Near-Field Auditory Localization

by

Douglas S. Brungart

B.S.C.E., Wright State University (1993)
S.M., Massachusetts Institute of Technology (1994)

Submitted to the Department of Electrical Engineering
and Computer Science
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

February 1998

© Massachusetts Institute of Technology 1998. All rights reserved.

Author .....
Department of Electrical Engineering and Computer Science
January 9, 1998

Certified by .....
Nathaniel I. Durlach
Senior Research Scientist
Thesis Supervisor

Accepted by .....
Arthur C. Smith
Chairman, Departmental Committee on Graduate Students

MAR 27 1998
ARCHIVES
LIBRARIES

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 1

Near-Field Auditory Localization

by

Douglas S. Brungart

Submitted to the Department of Electrical Engineering and Computer Science
on January 9, 1998, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

Although many researchers have examined auditory localization for relatively distant sound sources, very little is known about the spatial perception of sound sources near the head. A series of measurements and experiments have been performed to analyze auditory localization cues and localization performance in the near-field, defined as the region within 1 m of the head. First, a preliminary experiment compared four response methods that allowed subjects to indicate the perceived azimuth, elevation, and distance of a near-field source. The results indicate that a *direct-location* response, in which the subject simply moves a pointer to the perceived location of the source, is the most accurate response method even when the source is outside the visual field. Next, near-field head related transfer functions (HRTFs) were calculated using a rigid-sphere model of the head and measured using an acoustic manikin and a point source. The resulting HRTFs show that the interaural intensity difference (IID) increases dramatically as a lateral source approaches the head, while the interaural time delay (ITD) is roughly independent of distance. The HRTFs also suggest a possible low-pass spectral distance cue for nearby sources, but failed to show a strong distance-dependence in high-frequency pinna-based elevation cues. Finally, a series of psychoacoustic experiments were performed to evaluate near-field localization performance. An experiment with a broadband source showed that the overall directional error in the near-field (17°) was comparable to previously measured errors in far-field localization. Distance perception accuracy was strongly dependent on azimuth, with the stimulus-response correlation ranging from less than 0.4 in the median plane to 0.85 to the side. Additional experiments showed that distance perception is degraded substantially when listening monaurally or when the stimulus does not contain low-frequency (< 3000 Hz) energy. The results of the psychoacoustic experiments demonstrate that listeners are able to make reasonably accurate judgments about the distance of nearby sources, and suggest that these judgments are largely based on distant-dependent changes in the low-frequency IID.

Thesis Supervisor: Nathaniel I. Durlach
Title: Senior Research Scientist

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 2

Acknowledgments

While it is impossible to list everyone who has contributed to this research, there are a few specific individuals and organizations that have been particularly helpful. Financial support has been provided by the Air Force, both through its Palace Knight civilian training program and through the Air Force Office of Scientific Research. My colleagues at the Air Force Research Laboratory have also been extremely supportive. The Bioacoustics and Biocommunications branch graciously allowed me to borrow some of the equipment used in this research, and its personnel have provided excellent advise (and coffee!) on my many visits back home. Ted Knox, who has served both as my mentor and supervisor under the Palace Knight program, has been instrumental in keeping the administrative gremlins at bay. My subjects, many of whom were more volunteer than paid, deserve my extreme gratitude for their patience and understanding. Likewise, my UROPs, Guy Fisher and Michael DeSenna, were extremely tolerant of a tedious experimental procedure and were absolutely critical to the collection of data in these experiments. Jay Desloge wrote the initial Matlab code for implementing the rigid sphere model of the head, and has made some excellent suggestions about improving the presentation of data in this document. Jeanie Krause has provided support in a variety of ways, including subject, messenger, and experimenter. Bill Rabinowitz has spent many hours discussing my research and editing my manuscripts, and his insights and constructive criticisms have been invaluable. Steve Colburn and Bill Peake have also been exemplary committee members, and their comments have substantially improved the quality of this thesis. Nat Durlach's sage advise has provided just enough guidance to allow me to find my own way without straying too far from the path. I would be remiss if I didn't thank all my colleagues in the sensory communication group who, through passionate discussions and general camaraderie, have made my tenure at MIT enjoyable as well as educational. Finally, I want to thank Lisa, whose persistent inquiries about my return home have inspired the timely completion of this thesis.

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 3

Contents

1 Introduction 9

2 Background 14

2.1 The Nature of Auditory Localization 14

2.2 Directional Localization Cues 14

2.3 Distance Localization Cues 17

2.3.1 Anechoic Environment 17

2.3.2 Reverberant Environment 20

2.4 Perceptual Limits on Localization Performance 21

2.5 Localization vs. Lateralization 22

2.6 Directional Localization Performance 23

2.7 Distance Localization Performance 25

2.8 Dynamic Localization Cues 27

2.9 Virtual Audio Displays 28

3 The Near-Field for Auditory Localization 30

3.1 Enlarged Interaural Intensity Differences 32

3.2 Direction and Distance in Near-Field Localization 32

3.3 The Importance of Head Movements 33

3.4 The Physical Extent of the Source 34

3.5 Available Literature on the Near-Field 34

3.6 Physical Acoustics in the Near-Field 37

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 4

4 Evaluation of Response Methods for Near-Field Auditory Localization Experiments **41**

4.1 Introduction 42

4.2 Background 43

4.3 Experiment 1 46

4.3.1 Method 46

4.3.2 Results 51

4.3.3 Bias Correction 60

4.3.4 Locational Dependence of Response Errors 65

4.3.5 Memory Effects 66

4.4 Experiment 2 68

4.4.1 Method 68

4.4.2 Comparison to Experiment 1 69

4.4.3 Comparison of Profiled and Forward-Facing SH 70

4.5 Experiment 3 71

4.5.1 Method 71

4.5.2 Results and Discussion 72

4.6 Conclusions 74

5 Head-Related Transfer Function Measurements in the Near-Field **76**

5.1 Introduction 77

5.2 Background 77

5.3 The HRTF in the Near-Field 81

5.4 Sphere Model of the Head 84

5.5 Procedures for HRTF Measurements 86

5.5.1 Facilities 86

5.5.2 Measurement Procedure 88

5.5.3 Calibration 89

5.6 Monaural Transfer Functions 90

5.6.1 Evaluation of Monaural Transfer Function with Sphere Model 90

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 5

5.6.2 Measurements of Monaural Transfer Function with KEMAR Manikin . . . 97
5.7 Interaural Intensity Differences . . . 104
5.8 Interaural Time Delays . . . 105
5.9 Elevation Effects on Near-Field HRTFs . . . 109
5.10 Pinnae Effects for Very Near Sources . . . 111
5.11 Comparison of Sphere Model and KEMAR measurements . . . 113
5.12 Perceptual Implications of the Near-Field HRTFs . . . 115
5.13 Conclusions . . . 120

6 Localization of a Broadband Source in the Near-Field 122
6.1 Introduction . . . 122
6.2 Background . . . 123
6.3 Experimental Setup . . . 127
6.3.1 Subjects . . . 127
6.3.2 Apparatus . . . 127
6.3.3 Stimulus . . . 129
6.3.4 Procedure . . . 130
6.4 Directional Localization Results . . . 132
6.4.1 Data Analysis . . . 132
6.4.2 Angular Error . . . 134
6.4.3 Azimuth Error . . . 135
6.4.4 Front-Back Reversals . . . 142
6.4.5 Elevation Error . . . 145
6.4.6 Comparison of Azimuth and Elevation Errors . . . 145
6.4.7 Response Biases . . . 152
6.4.8 Discussion . . . 154
6.5 Distance Localization Results . . . 157
6.6 Comparison of Results to Near-Field HRTF Measurements . . . 168
6.7 Conclusions . . . 172

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 6

7 The Effects of Stimulus Characteristics on Near-Field Localization 174

7.1 Introduction 175

7.2 Background 175

7.3 Experiment 1 179

7.3.1 Method 179

7.3.2 Directional Localization 182

7.3.3 Elevation Performance 189

7.3.4 Response Bias 192

7.4 Discussion 192

7.5 Distance Localization 197

7.6 Experiment 2 208

7.6.1 Experimental Procedure 209

7.6.2 Results 210

7.7 Conclusions 214

8 Conclusions 217

8.1 Summary of Results 217

8.2 Other Possible Issues in Near-Field Localization 222

8.2.1 Psychoacoustic Issues 222

8.2.2 Virtual Display Issues 223

A Coordinate System Definitions 225

B Derivation of the Sphere Model 228

C Acoustic Point Source 231

C.1 Purpose 231

C.2 Background 233

C.3 Description, Manner, and Process of Making and Using Device 237

C.3.1 Description of Device 237

C.3.2 Characteristics of Device 239

C.3.3 Use of Device 241

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 7

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 8

Chapter 1

Introduction

In addition to its vital role in verbal communication, the auditory system plays an important part in our ability to sense and react to remote events. In many ways, the auditory system is an ideal complement to the visual system in providing information about the environment. The visual system provides high-resolution, three-dimensional images of distant objects, but has several limitations. Vision requires a direct line of sight and sufficient ambient light to illuminate the target. Furthermore, our eyes can only view a small fraction of the environment at one time. If we were forced to rely on vision alone, continuous scanning would be necessary to maintain environmental awareness.

Fortunately, our auditory system compensates for the inadequacies of vision. Sound diffuses more readily than light, so we can hear sounds from sources which are hidden behind objects. Hearing does not require illumination, so we can hear in the dark. And, perhaps most importantly, hearing is omnidirectional. We can not only hear sounds originating from any direction, but can also identify the location of each source. In fact, this sound localization ability allows us to direct our gaze toward the source and then use our visual system to evaluate interesting events without constant scanning.

The ability to identify the location of a sound source is called auditory localization. Scientists have studied the process of auditory localization for almost a century. This research has determined that localization is made possible by the acoustic effects of the

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 9

head, torso, and pinnae on the sounds that reach the left and right ears. Many studies have examined the mechanisms that allow human listeners to determine the azimuth and elevation of a sound source and the accuracy of directional auditory localization. Some data are also available on the accuracy of auditory distance judgments, but the mechanisms that allow audio depth perception are not fully understood. Chapter 2 gives an overview of the literature on auditory localization.

Conspicuously absent from the literature are studies examining the auditory localization of nearby sound sources. Virtually all of the available information on auditory localization focuses on sound sources located at distances greater than 1 m from the subject's head. In this region, the head and torso are in the "far-field" of almost all sources, and the sound impinging the head can be approximated by a uniform plane wave. Consequently, free-field directional auditory localization cues in the far-field are essentially independent of source distance[1]. In contrast, in the "near-field", which for sound localization will be defined as the region within 1 m of the head, the interaural intensity difference is expected to be highly dependent on distance. Chapter 3 discusses the differences between the near-field and the far-field in more detail.

The lack of research adequately addressing the issues of near-field localization is surprising because of the importance of the near-field to human listeners. It is the only region where a listener can manually interact with sound sources. Whenever we swat at a buzzing insect, for example, we are using near-field localization. Sounds in the near-field can also be threatening, and may require immediate attention. In addition, most interactions between young children and their mothers occur in the near-field, and are crucially important to human development. Finally, the distance dependence of interaural cues in the near-field may provide a robust binaural distance cue. The availability of binaural distance cues in the near-field would make it the only region where reliable distance judgments can be made without reliance on room reverberation or *a priori* information about the characteristics of the sound source.

[1]The free-field distinction is necessary because in reverberant environments the ratio of direct to reflected energy, which increases as distance decreases, can serve as a distance cue. Loudness may also serve as a distance cue, but only when there is some *a priori* knowledge about the intensity of the source

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 10

A working knowledge of near-field localization ability also has practical implications in the design of virtual audio displays. The ability to assign high priorities to auditory icons by placing them in the near-field or to generate virtual sound sources with robust natural distance cues would be useful in the creation of audio displays with applications in military and commercial aviation, communications, virtual environments and entertainment.

No previous study has systematically examined auditory localization in the near-field, so this research focuses on the basic aspects of auditory localization at distances less than 1 m. Two methodological developments facilitated this research: a novel response method for the near-field, and a compact sound transducer which approximates the behavior of an acoustic point source. Chapter 4 describes a preliminary experiment that evaluated four response methods for azimuth, elevation, and distance in the near field, and indicated that *Direct-Location* was the most appropriate response method for these experiments. Appendix C reports on the performance of a special compact, non-directional sound source.

With this methodology established, the acoustics of near-field cues were examined in depth. Chapter 5 discusses a mathematical analysis of near-field head-related transfer functions, in which the head is approximated by a rigid sphere, and compares the predictions of this model to near-field head-related transfer functions (HRTFs) measured with the Knowles Electronic Manikin for Acoustic Research (KEMAR).

The psychoacoustic localization experiments are described in Chapters 6 and 7. Chapter 6 looks at near-field localization with a broadband source. Enough trials were collected in this condition to allow a detailed analysis of localization accuracy as a function of source location in the near-field. Chapter 7 examines near-field localization under a variety of conditions, including high-pass and low-pass stimuli and monaural listening. Chapter 8 discusses the overall conclusions of the thesis and some possible areas for future work.

Chapters 4-7 were written as stand-alone papers for journal publication, so each of these chapters includes a review of the literature relevant to that particular chapter. The review of the literature in Chapters 2 and 3 provides a more detailed review of

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 11

localization in general, and should be viewed as a general introduction into the field.

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 12

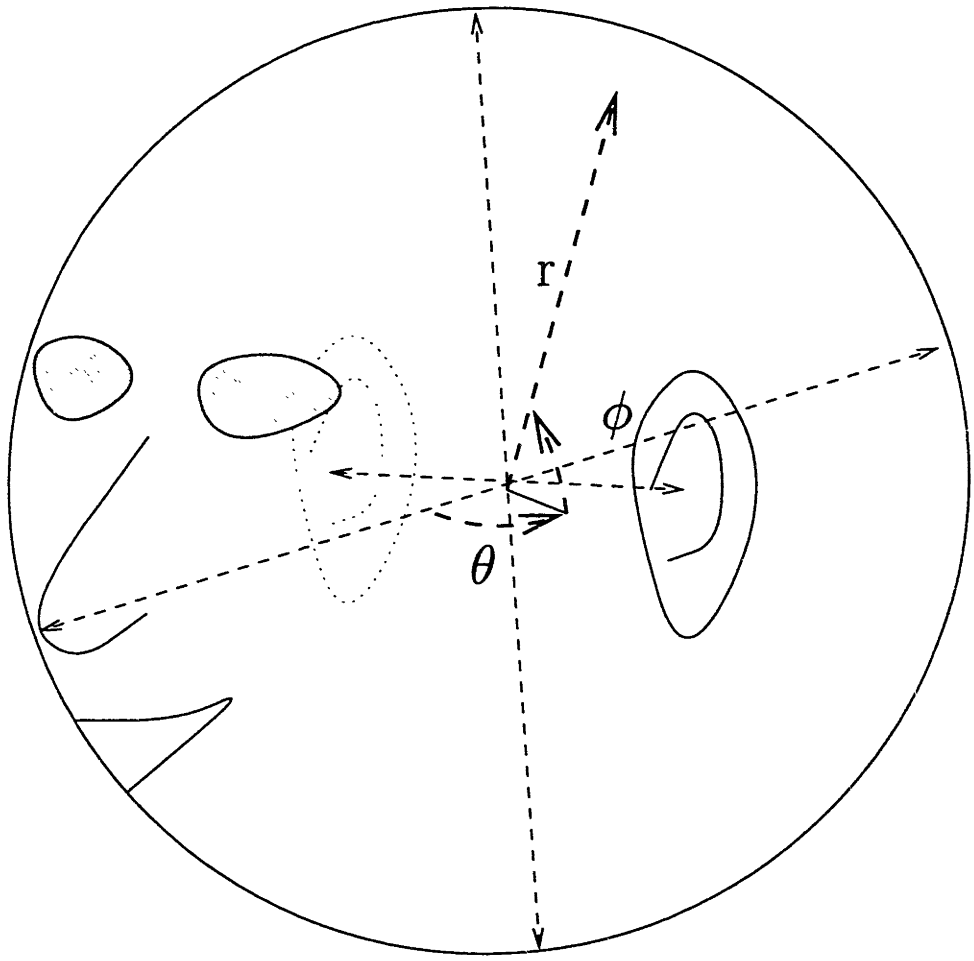


Figure 1-1: Definition of spherical coordinate system. θ is azimuth, ϕ is elevation, and r is distance from the center of the head.

This figure defines the spherical coordinate system used throughout this document. The azimuth is labeled by θ and is equal to 90° for sources to the left of the listener and $-90°$ for sources to the right of the listener. The elevation is labeled ϕ and ranges from 90° for sources above the listener to $-90°$ for sources below the listener. The distance to the source, measured from the center of the head, is r.

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 13

Chapter 2

Background

2.1 The Nature of Auditory Localization

Auditory localization is the process of determining the location of a sound source based solely upon the acoustic signals that reach the ears. Anyone with normal hearing is able to make reasonably accurate guesses about the location of a sound source. Therefore, acoustic cues that indicate the location of the sound source must be available to the auditory system. These cues are a direct result of the form of the human body. The eardrums are located at the ends of the ear canals, and the ear canal openings are separated in space by approximately 18 cm. At the openings are the flanges of the outer-ears, or pinnae. The pinnae are located on the surface of an approximately round head, which is connected to the torso by the neck. Together, all of these physical features form a complex acoustic system that interacts with sound waves as they approach the eardrums, adding the consistent directional information to the sounds reaching the ears that allows auditory localization to occur.

2.2 Directional Localization Cues

The nature of these cues was first explored in earnest by Lord Rayleigh in 1907 (1907). He noted that a sound located off to one side of the head has a direct path to the closer, or ipsilateral, ear, but must go around (or through) the head to reach the

Eight kHz, LLC
Exhibit 2013
IPR2023-01024

farther, or contralateral ear. Thus, the contralateral ear is effectively "shadowed" and the level of the sound at that ear will be lower than at the ipsilateral ear. This causes an interaural intensity difference (IID) between the ears. At low frequencies (in fact when the size of the head is small relative to the wavelength of the sound), the head-shadowing effect produces little attenuation at the contralateral ear and the IID (for distance sources) becomes negligible. Rayleigh noted that at these frequencies (below 1 kHz) the only possible directional cue for a sine wave source was the phase difference caused by the different propagation lengths from the source to the near and far ears. This delay causes an interaural time delay (ITD) between the signals at the left and right ears. At high frequencies, the wavelength of the sound is shorter than the difference in propagation lengths between the left and right ears, and the interaural phase difference becomes ambiguous and ceases to be a useful localization cue. Thus, Rayleigh formed his famous "Duplex Theory" of sound localization, which states that directional information for a sound source is derived from the ITD at low frequencies and the IID at high frequencies.

The duplex theory implies the existence of a transition region between the low-frequencies where ITDs dominate and the high-frequencies where IIDs dominate, and research has shown that localization is poor in this region. Stevens and Newman (1936) examined the accuracy of localization for sinusoidal sources mounted on a boom 4 m away. They found that the magnitude of errors (difference between reported angle and actual angle) is greatest at 3 kHz, and decreased at higher and lower frequencies. These results indicate that the transition frequency between ITDs and IIDs is somewhere in the region of 3 kHz, and that in the transition band neither ITDs nor IIDs are particularly effective as localization cues. Mills (1958) measured the ability to discriminate changes in the azimuth of a sinusoidal sound source and also found that resolution in azimuth perception is worst for frequencies around 3 kHz.

There is evidence that listeners can detect interaural delays in the envelopes of high frequency sounds. Yost and colleagues (1971) found that listeners could detect ITDs in high-pass filtered click trains, but that the minimum detectable change in

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 15

ITD was much greater for a pulse train high-pass filtered above 1.5 kHz ($162\mu s$) than for signals low-pass filtered below 500 Hz ($25\mu s$). They also found that the addition of low-frequency noise severely impaired the detection of interaural delays in the high-pass filtered click train, but that high-frequency noise had little effect on the detection of ITDs in the low-pass filtered click train. A later paper (Yost, 1976) examined the detection of ITDs in high-frequency band-pass filtered pulse trains. Yost reported that sensitivity to ITDs at high frequencies increased as the number of transient repetitions increased and as the bandwidth of the filter increased, but was roughly independent of the repetition rate of the pulse train. Similar results have been reported for band-pass filtered noise (Trahoitis & Bernstein, 1986) and sinusoidally amplitude-modulated tones (Henning, 1968, 1974). These results indicate that interaural delay information may be available at high frequencies.

Listeners are, however, less sensitive to ITD at high frequencies than at low frequencies, and when there are discrepancies between the high- and low-frequency ITD the low-frequency value dominates perception. Evidence of the dominance of ITDs at low frequencies is provided by Wightman and Kistler (1992), who found that that listeners faced with conflicting IID and ITD information base their localization judgments primarily on the ITDs as long as low frequency energy (below 2.5 kHz) is available.

The duplex theory is not sufficient to explain auditory localization completely. The IID and ITD are based on the location of the sound source relative to the ipsilateral and contralateral ears. Because of the approximate bilateral symmetry of the head, a particular ITD and IID pair does not correspond to a unique source location. In fact, if the head is approximated by a sphere and the source is assumed to be relatively distant, all of the points on the surface of a cone centered on the interaural axis with its apex at the center of the head will have the same interaural time delays and interaural intensity differences. These regions are commonly known as "cones of confusion" (Woodworth, 1938). In the horizontal plane, these cones of confusion cause "front-back" confusions, in which a listener perceives a sound source at the mirror image of its location across the frontal plane. The entire median plane is

Eight khz, LLC
Exhibit 2013
IPR2023-01024

in effect a cone of confusion, as interaural differences do not occur for sources in the median plane. Yet, despite the lack of interaural cues, listeners are generally able to resolve front-back confusions and perceive elevation in the median plane. There must be some acoustic information other than interaural differences that allows a listener to localize sound. In fact, the outer ear and torso provide these cues by spectrally shaping the sound that enters the ear. The outer ear, or pinna, in particular acts as a filter which has a transfer function that varies as a function of source direction. The characteristics of this transfer function include relatively sharp peaks or notches at high-frequencies that shift systematically as a function of source azimuth and elevation (Butler, 1987). When the listener has *a priori* information about the spectrum of the source, these spectral cues allow reasonable judgments about source location. Localization judgments for unfamiliar sources in the median plane are strongly influenced by the spectral content of the source (Blauert, 1983). Spectral cues also allow auditory localization to take place (with slightly reduced accuracy in elevation and greatly reduced accuracy in azimuth) with one ear blocked (Oldfield & Parker, 1986).

2.3 Distance Localization Cues

2.3.1 Anechoic Environment

In comparison to the cues that allow directional auditory perception, little is known about the acoustic cues that provide information about the distance of a sound source. Mershon and Bowers (1979) divided auditory distance cues into two major categories. *Relative* distance cues allow the listener to identify changes in the distance of a sound source, i.e., whether the sound source has moved closer or further away. *Absolute* distance cues allow the listener to estimate the distance to the sound source. Coleman (1963) has provided a comprehensive list of possible distance cues under free-field (anechoic) conditions. Coleman outlines four primary distance cues: amplitude, frequency spectrum in the near-field, frequency spectrum in the far-field, and binaural

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 17

cues.

Amplitude cues are the primary distance cue for determining relative changes in distance. In a free-field, objects propagate sounds spherically and the pressure of the spherical sound wave is inversely proportional to the distance from the source. Thus a sound increases in intensity as the source approaches a listener. The amplitude cue is very powerful as a relative distance cue, and in non-reverberant environments relative distance discrimination (for distant sources) can be adequately explained by loudness cues (Strybel & Perrott, 1984; Cochran, Throop, & Simpson, 1968).

Coleman also suggests that, in the near-field, the frequency spectrum may provide a useful distance cue. This assertion is based on the fact that the particle velocity of a sound wave propagating spherically is the sum of two components, one inversely proportional to distance (and directly proportional to pressure) and one inversely proportional to the square of the distance. In the far-field, the first component dominates, while in the very near-field, the second component dominates. The magnitude of the second component is also proportional to the wavelength of the sound, so at a short distance from the source the velocity will be relatively greater for the low frequency components of the signal than for the high frequency components of the signal. Von Bekesy (1960) proposed that listeners can perceive this relative increase in low frequencies in the near-field and use it to determine distance. In fact, he found that integrating a sound signal produced closer auditory judgments, and differentiating it caused more distant auditory judgments. Bekesy's assumption, however, is based on the ability of human listeners to directly perceive the particle velocity of a sound source, while most researchers model the ear as a pressure detector. While Bekesy's explanation is suspect, his conclusion that low-pass filtered sounds are perceived closer has some merit. Diffraction causes a relative increase in the low-frequency energy reaching the ears as a sound source approaches the head, as discussed in Chapter 5. Another possible explanation for the correlation between nearness and low-frequency content can be found in the familiar Fletcher-Munson equal-loudness curves (Begault, 1987). As the source approaches the head it becomes much more intense at all frequencies, but the perceived increase in loudness is greater

Eighth Wall, LLC
Exhibit 2013
IPR2023-01024
Page 18

at low frequencies.

The frequency spectrum may also serve as a distance cue in the far-field. Atmospheric absorption of sound is greater at high frequencies than at low frequencies. Therefore a sound 30 m away will be attenuated only 0.2 dB at 1 kHz by atmospheric absorption but, the attenuation at 10 kHz is about 4 dB. This absorption is certainly a salient distance cue for very loud, very distant sources (as demonstrated by the ominous distant rumble of thunder), but seems to be of little practical use for sources closer than 10 m. Another important spectral distance cue can result from the interaction of sounds with objects in the listening environment. Low frequency sounds diffuse around obstacles better than high-frequency sounds, so sounds propagating in a complex environment will be effectively low-pass filtered if no direct path exists between the source and the listener. Psychophysical data have shown that frequency spectrum strongly influences distance perception. Coleman (1968) and Mershon et al. (1989) found that increasing the low-pass cutoff frequency of a signal decreased the perceived distance. Lounsbury (1979) and Butler (1980) found that binaural recordings of sound sources were judged to be more distant when the sounds were low pass filtered than when they were high pass filtered. It is curious that the enhancement of low frequencies has been reported to indicate both a closer source in the near-field and a more distant source in the far-field.

The final distance cues proposed by Coleman are binaural distance cues. These cues result from the interactions between the head and the sound waves propagating from the source. Hartley and Fry (1921) and Stewart (1911a) examined these cues by modeling the head as a rigid sphere with ears at diametrically opposed points on the surface. They found that the interaural intensity difference increases dramatically as a nearby source approaches the head. Binaural cues are believed to be dominant for an unfamiliar sound source in the near-field. An updated version of the sphere model and its implications on distance perception is given in Chapter 5.

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 19

2.3.2 Reverberant Environment

When the listening environment is not anechoic, reverberation becomes an important audio distance cue. In a room, the direct sound path between the source and the listener's ears is augmented by a large number of reflections off the floor, ceiling, and walls. When the sound source is very close to the listener, the direct path is much shorter than the reflected paths and thus the direct sound will be much more intense than the reflected sounds. As the distance increases, however, the reflected sounds become increasingly important and when combined may be louder than the direct sound. A number of researchers have proposed that the ratio of direct sound intensity to reflected sound intensity is a salient distance cue, and that distance judgments increase as this ratio decreases. The importance of this ratio was shown by Mershon and King (1975) for live listeners in anechoic and reverberant environments. Mershon et al. (1989) later found that initial distance judgments increased with the reverberation time of the room. Lounsbury and Butler (1979) and Butler, Levy and Neff (1980) made binaural recordings of sound sources in anechoic and reverberant environments and asked subjects to listen to the recordings and estimate the distances of the sources. It was found that subjects consistently made greater distance judgments when the recordings were made in reverberant environments than when they were made in anechoic environments. Both studies found that distance accuracy was equivalent when the subjects listened binaurally and when they listened monaurally (with the contralateral ear disconnected), implying that binaural cues were not a factor. Clearly reverberation can strongly influence auditory distance perception.

In some cases, the listener's preconceptions about the intensity and other characteristics of the source greatly influence their perception of distance. This effect is most pronounced for speech. Gardner (1969) used live talkers for stimuli and found the subjects could accurately judge distances for conversational level speech, but systematically overestimated distances for shouted speech and underestimated distances for whispered speech. Mershon and Philbeck (1991) found similar effects for recorded speech played through a loudspeaker. McGregor et al. (1985) found that subjects

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 20

could more accurately judge the relative distances of two loudspeakers with a familiar speech stimulus than with the same stimulus played backwards. These studies indicate that subjects are able to use their experience with a certain type of sound stimulus, such as shouted, whispered, or conversational speech, to estimate the expected loudness of the signal and can then used this expected loudness to judge the distance of the source.

2.4 Perceptual Limits on Localization Performance

Many researchers have performed psychoacoustic experiments to measure the perception of individual localization cues. Perhaps the most studied directional cues are interaural differences. Hershkowitz and Durlach (1969), for example, measured the just noticeable differences (JNDs) in interaural time delay and interaural intensity differences for a 500 Hz tone under a variety of conditions. One major variable that affects these thresholds is the overall sound level of the stimulus. As the stimulus approaches the threshold of hearing, it becomes more difficult to derive interaural information from the signal. Hershkowitz and Durlach found that the JNDs for both time and intensity increased as the overall sound level decreased below 20 dB SL. It is also reasonable to expect that the ability to process binaural information degrades as interaural intensity differences become large. Hershkowitz and Durlach found that the ability to detect small changes in the ITD degraded significantly as the IID increased above 20 dB. ITD discrimination for a sound at 50 dB SL in the left ear and 20 dB SL in the right ear was significantly worse than ITD discrimination for sounds at 20 dB SL in both ears, so this effect was not purely a result of an inaudible signal in the shadowed ear. The thresholds for IID also decreased slightly as the baseline interaural intensity difference increased. The JND for the interaural time difference was approximately $15\mu s$ for ITDs up to $400\mu s$, but increased dramatically for ITDs above $400\mu s$. The JND for interaural intensity difference was essentially independent of the IID and the ITD and was approximately 0.8 dB over a broad range of conditions.

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 21

The JND for changes in the amplitude of broadband noise has also been extensively studied. Miller (1947) found the JND in intensity to be approximately 1 dB. The ability to determine changes in spectral shape, such as the peaks and notches in the audio spectrum caused by the pinnae, has been referred to as profile analysis in the literature (Green, 1988). In a typical profile analysis study, Malme (1959) measured the ability of listeners to detect notches and peaks in a broadband noise spectrum, and found that subjects could detect peaks as small as 5.2 dB and notches as small as 8.8 dB at 3 kHz, and that peak and notch detection was slightly better at 1 kHz, and slightly worse at 300 Hz. A similar experiment by Moore et al. (1989) found that subjects could detect 6 dB peaks and 10 dB notches at 8 kHz with a bandwidth of 2 kHz, and that randomizing the background level slightly inhibited peak detection but substantially degraded notch detection.

These perceptual limits are useful because they place an upper bound on the accuracy of auditory localization, but it is not clear that human localization performance will approach this upper bound even under ideal listening conditions.

2.5 Localization vs. Lateralization

When subjects listen to artificially created binaural signals through headphones, the location of the perceived auditory source is usually inside the head. Moreover, the auditory event can be projected to some point on the line connecting the two ears. This phenomenon is known as lateralization, rather than localization. The perceived location on the interaural axis can be manipulated by changing the interaural time delay or interaural intensity difference. Toole and Sayers (1965) showed that lateralization saturates (the sound is located at the left or right ear and cannot be lateralized further) for ITDs above 700 μs. Yost (1981) showed that lateralization caused by the IID saturates for interaural differences greater than 20 dB. Many researchers have shown that there is a trade-off between interaural time and intensity differences in the perceived position of a lateralized source.

When the binaural signals generated by the headphones are sufficiently similar to

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 22

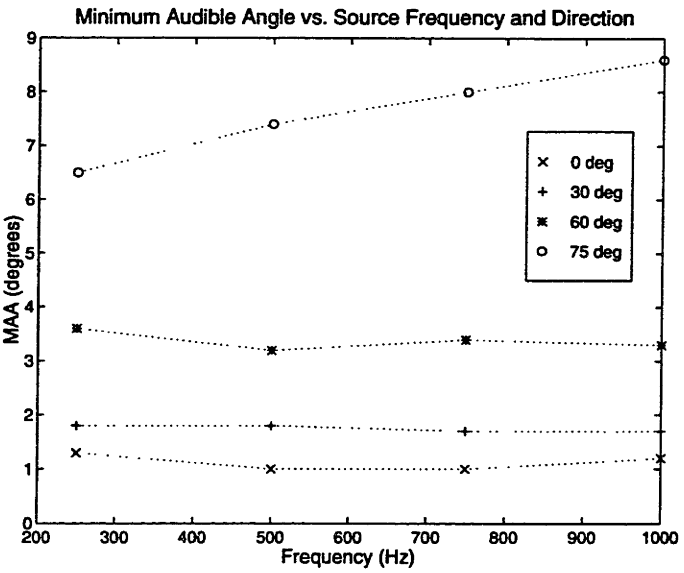


Figure 2-1: The Minimum Audible Angle (MAA) in degrees as a function of the source frequency (shown on the ordinate scale) and source position θ shown in the legend. At $\theta = 90°$ the MAA was too large to measure. The source is in the horizontal plane and the MAA values are for azimuth. Adapted from Mills (1958).

signals that might reach the ears from a free-field source, the acoustic event is often heard outside the head, or "externalized". A detailed discussion of the requirements for externalization to occur is given by Durlach et al. (1992). Important requirements for externalization seem to include the spectral shaping of the pinnae, the presence of reverberation, and the availability of cues from head motion. In fact, it is possible to achieve the "internalization" of a real sound source in anechoic chamber if the source follows the motions of the listener's head (Wallach, 1940). In general, relatively little is known about the phenomena of externalization and internalization.

