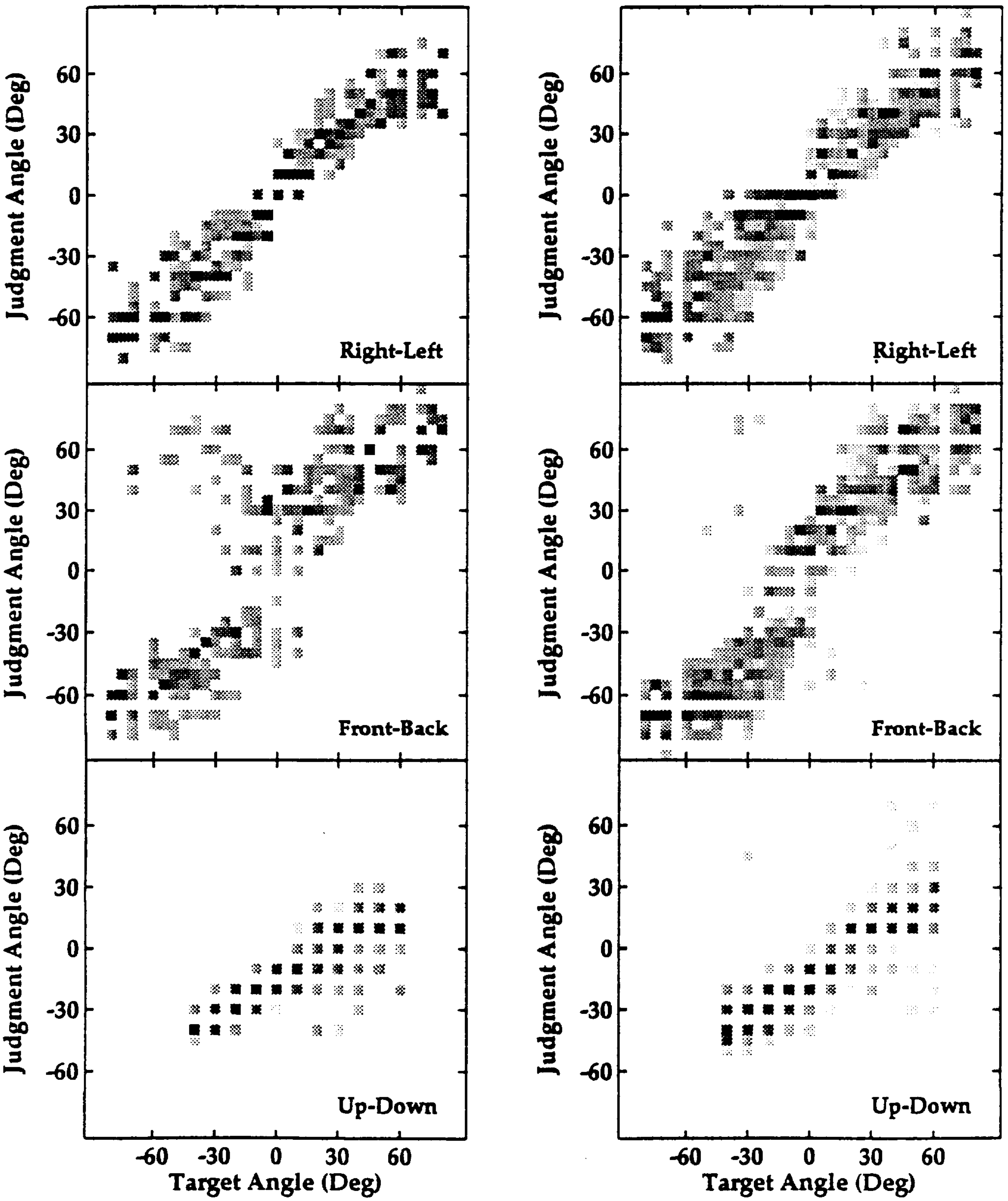

Judgment Angle (Deg)
60
30
0
-30
-60
Right-Left
Front-Back
Up-Down
-60
-30
0
30
60
Target Angle (Deg)

Figure 11