2.6 Directional Localization Performance

There are two basic methods for measuring auditory localization performance. The first method measures the ability of the subject to discriminate changes in the direction of the sound source. The subject listens to a sound source twice and must determine if the sound source has moved left or right. The Minimum Audible Angle

· Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 23

(MAA) is the smallest change in angle that allows the subject to correctly judge the movement of the source. The MAA is highly dependent on the location of the source. According to the data of Mills (1958) (see Figure 2-1), the MAA in azimuth for a source in the horizontal plane is approximately 1° directly in front of the listener. It increases as the source moves lateral to the head, and at 75° the MAA is 7.5° at 500 Hz. When the source is at 90°, the MAA task becomes a front/back discrimination task, and the MAA was larger than Mills was able to measure (> 14°).

The second method of measuring localization is a location estimation paradigm. Subjects listen to a stimulus, and are asked to estimate the location of the sound source, either through a verbal response or a pointing procedure. This method evaluates the ability to judge the direction of an isolated sound, rather than the ability to detect changes in the direction of a sound. Two recent studies have examined localization performance under these conditions. Makous and Middlebrooks (1990) mounted 36 loudspeakers every 10° on a ring 1.2 m in radius, and were able to rotate the ring around an axis coincident with the interaural axis of the subject to change the elevation of the sources. An electromagnetic sensor mounted on the head allowed the subjects to respond to a stimulus location by head pointing. For broadband (1.8-16 kHz) signals 150 ms in duration, Makous and Middlebrooks found mean unsigned errors for sounds directly in front and 5° below the listener of 1.5° in azimuth and 3.5° in elevation. Performance decreased as the azimuth and elevation of the source increased, and mean unsigned errors of up to 15° were found for sources behind the head. A notable feature of the data is that horizontal errors were smaller than vertical errors for sources at azimuths less than 30°, but for more peripheral sources the vertical errors were smaller than the horizontal errors.

Wightman and Kistler (1989b) performed a similar experiment using a vertical arc of 6 loudspeakers spaced 18° apart from −36° to 54° in elevation. This arc was rotated around the subject in an anechoic chamber, and the blindfolded subject was asked to verbally identify the location of the sound source. A string of five 350 ms Gaussian noise bursts spaced 300 ms apart was used as the source stimulus. The resulting total errors (great arc) ranged from 16.1° for middle elevations to the side

Eight kHz, HC
Exhibit 2013
IPR2023-01024
Page 24

of the listener to 29.8° for sources above and behind the listener.

The data from both of these experiments have excluded front-back confusions, where the subject perceived a sound at a location reflected across the frontal plane (a sound source at 45 degrees is heard at 135 degrees, for example.) These errors account for 6% of the total number of trials in the the Makous and Middlebrooks study, and 5% - 20% of the total number of trials in the Wightman and Kistler study.

The general trend in directional localization is that azimuth judgments are most accurate for objects in the horizontal plane directly in front of the listener, and elevation judgments are less accurate than azimuth judgments around $\theta \approx 0°$ and more accurate than azimuth judgments when $|\theta| \approx 90°$.

2.7 Distance Localization Performance

Auditory depth perception has been more difficult to measure than directional localization, in part because it depends on a large number of environmental factors. In distance localization experiments, the listener is usually interested in how well a listener can identify the location of a source with minimum *a priori* information about the source characteristics. Thus most distance localization studies do not provide feedback to the listener. Unless otherwise noted, all of the experiments described here were performed without feedback. Note that the majority of these studies examined distance perception only in the far-field. The handful of studies which have examined near-field distance perception (directly or indirectly) are described fully in the next chapter.

Visual cues can have a very compelling effect on audio depth perception, particularly in anechoic environments. Gardner (1968) reported that subjects who were able to see five loudspeakers in an anechoic room virtually always initially reported that the signal was coming from the closest speaker regardless of the actual origin of the sound, and attributed lower intensities to a lower level of output rather than a more distant speaker. He called this tendency to choose the closest rational location as the origin of the source the *proximity-image effect.*

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 25

When visual cues are removed, subjects often use their past experiences with a type of sound to interpret the available physical distance cues. Gardner (1969), for example, showed that listeners tended to underestimate the distance of a person whispering in an anechoic chamber, and overestimate the distance of a person shouting, indicating that their distance judgments are based in part on the type of stimulus. When speech was recorded and played back through a loudspeaker at a random level, judgments depended almost exclusively on the level of the source presentation and not the actual distance of the speaker. The subjects were apparently using the type of speech to make an estimate of the intensity of the sound, and using this estimate to judge the distance of the source. Similar results were later shown by Mershon and Philbeck (1991).

When the intensity of the source is not randomized and the source is unfamiliar, loudness dominates as a distance cue. This was shown by Strybel and Perrott (1984), who found that a reduction in distance of 3% to 6% was the minimum detectable change in distance when the source started more than 3 m away. This roughly corresponds to a just noticeable increase in the intensity of the sound at the listener's ears. At shorter distances the distance thresholds were much larger than the intensity model would have predicted, but later Ashmead et al. (1990) proposed that Strybel and Perrott's methodology was inappropriate for close distances and found that subjects could detect 5% changes in distance for sources as close as 1 m. Ashmead and colleagues also found the JND for distance for a source directly in front of the listener when amplitude was corrected for distance (held constant at head location) was approximately 16%, but this underestimates the true threshold because of a ceiling effect in their adaptive measurement procedure. Note that the Ashmead study, which used an adaptive two-alternative forced choice procedure, was the only study described here which provided feedback to the listener.

Subjects are not able to make accurate absolute judgments about source distance in the median plane when the amplitude is unknown and the environment in anechoic. Coleman (1968) and Mershon and King (1975) found that the distance judgments about the initial presentation of an unfamiliar sound source directly in front of the

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 26

subject were uncorrelated with actual source distance. Holt and Thurlow (1969) and Gardner (1969) found that subjects could not make accurate distance judgments for sources in front of a listener when the overall amplitude was corrected to eliminate loudness effects. However, Mershon and King found a mild correlation between source distance and response distance on the initial presentation of a sound in a reverberant environment. Apparently, reverberation is an important distance cue for distant sources in the median plane.

A final note should be made about the interaction of source direction and distance perception. Holt and Thurlow (1969) found that subjects could not accurately rank the distances of sound sources directly in front of the listener when amplitude was corrected for distance (a result consistent with other studies), but that they could do so when the sound source was directly to the right of the listener (rank order correlation of 0.93). Gardner (1969) informally reported a similar phenomenon. In contrast, there are a number of studies that report that source direction is not a significant factor when loudness is not corrected for distance effects. Cochran et al. (1968) report this for live human talkers, and Simpson and Stanton (1973) report this for pulse trains. These results imply that binaural difference cues, which occur primarily for sources outside the median plane, are important to distance perception only when loudness cues are not available, and that loudness cues dominate binaural cues when they are available.

2.8 Dynamic Localization Cues

If a sound's duration is sufficiently long, listeners have the opportunity to make exploratory head motions to improve their localization ability. These head motions can vastly improve localization performance. If the source is on continuously, the listener can turn to face the source and bring the sound into the region where localization accuracy is greatest. Therefore we would expect localization with unrestricted head motion to be at least as good as the best localization possible without head motion. Makous and Middlebrooks (1990) compared open-loop (no head motions) and closed-

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 27

loop (unlimited head motions) localization. They found, predictably, that open-loop performance was strongly dependent on source location and closed-loop performance was roughly independent of source location. The standard deviation error in azimuth was $1° - 2°$ in the closed-loop case, and ranged from 2° - 10° in the open-loop case. Head motion also improved localization in elevation, but less dramatically than in azimuth. Head motions clearly improve directional localization.

Head motions are also the most robust way to eliminate front-back confusions. When the head is turned, the magnitude of the interaural time and intensity differences will either increase or decrease, depending on the location of the source and the direction of head rotation. Thus, head rotations allow discrimination between sources in the front and rear hemispheres.

The effects of listener motion on audio depth perception are more controversial. Simpson and Stanton (1973) found that head motion did not significantly influence audio depth perception, but their experiment did not randomize intensity. It it possible that any contributions due to head motion were dominated by the available intensity cues. Holt and Thurlow (1969) eliminated intensity cues and found that distance judgments were significantly correlated with source distances when the sources were located at 90°, but not when the sources were directly in front of the listener (0°). This result showed that listener orientation has some effect on depth perception. More recently, Spiegel and Loomis (1993) and Ashmead et al. (1995) have found that walking towards a sound source can enhance the accuracy of auditory distance judgments. This increase in accuracy occurs because the derivative of loudness with respect to the position of the listener is greater when the source is nearby than when the source is distant.

2.9 Virtual Audio Displays

All auditory localization cues are based on the transfer functions from the location of the sound source to the eardrums of the listener. If these transfer functions are accurately measured for a source at a specific location relative to the listener, they

Eight Khz, LLC
Exhibit 2013
IPR2023-01024
Page 28

can be used to electronically synthesize signals that, when presented through stereo headphones, appear to originate from the same location relative to the listener. The headphone signal which generates a spatially positioned auditory event is known as a virtual audio source, and the system that generates these signals as a virtual audio display. Much recent research has focused on the measurement of these transfer functions, called head-related transfer functions or HRTFs, and the implementation of virtual audio displays.

Virtual audio displays are usually coupled to a head-tracking device which updates the relative position of the virtual source to maintain a fixed position in space. One such display, created by the Air Force's Armstrong Laboratory, allows subjects to localize virtual sounds in the horizontal plane with mean magnitude errors of 5° (McKinley, Ericson, & D'Angelo, 1994). In general, these systems do a reasonably good job of simulating sources in the horizontal plane. They are relatively poor at synthesizing elevation cues, and they often create sounds that are internalized rather than externalized. The origins of these problems are not fully understood, but probably result, in part, from the use of generalized HRTFs rather than transfer functions specifically measured for each individual subject. A detailed review of virtual audio displays is provided by Wenzel (1991).

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 29

Chapter 3

The Near-Field for Auditory Localization

The term "near-field" is often used in acoustics, but the exact definition of the "near-field" varies according to the specific application. The most widely accepted use is in physical acoustics, where the near-field is defined as the region of space within a fraction of a wavelength from a sound source. This definition is used primarily because the ratio of the pressure of a sound wave to its velocity changes significantly with distance in this region, while at greater distances (the "far-field") this ratio is constant. In recording studios, speakers located close to the listener are often referred to as "near-field" monitors. In this context, the near-field designation is used to indicate that the speakers are sufficiently close to the listener that the direct sound from the speaker dominates any reflected sound energy. The common feature of all definitions is that some important aspect of the sound field is significantly different when the source is nearby than when it is more distant, and thus it is useful to differentiate between the "near-field" and "far-field".

At the risk of further confusing the definition of the "near-field", we have defined the term in the context of human auditory localization as the region within 1 m of the center of the head. This definition follows from the distance dependent variation in the head-related transfer function in the free-field. At distances greater than 1 m (the "far-field") the HRTF is roughly independent of distance, while at distances

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 30

less than 1 m the HRTF changes significantly with distance. Since auditory localization is based on the HRTF, it follows that human localization performance should be substantially different in the near- and far-fields. In the far-field, only directional information is provided by the HRTF, and the distance of the source is largely irrelevant to localization performance. In contrast, the near-field HRTFs provide information about the both the direction and distance of the source, and the directional cues interact significantly with distance. Since auditory localization cues at distances less than 1 m are substantially different from those at distances greater than 1 m, it is appropriate to differentiate the two regions into the near- and far-fields.

The adaptation of the over used word "near-field" to this new application will undoubtedly receive some criticism, so it is important to clearly justify its use. Ideally, it would be more appropriate to use more descriptive definitions of the two regions. For example, the phrase "localization of a nearby source" or "localization at close range" would be strictly more accurate than "near-field localization". However, these uses do not facilitate a concise reference to localization at greater distances ("localization of a relatively distant source" is certainly awkward) and they do not provide labels for referring to the actual regions of space defined by the near-field and the far-field. While initially there may be some confusion about the exact meaning of "near-field", once the term is defined in the appropriate context its meaning is obvious and its usage is more concise than the alternatives. Thus we feel the advantages of the term "near-field" outweigh its drawbacks.

This chapter examines some important aspects of near-field localization and how they differ from the localization of more distant sources. The first four sections describe the properties that are unique to near-field localization, including enlarged interaural intensity differences, interactions between distance and directional cues, substantial changes in source position with small translational head-movements, and the requirement for a compact, non-directional source. The next section summarizes the results of the few studies which have indirectly examined near-field localization. Finally, the last section gives a brief introduction to near-field physical acoustics.

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 31

3.1 Enlarged Interaural Intensity Differences

The interaural intensity difference occurs because the sound reaching the closer, or ipsilateral, ear is generally greater in amplitude than the sound reaching the more distant, or contralateral ear. In the far-field, this difference occurs because the head and torso scatter the sound wave reaching the contralateral ear. This effect, known as head-shadowing, occurs primarily at high frequencies where the wavelength of sound is small relative to the dimensions of the head. In the near-field, the IID increases as distance decreases for two reasons. First, the head-shadowing effect is stronger for nearby sources. Second, the source is relatively closer to the ipsilateral ear. The second effect causes an increase in the IID because the pressure of a sound wave is generally inversely proportional to its distance from a source. If the head were removed, the IID would effectively be equivalent to the ratio of the distance from the source to the contralateral ear to the distance from the source to the ipsilateral ear. As source distance decreases, this ratio increases dramatically. Unlike head-shadowing, the proximity effect increases the IID at all frequencies. The combination of head-shadowing and source proximity causes the IID to increase dramatically as the source approaches the head, as will be shown in Chapter 5. The increase in IIDs with decreasing source distance should have important perceptual implications in near-field localization.

3.2 Direction and Distance in Near-Field Localization

One interesting aspect of near-field localization is the interaction between the cues providing information about the direction and distance of a source. In the far-field, the HRTF is roughly independent of distance and source distance does not effect judgments about the direction of a source. In the near-field, however, the interaural intensity difference, which is known to serve as an important directional cue at high frequencies, increases as distance decreases. Thus, in situations where a listener is

Eighth Wall, LLC
Exhibit 2013
IPR2023-01024

forced to rely on IID cues to determine the direction of a source, it is possible that the listener may confuse the distance and direction of the source. A particular IID could be the result of a relatively distant source near the interaural axis, a relatively close source near the median plane, or somewhere in between. Such confusions may, however, be resolved either using monaural localization cues or high-frequency interaural delay cues. No previous studies have examined this type of interaction between directional localization and distance localization. In fact, almost no studies have required subjects to simultaneously report both the distance and direction of a source.

3.3 The Importance of Head Movements

Most discussions of head motion in auditory localization focus exclusively on rotational head motion. Under ordinary circumstances, listeners do not translate their heads more than a few centimeters, and when a sound source is more than a meter away such movements cause only a negligible change in the position of the sound source relative to the head. The situation changes, however, when the sound source is close to the head. Consider a sound source 25 cm in front of the center of the head. If the listener moves his head 5 cm to the left, the angle of the source relative to the head moves 11° to the right. If the listener moves 5 cm forward, the overall amplitude of the signal reaching the ears increases by more than 1 dB. A sound source 1 m away would only shift 2° from a similar sideways head motion, and 0.4 dB from a similar forward head motion. This parallax effect certainly could be a viable distance cue if translational head motions were allowed while a near-field source is localized.

Rotational head motions may also be important in the near-field. A possible cue for distance perception in the near-field is the increase in loudness at the ipsilateral ear when the ear is rotated toward from the source. When the source is close, this increase in intensity is much larger than when the source is far away. The derivative of the loudness at the ipsilateral ear with respect to the rotation of the head may be a powerful monaural distance cue in the near-field.

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 33

3.4 The Physical Extent of the Source

Special care is required for the creation of near-field auditory stimuli. An ordinary loudspeaker is typically 10 cm or more in diameter. A source this size is fine for sources more than a meter away, but such a large speaker is not acceptable in the near-field. When the loudspeaker is close to the head, a speaker with a diameter larger than a few centimeters will cover an arc 10° or more in width from the center of the subject's head. The sound signal reaching the subject's ears will essentially be the average of the signal from an array of point sources covering the cone of the speaker. Since localization resolution is known to be as good as 1° directly in front of the listener, this averaging makes it difficult to study localization ability. Furthermore, some complicated near-field effects occur along the axis of a loudspeaker at very close distances, and the directionality of the speaker may affect the head-related transfer function when the source is close to the head. These complications can be avoided through the use of a point source that is non-directional and at most a few centimeters in diameter. A source of this type was constructed specifically for these experiments, and is discussed Appendix C.

3.5 Available Literature on the Near-Field

Three papers address auditory localization for source distances less than 1 m, and none is particularly useful. The first is a 1973 paper by Simpson and Stanton (1973) which examines the usefulness of head-motion as a distance cue for sources from 30 cm to 266 cm away from the subject's head. The sound sources were suspended from the ceiling at eye level, and a 5 Hz train of pulses from an 800 Hz sine wave generator was used as the stimulus. Subjects were in the corner of a sound treated room, 25 cm from the two walls, and the sound source was moved along the diagonal of the room. The amplitude of the source was fixed across all trials. Subjects were asked to identify the distance of the sound source using a magnitude scaling paradigm. The results showed that subjects could consistently rank order the source distances correctly, but

Eight kHz, LLC
Exhibit 2013
IPR2023-01024

that the use of head-motions did not significantly improve distance perception. These results are not consistent with the theory that binaural cues are useful for distance perception in the near-field. There are three reasons to believe that these results may not rule out this possibility, however.

1. Only 4 of the 10 distances used in the experiment were less than 1 m. Since no significant advantages are expected for head movements outside of 1 m, the fact that 60% of the trials were in the "far-field" may have diluted the effects of head motion to statistical insignificance.

2. The overall amplitude of the source was held constant across all trials. Consequently, intensity should have been the dominant distance cue in this experiment. Any information gained through binaural distance cues may have been overwhelmed by the dominance of the intensity cue.

3. The subject's head was in the corner of the room and only 25 cm from the two walls. Since the head was so close to the walls, it is likely that reflections from the walls were influencing the subject's perception in this experiment. The precise consequences of this influence are difficult to determine.

The second study describing near-field auditory localization is also a distance experiment, conducted by Ashmead et al. (1990). In this experiment, a subject was seated in an anechoic chamber and was asked to listen to two sounds and determine which source was closer. A 500 ms broadband (0.1-8 kHz) noise signal was used as the stimulus. The minimum change in distance (percentage decrease) necessary to accurately distinguish between the closer and further sources (in a 2 up, 1 down adaptive paradigm) was measured at reference distances of 1 m and 2 m. The measured thresholds were 5.73% at 1 m and 5.91% at 2 m when the amplitude of the source was fixed. When the amplitudes were adjusted to eliminate intensity cues, the thresholds jumped to 16%, and this underestimates the actual threshold because of ceiling effects in the adaptive procedure employed in the experiment. The sound source was directly in front of the listener, so the ability to determine changes in dis-

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 35

tance was not dependent on interaural differences. The mechanisms allowing distance discrimination without the intensity cue are not known.

The third and final study examining near-field localization is by Litovsky and Clifton (1992). This study was designed to compare auditory distance perception in infants and adults. We will only consider the results from the adult subjects here. A recording of shaken "jingle bells" was used as the stimulus, and the stimuli were presented from loudspeakers 45° to the left and right of the subjects. The loudspeakers were at mid-torso level on the adults. The sounds were presented at distances of 0.15 m and 1 m from the mid-torso of the subject, and at sound pressure levels of 67 dB SPL and 74 dB SPL at the mid-torso. In the control group, for which the sounds were always at 74 dB SPL (at the mid-torso) when they were at 0.15 m and at 67 dB when they were at 1 m, the adults were correct about the distance more than 80% of the time. In the experimental condition, when the intensity of each presentation was randomly selected, the adults were correct only 63% of the time. This large error was attributed to a strong bias which caused the adults to label 85% of the louder (74 dB) sound presentations as close, and 70% of the softer (67 dB) sound presentations as far. These results seem to indicate poor localization ability in the near-field, since the listeners could only tell the difference between a source 0.15 m away and 1 m away in 63% of the trials, while they should get 50% by guessing randomly. Reverberation in the room is likely the cause of these poor results. Litovsky and Clifton indicate that the sound level in the room dropped only 7 dB as the source moved from 0.15 m to 1 m away from infant's head. This indicates that the critical distance in the room, the distance where the direct sound from the source is equal in intensity to reverberant sound in the room, is on the order of half a meter. At ear level for the adults, both the 0.15 m condition and the 1 m condition were outside this critical distance, so the listeners may have been hearing primarily reverberant sounds which rise and fall with the intensity of the source and are roughly independent of distance. This would explain the poor performance in the task.

None of these studies was designed specifically to examine localization in the near-field, and their results are, at best, indirectly applicable to this research. The Litovsky

Eightree, LLC
Exhibit 2013
IPR2023-01024
Page 36

and Clifton and the Simpson and Stanton papers indicate that binaural distance cues in the near-field are relatively weak, but their experiments may have been tainted by reverberation and other extraneous factors. The Ashmead et al. paper indicates that there are some distance cues in the near-field that allow distance discrimination, but since their source presentations were all in the median plane, these results tell us nothing about near-field binaural localization cues. At this point, no known studies in the literature directly address auditory localization in the near-field.

3.6 Physical Acoustics in the Near-Field

The sensations we perceive as sounds are actually very small changes in the pressure of the atmosphere in the immediate vicinity of our eardrums. In relation to the normal atmospheric pressure, which is $10^5 \frac{N}{m^2}$, these pressure variations are extremely small (the smallest audible change is approximately $10^{-5} \frac{N}{m^2}$). The variable p will be used to represent this small deviation from the ambient pressure.

These small changes in pressure are propagated through the air in the form of acoustic waves. This propagation is governed by the mechanical properties of the medium of propagation. In fact, a fundamental equation governing the propagation of acoustic waves in air can be derived from the principle of conservation of mass, the elastic properties and density of air, and Newton's Second Law of Motion. This equation, known as the general acoustic wave equation, is:

$$\frac{\partial^2 p}{\partial t^2} = c^2 \nabla^2 p,$$

where ∇^2 is the symbolic *Laplacian* operator, and c is a constant related to the properties of the medium of propagation and the ambient atmospheric pressure.

If a spherical coordinate system system is used, and the acoustic wave is assumed to be spherically symmetrical, this wave equation has the general solution

$$p = \frac{1}{r} f_1(ct - r) + \frac{1}{r} f_2(ct + r).$$

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 37

This solution consists of two spherical waves, the first diverging away from the origin at velocity c, the second converging towards the origin at velocity c. Only the diverging portion of the solution is useful in acoustics, as converging acoustic waves do not commonly occur in nature. Note that the constant c represents the velocity of the propagation of the sound wave. For air, this propagation velocity is 331.6 m/s at 0 deg C and 1.013 Atm.

Propagating sinusoidal waves are commonly represented by harmonic functions. These functions represent pressures that vary sinusoidally with time and distance. They are represented in complex form as

$$\mathbf{p} = \frac{\mathbf{A}}{r} e^{j(\omega t - kr)}.$$

Here the function $f_1(x)$ in the general solution has been replaced by the function $e^{j\omega x}$. Note that ω is the frequency of the sinusoid in rad/s and $k = \frac{\omega}{c}$, the wave number, is measured in rad/m.

There is one more important acoustic variable which must be introduced before continuing. When the pressure p increases or decreases, the molecules in the air are condensing or expanding in that region. The velocity of this movement is called particle velocity, and is labeled as the variable u. The complex harmonic form of u, indicated by $\mathbf{u}$, is related to $\mathbf{p}$ by Newton's second law:

$$\mathbf{u} = -\frac{1}{j\omega\rho_0}\frac{\partial \mathbf{p}}{\partial r} = (\frac{1}{r} + jk)\frac{\mathbf{p}}{j\omega\rho_0},$$

where ρ_0 is the equilibrium density of the medium.

With these definitions of the harmonic forms of particle velocity and pressure for an acoustic spherical wave, it is possible to examine the differences in the acoustic waves generated by near and far sound sources. When the acoustic spherical wave is generated by a very distant sound source, the $\frac{1}{r}$ term in the particle velocity equation is negligible and

$$\mathbf{u} = \frac{k\mathbf{p}}{\omega\rho_0}.$$

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 38

In this case, the pressure and particle velocity are in phase and are related by a constant. Furthermore, the derivative of the magnitude of the pressure with respect to distance for $r \gg 1$ is

$$\frac{dp}{dr} = \frac{d}{dr}\frac{1}{r} = -\frac{1}{r^2} \approx 0.$$

Also, for $r \gg 1$ the curvature of a small area on the surface of the expanding spherical wave is negligible. Imagine a small area of space a great distance from the source of the spherical wave. The wave travels across this area as a flat plane, its pressure remains approximately constant, and the pressure and particle velocity are in phase. These characteristics are identical to those of a uniform acoustic plane wave propagating through space in a direction perpendicular to its surface. In this region, known as the acoustic far-field, an acoustic spherical wave can be approximated by an acoustic plane wave.

When $kr < 2\pi$ (where k is equal to 2π divided by the wavelength of the sound in meters) the $\frac{1}{r}$ term dominates the velocity equation. In this case, the phase of the velocity lags the pressure by up to 90°, and is much greater in proportion to the sound pressure than would occur at greater distances. This region, within $\frac{1}{2\pi}$ wavelengths of the origin, is called the acoustic near-field. Note that this region ranges from 16.5 m from the source for a 20 Hz sound, to 1.65 cm from the source for a 20 kHz sound. It is inconvenient to use the acoustic definition of the "near-field" in psychoacoustics because it varies with frequency. In this paper "near-field" designates the region of space within 1 m of the listener's head, where most of the perceptually relevant distance-dependent changes in localization cues occur.

The relationship between particle velocity and pressure in the near-field raises questions about auditory localization in this region. The ratio of velocity to pressure at a given distance increases as frequency decreases, as lower frequency sounds exhibit near-field effects more strongly than higher frequency sounds at the same distance. A velocity detector, therefore, would show an increase in low frequency output as the distance to the source decreased. If both the pressure and velocity spectra of a wideband source were available, it should be possible to determine the actual distance

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 39

of a nearby source with this information alone. Some researchers have argued that the human ear can act as a velocity sensor as well as a pressure sensor, and that spectral content can act as a distance cue in the acoustic near-field region, but most evidence indicates that the ears are unable to detect velocity. In addition, the interactions of the head, torso, and ears with the sound waves are probably different in the acoustic near-field. Therefore it is possible that the relative increase in the velocity of the sound wave for a nearby source is transformed by interactions with the head into an increase in pressure at the eardrums and a perceptible change in the characteristics of the sound.

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 40

Chapter 4

Evaluation of Response Methods for Near-Field Auditory Localization Experiments

Abstract

Four response methods that allow subjects to indicate perceived locations within one meter of their heads were evaluated experimentally. In the Direct-Location (DL) method, the subject moved a response pointer directly to the perceived target location. In the Large-Head (LH) method, the subject moved the response pointer to the perceived location, relative to a manikin head, that corresponded to the location of the target relative to their own head. The Small-Head (SH) method was similar to LH, except that a half-scale manikin head was used and the subjects were asked to scale down their response distances by a factor of two. In the Verbal Report (VR) response, subjects verbally indicated the spherical coordinates of the target location. Measurements with a visual target indicated that DL was relatively unbiased and considerably more accurate than the other three methods. The three indirect methods, LH, SH, and VR, were all roughly equivalent in performance. Correcting for bias improved accuracy in the LH, SH, and VR responses, but even with bias correction these methods were inferior to the uncorrected DL responses. When the visual target was replaced with an acoustic stimulus, the errors in the DL response were approximately doubled. The errors in the acoustic experiment were, however, roughly equivalent in the front and rear hemispheres, despite the expected difficulties of reaching behind the body and outside the visual field. The results suggest that DL is the most appropriate response method for near-field auditory localization experiments.

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 41

4.1 Introduction

In traditional psychoacoustic localization studies, subjects have been asked to estimate either the direction of a sound source or its distance. The perceived distance and direction of a sound source have been simultaneously measured only in a few limited cases. In planning an experiment to examine localization ability for nearby sound sources (less than 1 m away from the listener), a response method was required that would accurately measure localization accuracy in azimuth, elevation, and distance. Since some targets would be extremely close to the subject (5 cm or less from the head), a subject-based coordinate system was used to provide a consistent basis for expressing locations relative to the head in terms of spherical coordinates. The subject would be required to respond to source locations behind the head and out of the visual field, so a response method that was consistent for locations in front and behind the listener was desired. Furthermore, as a large number of trials was required, he response method had to be relatively fast. No response method found in the literature seemed adequate.

Four response methods were tested: Direct-Location (DL), Large-Head Transformation (LH), Small-Head Transformation (SH), and Verbal Report (VR). In the DL method, the subject moves an electromagnetic position sensor directly to the perceived location. In the LH and SH paradigms, the subject moves the position sensor to a location relative to a full-size or half-size manikin head that matches the perceived source location relative to his or her own head. In the VR response, the subject simply states the coordinates of the perceived sound in degrees azimuth, degrees elevation, and distance. These four response methods were tested by determining the accuracy of subject responses when the stimulus was a visual target. The results show that DL is superior to the other three methods, and that LH, SH, and VR are roughly comparable.

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 42

4.2 Background

In directional localization experiments, researchers have depended on two types of response methods— verbal report and pointing. Wightman and Kistler (1992) used verbal reports to collect azimuth and elevation responses. Subjects stated the perceived source azimuth and elevation in degrees and the investigator typed these coordinates directly into a computer. This procedure has two major drawbacks. First, the verbal response must be correctly interpreted and properly entered by the investigator; as a result, this procedure has a higher chance of error than an automated system where the control computer can directly read the subject's response. Second, the response method is slow; Wightman and Kistler collected only 2-3 responses per minute using this technique.

Makous and Middlebrooks (1990) used a head-pointing response method in their directional localization experiments. Subjects wore an electromagnetic head-tracking sensor, and responded by pointing their nose in the perceived direction of the source. This method is slightly faster than the verbal report method (3-4 responses per minute) and eliminates data entry errors. However, head-pointing is difficult for locations behind and above the subject, and it complicates methods for immobilizing of the subject's head.

Gilkey and colleagues (1995) examined an alternative to these methods called the God's Eye Location Pointing (GELP) method. The subject was seated with a plastic sphere (20 cm in diameter) located 22 cm in front of the subject and approximately 50 cm below ear level. The subject moved an electromagnetic sensor on the surface of the plastic sphere to the perceived direction of the sound. This method eliminated the need for the subject to speak or move his or her head, and permitted the use of a bite bar to restrict head motion. Furthermore, the subjects never had to move their hands away from the response sphere, so they responses could be made quickly (16-20 per minute). Gilkey assessed the GELP method with two experiments. First, subjects were asked to identify the directions of sound sources; the average angle error was 18.2°. This performance is comparable to the 20° average error reported by Wightman

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 43

and Kistler using verbal reports. The mean errors in azimuth and elevation were slightly lower than those reported by Makous and Middlebrooks using head-pointing. In a second assessment, the sound source was eliminated; the experimenter simply read verbal coordinates to the subject, who was then asked to respond at that location with the GELP method. This produced mean angular errors of approximately 9° (vs. 20° with the sound source). One concern about the GELP system is the possibility that the 20-cm rigid sphere might generate unwanted reflections that confound the auditory experiment. Overall, though, the GELP system seems to be a significant improvement over head pointing and verbal reporting for giving directional responses.

Without modification, the head-pointing and GELP response methods cannot be used to make distance judgments. Many experiments in audio distance perception have used verbal judgments of distance (Coleman, 1968; Mershon & Bowers, 1979). Studies that have simultaneously examined directional and distance perception have asked subjects to draw the location of the speaker relative to the listener on a sheet of paper (either in azimuth and distance or in both azimuth and distance and elevation and distance) (Butler et al., 1980). Gilkey suggested using a wire-frame sphere model in the GELP procedure to allow subjects to place the response sensor inside the sphere to indicate distance, but did not test this procedure. For a true three-dimensional localization experiment, drawing responses on paper will be slow and will also require additional time to digitize. Verbal report may be adequate, albeit slow, but no data are available on its accuracy in three dimensions.

Four response methods that were potentially appropriate for collecting three-dimensional locational responses at distances less than one meter were chosen for evaluation in a visual experiment.

1. Direct-Location (DL): In the DL method, a subject simply moves an electromagnetic position sensor directly to the location of the visual target. The location of the target and response can be measured using a spherical coordinate system based on the location of each subject's ears and nose, as described in the Appendix. *A priori*, this appears to be a natural response, since no mental transformation of the target location is required, and subjects can use their own

Eight kHz, LLC
Exhibit 2013
IPR2023-01024

anatomical reference points. Soechting and Flanders (1989) examined the accuracy of pointing to the remembered locations of visual targets with the tip of the finger. They found that when pointing in darkness, estimates of direction were quite accurate but that distance was underestimated. The RMS vector magnitude error from the tip of the finger to the actual target location was 11.7 cm. Surprisingly, performance did not improve when pointing with the fingertip in a lighted room, but did improve substantially when subjects used a pointer instead of the fingertip to indicate the target location. Although the DL response has many desirable properties, no data are available on the accuracy of pointing to targets that are outside the visual field. It is possible that pointing accuracy will decrease rapidly for sources behind the subject, both because no visual feedback is available and because such locations can be difficult to reach even with a pointer. A different technique that would be roughly equivalent for locations in front and behind the listener was desired.

2. Large-Head Transformation (LH): The LH technique requires the subject to remember the location of the target relative to their own head, and move an electromagnetic position sensor to that same location relative to a full-size Styrofoam manikin head. It was hoped that the anatomical features of this head would allow better judgments of distance and direction than the perfect sphere used for responses in the GELP technique. Because the experiment limited all target locations to the right hemisphere, the head was placed in profile, facing to the right of the subject; thus, response locations in front, to the right, and behind the head were equally accessible. In order to allow responses at distances up to 1 m on the right side, the head was placed 1 m away from the subject. Response locations were recorded using a coordinate system based on the locations of the ears and nose of the manikin head (see Appendix A).

3. Small-Head Transformation (SH): In the SH technique, the full-size Styrofoam manikin head used in the LH technique is replaced with a half-scale soft foam head. When making a response, the subject moves a position sensor to the

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 45

location relative to the manikin head corresponding to the location of the target relative to their own head, as in the LH response. The major difference between the two responses is that the subjects are expected to scale distance by a factor of one-half in the SH response, allowing the manikin head to be placed closer to the body.

4. Verbal Report (VR): Subjects were familiarized with the spherical coordinate system and asked to give verbal estimates of the azimuth and elevation (in degrees) and distance (in inches) of the target. The VR response is very similar to that used by Wightman and Kistler (1989b) in their localization studies.

4.3 Experiment 1

The first experiment was designed to compare the basic accuracy of the four response methods when there was little uncertainty about the actual target location. A visual target was used, and source locations were limited to the subject's field of vision. The results compare the fundamental accuracy of each of the four response methods.

4.3.1 Method

Apparatus

Figure 4-1 shows the overall setup used in the experiment. The experiment was conducted in a large, normally lighted listening booth. The four walls and ceiling of the booth were covered with acoustic foam, and the floor was carpeted. Subjects were seated on a wooden chair near the center of the room and asked to immobilize their heads with the help of a chin-rest constructed from plastic pipe mounted on a heavy base plate. The top of the chin-rest consisted of a lucite block and two plastic screws covered by a soft cloth; the screws provided a reference point allowing the subjects to maintain a consistent head position throughout each block of trials.

A half-scale model of a human head was mounted on the chin-rest in front and below the subject's head. The model was fabricated with soft packing foam and was

Eight Kind LLC
Exhibit 2013
IPR2023-01024

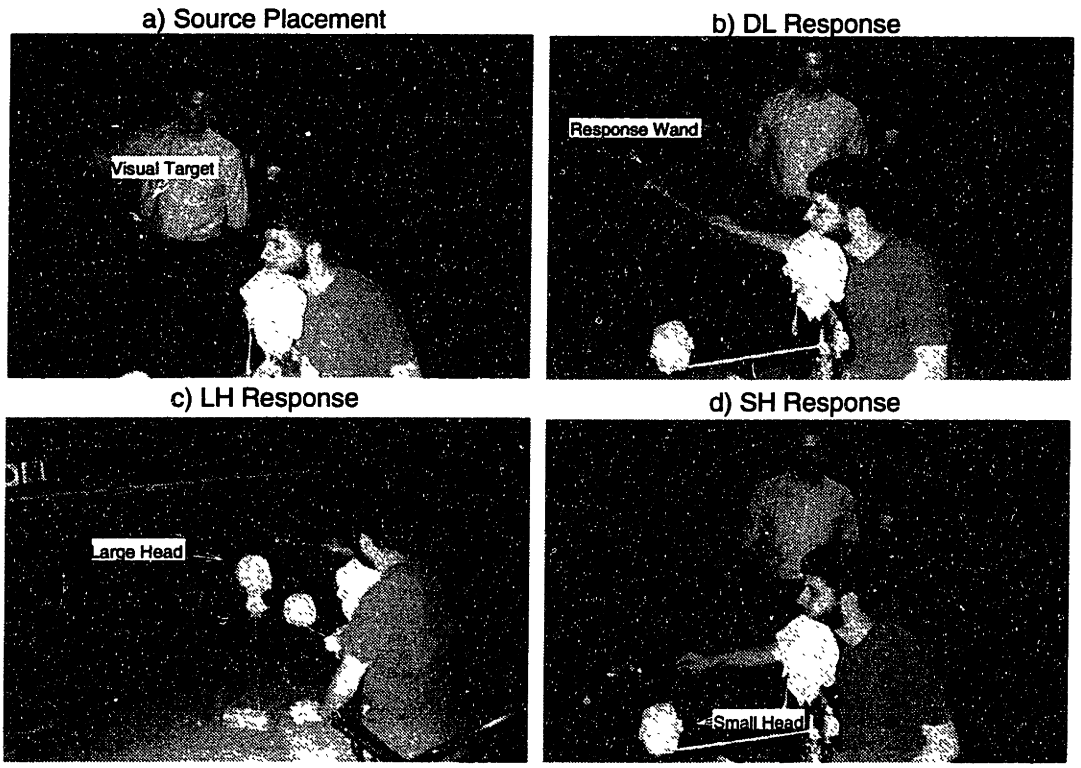


Figure 4-1: Setup and procedure used for the pointing experiments. The experimenter first places the sensor on the target wand at a random location in the front, right quadrant of the subject, makes sure the subject has seen the location, and presses a response button to have the control computer record that location (a). The target is then moved away and the subject estimates the location of the target by four methods: moving the sensor on the response wand to the location of the target (DL) (b), moving the sensor to the appropriate location relative to the large manikin head (LH) corresponding to the location relative to his or her own head (c), moving the sensor to the location relative to the small manikin head (SH) (d), and verbally reporting the azimuth, elevation, and distance of the source (VR). Note that soft cloth is draped over the chin-rest to enhance the comfort of the subjects.

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 47

roughly 10 cm wide at the interaural axis. Although a crude replica, the foam head exhibited most of the basic features of the human head. Eye-sockets and a prominent chin were carved into the solid foam, and a nose, ears and lips were fashioned separately and glued onto the head with rubber cement. A wooden dowel was affixed to the bottom of the head and attached to the chin-rest. This suspended the manikin head 25 cm below and 50 cm in front of the subject's chin. The head was in profile when viewed by the subject, and facing to the subject's right side.

A full-size replica of a human head was also positioned in front of the subject's chair. This manikin head was made of Styrofoam and purchased from a wig shop. The features of this head included ears, eyes, nose, mouth, and neck. The full-sized head is relatively small for a human, with an interaural axis length of approximately 16 cm. The head was mounted on a vertical dowel rod in a wooden base, and was placed in profile (facing to the subject's right) 100 cm in front and 25 cm below the subject's chin.

A Polhemus 3-Space Tracker was used to record stimulus and response positions. The source of the Tracker was attached to the chin-rest assembly approximately 20 cm below the subject's chin. One sensor was mounted on the end of a wooden rod, 33 cm long, that was used by the subject to make responses. The other sensor was attached to the end of a clear plastic tube, 58 cm long, which the experimenter used to place the stimuli. The experimenter also used a small hand-held switch to signal when stimulus and response positions should be recorded by a control computer. The Polhemus system is capable of measuring the three Cartesian coordinates of the sensors relative to the source within 0.5 cm in each dimension at distances up to 90 cm, and with slightly degraded accuracy at greater distances.

Calibration

Prior to each block of trials, the subject was asked to find a comfortable position in the chin-rest and immobilize his or her head. Then the Polhemus system was calibrated using nine reference positions. Specifically, the Cartesian coordinates of the left ear, right ear, and tip of the nose were recorded for the subject's head, the

Eightfold, LLC
Exhibit 2013
IPR2023-01024
Page 48

large manikin head, and the small manikin head. As discussed in Appendix A, these reference locations were used to establish the coordinate systems used for each of the three heads throughout the block of trials.

Stimulus

The target location in each trial was indicated by the position of a visual pointer. This pointer, a Polhemus head-tracking sensor on the end of a clear plastic wand, was moved by the experimenter to a random position in the front right quadrant of the subject at the beginning of each trial. To determine a random location for the target, the experimenter rolled three fair six-sided dice. One die designated the approximate azimuth of the target (ranging from 0° for a one to 90° for a six), one die designated the approximate elevation (ranging from +60° for a one to −60° for a six), and one die designated the approximate distance (15 cm for a one to 100 cm for a six).

Once the target was in position, the experimenter pressed a switch and the computer recorded the coordinates of the target sensor. Two consecutive measurements of target location were made, and the distance between the two measurements was used to ensure that the target was stationary when the switch was pressed. If the target moved more than 1.3 cm between the measurements, a warning tone alerted the experimenter to place the target again. If the target did not move, a different tone alerted the experimenter to move the target away and also informed the subject to begin his or her responses.

Responses

In each trial, the subjects estimated the location of the visual target using four different response methods, DL, LH, SH, and VR, as described in the previous section. In the DL, LH, and SH responses, the subject moved the tip of the response wand to the appropriate location corresponding to the location of the target relative to his or her own head. Once the response sensor was in place, the experimenter pressed the response switch and the coordinates of the response (relative to the manikin head) were recorded. Note that with the half-scale head the subjects scaled distance down

Eight KHz, LLC
Exhibit 2013
IPR2023-01024
Page 49

Table 4.1: Response Order in Each Block for Each Subject. V=VR; S=SH; L=LH; and D=DL.

Subject	Block 1	Block 2	Block 3	Block 4
LCL	VLDS	LSVD	DVSL	SDLV
WRD	LSVD	SDLV	VLDS	DVSL
LAR	DVSL	VLDS	SDLV	LSVD
KRJ	SDLV	DVSL	LSVD	VLDS

by a factor of two, and the Cartesian coordinates of the response (relative to the Small-Head) were doubled to allow direct comparison with the other methods.

The VR response did not require the use of the response wand. The subject simply stated the azimuth, elevation, and distance of the target and these were entered into the control computer by the experimenter.

Subjects and Response Ordering

Two male and two female volunteer subjects, ranging in age from 21 to 26, participated in the experiment. Each subject engaged in 4 blocks of 30 trials each, for a total of 120 trials per subject. Each block of trials lasted approximately 45 minutes, and the subjects were allowed 15-minute breaks between blocks. Subjects LCL and LAR completed all four blocks in one day. Subjects KRJ and WRD completed 2 blocks on each of two days.

The response order used in each of the four experimental blocks by each subject is shown in Table 4.1.

These sequences were selected to minimize the effects of response order on the overall results according to the following criteria:

- The position of each response method was different in every block. Each of the four response methods occurred first in one block, second in one block, third in one block, and fourth in one block.

- No pairs of consecutive response methods were repeated across blocks. The response immediately following a particular response type was different in each

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 50

of the four blocks.

- The same four response orders were used for each subject's four blocks of trials, but the order of the blocks was different for each subject.

Raw Data

Figures 4-2 to 4-4 show the stimulus-response pairs in azimuth (Figure 4-2), elevation (Figure 4-3), and log distance (Figure 4-4), respectively, for each subject and each response method. In each panel, "correct" (i.e., veridical) responses are indicated by a solid line. From the raw data shown in the left panels of these figures, three observations can be made. First, there are substantial differences between the accuracy of the response methods in each dimension. In azimuth, for example (Figure 4-2), the spread of responses is much greater for the LH method than for any other response method. Second, biases are evident in that the responses are often clustered away from the correct response, particularly in the LH, SH, and VR methods. For example, in distance with the VR method (Figure 4-4), the responses cluster below the solid line, indicating a bias to underestimate the target distance. Third, and of greatest significance, the DL method has the least response variability and the least bias of the four methods, with responses clustered closely around the correct response in azimuth, elevation, and distance.

4.3.2 Results

Five measures of response accuracy were used to quantify the results from the four response methods. These measures (Table 4.2) summarize the mean errors and standard deviations for the following error parameters:

- The signed error in *azimuth*: The difference between the azimuth location of the response and the azimuth location of the target. Recall that $-90°$ is directly right of the subject and all targets were on the right side, so a negative value for this error parameter implies the response was more lateral than the target.

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 51

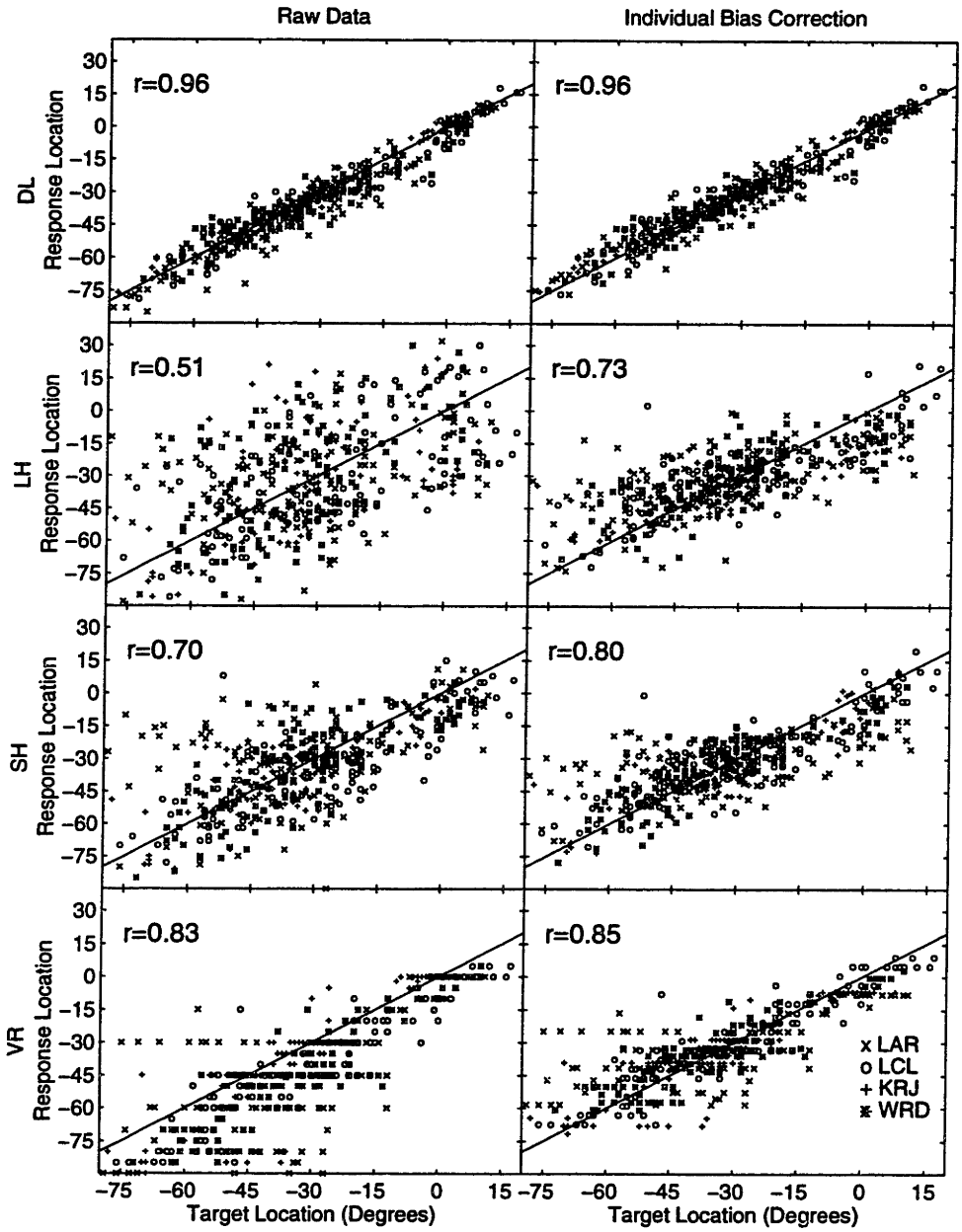


Figure 4-2: Stimulus-response pairs for azimuth. Each row shows results for a different response method (as labeled). Raw data are plotted in the left panels; data corrected for individual response biases are plotted in the right panels (see text for details). Different symbols are used to represent the responses of each subject. The solid line indicates "correct" responses. Note that the responses for verbal report are quantized (along the ordinate), representing subject biases in favor of particular response values. The azimuths range from −90° directly to the right of the subject to 0° directly in front of the subject.

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 52

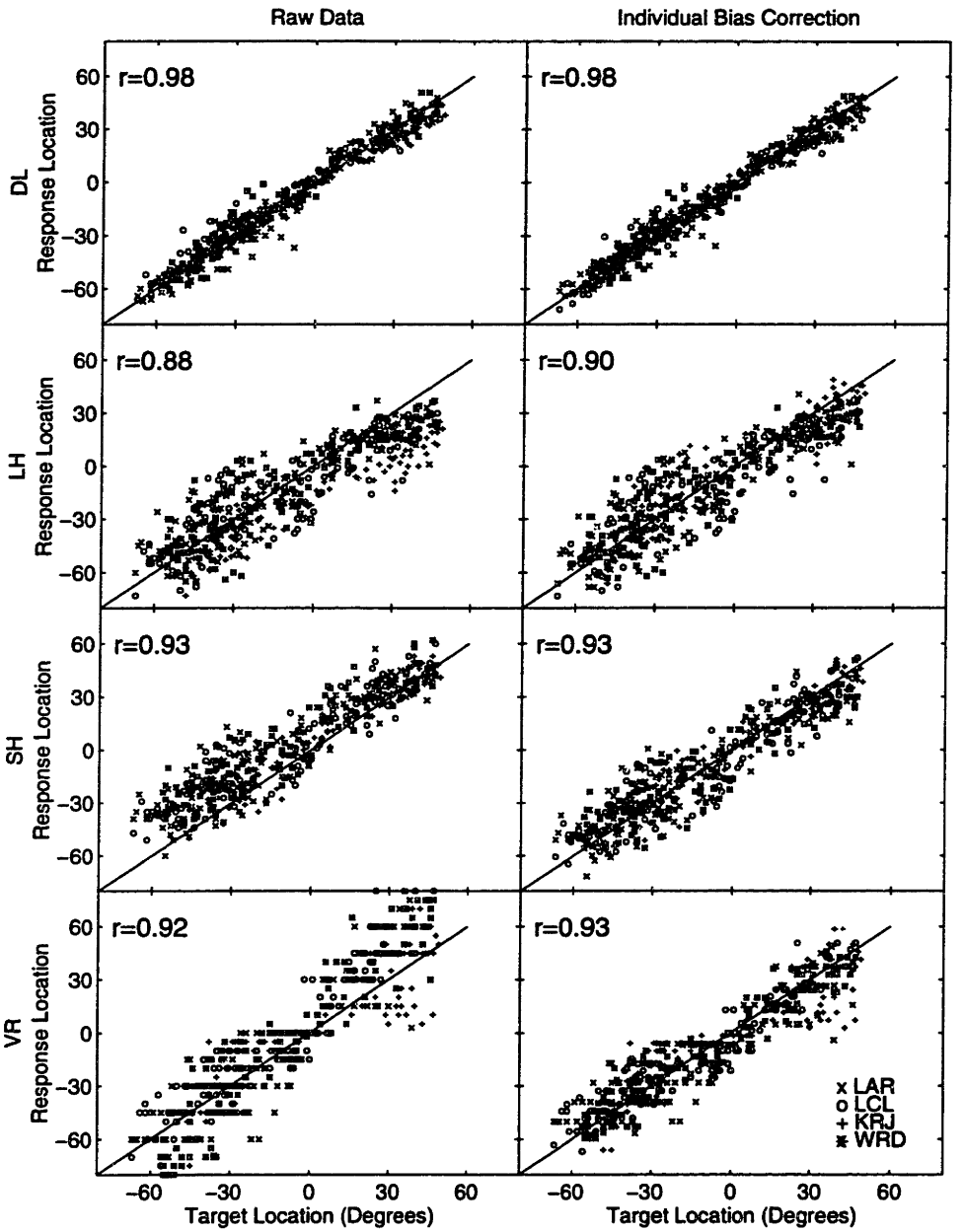


Figure 4-3: Stimulus-response pairs for elevation. Other details as in Figure 4-2. Elevations range from −90° directly below the subject to 90° directly above the subject.

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 53

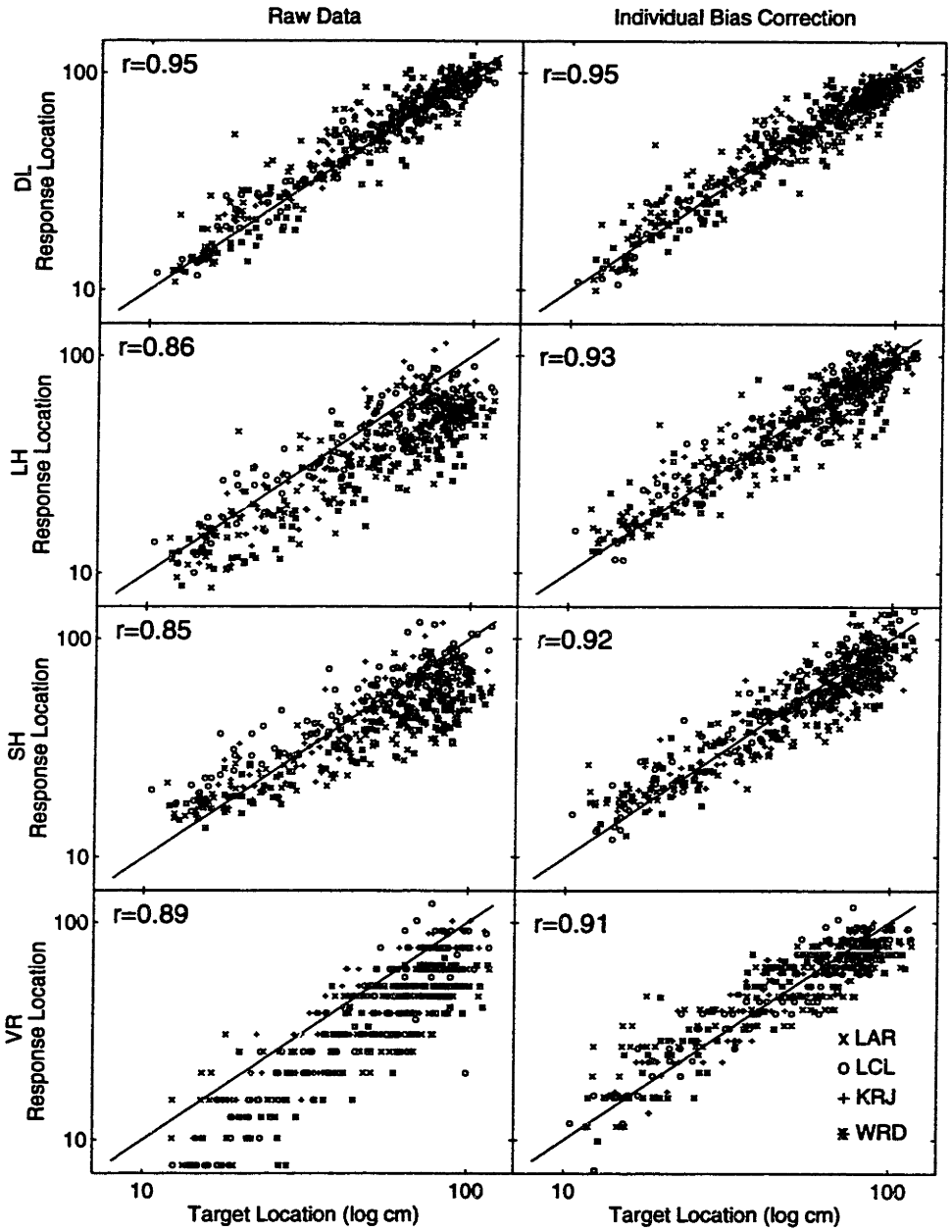


Figure 4-4: Stimulus-response pairs for distance. Other details as in Figure 4-2 but for distance rather than azimuth. Distances range from 10 cm to 100 cm from the center of the head.

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 54

- The signed error in *elevation*: The difference between the elevation location of the response and the elevation location of the target. Note that 90° is directly above the subject, so a negative value of signed elevation error implies the response is below the target.

- The signed percentage *distance* error: The percent difference between the target distance and the response distance $((\frac{r_{response}}{r_{target}} - 1) \cdot 100\%)$. A negative value indicates the subjects have underestimated distance.

- The *angle* error: The magnitude of the angle of the arc between the target and response locations on a great circle centered at the origin. Note that this error depends on both the stimulus azimuth and elevation and the response azimuth and elevation.

- The *vector-length* error: The ratio of the length of the vector going from the location of the response to the location of the target divided by the distance from the center of the head to the target, expressed as a percentage. The vector-length error includes both directional and distance components and measures overall performance in the task.

These five errors can be divided into two categories. The signed azimuth, elevation, and percentage distance errors can be positive or negative according to the direction of the error. The mean value of these quantities represents the bias of the responses, while the standard deviation measures consistency. In contrast, the angle and vector-length errors are strictly positive error measurements. The mean values of these parameters represent the total error, including any response bias and the spread of the responses around the mean, while the standard deviations are useful primarily for evaluating the significance of changes in the mean value. When responses are unbiased, or when bias is removed as described in the next section, the mean angle and vector-length errors measure only the spread of responses and become similar to the standard deviations of the signed errors.

A note should be made about the interactions between the azimuth, elevation, and angle errors. The azimuth error is potentially flawed, because it increases in

Eight KHZ, LLC
Exhibit 2013
IPR2023-01024
Page 55

Table 4.2: Mean Errors for Each Response Method

Error Type	DL		LH		SH		VR	
Azimuth	−1.8°	(5.2)	0.5°	(22.1)	−1.8°	(15.1)	−9.6°	(12.6)
Elevation	0.5°	(5.8)	−4.9°	(13.9)	10.6°	(11.6)	4.7°	(15.9)
Distance	7.6%	(19.9)	-21.8%	(23.7)	-9.6%	(26.1)	-26.2%	(22.4)
Angle	6.0°	(4.2)	21.7°	(9.5)	17.6°	(9.4)	18.7°	(10.6)
Vector-Length	21%	(16)	48%	(15)	41%	(18)	45%	(16)

Note: Standard deviations are in parentheses.

sensitivity as the target moves away from the horizontal plane. The interaction can be visualized by imagining the spherical coordinates of a world traveler. The angle, or great-circle, measure of the traveler's movements is directly proportional to the absolute number of miles he or she has moved from the starting point, and does not depend on the starting location. Similarly, the change in elevation is proportional to the number of miles north or south the traveler moves and is independent of starting location. The change in azimuth is trickier; it is proportional to the number of miles between the points on the equator due south (or north) of the traveler's starting and ending positions. The change in azimuth generated by moving one mile due west (or east) is inversely proportional to the cosine of the elevation, so the change in azimuth caused by moving one mile west near the north pole is much greater than the change in azimuth caused by an identical movement near the equator. As a result, azimuth accuracy may decrease as elevation magnitude increases. For this reason, the angle error is a better measure of accuracy than the azimuth error in spherical coordinates. Also, note that errors in elevation that cause the response to move across the top of the unit sphere will cause both the azimuth and elevation errors to be meaningless, but will not affect the angle error (although the range of elevations in this experiment is limited to −60° to 60° so polar transversals are not a problem here).

As can be seen in Table 4.2, the azimuth response is essentially unbiased except in the VR paradigm, where an average bias of almost 10° toward the right side occurs. For elevation, the largest bias is in the SH paradigm, where subjects on average responded more than 10° above the true target location. The other biases

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 56

in elevation were less than 5° in magnitude. In both azimuth and elevation, DL has less than 2° of bias. The signed distance errors show that, on average, subjects overestimated distance in the DL response, but underestimated distances when using the other response methods.

For angle error, an ANOVA on the raw data shows that the main effect of response method is highly significant ($F_{3,1650} = 425.12$, $\alpha = 0.0001$). Pairwise t-tests (at the $\alpha = 0.01$ level) show that DL is the best response angle response method (mean error=6.0°), SH and VR are not significantly different (mean errors=17.6° and 18.7°), and LH is the worst method (mean error = 21.7°).

As the final measure of accuracy, the vector-length error shows again that DL is superior to the other pointing methods (mean error = 21%), followed by SH (mean error=41%), then VR (mean error 45%) and finally LH (mean error = 48%). The DL method is clearly best, and the differences between the other three methods, though small, are also significant (one tailed t-test, α=0.01).

The overall results can be summarized as follows:

- The DL response method is greatly superior to all other methods, and is essentially unbiased in direction.
- The LH, SH, and VR responses all exhibit some type of directional bias (in azimuth, elevation, or both) and all underestimate distance.
- The overall angle errors for LH, SH, and VR are all large, although LH is slightly (but significantly) worse than the other two methods.

Inter-Subject Variability (Raw Data)

Although there were some differences in performance among the four subjects, they performed quite similarly in general (see Figure 4-5). For both the angle (great circle) error and the vector-length error (the top two panels of Figure 4-5), all four subjects exhibited the best performance (smallest mean error) in the DL method. For the other response methods, three of the four subjects performed similarly. The fourth

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 57

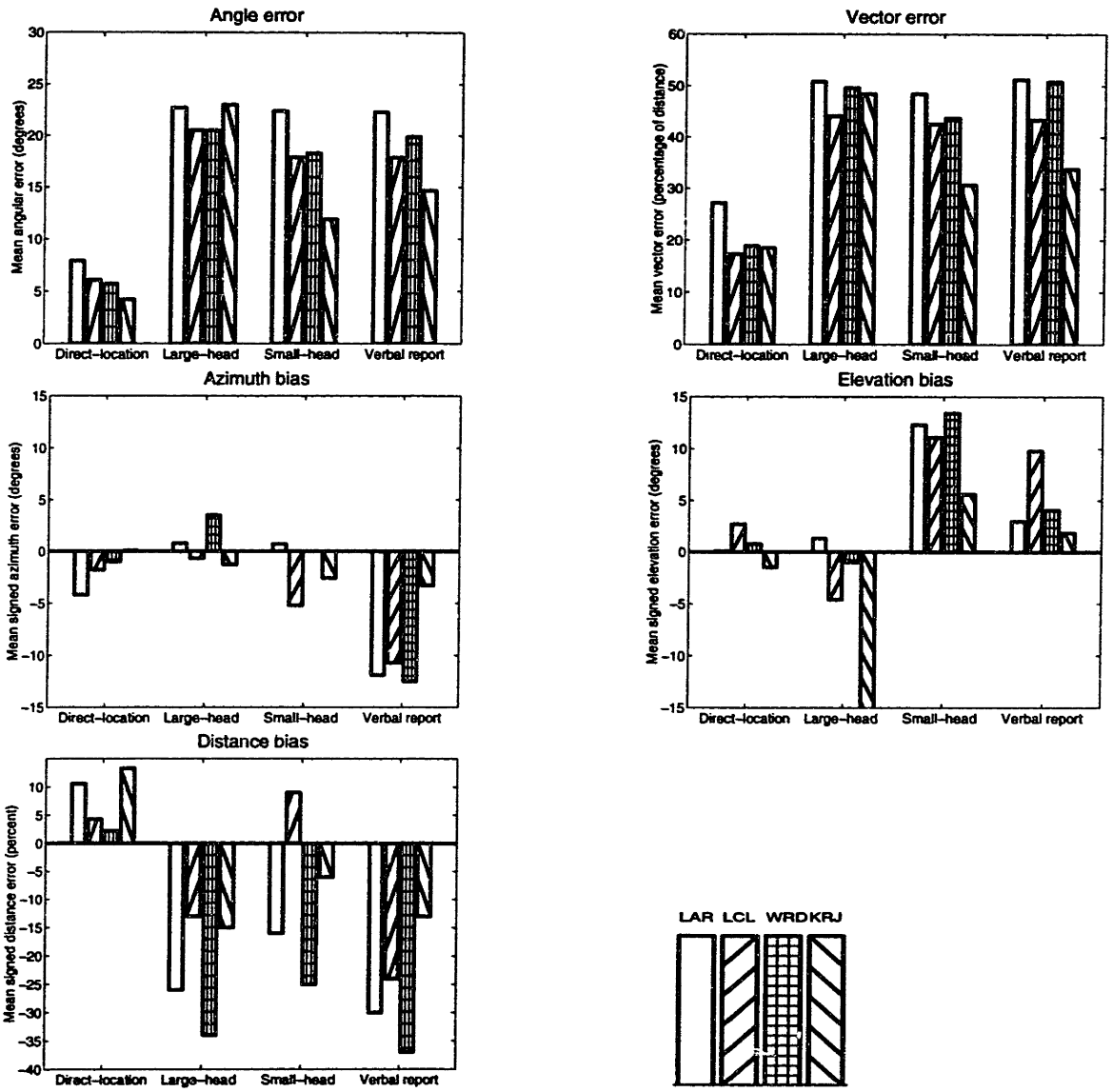


Figure 4-5: Mean errors for each subject and each response method

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 58

subject (KRJ) showed significantly better performance with the VR and SH methods than with the LH method.

The biases in each response method were also generally similar among the subjects (the bottom three panels of Figure 4-5). In every response method and every response dimension, at least three of the four subjects exhibit the same general bias trend (negligible bias, positive bias, or negative bias). In three cases, subject KRJ deviates from the general bias trend. KRJ has a much smaller bias in azimuth than the other subjects in the VR method, and his mean elevation response was significantly below that of the other subjects in the LH and SH methods. Since the other subjects exhibit small elevation biases in the LH response and large positive elevation biases in the SH response, KRJ's unique trends make him the least biased subject in the SH method and the most biased subject in the LH method. These differences in bias help explain why KRJ's mean angle and vector-length errors (which include bias effects) are smaller in the VR and SH methods than in the LH method.

A difference in head positioning during the experiment may explain the unusual elevation response biases exhibited by subject KRJ. An examination of the orientation of the subjects' heads during calibration indicates that KRJ's head was tilted slightly down (i.e., his nose was slightly lower than his interaural axis), while the other subjects' heads were tilted slightly up. The effect of this tilt is clearly seen in Table 4.3, which shows the mean stimulus target locations for the four subjects. While the average target azimuths and distances were similar for the three subjects, the average target elevation for KRJ was 13.3° higher than the average target elevation for the other three subjects. This difference suggests that KRJ's head was tilted down slightly during the experiment, and that the experimenter did not correct for this tilt when positioning the visual target. If a subject's internal representation of the target location did not account for the tilt of his head, he would position the target sensor at a lower elevation relative to the manikin head when his head was tilted downward than when his head was tilted upward, while the response bias in the DL method would be unchanged. The bias trends in Figure 4-5 are consistent, therefore, with KRJ's head having been tilted downward and the other subjects' heads tilted

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 59

Table 4.3: Mean Location of the Visual Target for the Stimuli Used With Each Subject

Subject	Azimuth	Elevation	Distance
LAR	−33.8°	−15.2°	54.5 cm
LCL	−29.3°	−10.8°	57.8 cm
WRD	−32.4°	−12.1°	57.4 cm
KRJ	−31.8°	0.6°	56.2 cm

Note: Values are given based on the coordinate system determined for each subject. With the exception of elevation of subject KRJ, the mean locations are roughly constant across subjects.

upward, although it is not clear why the VR response is apparently unaffected.

4.3.3 Bias Correction

Overall Correction

In every response method except DL, the subjects generally performed quite poorly. Angle errors were close to 20°, and vector-length errors were approximately 45%. In part, these overall errors were caused by biases in the responses. In VR, for example, responses were on average approximately 10° too far to the right, and 26% too close (Table 4.2). Each of these biases contributed to the average errors of 17.6° for angle and 45% for vector-length. If biases were systematic, the accuracy of a given response method could be enhanced by transforming the biased response locations into non-biased estimates of actual source location.

A linear transformation of each response coordinate was used to compensate for response biases. The coefficients of these linear transformations were found from the linear regression of the actual target coordinates (independent variable) on the response coordinates (dependent variable), separately for azimuth and elevation. The distance transformation was based on the regression of the log of the actual distance on the log of the response distance. The parameters of the overall bias corrections are shown in Table 4.4.

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 60

Table 4.4: Bias Correction Parameters and Correlation Coefficients

	DL			LH			SH			VR		
	m	b	r_{corr}	m	b	r_{corr}	m	b	r_{corr}	m	b	r_{corr}
Azimuth	0.97	0.9	0.96	0.42	-18.6	0.51	0.71	-8.6	0.70	0.69	-3.2	0.83
Elevation	1.00	-0.5	0.98	1.05	5.5	0.88	1.05	-10.6	0.93	0.73	-5.9	0.92
Distance	0.98	0.0	0.95	0.93	0.5	0.86	1.06	-0.1	0.85	0.70	1.1	0.89

Note: m gives the slope and b gives the y-intercept of the linear bias corrections. The transformed value of response x in each case is $mx + b$. The correlation (r_{corr}) between target location and response location is also shown.

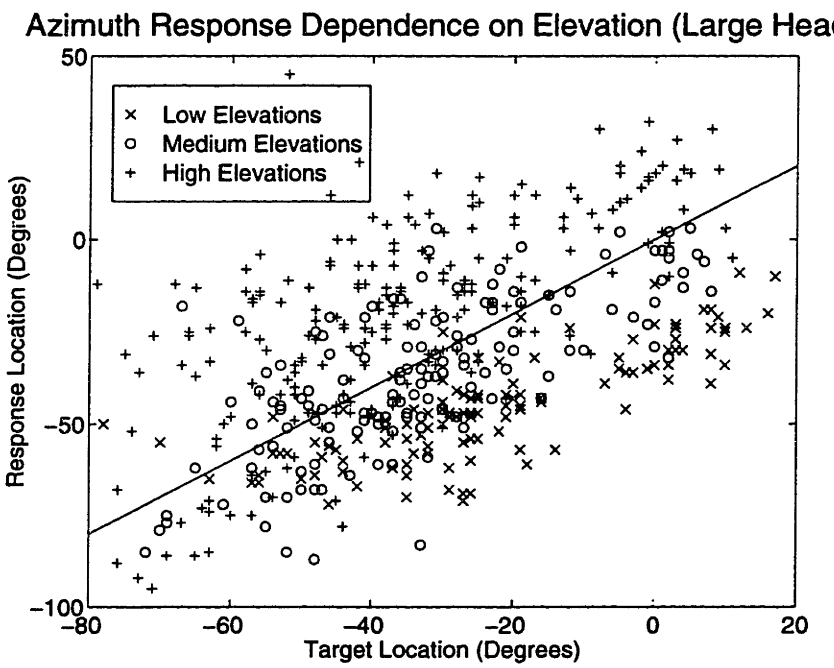


Figure 4-6: Stimulus and response azimuth for the LH method, with elevation as a parameter. The three symbols represent bias-corrected elevation responses less than $-15°$ (low), between $-15°$ and $15°$ (medium), and greater than $15°$ (high). Note that the azimuth response has a negative bias at low elevations (with x's falling below the zero-bias diagonal), is approximately unbiased at medium elevations, and is positively biased at high elevations.

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 61

Elevation Dependent Correction

The bias correction can be improved further by exploiting interactions that occur across response dimensions. An analysis of the biases in azimuth, elevation, and distance revealed that response biases for elevation were essentially independent of both distance and azimuth for all four response methods. The response biases in azimuth were independent of distance, but were strongly dependent on elevation for the LH response (see Figure 4-6) and slightly dependent on elevation for the SH responses. Similarly, the response biases for distance were dependent on elevation for the LH and SH responses. Since the elevation responses are independent of target azimuth and distance, it is possible to first estimate the target elevation by transforming the elevation response, and then transform the azimuth and distance responses based on this unbiased estimate of elevation. For simplicity, the estimate of elevation was used to divide the response region into three bins — one for corrected elevations below $-15°$, a second for elevations from $-15°$ to $15°$, and a third for elevations above $15°$. Linear regressions of target azimuth on response azimuth were performed in each of the three bins, and the resulting coefficients were used for bias correction in azimuth (Table 4.5). A similar procedure was used to determine the bias correction coefficients for the log of the response distance (Table 4.6). The elevation-dependent bias correction causes the correlation of stimulus and response azimuth to increase from 0.51 to 0.73 for the LH method and from 0.70 to 0.77 for the SH method (see Tables 4.4 and 4.5), and slightly increases the stimulus-response correlation in distance for each method.

Subject Dependent Correction

When the individual subject data were examined it was found that one subject sometimes deviated from the overall trend (see Figure 4-5). It is possible to correct the responses for these individual subject biases. This was done in the same way as the elevation-based corrections described in the previous section, except that different correction parameters were calculated for each individual subject. This results, of course,

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 62

Table 4.5: Elevation-Dependent Bias Correction Parameters for Azimuth

	LH			SH		
Elevation	m	b	r_{corr}	m	b	r_{corr}
$\leq -15°$	1.11	26.6	0.80	0.86	6.1	0.83
$\leq 15°$	0.70	-6.3	0.75	0.81	-2.3	0.82
$> 15°$	0.51	-25.7	0.65	0.66	-15.6	0.69
Corrected			0.73			0.77

Note: Elevations were corrected for linear bias before sorting into bins. See text for details.

Table 4.6: Elevation-Dependent Correction Parameters for Log Distance

	LH			SH		
Elevation	m	b	r_{corr}	m	b	r_{corr}
$\leq -15°$	0.96	0.2	0.92	0.96	0.1	0.85
$\leq 15°$	0.89	0.6	0.89	1.12	-0.2	0.89
$> 15°$	0.93	0.7	0.89	1.06	0.1	0.84
Corrected			0.91			0.88

Note: Elevations were corrected for linear bias before sorting into bins. See text for details.

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 63

in a large set of correction coefficients, and these coefficients will not be shown. The individually corrected data were used to calculate vector-length and angle errors for comparison with the non-individualized bias-corrected results. The correspondence between the subject responses when bias corrected for individual differences and the actual target locations is shown in the right panels of Figures 4-2, 4-3, and 4-4. In all cases the corrected results are clearly superior to the raw data shown in the left panels, with the responses more tightly clustered around the diagonal representing correct responses.

Bias Correction Summary

The angle and vector-length errors were calculated with each level of bias correction (Table 4.7). The errors decrease to varying degrees as each additional level of complexity is added to the bias correction scheme:

- The DL response method changes insignificantly with bias correction, indicating that these responses were essentially unbiased without any corrections.
- The LH, SH, and VR responses are all improved by bias correction, but remain roughly comparable in overall performance.
- Even with bias correction, the accuracy of the LH, SH, and VR methods remains substantially inferior to the DL method.
- The majority of the performance improvement is obtained with the overall and elevation-based bias correction. This joint correction captures more than 80% of the total improvement in angle error, and at least 50% of the total overall improvement in vector-length error. The additional improvement provided by correction for individual subject biases is small.

These results indicate that a linear bias correction can improve the accuracy of the indirect response methods. Furthermore, it appears that a general bias correction scheme chosen for all subjects will work almost as well as a more complex scheme based on the individual peculiarities of each subject's responses.

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 64

Table 4.7: Response Errors for Different Levels of Bias Correction

Error	Method	No Correction	Overall	Overall + Elevation	Individual + Elevation
Angle	DL	6.0°	6.0°		5.6°
	LH	21.7°	18.6°	16.5°	15.4°
	SH	17.6°	14.6°	13.6°	13.0°
	VR	18.7°	13.2°		12.4°
Vector-Length	DL	20%	19%		19%
	LH	48%	46%	39%	35%
	SH	41%	41%	37%	33%
	VR	45%	35%		33%

4.3.4 Locational Dependence of Response Errors

A priori, one might expect the accuracy of the response to depend on the location of the visual target. For example, the angle error might be larger for targets off to the side of the subject than for targets directly in front of the subject. In order to investigate this possibility, the correlation coefficients for the response error and the azimuth, elevation, and distance of the stimulus location were computed. The data were first corrected for bias using the overall elevation-adjusted bias correction discussed previously. The results (Table 4.8) are the correlation coefficients between the error variable (angle and vector-length) and the azimuth, elevation, and distance of the target. The angle errors are essentially uncorrelated with target position: the largest correlation coefficient for angle has magnitude 0.17, and the average magnitude of the correlation coefficient is 0.09. The vector-length errors are also essentially uncorrelated with elevation and azimuth ($|r_{corr}| \leq 0.11$, mean magnitude 0.07), but negatively correlated with distance for all four response methods (average -0.28). This indicates that the vector length error is not simply proportional to distance and may include a constant term that contributes more to the error ratio at close distances than at far distances. Overall, however, it appears that the specific location of the target within the visual field has little influence on performance.

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 65

Table 4.8: Correlation Coefficients Between Response Errors and Target Location

Error	Variable	DL	LH	SH	VR
Angle					
	Distance	-0.13	-0.11	0.03	-0.16
	Elevation	0.02	0.06	-0.14	0.08
	Azimuth	-0.17	-0.08	-0.10	-0.03
Vector-Length					
	Distance	-0.27	-0.34	-0.19	-0.32
	Elevation	0.03	0.02	-0.08	0.06
	Azimuth	-0.11	-0.10	-0.09	0.09

4.3.5 Memory Effects

On each trial, subjects were required to use all four response methods for each stimulus. While this ensures that each response method used exactly the same stimulus set, it presents the possibility that the accuracy of a response could depend on its position in the response sequence. *A priori*, one expects the subject's mental image of the target location to degrade over time; hence, the first response given after each stimulus might be the most accurate, and the last response the least accurate. In our experiment, four different response orders were used, and each response method was first, second, third, and fourth in an equal number of trials.

As a simple assessment of response order effects, we compared the accuracy of each response method in trials when it was the first response and trials when it was the last response (Table 4.9). The DL response degrades the most when it follows the other response methods. The average DL angle error increased 80% and the vector-length error increased 130% when DL was the last response versus the first response. For LH, the average angle and vector errors also increased significantly, but not as dramatically as in the DL response. None of the other errors increased significantly. A likely explanation for these results is that the DL response relies on explicit memory of the source's visual location, and this memory degrades when one is distracted by making other intervening responses. In contrast, the three indirect response methods require one to encode the location of the source and mentally

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 66

Table 4.9: Accuracy of First and Last Responses

Error	Method	First Response		Last Response	
Angle					
	DL†	4.0°	(3.0)	7.2°	(4.8)
	LH†	19.8°	(9.5)	24.8°	(9.4)
	SH	17.0°	(9.2)	17.2°	(9.8)
	VR	19.0°	(10.9)	19.3°	(10.9)
Vector-Length					
	DL†	12%	(7)	28%	(20)
	LH†	45%	(14)	51%	(15)
	SH	41%	(15)	42%	(19)
	VR	47%	(15)	46%	(15)

Note: The data were not corrected for biases. Standard deviations are given in parentheses. The † indicates differences significant at the $\alpha = 0.001$ level (two-tailed t test); no other differences are significant at the $\alpha = 0.05$ level.

transform this location into either verbal coordinates or coordinates relative to the manikin heads. This encoded memory may be less volatile than the explicit visual memory. By analogy to the memory model developed by Durlach and Braida (1969), the memory model used for DL is a *trace memory* that degrades over time and that used for the three indirect responses is a *context coded memory* that is temporally invariant. Suppose the same encoded version of the source location was used for all three indirect response methods. Then one would expect the same types of errors in all three methods within a given trial. This hypothesis was tested by examining the correlations between the errors of the three indirect response methods in each trial. Correlation coefficients between the LH and SH response errors were found to be very high – 0.79 for azimuth errors, 0.78 for elevation errors, and 0.76 for log-distance errors. Correlation coefficients between the errors of the other response methods did not exceed 0.47 (and the average of the other correlations was only 0.31). It is therefore likely that the same mental encoding of location was used for both the LH and SH responses (not surprisingly, considering the similarities between the tasks), but this encoding does not appear to be the same as that used for the VR responses.

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 67

4.4 Experiment 2

Overall, none of the three indirect response methods was found to be particularly attractive. The accuracy of the LH, SH, and VR responses was similar when (subject-independent) bias correction was applied. Other factors do differentiate the methods, however. The VR method was found to be too slow to be practical (often taking as long as the other three responses combined), and the LH response showed marginally larger angle errors than the other indirect methods. Consequently the SH response method was tentatively chosen as the best of the indirect methods. Since the small manikin allows subjects to scale distances down by a factor of two, it is not necessary to place the manikin in profile (as was done in Experiment 1) in order to allow the subjects to reach all locations in the right hemisphere at distances up to 1 m. That is, the head could be placed facing in the same direction as the subject. We hypothesized that this might yield improved performance by simplifying the internal transformations made by the subject, since a 90° reference frame rotation is eliminated. A simple experiment was designed to test this hypothesis.

4.4.1 Method

The second experiment was similar to the first experiment with only a few exceptions. First, testing was performed in an open laboratory space and not a small listening booth. Second, the responses were restricted to the SH technique. The small manikin head was no longer attached to the chin rest, but was mounted on the vertical stand previously used for the large manikin head. This allowed the Small-Head to be rotated to face either the same direction as the subject or directly to the right of the subject (i.e. in profile as in Experiment 1).

Four subjects (three male, one female, ranging in age from 23-30) were paid to participate in the experiment. Four blocks of 50 trials were collected from each of the subjects. For two of the subjects, the head was in profile for the first and third blocks and facing forward in the second and fourth blocks. For the other two subjects, the opposite ordering was used. Each block took about 20 minutes, and subjects were

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 68

Table 4.10: SH Results from Experiment 1 and Experiment 2

Error Type	Experiment 1		Experiment 2	
Azimuth	−0.8°	(12.1)	−3.4°	(12.7)
Elevation†	12.5°	(10.5)	9.7°	(9.1)
Distance††	-16.5%	(23.5)	-0.6%	(21.2)
Angle	16.9°	(9.1)	15.6°	(7.7)
Vector-Length††	41%	(15)	35%	(15)

Note: Comparison of the average response errors for the SH response in Experiment 1, when it was the first response in a trial, to the response errors for the profiled head condition of Experiment 2. Standard deviations are in parentheses. The † and †† indicate differences significant at the $\alpha = 0.05$ and $\alpha = 0.001$ level respectively (two-tailed t test).

given a break between blocks.

4.4.2 Comparison to Experiment 1

The data collected with the head in profile from Experiment 2 were first compared to the results from Experiment 1 for the trials in which the SH response was the initial response (Table 4.10). The similarity between the angle biases and the overall angle error (and their associated standard deviations) is striking. In each experiment, the subjects showed a small negative bias in azimuth and a positive bias of approximately 10° in elevation. In both cases, the overall angle error was near 16°. Only the difference in elevation bias was statistically significant ($\alpha = 0.05$, using a two-tailed t test).

In contrast, the bias for distance and the mean vector-length errors were both significantly smaller in Experiment 2 than in Experiment 1. This could be a result of reduced response complexity since subjects only needed to think about one response, or from variations among subjects.

The availability of SH response data from a second group of subjects allows us to compare the best overall bias correction of the two groups. As before, this correction was calculated from the linear regression of the target location on the response location. The results (see Table 4.11) are very similar for both experiments. Although the slope of the bias correction in elevation was slightly higher in the first experiment,

Eightkhz, LLC,
Exhibit 2013
IPR2023-01024
Page 69

Table 4.11: Bias Correction Parameters and Correlation Coefficients in Experiments 1 and 2.

	Experiment 1			Experiment 2		
	r_{corr}	m	b	m	b	r_{corr}
Azimuth	0.71	-8.6	0.70	0.72	-6.8	0.75
Elevation	1.05	-10.6	0.93	0.92	-9.3	0.95
Distance	1.06	-0.1	0.85	1.06	-0.1	0.92

Note: The transformed value of response x in each case is $mx + b$, where m is the slope and b is the y-intercept. The correlation (r_{corr}) between target and response locations is also shown.

the corrections are otherwise nearly identical. Interestingly, the optimal bias correction in distance is identical even though distance accuracy was significantly better in Experiment 2.

The individual subject biases (not shown) were not as consistent as the average biases for the two groups. Individual biases in azimuth for the SH response (in both experiments) ranged from $-10.6°$ to $2.7°$, and the biases in elevation ranged from $5.6°$ to $19.7°$. Standard deviations for both azimuth bias and elevation bias were $4.9°$. Despite these individual differences, the mean results from the two experiments support the idea that the SH response can be improved substantially with a subject-independent bias correction scheme.

4.4.3 Comparison of Profiled and Forward-Facing SH

The principal hypothesis motivating Experiment 2 was that the forward-facing condition would yield better performance than the profiled condition. The results clearly reject this hypothesis (see Table 4.12). In fact, the profiled head condition was found to be slightly, but significantly, more accurate in vector-length error. Overall, however, the difference between the two manikin orientations appears negligible, and this difference diminishes further when the data are corrected for overall biases.

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 70

Table 4.12: Mean Responses for Forward-facing and Profiled SH

Error Type	Forward-Facing		Profiled	
Azimuth††	0.6°	(11.6)	−3.4°	(12.7)
Elevation††	7.0°	(9.7)	9.7°	(9.1)
Distance	-1.8%	(26.4)	-0.6%	(21.2)
Angle	15.9°	(7.9)	15.6°	(7.7)
Vector-Length†	38%	(18)	35%	(15)

Note: Standard deviations are in parentheses. The † and †† indicate differences significant at the $\alpha = 0.05$ and 0.001 levels, respectively (two-tailed t test).

4.5 Experiment 3

In Experiments 1 and 2, the DL response was by far the best response method. The SH response, which was suggested as the best alternative to DL, produced angle and vector-length errors 2-3 times as large as the DL response even when corrected for bias. The DL method was so much better than the indirect response methods in the front hemisphere that it seemed possible that it might also be superior in the rear hemisphere, despite the difficulties of reaching behind the body and outside the visual field.

In order to test response accuracy for locations outside the visual field, a different (non-visual) stimulus was necessary. Since the primary intent of these experiments was to evaluate response measures appropriate for a near-field psychoacoustic experiment, an auditory stimulus was chosen. While such a stimulus was expected to generate a noisier response estimate than a visual target, it would allow useful comparisons of the accuracy of DL inside and outside the visual field.

4.5.1 Method

There were two major differences between Experiment 3 and Experiments 1 and 2. First, the stimulus range was expanded from the front right quadrant to the entire right hemisphere. An approximate azimuth location of the stimulus was still generated by rolling a six-sided die on each trial, but the range was increased to 0° to −180°. Second, in Experiment 3 an acoustic point source was used to indicate the

Eight KHz, LLC
Exhibit 2013
IPR2023-01024
Page 71

Table 4.13: Comparison of Response Accuracy in the Front and Rear Hemispheres for the DL and SH Responses in Experiment 3

	DL				SH			
Error	Front		Rear		Front		Rear	
Azimuth	−4.9°	(10.7)	−0.1°	(11.1)	−5.7°	(17.1)	−14.8°	(13.6)
Elevation	0.2°	(11.4)	4.4°	(11.5)	8.3°	(13.2)	1.5°	(13.0)
Distance	11.9%	(32.5)	8.9%	(29.7)	-1.5 %	(33.8)	1.6%	(34.4)
Angle	11.9°	(8.4)	13.9°	(8.0)	16.9°	(11.7)	20.7°	(9.6)
Vector-Length	37%	(24)	38%	(21)	43%	(23)	48%	(24)

Note: The stimulus was the location of a short sound signal, a 125-ms burst of white noise.

target location. Subjects were asked to keep their eyes closed while the source was being placed prior to each trial. Then a 125-ms burst of white noise was generated at the random target location. The point source was moved out of the way, and the subjects opened their eyes and gave estimates of the source location with the DL and SH responses. The subjects used the chin-rest to keep their heads stationary throughout the experiment.

Two subjects were tested, one male and one female, both of whom had previously participated in Experiment 1. Four blocks of 50 trials were collected from each subject, but a few trials had to be discarded because of head-tracker failures, so a total of 392 trials are usable. Each block of trials took approximately 30 minutes. In some trials subjects misperceived sound sources in the front hemisphere as arising from the rear hemisphere, a phenomenon that has been reported frequently by others. Whenever both the DL and SH response methods indicated a reversal had occurred, the responses were 'corrected' by reflecting the subject response across the front-back plane, as discussed by Wightman and Kistler (1989b).

4.5.2 Results and Discussion

Results were analyzed separately for stimuli in the front and rear hemispheres of the subject (trials with azimuth $> -90°$ and trials with azimuth $< -90°$). The trials were about evenly divided between the front and rear (208 front, 184 rear). The results

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 72

can be summarized as follows (see Table 4.13):

- In the front hemisphere, the errors with the DL response and an acoustic stimulus are much larger than the corresponding DL errors seen earlier with a visual stimulus. Specifically, the angle error of 12° and vector length error of 37% in Experiment 3 are both approximately double the values in Experiment 1 (6° and 21%, Table 4.2). For the SH method, however, the errors are approximately the same in Experiments 1 and 3. These results suggest that the response error due to uncertainty about the location of the auditory target dominates the DL errors, but the inaccuracies inherent to the SH response method contribute substantially to its overall variability.

- The vector-length and angle errors with the DL method and an auditory target (Experiment 3) are both comparable to the best performance seen in the SH response with a visual target (Experiment 2, see Table 4.10). In other words, the combination of variability associated with auditory localization and the noise in the DL task is about the same as the noise in the SH response alone.

- For each response method, performance in the front and rear hemispheres is similar, both in terms of mean errors and standard deviations. Although a few error measures do increase significantly (SH angle error is 22% higher in the rear hemisphere), overall differences are small. This result is surprising. *A priori* one would expect the DL task to be substantially less accurate when the subject is forced to respond outside the visual field, and at locations in the rear hemisphere that are relatively difficult to reach. However, these data indicate that average performance is virtually identical in the front and rear hemispheres. Thus, any degradation associated with positioning the response sensor out of the visual field is small compared to the overall noisiness of the auditory localization task.

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 73

4.6 Conclusions

- The Direct-Location (DL) response is markedly superior to any of the alternative response methods tested. When the DL response was the first response made with a visual target, the mean angle error with DL was only 4° and the mean vector-length error was only 12%. In contrast, the next-best response method produced angle and vector-length errors roughly three times as large. DL was also the only response without significant biases in azimuth, elevation, and distance. Clearly the DL response is the most appropriate for a near-field localization experiment.

- Replacing the visual target with an acoustic source approximately doubled the errors in the DL response method. The errors in the acoustic experiment were, however, approximately equal in the front and rear hemispheres. Although it is possible that the DL response was noisier in the rear hemisphere than in the front hemisphere, it appears that the DL response errors in either hemisphere are small compared to the uncertainty in the auditory localization task.

- Indirect location with a large or small manikin head (LH and SH) and the verbal report of coordinates (VR) were all similar in performance once the responses were corrected for overall biases. Angle errors ranged from 13.2° to 16.5°, and vector-length errors ranged from 35% to 39%. Although VR was slightly more accurate, it was slower than the other response methods. The SH response was always at least as accurate as the LH response, indicating that scaling distance down by a factor of two did not adversely influence response accuracy.

- Although variations in response biases across subjects were evident, bias trends common to all subjects were substantial. In particular, a bias correction that depended only on response elevation improved performance at least 60% as much as a more complicated correction scheme based on the individual subject biases.

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 74

- When giving all four responses on each trial (Experiment 1), response accuracy was affected by response order only for DL, and not for any of the three "indirect" response methods (LH, SH, and VR). The DL errors were about twice as large when it was the last response than when it was the first response.

- Response errors in the LH and SH methods were highly correlated, implying that both responses may be derived from the same encoded memory of the target location. Correlations between the other response methods were small.

- Once the responses were bias corrected, the errors were roughly independent of target location. Vector-length error was negatively correlated with distance for all four response methods, but the correlation was relatively weak (average r=-0.28).

- The SH response was not significantly different in angular error when the manikin head was profiled than when it faced the same direction as the subject.

In summary, the DL was superior to the three indirect response methods in terms of having the best overall accuracy, the smallest biases, and the shortest response time. These advantages promote DL as an attractive response metric for future localization studies. One caveat to this recommendation is that the present work utilized stimulus locations restricted to one side of the head. This restriction allowed positioning of the response sensor with a wand without the need for passing the wand from one hand to another. Such passing or some other modification might be required if stimulus distances up to a meter at any angle about a subject were to be included for testing. Evaluations of such response modifications await future research.

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 75

Chapter 5

Head-Related Transfer Function Measurements in the Near-Field

Abstract

For more than a century, scientists have recognized the vital role that the head-related transfer function (HRTF) plays in the spatial perception of sound. Although many HRTF measurements are available in the literature, almost all of these measurements have been made a sound source located 1 m or further from the head. There are, however, some important changes in the HRTF at closer distances. A mathematical model of near-field HRTFs, which represents the head as a rigid sphere with ears at diametrically opposite points on its surface, has been used to calculate HRTFs at distances less than 1 m. In addition, acoustic measurements of the HRTF have been made at distances of 0.12 m, 0.25 m, 0.50 m, and 1.0 m with a KEMAR manikin and an acoustic point source. Both the KEMAR measurements and the sphere model calculations indicate that the interaural intensity difference increases substantially for lateral sources as distance decreases below 1 m, while the interaural time delay remains roughly independent of distance. In the KEMAR measurements, the high-frequency responses were found to be somewhat compressed in azimuth around the ipsilateral ear when the source was close. Little interaction between distance and elevation was found in the HRTFs, and the response of the pinna was found to be independent of distance except when the source was within a few centimeters of the ear. The measurements indicate that binaural source-distance cues, which occur only with nearby sources, may allow listeners to make judgments about the distances of nearby sound sources.

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 76

5.1 Introduction

Over the past 100 years, considerable efforts have been made to characterize the relationship between the location of a sound source in space and the sound-pressure generated by that source at the eardrums of a human listener. This relationship has generally been described in the frequency domain as the head-related transfer function, or HRTF. Systematic variations in the HRTF with azimuth and elevation have been studied extensively and are well documented. These results indicate that the HRTF is roughly independent of distance when the source is more than 1 m away from the head. Despite the recognition by earlier researchers that the HRTF varies significantly with distance at distances less than 1 m (Stewart, 1911a; Hartley & Frey, 1921; Firestone, 1930; Blauert, 1983), the dependence of the HRTF on distance remains largely unexplored. This paper determines the changes that occur in the head-related transfer function at distances less than 1 m. The first two sections provide background on HRTF measurements and the changes in the HRTF at distances less than 1 m. These sections are followed by a description of a mathematical model and of the procedures used to make the acoustic measurements. The next five sections show the results of the model and KEMAR measurements in terms of the monaural HRTF in the horizontal plane, the interaural intensity difference, the interaural time delay, the interaction between distance and elevation in the monaural HRTF, and the effects of the pinna in the near-field. Finally, the perceptual implications of the distance-dependent features of the HRTF are briefly discussed.

5.2 Background

The ability of human listeners to identify the location of a sound source has been studied extensively in the past century. Lord Rayleigh (1907) first observed that sound localization is made possible by the geometric properties of the human head and, in particular, by the location of the ears at opposite sides of the head. This ear placement results in two important binaural cues for localizing the azimuth of a sound

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 77

source: a delay between the arrival of the sound at the left and right ears, known as interaural time delay (ITD), and a difference in pressure level at the near and far ears, known as interaural intensity difference (IID). Rayleigh reasoned that the interaural time delay provided an unambiguous localization cue only when the ITD was less than one full period of the sound wave, i.e., at frequencies below approximately 2 kHz. Similarly, the IID is only a useful cue at high-frequencies (above 3 kHz) where there is significant head-shadowing at the contralateral ear. The notion that the ITD is the primary localization cue at low frequencies and IID is the primary localization cue at high frequencies is known as the Duplex Theory.

Later researchers recognized some important limitations in the Duplex Theory. Interaural differences, on which the theory is based, do not correspond to a unique source location. To a first order approximation, both the ITD and IID are determined only by the angle between the interaural axis and the sound source. Consequently, ITD and IID information cannot distinguish between source locations on the surface of a cone centered on the interaural axis with its apex at the center of the head, a locus of points known as the "cone of confusion" (Wallach, 1939). Yet human listeners typically do not make errors along the cone of confusion during auditory localization, at least for wideband stimuli. Directionally dependent changes in the sound spectrum reaching the ear, produced both by the diffraction of sound by the head and torso and by the intricate shape of the outer ear or pinna, allow listeners to make accurate front-back and up-down judgments about sound location (Musicant & Butler, 1984; Oldfield & Parker, 1984). [For a thorough review of auditory localization cues, see Middlebrooks & Green (1991).]

The relationship between the sound originating from a point source in space and the sound actually reaching the eardrum of a listener is expressed by the head-related transfer function, or HRTF. The HRTF includes both magnitude and phase information and includes the effects of the locations of the ears, diffraction by the head and torso, spectral shaping by the pinnae, and the resonance of the ear canal. The precise definition of the HRTF varies. Any direct measurement typically includes the frequency response of the loudspeaker generating the stimulus and the microphone used

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 78

to make the measurement, as well as the desired transfer function from the source to the eardrum. In the literature, measured HRTFs have been presented in a variety of ways, including:

1. The ratio of the output of a probe microphone location 1-2 mm from the eardrum of a human subject to the input of the loudspeaker (Wightman & Kistler, 1989b).

2. The ratio of the output of a probe microphone near the eardrum to the free-field pressure at the location of the probe microphone with the head removed (Pralong & Carlile, 1994; Mehrgardt & Mellert, 1977).

3. The ratio of the sound pressure at the opening of a blocked ear canal to the free-field sound pressure at the center of the head with the head removed (Moller, Sorensen, Hammershoi, & Jensen, 1995).

4. The ratio of the sound pressure at the eardrum to the free-field sound pressure at the center of the head with the head removed (Gardner & Martin, 1994).

5. The ratio of the sound pressure in the ear canal to the sound pressure in the canal with the source is directly in front of the listener (Shaw, 1974).

6. The ratio of the sound pressure in the left ear canal to the sound pressure in the right ear canal (the interaural HRTF), (Firestone, 1930; Carlile & Pralong, 1994).

7. The ratio of the sound pressure in the ear canal to the maximum sound pressure (over all locations) at that frequency (used to show the directionality of the HRTF) (Middlebrooks, Makous, & Green, 1989).

These are just a few of the definitions of the HRTF which have been used in the literature. Although the actual location of the microphone within the ear-canal has a strong effect on the measured HRTF, this effect is largely independent of the sound location (Middlebrooks et al., 1989). In other words, the HRTFs measured at

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 79

different locations in the ear canal will vary only by a linear transformation which is independent of direction. The relative changes in the HRTF with direction are preserved regardless of the position of the microphone within the ear canal. It should be recognized, however, that the sound pressure measured in the ear canal is not equivalent to the sound pressure at the eardrum, and some investigators have attempted to estimate the transformation from ear-canal to eardrum to evaluate the true source-to-eardrum HRTF (Chan & Geisler, 1990; Moller et al., 1995). Note that the last three HRTF definitions are defined only relative to the HRTF at some other location and, therefore, they do not reflect any direction-independent features of the HRTF. In the measurements described in this paper, the "monaural" HRTF is defined as the ratio of the sound pressure at the eardrum to the free-field sound pressure at the location of the center of the head.

Another major difference among HRTF measurements in the literature is the use of humans and manikins in the HRTF measurements. Several studies have used acoustic manikins (Kuhn, 1979; Firestone, 1930; Gardner & Martin, 1994), which have some important advantages: they allow microphone placement at the exact location of the eardrum, do not require immobilization during measurements, and can remain in place indefinitely during a long series of measurements. The majority of studies, however, have used human subjects, both because they eliminate uncertainty about the similarity between human and manikin ears and because they allow comparison of the HRTFs across a population of subjects. In the past decade, technological advances including automated source placement systems (Wightman & Kistler, 1989b; Middlebrooks et al., 1989), and fast, high signal-to-noise ratio measurements (Foster, 1986) have allowed researchers to measure HRTFs on human subjects more quickly. Furthermore, much research in recent years has focused on the use of HRTFs in virtual audio displays (see Wenzel (1991) for review), and on the importance of the detailed features of one's own HRTFs in producing realistic virtual sounds (Wightman & Kistler, 1989b). Each of these developments has increased the emphasis on making HRTF measurements with human subjects.

Despite the wide variety of procedures used in previous HRTF measurements,

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 80

there has been no serious effort to study the effects of distance on the HRTF. With the exception of the early study by Firestone (1930), nearly all of the HRTF measurements presented in the literature were made with sound sources located 1 m or further from the listener. The importance of source distance has not been emphasized in these measurements because the HRTF is roughly independent of distance in this region. The next section briefly discusses the unique properties of the near-field in acoustics, and the changes that are expected to occur in the HRTF at distances less than 1 m.

5.3 The HRTF in the Near-Field

In acoustics, the definitions of the terms "near-field" and "far-field" depend on context. The distance-dependent changes in the auditory localization cues occur when the source approaches within one meter of the head. Therefore, in terms of human sound localization, we will designate the "near-field" as the region of space within 1 m of the center of the listener's head, and the "far-field" as the region at distances greater than 1 m.

The fundamental differences between the near-field and the far-field are illustrated in Figure 5-1. If diffraction by the head is ignored, there are two primary differences between the HRTF in the near- and far-fields. First, the decrease in the intensity of the radiating sound wave with distance (illustrated by the decreasing thickness of the lines) is large over the region occupied by the head in the near-field, but small over the region occupied by the head in the far-field. In this example, the intensity decreases by 10 dB from the nearest to the furthest point on the closer head, but only by 1.75 dB on the more distant head. As a result of this intensity effect, the amplitude of the sound at the ipsilateral ear increases more rapidly than the amplitude at the contralateral ear as a near-field source approaches the head. Therefore the interaural intensity difference is larger for nearby sources than for distant sources. In contrast, at distances greater than 1 m, the intensity of the sound wave is not significantly different at the locations of the ipsilateral and contralateral ears, and the the IID is essentially independent of distance.

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 81

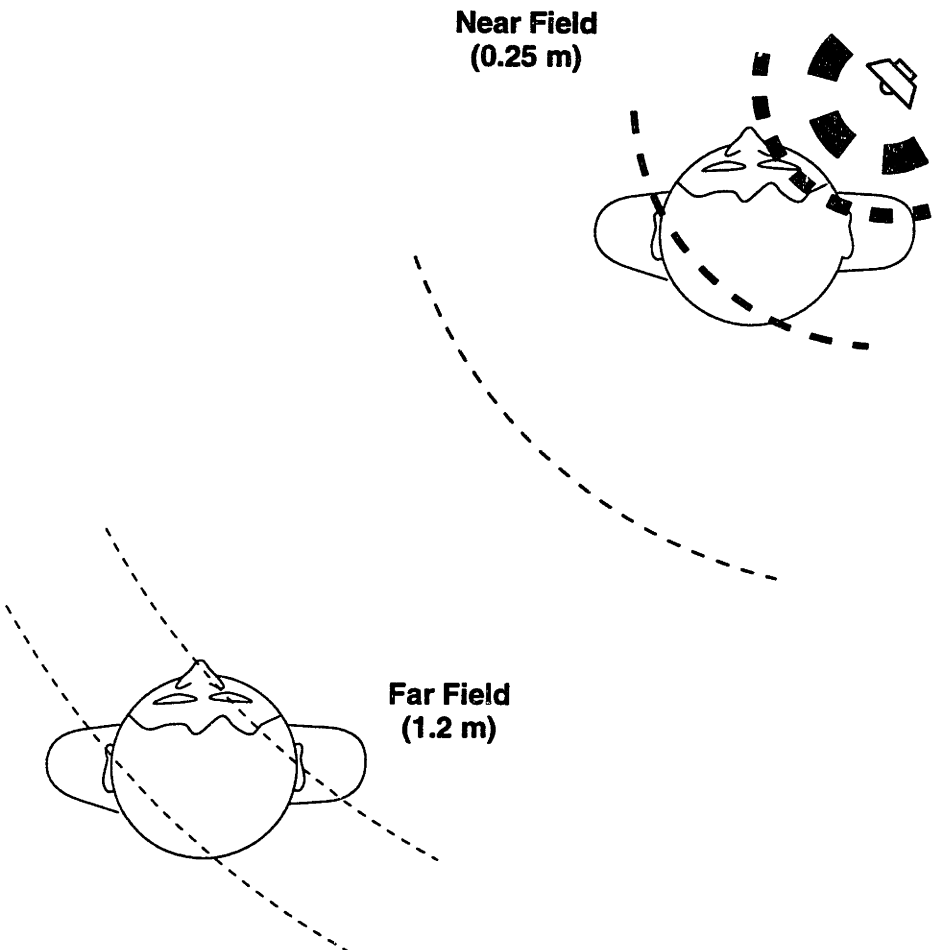


Figure 5-1: Comparison of near-field and far-field localization cues. The intensity of the spherically radiating sound wave is indicated by the thickness of the lines.

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 82

Second, the orientation of each point on the surface of the head relative to the point source varies significantly for the nearby source, but is roughly constant for the more distant source. In this figure, the angle from the nose to the source differs from the angle from the ipsilateral ear to the source by approximately 50° for the nearby source, but only by 9° for the more distant source. Since diffraction by the head depends, in part, on the angle of incidence of the sound wave impinging on the surface of the head, changes in the orientation of the source over the surface of the head can significantly influence the near-field HRTF.

Note that the interaural time delay, which depends on the the absolute propagation delay between the ipsilateral and contralateral ears and not on the ratio of the distances between left and right ears and the source, is much less dependent on distance in the near-field then the IID.

There are some practical issues which make the measurement of HRTFs in the near-field relatively difficult. The most important issue is the size of the sound source used to make the measurement. Traditionally, HRTF measurements have used conventional loudspeakers, 7 cm or larger in diameter, to generate a free-field signal. At distances of 1 m or more, loudspeakers are perfectly adequate. At close distances, however, there are serious problems associated with loudspeaker measurements:

- The precise location of a loudspeaker is not well defined in the near-field. The stimulus is generated by the entire diaphragm of the loudspeaker, and at close distances this may extend over a large region of space: at 12 cm, for example, a 7 cm loudspeaker covers an arc in excess of 30°. The HRTF measured will be, in effect, the average HRTF over the entire region covered by the loudspeaker.

- The directional properties of the loudspeaker may taint the HRTF. When the speaker is near the listener, the high-frequency directionality of the speaker will cause the sound pressure reaching the head and torso to vary according to the orientation of that region relative to the speaker. This may significantly effect the measured HRTF.

- The axial response of a loudspeaker is complicated by its distributed geometry

Eighteen, LLC
Exhibit 2013
IPR2023-01024
Page 83

at very close distances. At distances less than $2\frac{a^2}{\lambda}$, where a is the radius of the loudspeaker and λ is the wavelength of the sound, the intensity along the axis of the loudspeaker does not decrease monotonically with distance, but rather passes through a series of maxima of constant amplitude with intervening nulls (Kinsler & Frey, 62). For a 15 kHz sound generated by a 7 cm loudspeaker, this effect complicates measurements at distances less than 10 cm from the surface of the head (approximately 20 cm from the center of the head).

- A large loudspeaker may reflect the sound diffracted by the sphere, generating a standing wave between the speaker and the head and corrupting the HRTF measurements.

For these reasons, a loudspeaker cannot be used effectively to make near-field HRTF measurements. Therefore an approximation to an acoustic point source has been developed for this set of experiments, as described in the measurement procedure section.

Accurate placement of the head during the HRTF measurement is also more difficult in the near-field. An error in placement of a few centimeters is irrelevant for a source 1 m away from the head, but critical for a nearby source. Finally, the automated source placement systems widely used to make HRTF measurements are not easily adaptable to three dimensions. These difficulties may explain the absence of near-field HRTF measurements from the recent literature.

5.4 Sphere Model of the Head

Approximate interaural differences associated with a nearby sound source can be obtained from mathematical descriptions of the acoustic properties of rigid spheres. This approach to modeling near-field HRTFs is not new, and in fact was used to make manual calculations about near-field IIDs by Stewart (1911a). Hartley and Frey (1921) manually tabulated interaural amplitudes and time delays at a variety of distances and directions using Stewart's derivation. The model described in this

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 84

section was adapted from the work of Rabinowitz et al. (1993), who examined the frequency scalability of head-related transfer functions for an enlarged head. This variation of their model maintains a fixed head size and varies distance, rather than varying head size at a fixed distance (Brungart & Rabinowitz, 1996). Duda (1997) has compared the predictions of this model to acoustic measurements made on the surface of a bowling ball.

The model approximates the head as a rigid sphere, of radius a, with "ears" located at diametrically opposed points on the surface of the head. The sound source is a point velocity source radiating spherical acoustic waves, and is located at distance r from the center of the head, and at angle α from the perpendicular bisector of the interaural axis. The complex expression for the sound pressure at the ear, denoted by $\mathbf{P}_s$, is given by

$$\mathbf{P}_s(r, a, \alpha, f) = \frac{c\rho_0 u_0}{2\pi a^2} \sum_{m=0}^{\infty} (m + \frac{1}{2}) L_m(\cos\alpha) \frac{H_m(\frac{2\pi f r}{c})}{H'_m(\frac{2\pi f a}{c})} e^{j(-\frac{\pi}{2})}, \tag{5.1}$$

where a is the radius of the sphere, f is the sound frequency (in Hz), c is the speed of sound, u_0 is the volume velocity of the (infinitesimal) source, L_m is the Legendre polynomial function, and H_m is the spherical Hankel function.

In order to calculate the monaural transfer function from the model, the pressure at the ear must be divided by the reference pressure at the center of the head, which is simply the output of a point source of strength u_0 at distance r, or

$$\mathbf{P}_{ff}(r, f) = \frac{2\pi f \rho_0}{u_0} 4\pi r e^{j(-\frac{2\pi f r}{c} + \frac{\pi}{2})}. \tag{5.2}$$

A remaining complication is the transformation of the angle α between the location of the source and the location of the ear to the more standard spherical coordinates used to define HRTFs. Throughout this paper, the coordinate system has its origin at the midpoint of the interaural axis. Azimuth (θ) will be defined as 0° directly in front of the head, 90° directly to the left, and −90° directly to the right. Elevation (ϕ) will be 0° in the horizontal plane, 90° directly above, and −90° directly below. In this coordinate system, the monaural transfer function at the left ear is

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 85

$$\mathbf{H}_L(r, a, \theta, \phi, f) = \frac{\mathbf{P}_s(r, a, \arccos(\sin(\theta)\cos(\phi)), f)}{\mathbf{P}_{ff}(r, f)}. \quad (5.3)$$

5.5 Procedures for HRTF Measurements

5.5.1 Facilities

The HRTF measurements were made inside the anechoic chamber of the Armstrong Laboratory at Wright Patterson Air Force Base, a large chamber (10 m x 10 m x 10 m) which currently contains the Auditory Localization Facility (ALF). The ALF is a large, wire-frame geodesic sphere used in localization experiments at the Armstrong Laboratory (McKinley et al., 1994). Because of the ALF, it was necessary to make the HRTF measurements in a corner of the anechoic chamber. Acoustic measurements with a free-field microphone verified that the presence of the ALF did not significantly impair the anechoic conditions in the corner where the measurements were made for source distances up to 1 m. All of the HRTF measurements were made with a Knowles Electronic Manikin for Acoustic Research (KEMAR). The KEMAR manikin consists of an anthropomorphic rigid plastic head and torso. The left and right pinnae are constructed of soft rubber and mounted in removable panels on the sides of the manikin head. Inside each manikin ear, a Zwislocki coupler simulates the acoustic properties of the ear canal and the middle ear impedance, and a Bruel & Kjaer 1.2 cm (0.5 in) pressure microphone attached to the coupler measures a pressure approximately equivalent to that at the eardrum of a human listener. The output of the left and right microphones were connected to a Bruel & Kjaer 5935 dual microphone power supply, and then passed through a patch panel into the control room.

The KEMAR was mounted on a metal stand equipped with optically encoded stepper motors which allow electronic control of the azimuth and elevation of the manikin within a fraction of a degree.

The sound source used in the measurements was an approximation to a point

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 86

source. A sound driver was connected to a 3 m long piece of Tygon tubing, with an internal diameter of 1.2 cm and 1.5 mm thick walls. For convenience, the end of this tube was mounted in a PVC pipe sleeve, 2.5 cm in diameter and 64 cm in length, with the end of the tube projecting 2 cm from the end of the pipe and foam material sealing the space between the tube and the interior of the sleeve. The sleeve was used to clamp the point source to a tripod stand which was used to position the source during the measurements.

The relatively small opening of the tube allows this source to act approximately as an acoustic point source in that it is essentially non-directional even at high frequencies. At 15 kHz, for example, the 3-dB beam-width of the source was found to be 120°. Furthermore, the small size allows a precise placement of the source and eliminates the potential problem of secondary reflections off the source. Although the frequency response of this source is highly irregular, its shape is consistent and easily removed from the transfer function measurements, and its effective frequency range is from 200 Hz to 15 kHz.

In order to measure the HRTF, reference measurements were made with a Bruel & Kjaer 1.2 cm (0.5 in) free-field microphone located at the position of the center of the manikin head with the manikin removed. Measurements were made at 0.125 m, 0.15 m, 0.25 m, 0.50 m, and 1.00 m before and after the HRTF measurements. Changes in the frequency response of the source over the course of the measurements were found to be negligible (within ± 1.5 dB) and the signal measured at the response microphone was essentially independent of distance except for the inverse relation of overall amplitude to distance.

The measurements were controlled by a computer located in a small room adjacent to the anechoic chamber. The two rooms were connected by a patch panel, which passed the left and right microphone signals, the sound source signal, and the motor controller signal. The computer was also connected via GPIB bus to a Hewlett-Packard HP35665A dynamic signal analyzer, which was used both to generate the source signal and measure the transfer functions. The microphone signals were connected to the analyzer inputs. The source signal was amplified by a Crown

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 87

D-75 Power Amp before being passed through the patch panel to the sound driver.

5.5.2 Measurement Procedure

In all of the measurements the HP35665A was operated in transfer-function mode, which measures the ratio of the power spectrum on the second channel to the power spectrum on the first channel. In this mode, one of the ear microphones is connected to the second channel, and the first channel is connected directly to the source output of the analyzer. A periodic chirp source signal was used, in conjunction with a uniform window, to maximize the signal to noise ratio. At each source position, 64 FFT measurements were averaged using RMS averaging.

All measurements were made at two frequency ranges: from 100 Hz to 12.9 kHz, and from 7.78 kHz to 20.6 kHz. Each measurement consisted of a 400 point FFT with 32 Hz resolution. The measurements were then combined, with the first 240 points of the low frequency measurement and the first 360 points of the high frequency measurement, to give an overall 600 point transfer function with 32 Hz resolution from 100 Hz to 19.2 kHz. The two-part measurement was used for two reasons. First, the measurement allowed higher resolution over the entire range of human hearing than a single 400 point measurement with 64 Hz resolution. Second, and more importantly, the dual measurements allowed independent ranging of the HP35665A input channel at low and high frequencies. The transfer function of the point source has an approximately 20 dB drop-off in frequency response around 7.5 kHz, and in a single measurement the analyzer either overloaded at low frequencies or approached the noise floor at high frequencies. By dividing the measurement into two parts, it was possible to adjust the analyzer to maximize the signal to noise ratio in both frequency bands without overloading. Proper ranging in each frequency band was ensured by the control computer, which forced the signal analyzer to auto-range prior to each measurement in each frequency band, and repeated any measurements where an overload occurred with a slightly higher input range.

After each measurement, the amplitude and phase of the transfer function at each frequency value were saved into separate ASCII files. The monaural HRTF at

Eight Khz, LLC
Exhibit 2013
IPR2023-01024
Page 88

each location was found by dividing the amplitude spectrum at that location by the calibration measurement corresponding to the source distance. The phase files were used to calculate interaural time delays. First, the phase spectrum at the right ear was subtracted from the phase spectrum at the left ear for a given source direction and distance. Then the time delay was calculated by finding the constant time delay closest to the measured phase difference (in the least squared-error sense) in the frequency range of 100 Hz to 6500 Hz. Note that the squared error is based on the angular error which is restricted to the range $-180°$ to $180°$. The calculated time delay may also be viewed as the slope of the line which best fits the unwrapped phase difference between the left and right ears. This time delay measurement has been found to be repeatable within 1-2 μs.

5.5.3 Calibration

Prior to each set of measurements, the KEMAR manikin was carefully positioned to place the center of the interaural axis directly over the axis of rotation of the stand. In near-field measurements, correct placement is particularly important because even a small deviation between the center of the head and the center of rotation will cause the distance from the source to the center of the head to change as a function of azimuth and severely corrupt the transfer function measurements at very close distances. The centering of the KEMAR head was accomplished automatically with a series of acoustic measurements with the source placed in front of the manikin at $0°$ elevation. First, the manikin was centered in azimuth by rotating the head until the magnitude of the interaural time delay was reduced below 2 μs. Then the yaw of the manikin was verified by rotating the KEMAR $180°$ and again verifying that the time delay was approximately 0 μs. With a nearby source, any left or right tilt of the manikin will prevent the source from falling on the median plane at both $0°$ and $180°$ in azimuth. Finally, the manikin head was centered in elevation by verifying that the low frequency time delay from the source to the left ear at $0°$ in azimuth was equivalent (within 5 μs) to the delay from the source to the left ear at $180°$. If the manikin were tilted forward or backward, the distance (and therefore delay) from the

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 89

source to the ear would be different when the manikin was facing 0° than when facing 180°. After the elevation was adjusted, the centering measurements were repeated until the manikin was acceptably centered in azimuth, elevation, and yaw. Note that, while the adjustments in azimuth and elevation were completely automated, the adjustments in yaw were made manually by inserting material between the base of the manikin and the motorized stand. This was not a serious limitation, however, because yaw only required adjustment once prior to the measurement procedure.

5.6 Monaural Transfer Functions

5.6.1 Evaluation of Monaural Transfer Function with Sphere Model

First consider the monaural transfer function of the left ear predicted by the Sphere-based mathematical model of the head. The important characteristics of these transfer functions (Figure 5-2) can be summarized by four observations:

1. The magnitude of the HRTF increases across all frequencies as the ear rotates towards the source, and decreases as the ear rotates away from the source.

2. The magnitude of the HRTF increases with frequency when there is a direct path between the sound source and the ear, and decreases with frequency when the ear lies in the acoustic shadow of the head.

3. The magnitude of the HRTF decreases with distance when there is a direct path from the source to the ear, and increases with distance when the ear is shadowed by the head.

4. The monaural HRTFs change rapidly as distance decreases below 0.5 m, but change by no more than 1 dB as distance increases from 1 m to 10 m.

Many of the features of the sphere-model HRTFs can be explained intuitively with relatively simple acoustic concepts. Head shadowing effects are primarily responsible

Eight kHz, LLC
Exhibit 2013
IPR2023-01024

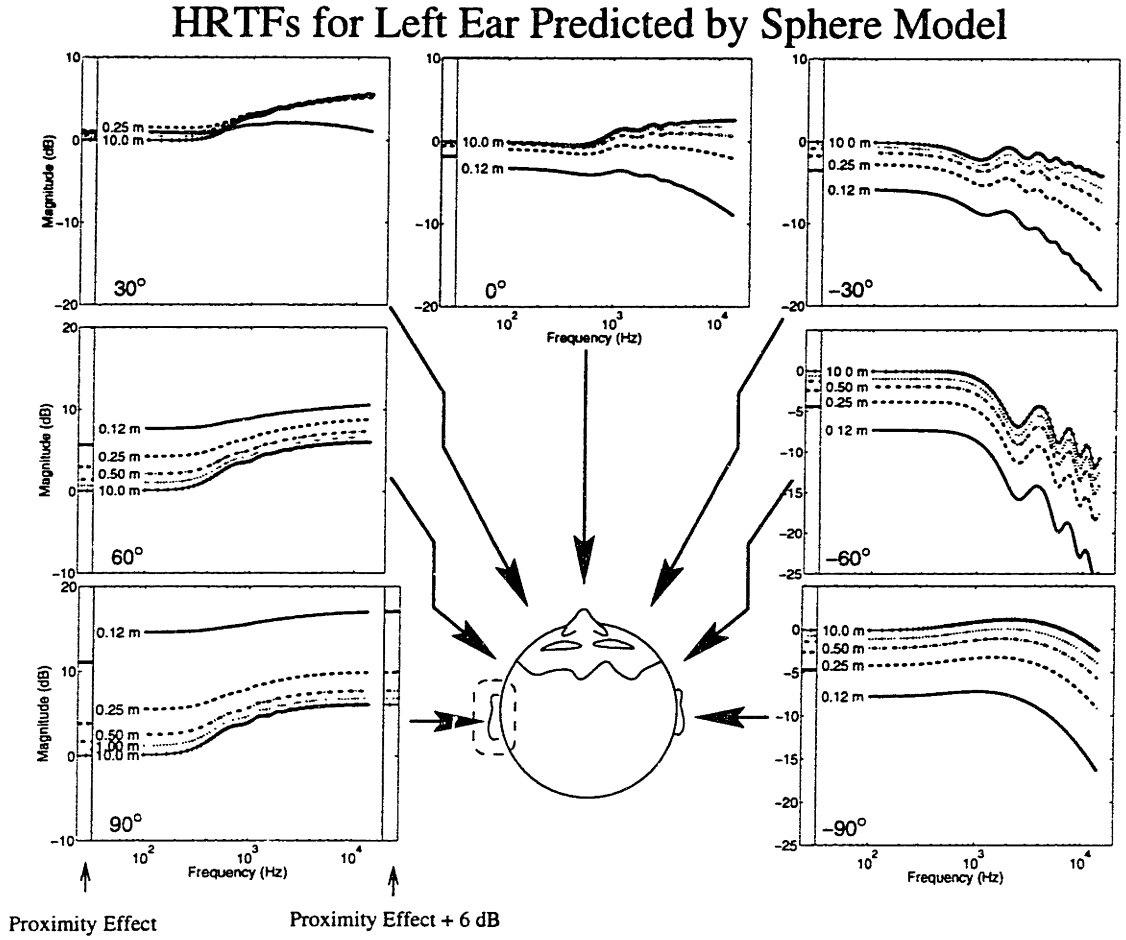


Figure 5-2: The monaural HRTF for horizontal-plane sources from 0.125 m to 10 m when the head is a rigid sphere 18 cm in diameter. The HRTFs were calculated by dividing the pressure at the left ear by the free-field pressure at the center of the head (See Equation 5.3). Results are provided at 30° intervals in the front hemisphere only, as the sphere model is symmetric across the frontal plane. Frequency is shown at 100 Hz intervals from 100 Hz to 1 kHz, and at 1/12 octave intervals from 100 Hz to 15 kHz. The bars at the left side of each graph show the source proximity effect, which is the gain of the HRTF ignoring diffraction by the head. In the 90° graph, the bar at the right side of the figure shows the source proximity effect plus the 6 dB high-frequency pressure doubling effect, and illustrates that the combination of these two effects fully explain the high-frequency asymptotes of the HRTFs at this location. See text for details.

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 91

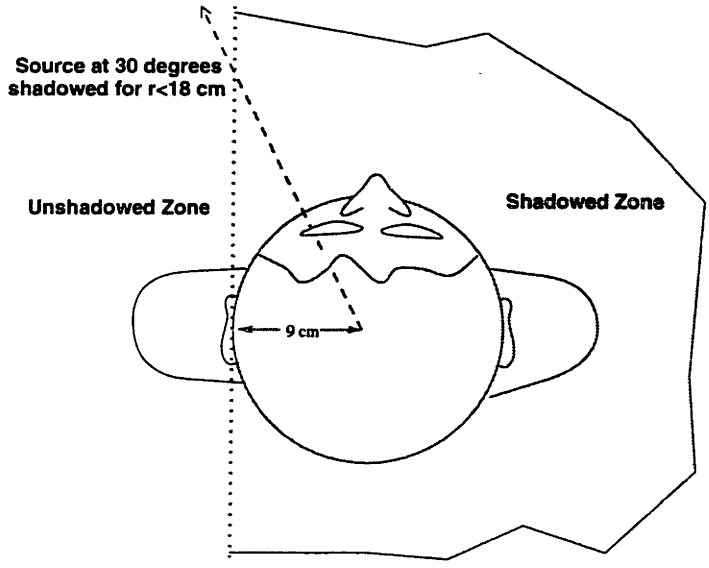


Figure 5-3: Head shadowing of a source at the ipsilateral ear. The boundary between the shadowed and non-shadowed regions for the ear is plane tangent to the sphere at the location of the ear. In this illustration of the shadowed region for the left ear, sources in the shaded region are shadowed, and sources in the unshaded region have a direct line of sight to the ear. The line shows a source at $-30°$, where a source is shadowed when it is closer than 18 cm to the center of the head. Note that sources at all distances are shadowed when directly in front of the listener.

for low-pass filtering the signal at the contralateral ear. Source proximity effects contribute to the change in the magnitude of the transfer function with azimuth at low-frequencies, and to the changes in the magnitude of the transfer function with distance. High-frequency pressure doubling causes the magnitude of the ipsilateral ear transfer functions to increase with frequency. And the acoustic "bright-spot" causes the attenuation at high-frequencies to decrease when the ear is pointing directly away from the source. Each of these concepts is explained in more detail below.

Head Shadowing

Head shadowing is simply the attenuation in the HRTF that occurs when the head obscures the direct path from the sound source to the ear. The effect of the shadow is related to the size of the head relative to the wavelength of the sound, so the attenuation at the shadowed ear increases with frequency. As a result, the HRTFs

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 92

for the contralateral ear resemble low-pass filters (see the HRTFs for $-30°$, $-60°$, and $-90°$ in Figure 5-2).

As the source approaches the head, both the size and attenuation of the shadowed zone increase. The size increases because of the convexity of the spherical head. No unoccluded path exists from a point on a convex surface to the region inside the plane tangent to the surface at that point. In the case of the spherical head, the left ear is shadowed for all sources located to the right of the plane tangent to the head at the left ear (Figure 5-3). This region includes 0° azimuth at all distances, and 30° azimuth when the source is closer than 18 cm. Thus, at 30° in Figure 5-2, the ear is in the acoustic shadow of the head only at 0.12 m, and the high-frequency response of the HRTF at 0.12 m is attenuated by the head-shadow.

The amount of attenuation due to the head-shadow increases as the ear is located further inside the shadowed region. Since the size of the shadowed region increases as source distance decreases, this results in increased high-frequency attenuation for nearby sources at the shadowed ear. The increase in high-frequency attenuation with decreasing source distance is seen at 0°, $-30°$, and $-60°$ in Figure 5-2.

Source Proximity

The inverse relationship between pressure and distance for a spherically radiating sound wave can significantly influence the HRTF when the source is near the head. This *source-proximity* effect can be viewed as the portion of the HRTF which is not a result of diffraction by the head, i.e., as the HRTF for a pressure sensor suspended in free-space at the location of the ear. Since the HRTF is defined as the ratio of the pressure at the ear to the pressure at the center of the head, the source-proximity effect is simply the ratio of the distance from the source to the center of the head to the distance from the source to the ear.

The magnitude of the source-proximity effect is shown along the left side of each panel in Figure 5-2. At 10.0 m, the effect is negligible (near 0 dB) for all source directions. The effect increases as source distance decreases, and at 0.12 m the source proximity effect produces more than 10 dB of gain at 90° and almost 5 dB of atten-

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 93

uation at −90°. The source proximity effect can explain some, but not all, of the low-frequency behavior of the HRTF. In general, the ordering of the low-frequency HRTFs by distance is consistent with the source proximity effect at each azimuth location, but the magnitude of the low-frequency gain or attenuation is greater than that predicted by source-proximity.

High-Frequency Pressure Doubling

The magnitude of the HRTF at the ipsilateral ear generally increases with frequency due to high-frequency reflections off the surface of the sphere. When the source is located at 90°, the sound waves impinging on the ear are perpendicular to the surface of the sphere, and at sufficiently high frequencies the sound wave is specularly specularly off of the surface of the sphere back in the direction of the source. At the surface of the sphere, the direct and reflected sound waves combine in phase to produce a 6 dB pressure gain. This 6 dB increase in high-frequency gain is evident in the 10 m HRTF at 90° in Figure 5-2. In fact, the high-frequency magnitude of the HRTF at 90° is exactly equivalent to the combination of the source proximity effect and the high-frequency pressure doubling effect, as shown at the right side of the panel in Figure 5-2. As the source rotates away from 90°, the sound waves from the source are no longer perpendicular to the surface of the sphere and only a portion of the sound wave is reflected at the surface, resulting in a high-frequency gain less than 6 dB at 60° and 30°. Note that high-frequency pressure doubling does not occur in the contralateral HRTFs.

Acoustic Bright Spot

When the ear is located directly opposite the source (−90° in Figure 5-2), all of the possible sound paths from the source to the ear are cylindrically symmetric and, consequently, all of the components of the diffracted sound wave combine in phase at the ear. This in-phase combination results in a local maximum in the HRTF at all frequencies. The resulting phenomenon, known as the acoustic "bright spot", is clearly seen in the high-frequency responses of the HRTFs at −90°, which are

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 94

substantially greater than in the HRTFs at $-60°$ at each source distance.

The effects of constructive and destructive interference on the contralateral hemisphere of the head are also seen in the ripples of the high-frequency HRTF response at $-60°$. The interference patterns produce a series of circularly symmetric peaks and nulls around the location of the bright-spot, and the HRTFs at $-60°$ include four of these frequency dependent nulls from 1 kHz to 15 kHz.

Low Frequency Diffraction

The low-frequency response of the sphere-model HRTFs cannot be explained intuitively with the simple concepts described above. As noted before, the source-proximity effect only partially explains the low-frequency responses of the HRTF. The rest of the low-frequency responses are a result of diffraction by the head. Note that the diffraction effects tend to increase the low-frequency gain at the ipsilateral ear, and increase the low-frequency attenuation at the contralateral ear. The magnitude of these diffraction effects increase as distance decreases.

Low-Pass Filtering as a Possible Spectral Distance Cue

The combination of the low-frequency diffraction effects in the ipsilateral hemisphere and the head-shadowing effects in the contralateral hemisphere combine to produce a consistent relationship between the shape of the HRTF and source distance. At all source directions, the high-frequency response of the HRTF is lower relative to the low-frequency response of the HRTF when the source is close than when the source is more distant. The high-frequency response is generally 4-6 dB lower relative to the low-frequency response when the source is at 0.12 m than when the source is at 10 m (Figure 5-4). This relationship implies that a sound source at a fixed location relative to the head will appear to be low-pass filtered as the sound source approaches the head. Although this effect is modest, it could be used as a monaural distance cue in the near-field and it is consistent with previous observations that sound sources appear to "darken" in timbre as they approach the head.

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 95

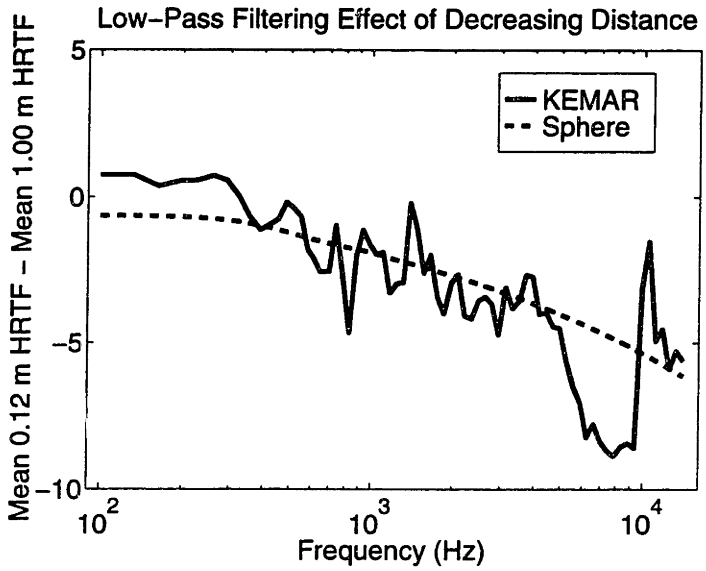


Figure 5-4: The difference between the mean HRTFs at 0.12 m and 1.0 m. The sphere model and KEMAR HRTFs at 0.12 m and 1.0 m were averaged (in dB units) across all locations in the horizontal plane. The difference between the mean HRTFs at 0.12 m and 1.0 m illustrates that the monaural HRTF, on average, decreases in magnitude more quickly at high frequencies than at low frequencies as the source approaches the head. This effect is more pronounced in the KEMAR measurements than in the sphere model. The decrease in spectral content at high frequencies as distance decreases could potentially serve as a spectral distance cue in the near field.

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 96

Contour Plots of Sphere-Model HRTFs

The left panels of Figure 5-5 provide a different perspective of the sphere-model HRTFs. Note that the more traditional HRTF plots shown in Figures 5-2 and 5-6 are essentially slices across frequency in these contour plots.

These contour plots are particularly useful for viewing the periodic nature of the acoustic bright spot in the contralateral HRTFs. At 2500 Hz, there is a single peak around $-90°$ (location B). As frequency increases, this peak decreases in width and additional peaks form on either side, until at high frequencies the central peak is very sharp and it is surrounded by multiple ridges on either side (location A). The increases in the high-frequency response of the ipsilateral HRTFs due to pressure doubling are also apparent in the contour plots (location C).

5.6.2 Measurements of Monaural Transfer Function with KEMAR Manikin

The monaural HRTFs measured with the KEMAR manikin are shown in Figure 5-6. Numerous studies examining HRTFs, both from manikins and from human listeners, are available in the literature, and the directional features of the far-field HRTFs are well documented. This discussion will focus on a comparison between the HRTFs calculated with the sphere model and measured with KEMAR, and on the distance-dependent features of the KEMAR HRTFs. Two important observations about the KEMAR HRTFs are detailed in the following sections:

- The overall shapes of the HRTFs are generally similar to those of the HRTFs calculated with the sphere model. At low frequencies (below 1 kHz) the sphere HRTFs are nearly identical to the KEMAR HRTFs. At higher frequencies, the KEMAR HRTFs diverge from the sphere model HRTFs, but the general direction and distance dependencies of the transfer functions are similar. The acoustic bright spot near $-90°$ shown in the sphere model is also apparent in the KEMAR transfer functions.

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 97

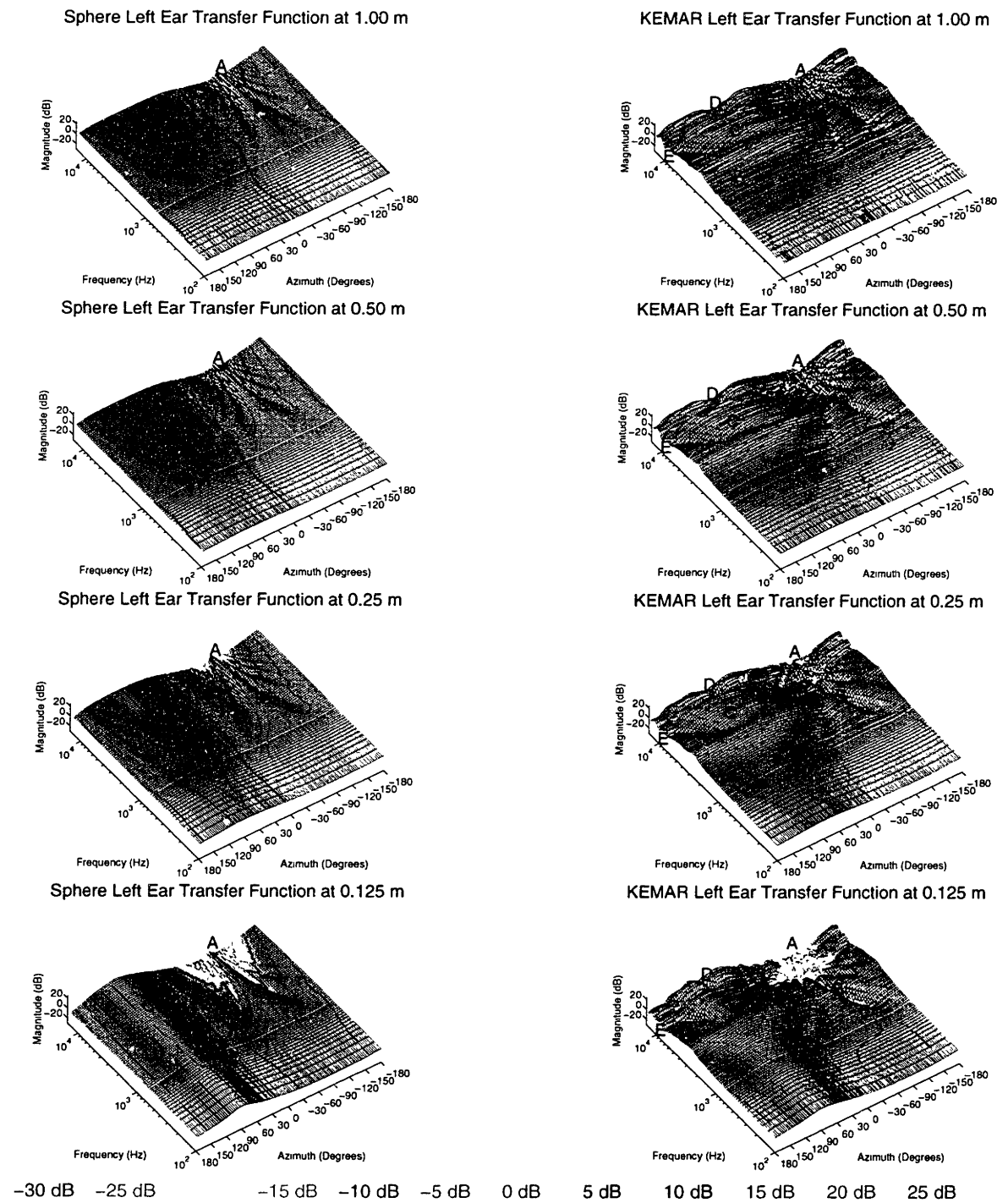


Figure 5-5: Surface contour plots of the monaural HRTFs predicted by the sphere model and measured with KEMAR. Azimuth is shown at 3° intervals, and frequency is shown at 100 Hz intervals from 100 Hz to 1 kHz, and at 1/12 octave intervals from 100 Hz to 15 kHz. The magnitude of the transfer function at each point is represented both by the height of the surface, shown on the Z-axis, and by the color, as shown by the legend across the bottom of the figure. In addition, contour lines are provided at 5 dB intervals ranging from -20 dB to 15 dB. Six reference points are present on the contour plots. Points A (-90°, 15 kHz), B (−90°, 2.5 kHz), and C (90°, 2500 Hz) are shown in all the contour plots. Points D (75°, 15 kHz), E (180°, 6.5 kHz), and F (0°, 7 kHz) are only on the KEMAR plots. Points B, C, and F are each located just above the surface of the plot.

Eight kHz, LLC
Exhibit 2013
IPR2023-01024

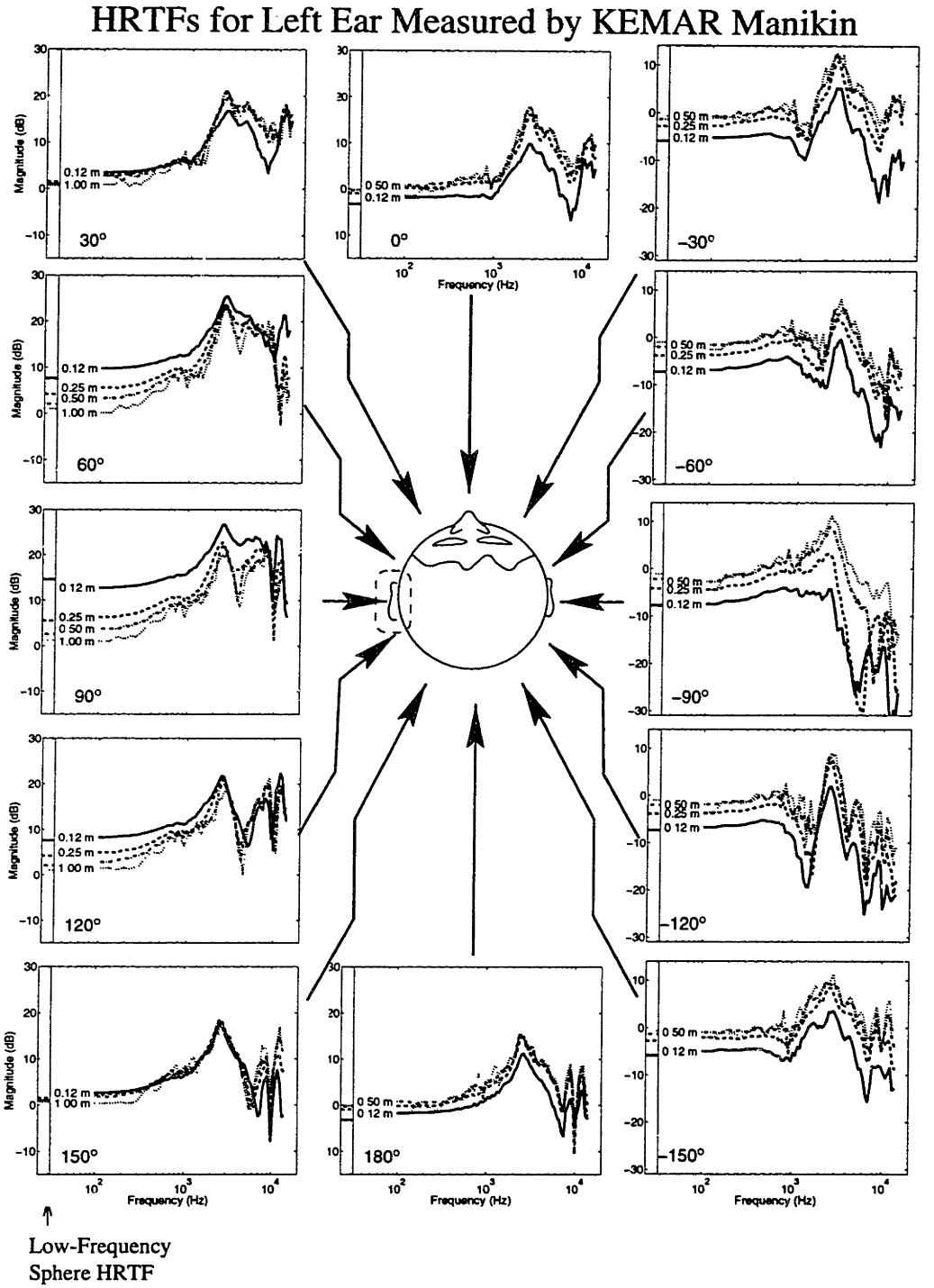


Figure 5-6: The monaural HRTF for horizontal-plane sources from 0.125 m to 10 m measured with the KEMAR manikin. The HRTFs were calculated by dividing the pressure at the left ear by the free-field pressure at the center of the head (See Equation 5.3). Results are provided at 30° intervals, and and frequency is shown at 100 Hz intervals from 100 Hz to 1 kHz, and at 1/12 octave intervals from 100 Hz to 15 kHz.

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 99

- The high-frequency features of the HRTFs are complex, particularly at the ipsilateral ear, but they appear to change systematically with source distance. In general, these features are compressed around the interaural axis as source distance decreases. The most likely results from the discrepancy between the location of the source relative to the ear and the location of the source relative to the center of the head.

Comparison of KEMAR measurements with sphere model

The sphere model best fits the KEMAR measurements at low frequencies (below 1 kHz). For comparison, the magnitude of the sphere model HRTF at 100 Hz is shown alongside the KEMAR measurements in Figure 5-6. The fit of the model to the measurements is best at the contralateral ear, and worst near the boundary between the shadowed and unshadowed zones (30° and −150°). There are also some discrepancies between the model and measurements at 0.12 m near 90°.

As frequency increases, the KEMAR HRTFs begin to diverge from the sphere model. At 2.9 kHz, the quarter-wavelength resonance of the ear canal causes a large peak in the KEMAR HRTFs at all directions and distances. At higher frequencies, the KEMAR transfer functions exhibit a complex series of direction-dependent peaks and notches derived from the geometry of the pinnae which are not reflected in the sphere model. Five major features of the sphere HRTFs are preserved in the KEMAR HRTFs.

- The magnitude of the HRTFs generally increase with frequency when there is a direct path from the source to the ear. In part, this feature is probably a result of the reflections that cause the 6 dB gain in the sphere model. The pinna, which is shaped like a cone, also provides some gain at high-frequencies. Note that the overall high-frequency gain at the ipsilateral ear is greater in the KEMAR measurements than in the sphere model.

- The high-frequency responses of the HRTFs are attenuated when the ear is in the acoustic shadow of the head.

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 100

- The overall gain of the HRTFs increases as distance decreases when a direct path exists between the source and the ear, and the overall attenuation of the HRTFs increases as distance decreases when the ear is shadowed by the head. Note that, as in the sphere model, the ear is first shadowed by the head at 30° and 150° when the source is at 0.12 m, and that the ordering of the HRTFs at high frequencies reverses at these locations.

- Overall, the magnitude of the HRTF increases more rapidly at low-frequencies than at high-frequencies as the source approaches the head (Figure 5-4). Thus, the sound reaching the eardrums is effectively low-pass filtered as the source approaches the head. This effect is more dramatic in the KEMAR HRTFs than in the sphere model, and may serve as a monaural distance cue in the near field.

- Although its structure is more complicated, the acoustic bright spot seen in the sphere model is also found in the KEMAR measurements. This is best seen in the contour plots of the KEMAR measurements (Figure 5-5). The peak at intermediate frequencies occurs slightly to the left of $-90°$ in the KEMAR measurements due to the asymmetries of the manikin head (location B). At higher frequencies, the periodic interference pattern around the bright spot is seen in the KEMAR measurements, but it is more erratic than in the sphere model. Note that one notch from the interference pattern appears to extend into the ipsilateral hemisphere at 6.5 kHz (locations E and F).

Distant Dependent Changes in the High-Frequency Fine Structure at the Ipsilateral Ear

The high-frequency features of the KEMAR transfer function are complex and are related to the geometric features of the manikin head and torso and, particularly at high frequencies, the folds and cavities of the external ear. Some of these features are present in the HRTFs of a wide variety of humans and manikins, and change consistently with the direction of the source. Shaw (1974) provides an excellent analysis of far-field HRTFs that describes several such features. In this analysis, two

Eighty Hz, LLC
Exhibit 2013
IPR2023-01024
Page 101

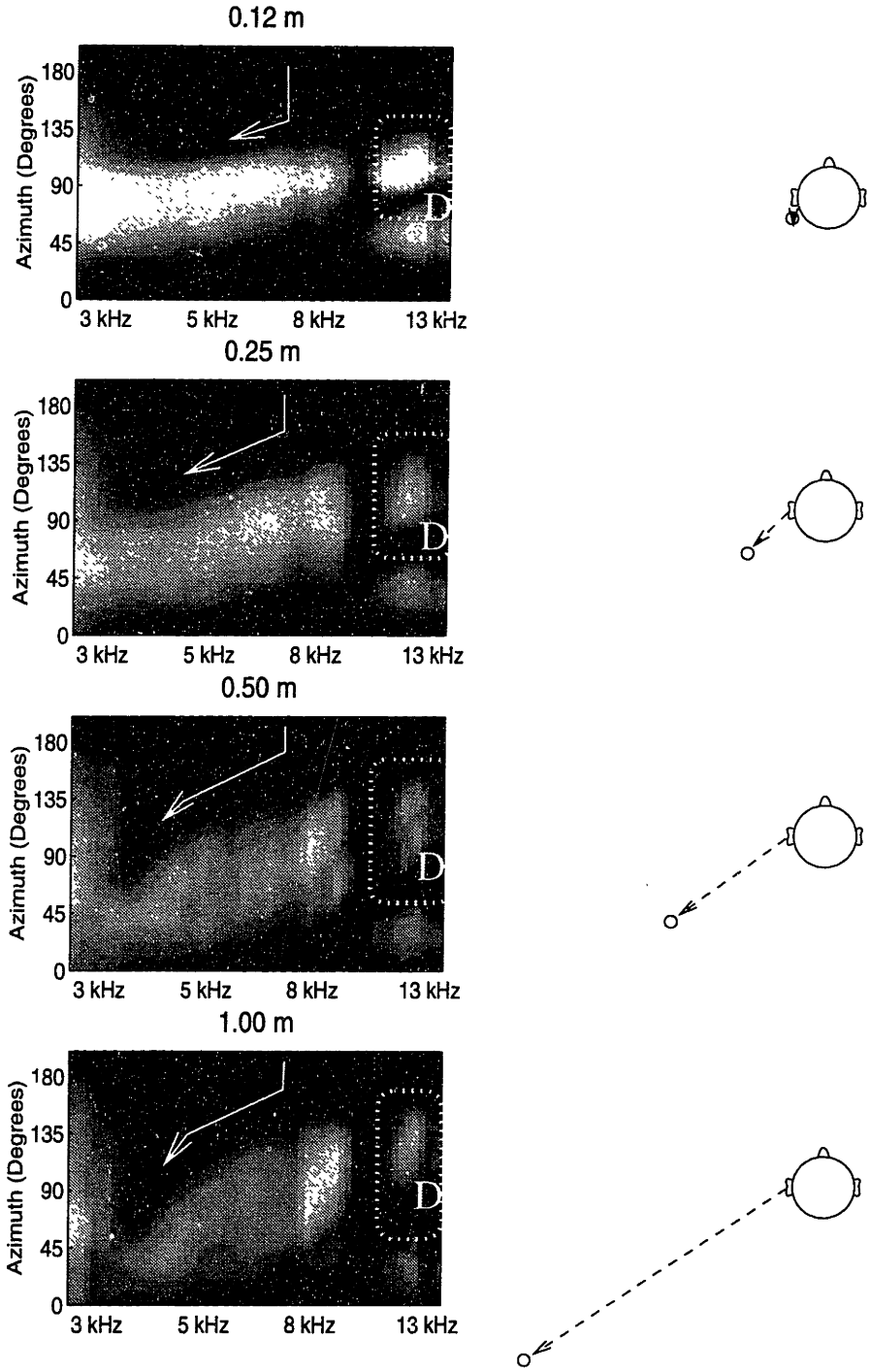


Figure 5-7: Detailed view of ipsilateral HRTF features at high frequencies. In these plots, darker features indicate a lower magnitude in the HRTF. The notch described in the text is shown by the white arrows, and the larger peak discussed in the text is surrounded by the dotted line. The right panels show the orientation relative to the ear of a source located at 120° relative to the head at each distance.

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 102

features of the high-frequency HRTF will be analyzed as a function of distance to provide insight into the behavior of the HRTF in the near-field.

Both features are visible in the contour plots of Figure 5-5. The first is a sharp notch located just to the left of Location C. This notch extends from 180° to 60° in the 1.0 m measurements, and decreases in frequency from 7 kHz to 4 kHz as azimuth increases. As the source distance decreases, note that this notch becomes shorter and decreases in frequency with azimuth more rapidly. This notch is visible at approximately 7 kHz at 180° in Figure 5-6 and decreases in depth and in frequency as azimuth moves to 60°.

The second feature is a pair of peaks in the HRTFs at 14 kHz, located at 90° and 125° at 1 m. These peaks are bordered by a deep notch across all ipsilateral locations at 9 kHz, and separated by a deep notch at approximately 60° (Location D). As distance decreases, these peaks are compressed around 90°.

A better view of these features is provided by Figure 5-7. In this figure, the notch is marked by a white arrow, and the larger of the two peaks is encircled by a dotted white line (the smaller peak is just below the dotted box). From this figure, note again that the notch decreases dramatically in length as distance increases from 1.0 m to 0.12 m. In addition, the rate at which the frequency of the notch increases with angle is greatest at 0.12 m. Note in Figure 5-6 that, at 120° and 150°, the notch is consistently at a higher frequency at 0.12 m than at the other measured distances. In effect, the notch has been pushed away from 90° at the closest distance.

There is no obvious explanation for the behavior of this notch, but it most likely results from a reflection from the head or torso destructively interfering with the direct signal at the location of the ear. The notch would be pushed away from 90° because, as the ear rotates toward the source, the ratio of the lengths of the direct path and reflected path decreases and the direct signal becomes stronger than the reflected signal and, consequently, less susceptible to destructive interference. The changes in frequency result from changes in the path length of the reflection due to the geometrical configuration of the head and a nearby source.

The second feature, the pair of peaks in the HRTF at 13 kHz, also changes sys-

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 103

tematically with distance. At 1.0 m, the peaks are relatively broad, extending from 0° to nearly 180° in azimuth. As distance decreases, the peaks progressively narrow and, at 0.12 m, they extend only from 20° to 140°. The sharpness of the peaks also increases as distance decreases. The deep notch separating the peaks moves from approximately 60° at 1.0 m to nearly 80° at 0.12 m. It appears that these peak features are compressed around 90° as distance decreases.

The distance dependencies of the high-frequency peaks are easily explained geometrically by the discrepancy between the angle of the source relative to the ear and the angle of the source relative to the head. At high-frequencies, the shape of the ear is the primary determinant of the features of the HRTF, and the response of the ear is governed by the direction of the sound waves impinging on the pinna. The right side of Figure 5-7 shows how a source located at 120° relative to the head changes in orientation relative to the ear as distance decreases from 1.0 m to 0.12 m. At the furthest distance, the direction relative to the ear is approximately equal to the direction relative to the head, but as distance decreases, the angle relative to the ear increases substantially and, at 0.12 m, the source at 120° is located at 180° relative to the ear. As a result, the high-frequency features based on orientation relative to the ear are compressed around the interaural axis.

5.7 Interaural Intensity Differences

Interaural intensity differences were calculated at 0.125 m, 0.25 m, 0.50 m, and 1.0 m with the KEMAR HRTFs, and at each of these distances plus 10.0 m with the sphere model HRTFs. The data are shown in polar form in (Figure 5-8). The important characteristics of the IIDs can be summarized as follows:

- The IID is always 0 dB in the median plane, and generally increases as the source moves lateral to the head. This result follows directly from the directional dependence of the monaural HRTFs, which increase in magnitude as the ear rotates towards the source and decrease in magnitude as the ear rotates away from the source.

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 104

- The IID generally increases as frequency increases. This behavior results from the tendency of the monaural HRTF to increase with frequency when there is a direct path from the source to the ear, and decrease with frequency when the ear is shadowed by the head. Both effects contribute to the enlarged IID at high frequencies.

- The IID increases as distance decreases, and increases dramatically as the source distance drops below 0.5 m. This distance dependence occurs because the magnitude of the monaural HRTF increases as distance decreases at the ipsilateral ear, and decreases as distance decreases at the contralateral ear.

- The acoustic bright spot directly opposite the location of the source causes a local minimum in the IID near $\pm 90°$ at intermediate frequencies (1500 Hz and 3000 Hz). In the sphere plots this minimum also occurs at higher frequencies, but in the KEMAR HRTFs the irregular shape of the head causes the bright spot to break into an erratic series of peaks and nulls which influence the IID around 90°. At 500 Hz, the bright spot does not significantly influence the IID.

As with the monaural HRTFs, the sphere model most accurately reflects the behavior of the KEMAR measurements at low frequencies. The IID at 500 Hz is a very smooth function both for the sphere model and the KEMAR measurements. At higher frequencies, the asymmetries of the KEMAR head cause the KEMAR IID to deviate from the sphere-model IID. At these frequencies, the KEMAR IIDs tend to be significantly larger than the IIDs predicted by the sphere model. This discrepancy is primarily the result of the directional properties of the pinna, which provides a significant amount of mid- to high-frequency gain at the ipsilateral ear.

5.8 Interaural Time Delays

The interaural time delay for a source in the horizontal plane was calculated at 0.125 m, 0.25 m, 0.50 m, 1 m, and 10 m with the sphere model and measured at 0.12 m, 0.25 m, 0.50 m, and 1 m with the KEMAR manikin (Figure 5-9). In each case

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 105

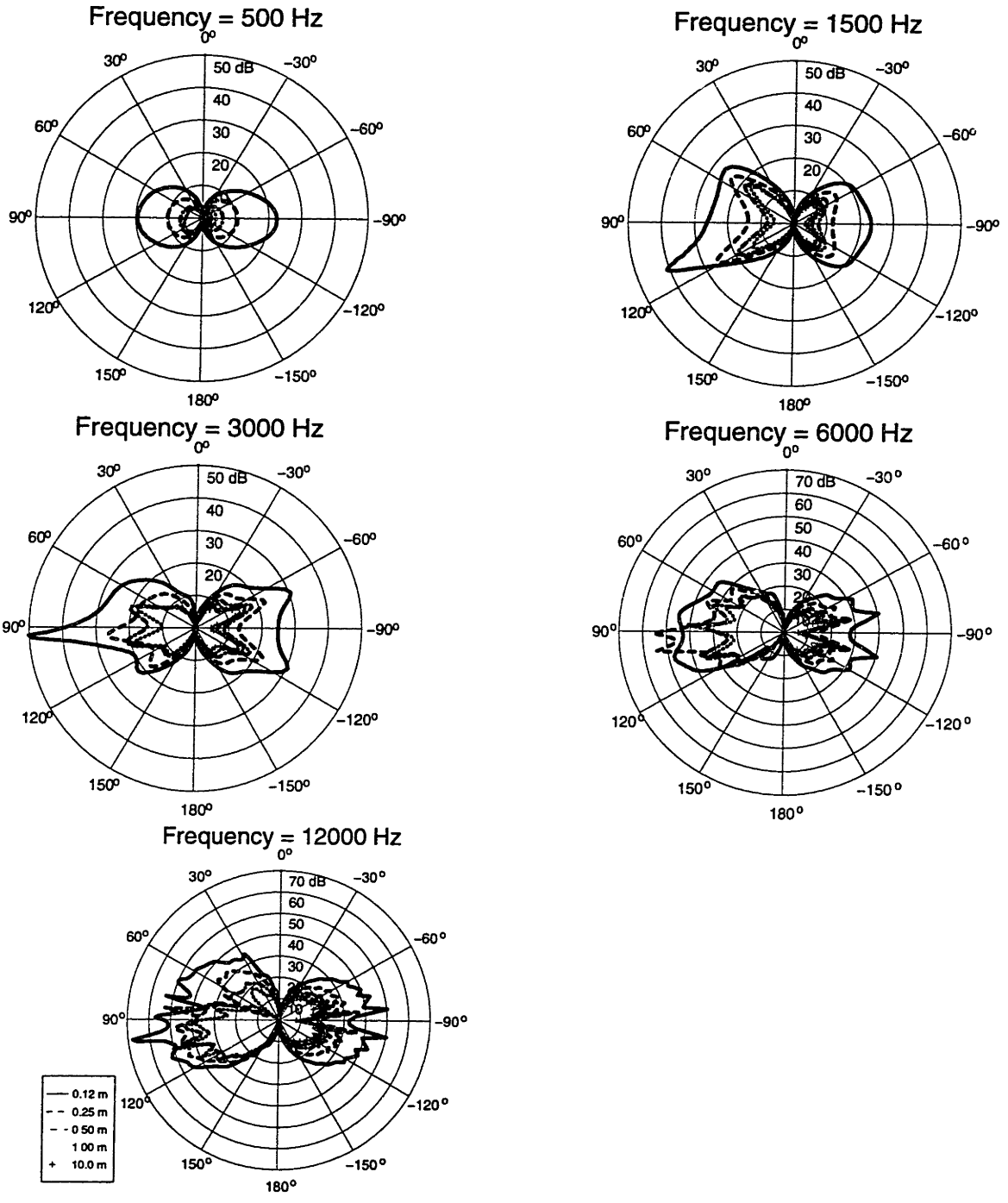


Figure 5-8: Interaural Intensity Differences. In these polar plots, the location of the source in azimuth is represented by angle, and the magnitude of the IID (in dB) is represented by the radius at each angle. Results are provided at five frequencies, ranging from 500 Hz to 12 kHz. The left side of each plot shows the IID from the KEMAR measurements, while the right side shows the IID calculated from the sphere model. Note that the scale in the plots at 6000 Hz and 12000 Hz is larger than in the lower-frequency plots.

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 106

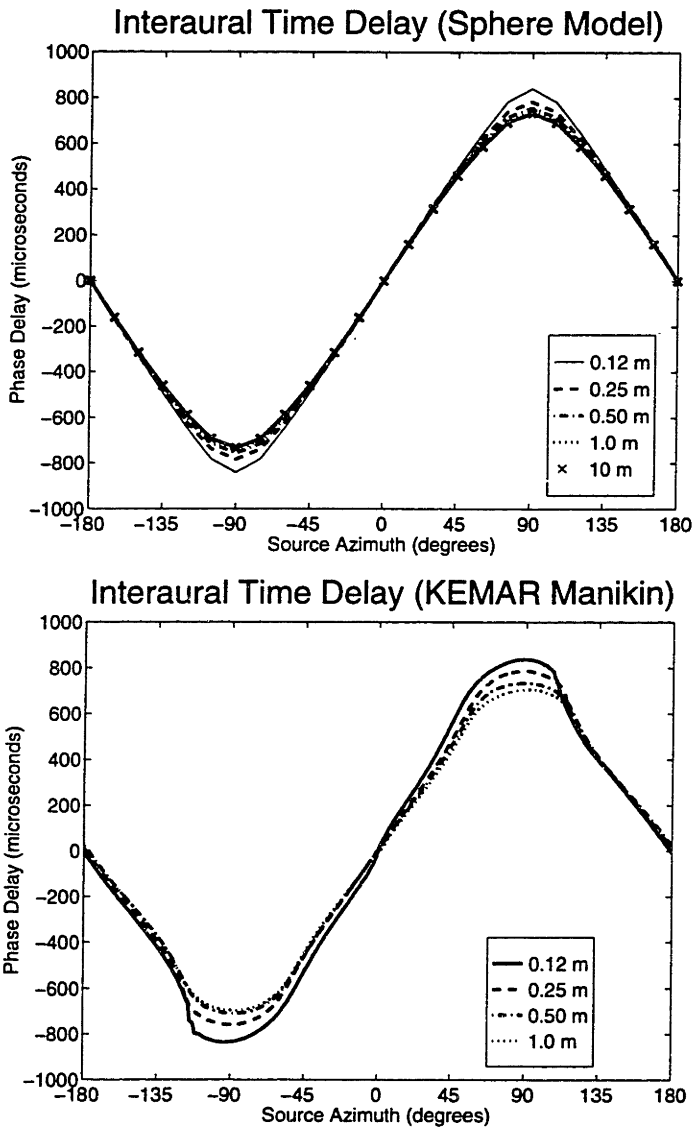


Figure 5-9: Interaural time delays. The delay was determined from the best linear fit of the unwrapped phase difference between the left and right ears (see text for details). Positive delays indicate a lag at the right ear, and negative delays indicate a lag at the right ear.

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 107

the time delay was calculated from the unwrapped phase of the difference spectrum between the left and right ears. The time delay was determined from the slope of the line best fitting the unwrapped phase of the difference spectrum from 100 Hz to 6.5 kHz. Positive time delays indicate a phase lag at the right ear, and negative time delays indicate a phase lag at the left ear.

The time delays from the sphere model are necessarily symmetric across the both the median and frontal planes. At 10 m, the time delay peaks at approximately 700 μs at 90°. As the distance decreases, the magnitude of the time delay increases slightly. This increase is most dramatic at 90° and −90°, where the time delay increases by about 100 μs as the distance decreases from 10 m to 0.125 m. The majority of this increase occurs as the source moves from 0.25 m to 0.125 m, and the remainder from 0.50 m to 0.25 m. There is virtually no dependence between the time delay and distance beyond 0.50 m. When the source is not near the interaural axis, the time delays do not vary significantly with distance.

The KEMAR time delay measurements are similar, except that the asymmetrical shape of KEMAR's head is readily apparent. This asymmetry causes the time delay to drop off more rapidly in the rear hemisphere than in the front hemisphere. Also, the time delays with the KEMAR manikin exhibit a much broader peak when the source is near the interaural axis. As with the time delays predicted by the sphere model, the KEMAR time delays increase by approximately 100 μs as distance decreases from 1 m to 0.12 m. One slight difference between the KEMAR time delays and those predicted by the sphere model is the mild increase in the magnitude of the time delay in the front hemisphere at 0.12 m. The sphere model predicts almost no increase in the delay except in the immediate vicinity of the interaural axis.

Both the sphere model and the KEMAR measurements indicate that the dependence of the interaural time delay on distance is substantially weaker than the dependence of the interaural intensity difference on distance. At the closest distances, the magnitude of the IID increases dramatically, while the ITD never increases by more than 10-12 %. The reason for this discrepancy is simple: the time delay depends on the arithmetic difference between the distance from the source to the ipsilateral

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 108

ear and the distance from the source to the contralateral ear. Ignoring the effects of diffraction by the head, the IID depends on the ratio these distances. As the source approaches the head, the ratio of distances to the ipsi- and contralateral ears increases much faster than the absolute difference between the distances, so the IID increases more dramatically than the ITD. As discussed later, this disparity between the distance dependence of IID and ITD is even greater if perceptual considerations are considered.

5.9 Elevation Effects on Near-Field HRTFs

The high frequency features of the monaural HRTF at the ipsilateral ear change substantially with elevation. This is illustrated in Figure 5-10, which shows the high-frequency features of the monaural HRTF at three elevation locations. The pattern of peaks and notches in the transfer functions are significantly different at each elevation location. At +30° elevation, for example, there is a notch at approximately 7 kHz in the HRTF that stretches across the entire ipsilateral hemisphere which is not found at either of the other two elevations. At +30° elevation, there is a wide null near 0° azimuth at 8 kHz. Similar patterns have been reported in previous HRTFs studies (Carlile & Pralong, 1994; Shaw, 1974) and they will not be discussed in detail here. It is apparent, however, that these patterns do vary significantly with elevation, and that they could provide a salient cue for evaluating the elevation of a sound source.

These patterns are apparently relatively independent of source distance. Careful observation reveals that the features of the HRTFs are considerably more consistent across the rows of Figure 5-10, which represent different elevation values, than across the columns, which represent different distances. The general pattern of features at each elevation is clearly recognizable at all three measured distances. If this result is generally true at all elevations, it would imply that elevation cues are roughly independent of distance, and that the same mechanisms that allow elevation perception in the far-field may also be used in the near-field.

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 109

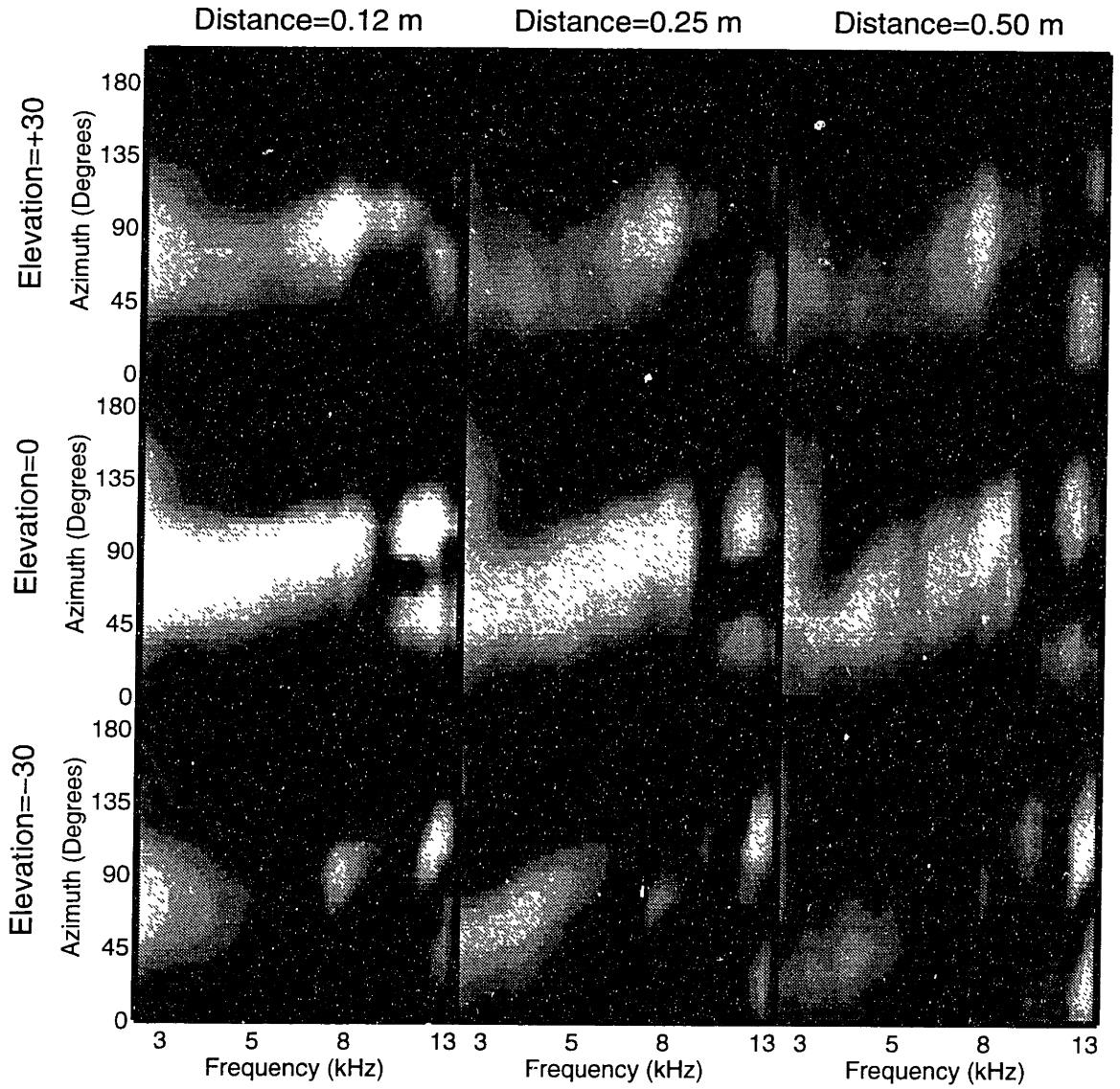


Figure 5-10: Surface plots of the left ear monaural HRTFs measured with the KEMAR manikin at +30°, 0°, and −30° in elevation. As in Figure 5-7, the darkness of the plots indicates the magnitude of the HRTF at each point: brighter areas indicate greater gain in the HRTF. The plots are limited to high-frequencies at the ipsilateral ear, where the most salient elevation cues occur, and are shown with 1/12th octave resolution in frequency and 15° resolution in azimuth (interpolation has been used to smooth the plots).

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 110

5.10 Pinnae Effects for Very Near Sources

The geometry of the pinnae is complex, and it is difficult to predict how the response of the pinnae will change when a source is close to the ear. In order to isolate the contribution of the pinnae to the near-field HRTF, the transfer function at the right ear was measured for sources located just outside the ear (at $-90°$ azimuth) with and without the pinnae. Measurements were made at distances of 2.5 cm, 3.75 cm, 5 cm, 7.5 cm, and 10 cm, measured from the opening of the ear canal rather than the center of the head. At each location, an initial measurement was made with the standard KEMAR pinna, and then a second measurement was made with the removable pinna of the KEMAR manikin replaced by a flat rubber sheet flush with the surface of the head and the opening of the ear canal.

The HRTF at each position was calculated by dividing the frequency response measured at the ear by the free-field pressure at the center of the head if the manikin were removed. Calibration measurements were unavailable at these distances, so the free-field signal was calculated by scaling the free-field transfer function measured at 12 cm in accordance with the inverse relation between pressure and distance.

The features of the monaural HRTFs with the pinna attached (upper panel of Figure 5-11) are similar to those measured for the left ear at 90°. The gain of the transfer function generally increases as distance decreases, but the increase is considerably larger at low frequencies than at high frequencies. Note the large discontinuity between the HRTFs at 3.75 cm and 2.5 cm at high frequencies. At this point, the source is surrounded by the concha and the HRTF changes dramatically.

When the pinna is removed, the variations in the HRTF with distance are more systematic (middle panel of Figure 5-11). Notice that the increase in gain as the source approaches the head decreases decreases as frequency increases from 100 Hz to 2 kHz, but is roughly independent of frequency beyond that point. This pattern is similar to the HRTF predicted by the sphere model for a source at 90° (Figure 5-2). Note that the elimination the pinna effectively decreases the length of the ear canal by about 20%, resulting in a 500 Hz increase in the quarter-wavelength ear-canal

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 111

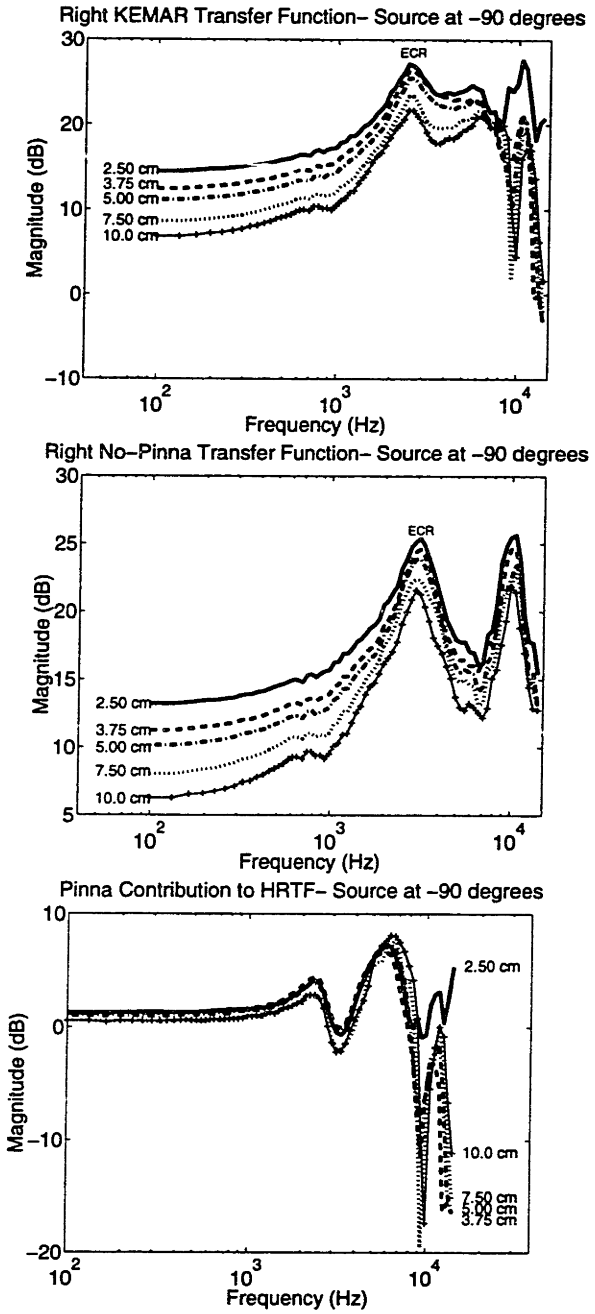


Figure 5-11: The contribution of the pinna to the HRTF for a nearby source at −90°. Measurements were made with the right ear of the KEMAR manikin both with the standard pinna attached (shown in the first panel) and with the pinna replaced by a flat rubber sheet with an opening at the ear canal (shown in the second panel). The third panel shows the ratio of the transfer function with the pinna to the transfer function without the pinna. All distances were measured from the surface of the head, rather than the center of the head. The label ECR represents the location of the ear canal resonance.

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 112

resonance when the pinnae is removed (label ECR on the graphs).

The bottom panel of Figure 5-11 shows the ratio of the HRTFs with and without the pinna, which represents the contribution of the pinna to the HRTF. The pinna contribution is roughly independent of distance at frequencies up to 8 kHz. The peak and notch near 3 kHz reflect the 500 Hz increase in the frequency of the ear-canal resonance when the pinna is removed. At higher frequencies the contribution of the pinna changes only modestly with distance, with the exception of the large increase in the high-frequency response of the pinnae as the source moves from 3.75 cm to 2.5 cm.

These results contrast with those of Shaw and Teranishi (1968). They measured the frequency response of an outer ear replica mounted in a rigid plate for a nearby point source, and found that the low-frequency gain provided by the outer ear increased with decreasing distance, and that the high-frequency gain of the ear decreased with decreasing distance. The low-frequency effect described by Shaw is not seen in our data, and the high-frequency effect is found only at distances greater than 2.5 cm.

Although data are only available for sources at 90° azimuth and 0° elevation, both Shaw and Teranishi'ss measurements and our measurements indicate that the response of the ear is roughly independent of distance at when the source is located more than 4 cm from the ear. Thus it appears that most of the distance dependent changes in the near-field HRTFs result from sound diffraction by the head and torso and the geometric orientation of the source relative to the ear, and not from changes in the acoustic behavior of the pinna in the near-field.

5.11 Comparison of Sphere Model and KEMAR measurements

The accuracy of the sphere model at low frequencies is displayed in Figure 5-12, which shows the distance dependences of the IID and ITD from the sphere model and from

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 113

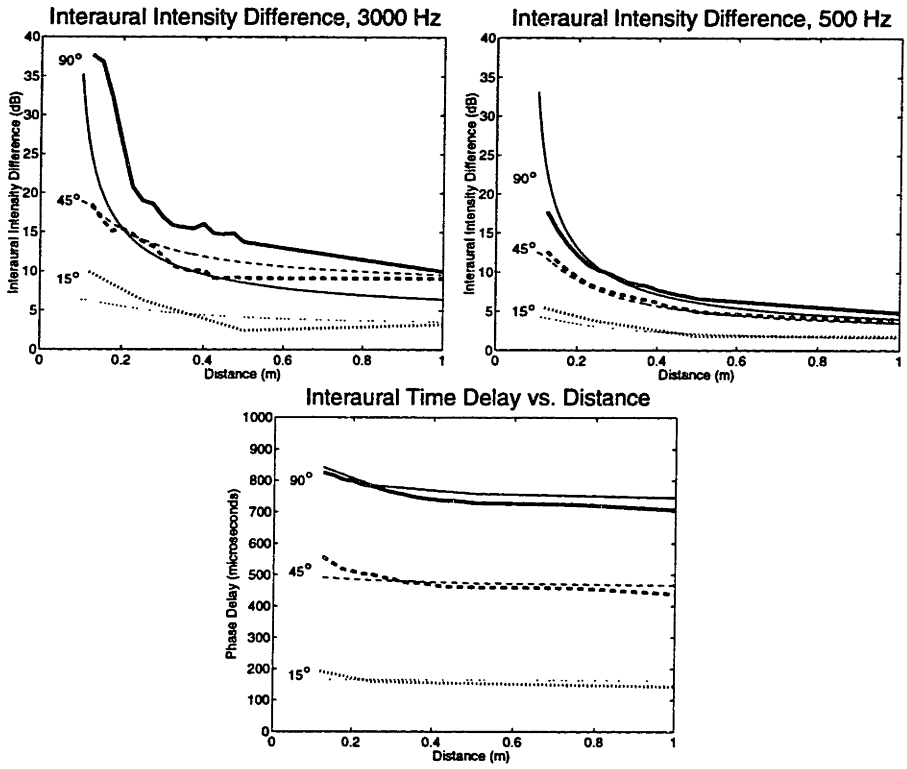


Figure 5-12: Comparison of Sphere Model and KEMAR Measurements. The bold lines are the KEMAR measurements, and the thin lines are the corresponding predictions by the sphere model. The KEMAR measurements at 45° and 90° are shown every 2.5 cm from 0.12 cm to 0.50 m, and at 1 m; the other data are shown only at 0.12 m, 0.25 m, 0.50 m and 1 m.

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 114

the KEMAR measurements. In general, the measured data fit the predictions of the sphere model well. The only exception is at 90° in the 3 kHz plot, where the IID is significantly larger in the KEMAR measurements. Two factors contribute to this result. First, the resonance of the ear canal is greater on the ipsilateral side than on the contralateral side, resulting in a net increase in the IID with the manikin. Second, the acoustic "bright spot", which causes an increase in pressure at the contralateral ear and decreases the IID, is less pronounced with the irregularly shaped head of the manikin than with a perfectly spherical head. Although the model is considerably less able to predict the actual HRTF at higher frequencies, the model is a valuable tool for predicting many of the features of the near-field HRTF.

5.12 Perceptual Implications of the Near-Field HRTFs

Figure 5-12 is also useful for analyzing the perceptual implications of the distance-dependent attributes of the HRTF in the near-field. The IIDs increase rapidly as distance decreases, especially at distances less than 0.5 m, while the ITDs increase only slightly at distances less than 1 m. The disparity between the distance dependence of the ITD and IID in the near-field is even larger when perceptual issues are considered. Hershkowitz and Durlach (1969) found that listeners could discriminate changes in IID on the order of 0.8 dB, so the changes in IID from 0.125 m to 1 m span a range of up to 15 JNDs at 500 Hz and 30 or more JNDs at 3 kHz. The JND for ITD was approximately 15 μs at ITDs below 400μs and increased rapidly for ITDs greater than 400μs, so the changes in ITD in the near-field span, at most, a few JNDs. Therefore subjects can be expected to perceive large changes in IID as the distance of a nearby source changes while the perceived ITD remains relatively constant.

The combination of perceptually invariant ITDs and strongly distance-dependent IIDs suggests a possible strategy for determining the distance of a nearby source in the horizontal plane based on binaural information from the HRTFs. The ITD information, which is relatively independent of distance, could be used to identify the azimuthal direction of the sound source. Once the source direction in azimuth is

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 115

known, the systematic dependence of IID on distance could be used to estimate the distance of a sound source, provided the source is outside the median plane (where the IID is near zero at all distances). This model of near-field binaural distance perception would predict the following characteristics in near-field distance estimation performance:

1. Distance accuracy would be greatest for sources on the left or right sides of the listener, and worst for sources directly in front or behind. Since the variation in IID with distance is greater for lateral sources than for sources in the median plane, listeners would have less resolution in distance perception near 0° and 180°.

2. The percentage JND in distance at a fixed azimuth would increase as distance increases. In Figure 5-12, the slope of the curve relating IID to distance increases substantially as the source approaches the head. Thus the percent decrease in distance necessary to produce a fixed increase in IID decreases as distance decreases. If the JND in distance is defined as the percent decrease in distance necessary to produce a single JND in interaural intensity, the distance JND will decrease with distance.

3. Provided the source is sufficiently broad-band to allow the listener to perceive both ITD and IID information, distance perception would not depend on the spectral shape or intensity of the source. The key advantage to binaural depth perception, in contrast to depth perception based on intensity, spectral cues, or reverberation, is that the binaural information is derived from the difference signal between the left and right ears and does not depend on the characteristics of the source (except its frequency range).

Note that this model is similar in concept to one suggested by Hirsch (1968), further explored by Greene (1968) and expanded by Molino (1970). This model demonstrated the possibility of determining the distance of a sound source based on the relationship between the IID and ITD. Hirsch's model, which ignored diffraction

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 116

by the head and assumed the ears were detectors in free space, predicted that distance could be calculated directly from the ratio of ITD to IID. Molino's expanded model, based on a spherical head, required that the azimuth location of the source to be known *a priori*. Greene and Molino used threshold data for ITD and IID to calculate the predicted accuracy of distance perception using this model and, predictably, found that distance perception in the far-field would be very poor. Molino noted that predicted accuracy would be much greater in the near-field, due to the dramatic increase in IIDs in that region. The present data indicate not only that the changes in IID in the near-field are easily perceptible, but also that the relative invariance of the ITD in the near-field may allow listeners to determine azimuth directly from the ITD without external knowledge about the direction of the source.

The situation becomes more complex outside the horizontal plane, but it is still possible to determine the azimuth, elevation, and distance of a sound source from the HRTF. Recall that the KEMAR measurements at 30° and −30° indicate that the high frequency, elevation-dependent features in the HRTF, which are believed to be important in localizing elevation, are roughly independent of distance. If these high-frequency cues could be used to determine the elevation of the source, and compensate for the elevation-dependent changes in the IID and ITD, the azimuth and distance of the source could still be accurately determined from interaural cues. This model would imply greater distance accuracy in the horizontal plane than the median plane, as IIDs decrease as a source moves directly above or below a listener.

Another possible strategy for determining the distance of a nearby source involves the disparity between the orientation of the source relative to the head and the orientation of the source relative to the ear. The ITDs, as well as the low to mid frequency IIDs, are determined primarily by the orientation of the source relative to the center of the head. However, the high-frequency features of the HRTF are largely a result of the geometric properties of the pinnae, and therefore are governed by the orientation of the source relative to the ear. The causes a compression in the spatial locations of high-frequency HRTF features in the near-field (Figure 5-7). If listeners were able to determine the azimuth of a sound source relative to the ear with high-frequency

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 117

spectral cues, and relative to the head with low-frequencies IIDs, they conceivably could use the two values to triangulate the distance of the source.

There is some evidence that listeners can use pinna based cues to determine the direction of a sound source. Studies examining monaural localization ability have shown that listeners have some ability to identify the location of a broadband sound when one ear is occluded by an ear-plug and muff (Wightman & Kistler, 1989b; Butler, Humanski, & Musicant, 1990; Oldfield & Parker, 1986) or congenitally impaired (Slattery & Middlebrooks, 1994). The apparent position in azimuth of monaural narrow-band stimuli is related to the direction-dependent gain of the pinna at the center-frequency of the signal (Rogers & Butler, 1992; Butler, 1987; Musicant & Butler, 1984; Butler et al., 1990; Belendiuk & Butler, 1977). Azimuth localization based on pinna cues does not, however, appear sufficiently accurate to allow accurate perception of distance via triangulation. Therefore, it is unlikely that subjects are able make distance judgments based on the geometric location of the source relative to the head and ear.

A final possible distance cue indicated by the HRTF measurements is a slight increase in the relative low-frequency gain as distance decreases. As distance decreases, the gain of the HRTF increases more at low-frequencies than at high frequencies at the ipsilateral ear, and the attenuation due to head shadowing increases more at high-frequencies than at low frequencies at the contralateral ear. As a result, the signal reaching the eardrums from a nearby source is effectively low-pass filtered as the source approaches the head (Figure 5-4). This low-pass filtering may explain the "darkening" of a very near sound source reported by Von Bekesy (1960). Bekesy observed that the particle velocity of a spherically radiating sound wave is increased relative to the pressure of the wave at distances less than a fraction of a wavelength from the source, and suggested that this might produce an increase in low frequency energy for a nearby source (since the low-frequency components of the sound are fewer wavelengths distant from the source than higher frequency components) (Coleman, 1963). Bekesy's subjects reported that integrated (effectively low-pass filtered) sound bursts were perceived closer than differentiated (high-pass filtered) stimuli. There is

EightKHz, LLC
Exhibit 2013
IPR2023-01024

no evidence, however, that the ear can directly perceive the velocity of a sound source, so Bekesy's explanation of this effect is suspect. Begault (1987) also noted the "darkening" in the timbre of very near sound sources, and suggested that the tendency to perceived larger increases in loudness at low frequencies than at higher frequencies from an equivalent increase in sound pressure level (the so-called Fletcher-Munson curve) might provide an explanation. Since the pressure level increases at all frequencies as the source approaches the ear, the Fletcher-Munson curve suggests that the perceived increase in loudness would be greater at low frequencies. The "darkening" of near-field stimuli reported in the literature is probably a combination of the boost in low-frequency gain due to near- field acoustic effects shown in the HRTFs and the non-uniform perception of increasing loudness across frequencies.

The implications of the near-field properties of the HRTF in directional localization are less obvious than those in distance localization. There is evidence that low-frequency time delays (Wightman & Kistler, 1992) dominate the perception of azimuth when they are available. Since time delays vary only slightly as distance decreases, it is not likely that azimuth perception will be significantly different in the near-field than in the far-field when the spectrum of the source extends into low frequencies. If time delay information is not available from low-frequency time delays or high-frequency envelope delays (Middlebrooks et al., 1989), it is possible that localization ability may degrade substantially. Without time delay information, there is no obvious mechanism for determining the relative contributions of distance and direction to the IID. In other words, there is no way to determine whether a certain IID is the result of a distant sound source near the interaural axis or a nearby sound source near the median plane. The consequences of this confusion on localization performance are unclear. The increase in IIDs at close distances will, however, decrease the JND in azimuth to the extent that the JND is limited by the change in IID. It is also possible that the change in IID with head orientation could provide a strong dynamic distance cue when exploratory head-motions are allowed. In addition, the spatial remapping that occurs in the high-frequency pinnae cues in the near-field may introduce a lateral bias if the auditory system is using these cues in the perception

Eight KHz, LLC
Exhibit 2013
IPR2023-01024
Page 119

of azimuth.

The high-frequency pinna cues that are believed to determine elevation were found to be roughly independent of distance over the limited range of elevations measured. These cues are believed to dominate the perception of elevation, so it is unlikely that the localization of elevation is strongly dependent on distance in the near-field.

5.13 Conclusions

The major conclusions of this study can be summarized as follows:

1. The dominant distant-dependent feature of both the sphere-model and KEMAR HRTFs in the near-field is an increase in the interaural intensity difference with decreasing source distance across all frequencies. When a source is near the interaural axis, the change in IID can span 15-30 JNDs as distance moves from 1 m to 0.12 m, providing a potentially strong binaural distance cue in the near-field. Note that significant IIDs at low frequencies occur exclusively in the near-field. In the far-field, the low-frequency IID is small at all source directions.

2. In contrast to the IID, the ITD is roughly independent of distance in the near-field.

3. Both the sphere model and the KEMAR measurements indicate that the average low-frequency gain of the HRTF increases more rapidly than the average high-frequency gain as the source approaches the head. Thus the sound reaching the ears is effectively low-pass filtered as the source approaches the head. This filtering may serve as a spectral distance cue in the near-field.

4. The HRTF at the contralateral ear is dominated by a complex interference pattern from sound propagating around the head by different paths. Up to about 2 kHz, this effect causes an increase in amplitude in the HRTF (a "bright spot") when the ear is located directly opposite the source. At higher frequencies, the interference pattern results in a complex series of ridges and notches which

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 120

change with frequency and azimuth. This interference pattern at the contralateral ear tends to dominate the detail of the IID at high frequencies.

5. The discrepancy between the orientation of the source relative to the ear and the orientation of the source relative to the head causes a remapping of the high-frequency azimuth cues at the ipsilateral ear. In general, the features of the transfer function tend to be shifted laterally as the source approaches the head.

6. HRTF measurements at three elevations and three distances indicate that the high-frequency features of the HRTF which vary systematically with elevation are not strongly dependent on distance. These data indicate that elevation localization may not be significantly different in the near- and far-fields.

In summary, it is clear that the distance-dependent attributes of the HRTF in the near-field provide possible binaural distance cues which are unavailable for more distant sources. The changes in azimuth and elevation cues are less dramatic, and their effect on near-field localization is more difficult to predict. The experiments described in the next two chapters measure localization accuracy in the near-field and determine whether listeners are able to make use of the available distance cues in that region.

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 121

Chapter 6

Localization of a Broadband Source in the Near-Field

Abstract

Although many researchers have examined auditory localization for relatively distant sound sources, little is known about the spatial perception of nearby sources. In the region within 1 m of a listeners head, which we define as the near-field, the interaural intensity difference increases dramatically as the source approaches the head, while the interaural time delay is roughly independent of distance. An experiment has been performed to evaluate near-field localization performance. In this experiment, an auditory point source was moved to a random position within 1 m of the subject's head, and the subject responded by pointing to the perceived location of the sound with an electromagnetic position sensor. The mean angular directional error (17°) was roughly comparable to previously measured results in far-field experiments. The error in azimuth increased slightly as distance decreased, but most of the degradation was in the region directly in front of the face. Elevation performance was not strongly dependent on source distance. Distance localization performance was generally better than has been reported in far-field experiments, and was strongly dependent on azimuth, with the stimulus-response correlation ranging from 0.85 to the side to less than 0.4 in the median plane. The results suggest that binaural cues are important to auditory distance perception in the near-field.

6.1 Introduction

Although human sound localization has been studied extensively in the past century, little is known about the spatial perception of nearby sources. The majority of ex-

Eight-KHZ, LLC
Exhibit 2013
IPR2023-01024
Page 122

periments examining directional sound localization have been conducted at distances greater than one meter. In this region, the overall amplitude of the sound reaching the ears varies with distance, but the binaural and spectral cues which are used for directional localization are roughly independent of distance. At distances less than one meter, however, there are important distance-dependent changes in the binaural and spectral characteristics of the sound reaching the ears. It is possible that these systematic changes allow listeners to make accurate judgments about source distance for nearby sources. Since nearly all of the perceptually relevant distance-dependent changes in auditory localization cues occur at distances less than 1 m, we will define this region as the psychoacoustic "near-field", and the region at distances greater than 1 m as the "far-field". This study will examine localization accuracy in the near-field in azimuth, elevation, and distance, and attempt to relate the findings to the near-field head-related transfer function.

6.2 Background

The most basic mechanisms of directional sound localization are well documented. In the horizontal plane, interaural difference cues have long been recognized as the dominant localization cues. Lord Rayleigh, in his famous "Duplex Theory" (1907) observed that interaural time differences (ITDs) and interaural intensity differences (IIDs) provide salient information about the lateral position of a sound source. According to the duplex theory, ITDs dominate low-frequency sound localization, while IIDs dominate high-frequency sound localization. The ITD and IID are important localization cues, but they cannot distinguish between sources at symmetric locations across the frontal plane without additional information provided by the geometry of the outer ear. The system of folds in the pinnae filters the sound reaching the ear with a directionally dependent transfer function at high frequencies (above approximately 4 kHz). When some *a priori* information about the spectrum of the source is available, pinna filtering allows listeners to resolve front-back confusions (Muscicant & Butler, 1984; Oldfield & Parker, 1986), and provides some information about the azimuth of

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 123

a sound source when binaural cues are eliminated by occluding one ear (Wightman & Kistler, 1987). Perhaps most importantly, pinnae cues allow listeners to judge the elevation of the sound sources (Roffler & Butler, 1968). All of the localization cues believed to be relevant to directional localization in the near-field are included in the head-related transfer function (HRTF), which is the transfer function from a sound source to the eardrums of the listener. The HRTF includes the effects of diffraction by the head, neck and torso, as well as the spectral shaping by the pinna.

The mechanisms that allow listeners to determine the distance of a sound source are less understood than those that allow directional localization. The most salient auditory distance cue under most conditions is the amplitude cue: the pressure of a spherically radiating sound wave is inversely proportional to the distance from the source. Spectral cues also play a role. Atmospheric absorption effectively low-pass filters sounds that propagate great distances, and low-frequency sounds propagate more effectively around obstacles in a room than high-frequency sounds. Both of these effects tend to cause more distant sound sources to appear low-pass filtered relative to closer sound sources, and may provide a spectral distance cue (Little, Mershon, & Cox, 1992). Amplitude and spectral based distance cues are sufficient for judging changes in the relative distance of a sound source, but can only be used to make absolute distance judgments when the listener has *a priori* knowledge about the characteristics of the source. The ratio of direct to reverberant energy has been proposed as a possible absolute distance cue for localization in rooms (Mershon & King, 1975; Lounsbury & Butler, 1979; Butler et al., 1980), and distance judgments in a reverberant environment are mildly correlated with source distance (Mershon & Bowers, 1979). Under free-field conditions with an unfamiliar source, distance perception is extremely inaccurate, and several researchers have reported that distance judgments in these conditions are effectively uncorrelated with the actual source position (Coleman, 1963; Mershon & Bowers, 1979; Holt & Thurlow, 1969; Gardner, 1969). A comprehensive review of far-field localization is provided in Middlebrooks and Green (1990).

One aspect of auditory localization that has received almost no attention is the

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 124

localization of sources close to the head. As early as 1911, Stewart recognized that interaural intensity differences increase significantly when a source approaches within a few cm of the head, while the interaural time delay is roughly independent of distance (Stewart, 1911a, 1911b). Stewart modeled the head as a rigid sphere with ears at diametrically opposed locations on its surface and used theoretical predictions of the sound pressure on the surface of a sphere to predict the IID and ITD as a function of source distance and direction. Hartley and Frey (1921) manually tabulated these values at a variety of locations, and Coleman (1963) cited the increased IIDs for nearby sources as a potential auditory distance cue in the near-field. Brungart and Rabinowitz (1996) have published a formula for evaluating near-field HRTFs using a sphere model and re-examined the possible use of near-field IIDs as a distance cue, and Duda and Martens (1997) have measured the range-dependence of the HRTF for a model of the head based on a bowling ball. Each of these studies found that interaural intensity differences increase dramatically when the source is near the head, while interaural time delays increase only slightly for nearby sources.

In the past year, near-field HRTFs have been measured with a KEMAR manikin and a compact, non-directional point source speaker (Chapter 5). Many of the features of the measured HRTFs were similar to those predicted by the rigid sphere models of Harley and Fry and of Brungart and Rabinowitz. The important aspects of the measured HRTFs can be summarized as follows:

1. The interaural intensity difference increases dramatically as the source approaches the head when the source is outside the median plane. This increase occurs even at low frequencies where head-shadowing is negligible in the far-field. At 500 Hz, for example, the IID increases from 4 dB to 19 dB at 500 Hz as a source at 90° decreases in distance from 1 m to 12 cm.

2. The interaural time delay is roughly independent of distance in the near-field. Although the time delay can increase by as much as 100 μs as the source approaches the head, this increase only occurs near the interaural axis, where the ITD is large and sensitivity to changes in the ITD is low.

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 125

3. The magnitude of the HRTF is relatively greater at low frequencies than at high frequencies when the source is near the head. This effective low-pass filtering of near-field sources results from a combination of diffraction at the ipsilateral ear and increased head-shadowing at the contralateral ear.

4. The high frequency features of the HRTF that are dependent on elevation are relatively insensitive to source distance. The features of the HRTF that changed significantly with elevation were not strongly dependent on source distance.

5. The azimuth-dependent high-frequency features of the HRTF appeared to be compressed around 90°. This is believed to be a result of the discrepancy between the location of the source relative to the pinnae and the location of the source relative to the center of the head — to the extent that these high-frequency features are a result of the direction-dependent properties of the pinnae, they are governed by the direction of the source relative to the ear rather than the direction of the source relative to the head.

These results indicate the existence of unique physical acoustic cues in the near-field which should be relevant to near-field localization. Yet, despite the recognition that localization cues are substantially different in the near- and far-fields, no studies in the literature have systematically measured near-field localization performance. The experiments which follow examine auditory localization in the near-field with a broadband source. In particular, it focuses on how localization accuracy changes as a function of azimuth, elevation, and distance in the near-field. The next section discusses the experimental setup. Directional and distance localization are discussed separately in the following two sections. The last two sections compare the results to the localization cues found in the HRTFs discussed in Chapter 5, and summarize the overall conclusions of this experiment.

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 126

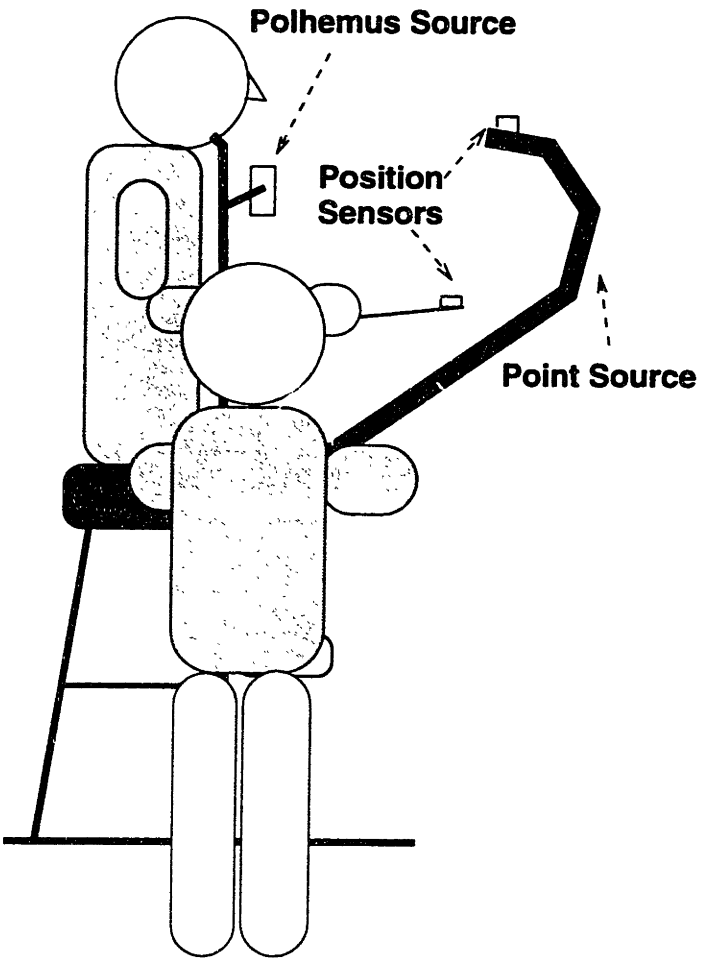


Figure 6-1: Experimental setup

6.3 Experimental Setup

6.3.1 Subjects

Four right-handed male subjects, ages 20-25, participated in the experiment. Three of the subjects were paid volunteers, and the fourth was the author. All reported normal hearing in both ears. Although three of the subjects had participated in psychoacoustic studies before, only the author had participated in localization experiments.

6.3.2 Apparatus

A simple diagram of the setup for the experiment is provided in Figure 6-1. The

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 127

experiments were conducted in MIT's anechoic chamber. The subjects were seated on a wooden stool located in the center of the chamber, which was supported on the wire-frame floor by a foam-covered plywood platform. The subjects were provided with a chin-rest which allowed them to immobilize their heads in a comfortable rest position during the experiments.

An experimenter, who stood approximately 1.5 m to the right of the subject, manually placed the sound source during each trial. The sound source consisted of an Electro-Voice DH1506 compression horn driver connected to 4 m of tubing with an internal diameter of 1.2 cm (see Appendix C). The end of the tube was enclosed in a rigid wand, constructed of PVC pipe. Three 45° elbow joints gave the rigid sleeve a curved shape which allowed the stationary experimenter to place the source at any location in the right hemisphere of the subject with the opening of the source pointing towards the subject's head. The small internal diameter of the tubing provided a relatively non-directional sound source at the opening of the tube: the measured 3 dB beam-width of the source was approximately 120° at 15 kHz.

A Polhemus Navigation 3-Space Tracker position sensing system measured the stimulus and response location during each trial. The electromagnetic source of the tracking system was mounted on the chin-rest approximately 15 cm below and 15 cm to the left of the subject. Although the chin-rest was not completely rigid, this arrangement fixed the relative positions of the Polhemus source and the subject's head and, therefore the coordinate system of the experiment was stable relative to the subject's head. One of the position sensors was mounted on the end of a 30 cm wooden rod, which the subject used to make responses. The second position sensor was mounted on the end of the experimenter's wand nearest the opening of the tube. Since it was impossible to place the sensor directly at the opening of the tube without interfering with the sound field, the orientation of the sensor and the offset between the sensor and the tube opening were used to calculate the location of the sound source on each trial. The Polhemus system is accurate within 0.25 cm in the X, Y, and Z coordinates up to approximately 1 m. In order to measure the effect of the correction on the accuracy of the location recording system, the response sensor was

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 128

placed directly at the tube opening and the location of each sensor was measured by the Polhemus system. These two measurements of location differed by 2-3 cm, which can be considered an upper bound on the vector error of the system.

The control computer was a 386-based PC equipped with a Digital Audio Labs CARDD 16-bit stereo sound card. One channel of the sound card was connected to a small ear-piece headphone worn by the experimenter. This channel was used to read source location coordinates to the experimenter on each trial. The other channel was connected to a Crown power amplifier, which was connected to the driver of the sound source. The Polhemus head-tracker was connected to the PC through the RS-232 serial port, and a response switch was connected through the parallel port. The control computer automated all the data recording and stimulus generation tasks in the experiment, and provided timing information to the subject and operator through its internal speaker.

6.3.3 Stimulus

The stimuli were sequences of five rectangularly gated 150 ms pulses of noise, separated by 30 ms intervals of silence. The noise waveforms were constructed from white Gaussian noise that was filtered by an FIR filter to flatten the irregular frequency response of the point source. In addition, the noise was band-limited to the frequency range 200 Hz - 15 kHz (120 dB/decade roll-off out of band) and low-pass filtered with a 6 dB/octave roll-off above 200 Hz. This roll-off was used to maximize the non-distorted output level of the point source. Five different noise waveforms were stored on the control computer, and one waveform was randomly chosen prior to each trial. This waveform was scaled, and then repeated five times to generate the stimuli for each trial.

The source was randomly located in the right hemisphere of the subject. Prior to each trial, the control computer read three random numbers, each ranging from one to six, to the experimenter through an earphone. The experimenter used these three numbers to choose the approximate sound source location in azimuth (from near 0° for a 1 to near 180° for a 6), elevation (from near +90° for a 1 to near −90° for a 6), and

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 129

distance (from 10-15 cm for a 1 to 1 m for a 6). Although the exact placement of the source varies across experimenters and some source locations were inaccessible due to interference by the subject's body or the chin-rest apparatus, this source placement system generated a reasonably uniform distribution of source locations throughout the right hemisphere.

Once the source was placed, the control computer recorded the location of the source through the Polhemus tracker, and crudely normalized the amplitude of the stimulus signal to eliminate amplitude-based distance cues. The normalization was based on the distance of the source from the left and right ears of the subject. The correction normalized the amplitude so the maximum output would occur at a distance of one meter. The scaling factor for this correction was

$$\frac{1}{\frac{50}{\textit{Distance to left ear (cm)}} + \frac{50}{\textit{Distance to right ear (cm)}}}.$$

The distance to the right ear dominates the scaling factor when the source is near the ear, but the scaling factor also considers the contribution of the left ear to perceived loudness when the source is in the median plane or is relatively distant. In addition to correction for distance, the source amplitude was randomized an additional 15 dB (from 0 dB to 15 dB in 1 dB steps). The amplitude scaling was accomplished by multiplying the noise waveform file by a scaling factor prior to playback by the CARDD. The maximum amplitude of the stimulus was approximately 59 dBA SPL (as measured by a B&K 4131 microphone) at 1 m, so with randomization and correction the effective stimulus amplitude ranged from 44 dBA- 59 dBA.

6.3.4 Procedure

The experiment was divided into blocks of 100 trials, with each block taking approximately 20 minutes. At the beginning of each trial, the subject placed his head in a comfortable position in the chin-rest and the locations of three reference points were recorded using the response sensor: the opening of the left ear canal, the opening of the right ear canal, and the tip of the nose. These locations were used to correct for

Exhibit 2013
IPR2023-01024
Page 130

stimulus distance and to define a spherical coordinate system based on the subject's head (see Appendix A).

Each trial was initiated when the control computer read the three source coordinates to the operator through the ear-piece headphone. A beep then instructed the subject to close his eyes while the operator moved the source to the appropriate location. Once the source was positioned, the operator pressed a response switch and the control computer initiated the stimulus. First the location of the source was read to allow for amplitude correction, then the stimulus was scaled and played through the sound source, and finally the source position was read again to verify that no movement had occurred during the stimulus presentation. If the source was stationary during stimulus presentation, the operator moved the source to a rest position and pressed the response switch again. The control computer then generated a second beep prompting the subject to move the response sensor to the perceived location of the stimulus. The subjects were permitted to open their eyes during the response process, but usually chose not to do so. Once the subject had selected a response location, the operator once again pressed the response switch, and the control computer read the response location, generated three new coordinates for the next stimulus location, and beeped to tell the subject to close his eyes and prepare for the next stimulus. Each trial lasted approximately 12 seconds.

The response method used in the experiment, which we refer to as "direct location," was the method determined to be the least biased and most accurate among a number of three-dimensional near-field response methods considered in an earlier study (Chapter 4). Using the direct location method to identify the position of a visual target in the front hemisphere, the mean angular error was 4°. The subjects were also equally accurate at localizing sound sources in the front and rear hemispheres using direct location, indicating that precision does not fall off rapidly outside the visual field. The localization errors with a visual target using direct location were much smaller than those found when localizing sound sources, indicating that the response method probably contributed only a small fraction of the response errors in this experiment.

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 131

Although the subjects were asked to keep their eyes closed during the placement of the source, there were some extraneous cues (shadows visible through the closed eyelids, sounds generated by the experimenter, air movement during source placement, etc.) that may have allowed subjects to make judgments about the source location independent of the available audio information. In order to verify the insignificance of these cues, 100 trials were collected for each subject with the sound source disabled. The mean angular error in this condition was more than 50°, three times as large as when the sound source was enabled. The errors in azimuth, elevation, and distance were also much larger than in the audio experiment. Thus, although subjects received some information about source location from extraneous cues, this information was insignificant compared to the information provided by the intended auditory stimulus.

The data collection was divided into two-hour sessions, each consisting of four or five 100-trial blocks separated by short breaks. A total of 2000 trials per subject were collected over four or five two-hour sessions. Subjects participated in several training sessions prior to formal data collection in order to familiarize them with the experimental procedure. They were not, however, given feedback during these practice sessions.

6.4 Directional Localization Results

6.4.1 Data Analysis

Determining directional accuracy is challenging in a three-dimensional spherical coordinate system. The angular error, the azimuth and elevation errors, and the judgment centroid and dispersion factor are possible methods for characterizing the accuracy of localization judgments. The angular error, which is simply the angle separating the vectors from the origin to the stimulus and response locations, is useful for determining the magnitude of the error and is consistent across all locations. Unfortunately, it cannot distinguish between systematic response biases and response variability. If a subject's responses were consistently within 1° of a location 5° to the left of the stimu-

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 132

lus location, the mean angular error would be the same as if a subject's responses were uniformly distributed over the region within 10° of the stimulus location. Thus the angular error combines the bias and variability of the subject responses, and cannot be used to determine the magnitude and direction of systematic response biases.

A second possible strategy is independent analysis of azimuth and elevation. The mean of the signed azimuth and elevation errors can be used to estimate directional bias, and the standard deviations are reasonable estimates of response variability. These measures are also problematic, however, because they ignore any systematic interactions between azimuth and elevation, and because azimuth is by definition more sensitive at high and low elevations than in the horizontal plane.

A third method, proposed by Wightman and Kistler (1989a), is to determine the directional bias from the judgment centroid and the dispersion factor κ^{-1}. This system has a number of advantages, but can only be used when multiple data points are available at a single stimulus location. In this experiment the stimulus locations are continuously distributed, and the centroid and dispersion measures are not applicable.

We analyzed directional responses both in terms of the angular error and in terms of azimuth and elevation. The angular error was used only to provide a general overview of performance, while the azimuth and elevation data were used to determine response bias and response variability as a function of source location. These measurements were complicated by the continuous distribution of stimulus locations in azimuth, elevation, and distance. Response biases change as a function of source location, so it is not appropriate simply to average the signed response error across all source locations. The standard deviation of this mean signed error would include the effects of both bias and response variability. In order to isolate response variability, the *bias-corrected* unsigned error was used to measure the spread of responses around the mean response at each location. The bias-corrected unsigned error measures the absolute difference between the response location and the quadratic regression curve representing the best second-order fit of the stimulus locations to the response locations (Figure 6-6). The motivation for this method of analysis will be clearer once the data are presented, so we defer further discussion of this technique until later.

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 133

The random distribution of source locations required sorting the data into bins according to location in order to examine the relationship between source location and localization accuracy. This sorting was done in two ways. In most of the presentations of raw data, the locations were sorted into unequal sized bins according to specific cutoff points. An example would be a bin containing all data points at source elevations above 20°. In most of the analyses, however, equal sized bins were desired to allow direct comparison of statistics across bins. In this case, the data were sorted into n equal sized bins, sometimes overlapping, and the "location" of the bin was defined as the mean location of the points inside the bin. Thus, in the sorting of data it is possible either to have equally spaced locations with unequally sized bins or equally sized bins with unequally spaced locations.

Front-back reversals occurred in about 10% of the trials. A front-back reversal was defined as a response which was at least 10° closer to the mirror image of the source location across the frontal plane than to the actual source location. Note that this definition is more conservative than the more traditional definition of front-back reversals, which includes all trials where the stimulus and response locations are on opposite sides of the frontal plane. This definition helps to reduce the number of reversals near $-90°$, where it is difficult to distinguish between localization error and true perceptual front-back reversals. Trials in which front-back reversals occurred were eliminated from analyses of directional accuracy, but were included in analyses of distance performance.

6.4.2 Angular Error

The angular error, which includes the effects of both bias and response variability, varied significantly with source location (Table 6.1). The error was significantly larger behind the listener (azimuth $< -120°$) than in front or to the side (one-tailed t-test, $\alpha = 0.005$). Angular error was most sensitive to distance in front of the listener, where it was largest for close sources. The worst errors occurred at locations behind and above the head.

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 134

Table 6.1: Mean angular errors. The data at each location is the average across all four subjects, and the standard deviation (in parentheses) is the average of the standard deviation values calculated separately for each subject. Trials where front-back confusions occurred have been excluded from these calculations.

		Azimuth					
Elevation	Distance	$< -120°$		$-120°$ to $-60°$		$> -60°$	
$> 20°$	< 25 cm	25.4°	(14.6°)	14.9°	(8.0°)	19.8°	(9.3°)
	25 cm - 50 cm	21.3°	(11.9°)	15.5°	(7.3°)	11.8°	(7.5°)
	> 50 cm	22.2°	(11.1°)	13.0°	(8.6°)	12.5°	(6.9°)
$-20°$ to $20°$	< 25 cm	19.5°	(8.3°)	15.1°	(7.5°)	19.5°	(9.8°)
	25 cm - 50 cm	18.9°	(10.6°)	13.4°	(6.2°)	14.1°	(6.9°)
	> 50 cm	19.2°	(11.6°)	12.8°	(6.0°)	13.6°	(6.6°)
$< -20°$	< 25 cm	19.1°	(7.7°)	14.0°	(7.3°)	19.5°	(10.0°)
	25 cm - 50 cm	16.6°	(8.4°)	11.4°	(5.5°)	17.8°	(8.8°)
	> 50 cm	15.3°	(8.8°)	12.6°	(7.0°)	15.5°	(8.2°)

6.4.3 Azimuth Error

The raw data give an indication of both the accuracy of azimuth localization and of any major response biases (Figures 6-2 to 6-5). Several observations about individual subject performance can be made from these results. Most notably, there is a tendency for the subjects to respond to the left of the actual source location when the source is directly in front. This bias is seen in subjects KMY, DTD, and CLL at middle and low elevations in the form of an increase in the slope of the responses at about $-20°$ in azimuth. Overall, subjects DSB, KMY, and DTD performed comparably, but subject CLL experienced front-back confusions more frequently than the other subjects.

A more direct comparison of azimuthal response variability is provided in Figure 6-7. This figure shows the bias-corrected unsigned error in azimuth as a function of azimuth, elevation, and distance for each subject. The data in this figure represent the mean bias-corrected error in ten overlapping bins representing different locations in azimuth. The definition of the bias-corrected error for a single bin is illustrated in Figure 6-6. Note from Figures 6-2 to 6-5 that the raw azimuth responses are

Facebook, Inc.
Exhibit 2013
IPR2023-01024
Page 135

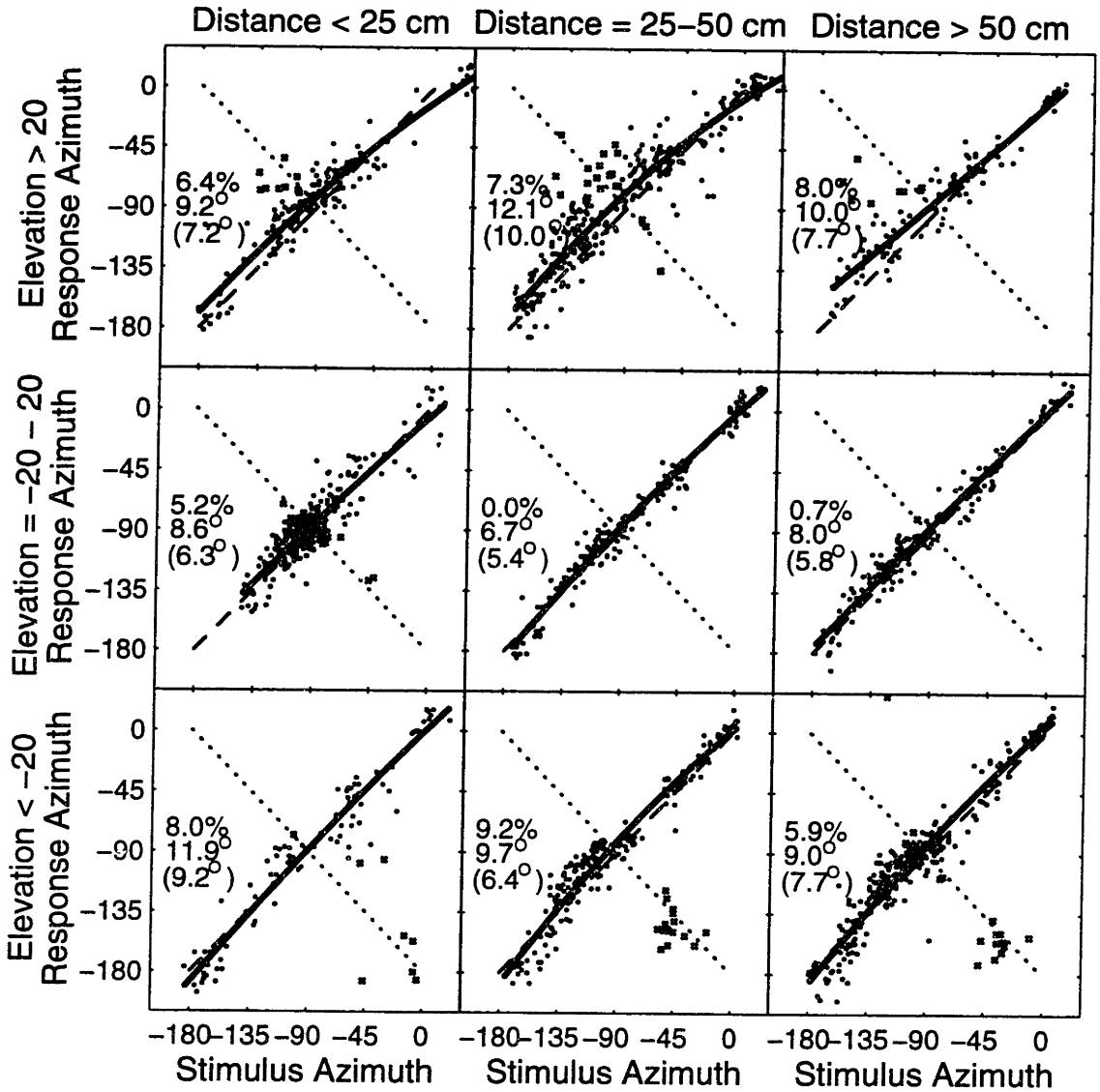


Figure 6-2: Raw Azimuth Data for Subject DSB. The nine panels show raw azimuth stimulus and response data for subject DSB. The data were sorted according to source distance into three regions: closer than 25 cm; from 25-50 cm; and further than 50 cm. Similarly, the data were sorted by elevation into three regions: above 20°; between −20° and 20°; and below −20°. Columns represent distance, increasing left to right, and rows represent elevation. The dots represent normal data, while the x's represent front-back confusions (see text). The solid line is the second-order polynomial function of the stimulus location that best fits the response location. At the left of each panel are three numbers. The top number is the percentage of front back reversals in the region. The middle number is the mean bias-adjusted unsigned error. The mean standard deviation of the unsigned error is shown in parentheses. The dashed line represents "correct" responses, while the dotted line represents "perfect" reversals.

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 136

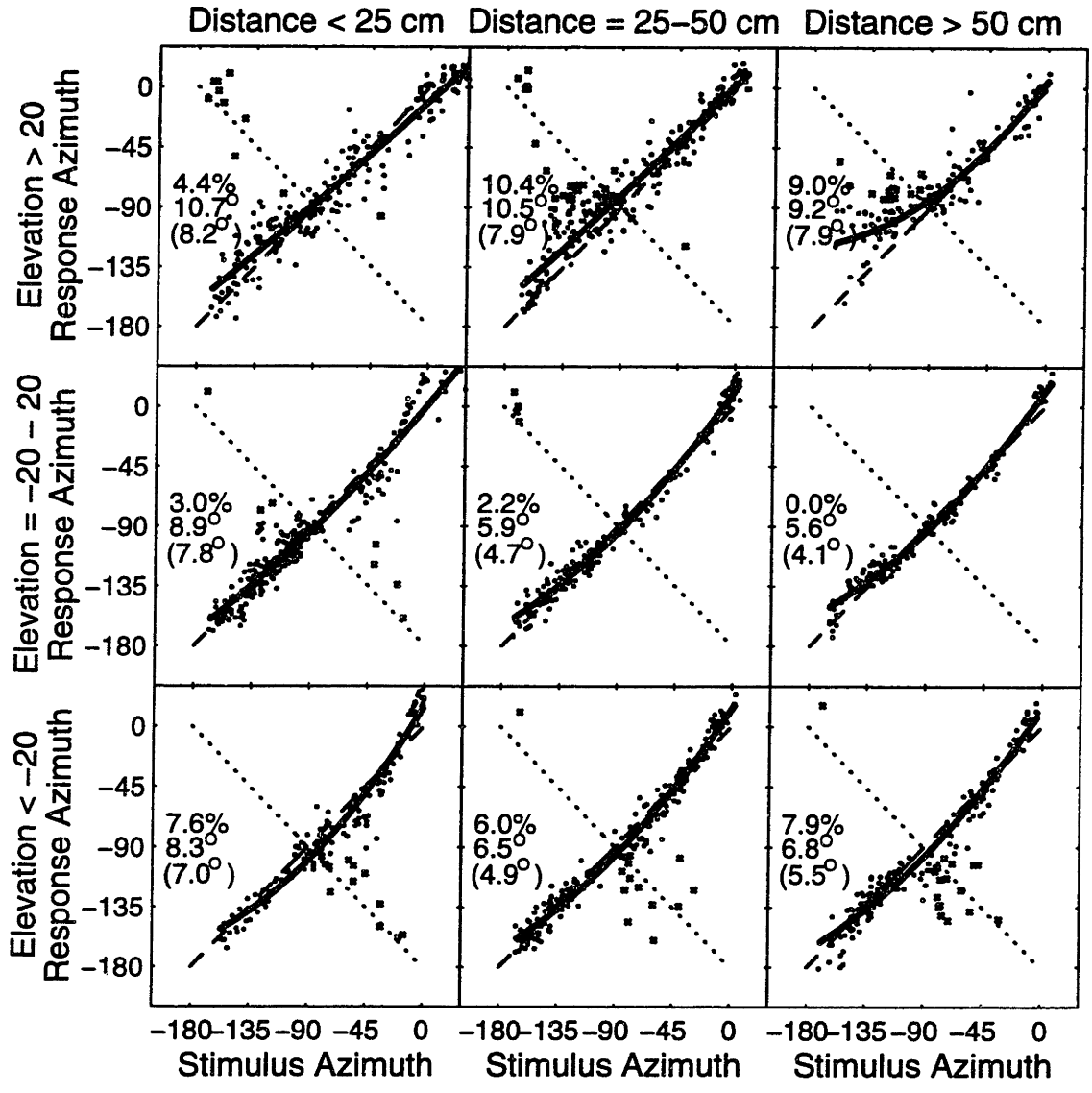


Figure 6-3: Raw Azimuth Data for Subject KMY. Format is as described in Figure 6-2.

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 137

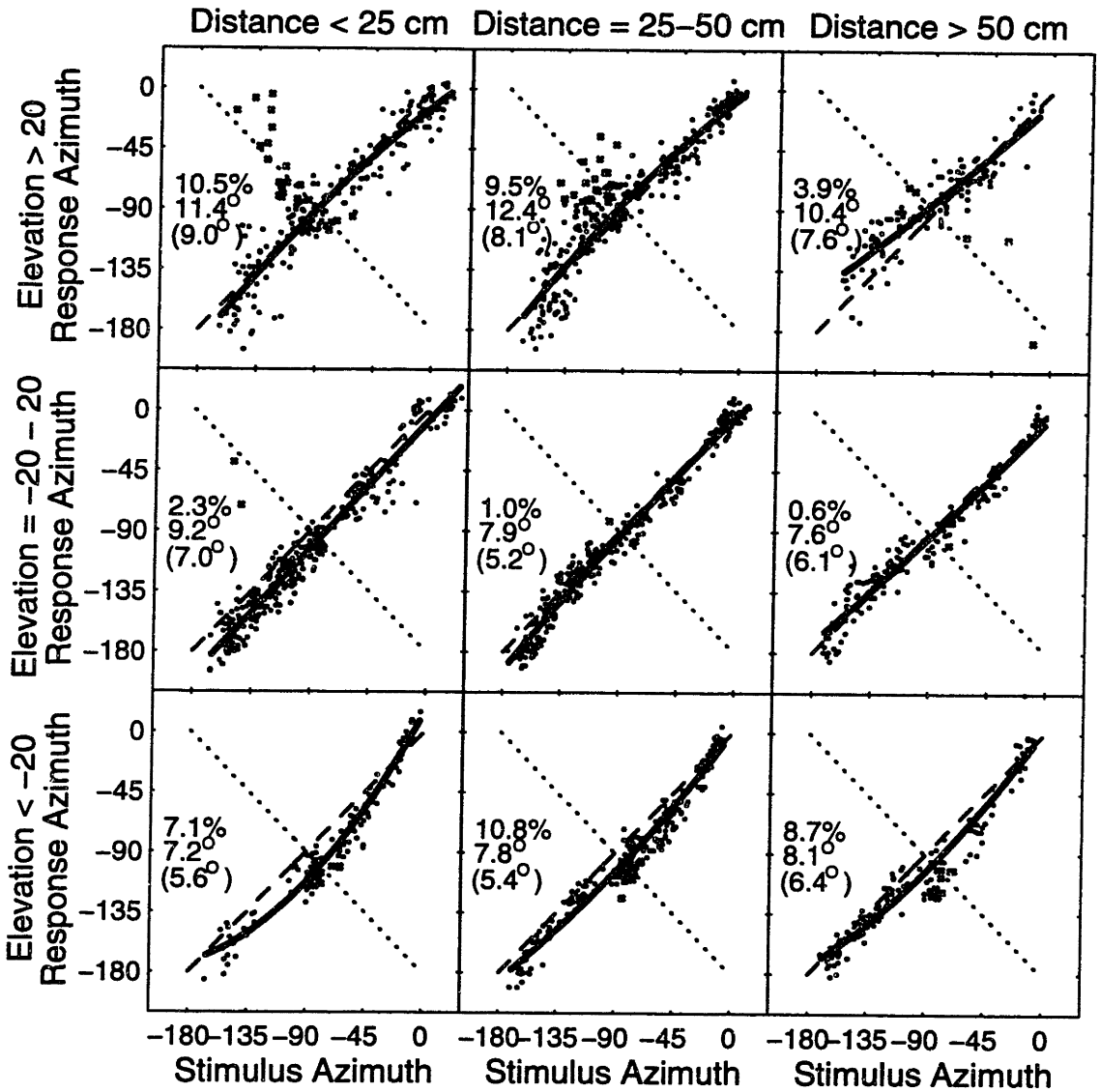


Figure 6-4: Raw Azimuth Data for Subject DTD. Format is as described in Figure 6-2.

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 138

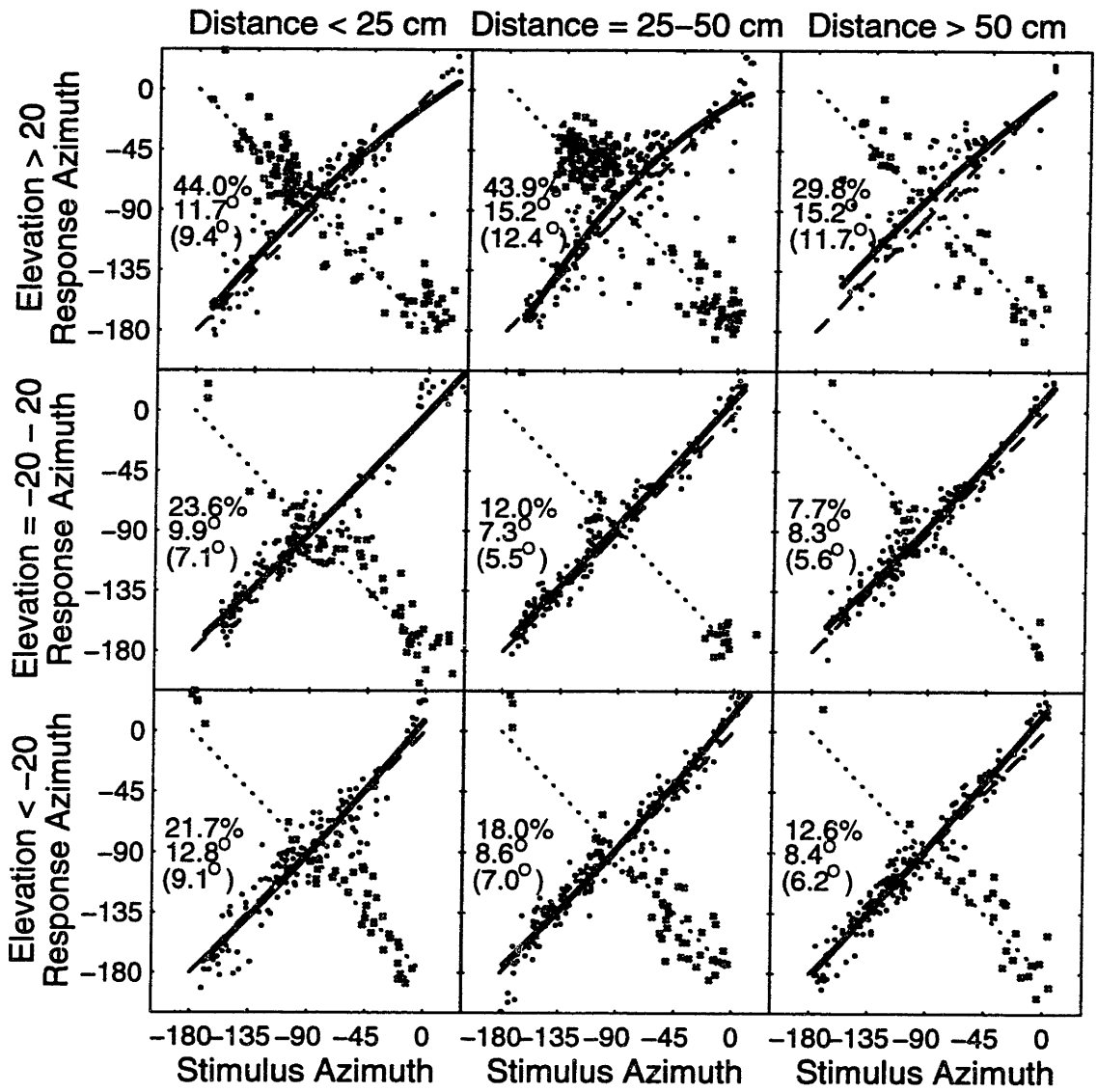


Figure 6-5: Raw Azimuth Data for Subject CLL. Format is as described in Figure 6-2.

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 139

not generally linearly related to the stimulus azimuth, but rather they follow a slight curve. For this reason, the bias-corrected error was defined as the mean unsigned error between each response location and the quadratic curve best fitting the stimulus data to the response data. The mean bias-corrected unsigned error can be viewed as a measure of deviation from the mean, similar to the standard deviation of the signed azimuth and elevation errors.

- The variation in performance across subjects was minimal. The bias-corrected azimuth errors of all four subjects typically fell within a narrow range and exhibit similar dependencies on azimuth, especially at middle elevations. Subject CLL exhibits some unusually large errors for lateral sources at high elevations, but the general trends in performance were consistent across subjects.

- The responses were most variable at high elevations. The mean error was lower at low and middle elevations than at high elevations for each of the four subjects (One-tailed t-tests, $\alpha = 0.005$). Note that this increase in error is at least in part a result of the increased sensitivity of azimuth at high elevations. The same effect is not as pronounced at low elevations because more trials were collected at very high elevations ($> 45°$) than at very low elevations ($< -45°$).

- Response variability was surprisingly large when the stimulus was located directly in front of the listener at distances less than 25 cm. All four subjects showed a substantial increase in response variability in this region as the source distance decreased below 25 cm, as shown by the upward sloping curve in the left-middle panel of Figure 6-7. The subjects apparently could not reliably locate sound sources in the region directly in front of their face.

- Other than the anomaly in front of the face, the bias-corrected azimuth error was not strongly dependent on distance. Although two of the subjects (KMY and CLL) showed significantly larger errors at distances less than 25 cm than at distances greater than 50 cm (one-tailed t-tests, $\alpha = 0.02$), almost all of this increase at close distances was a result of the larger errors directly in front of the face.

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 140

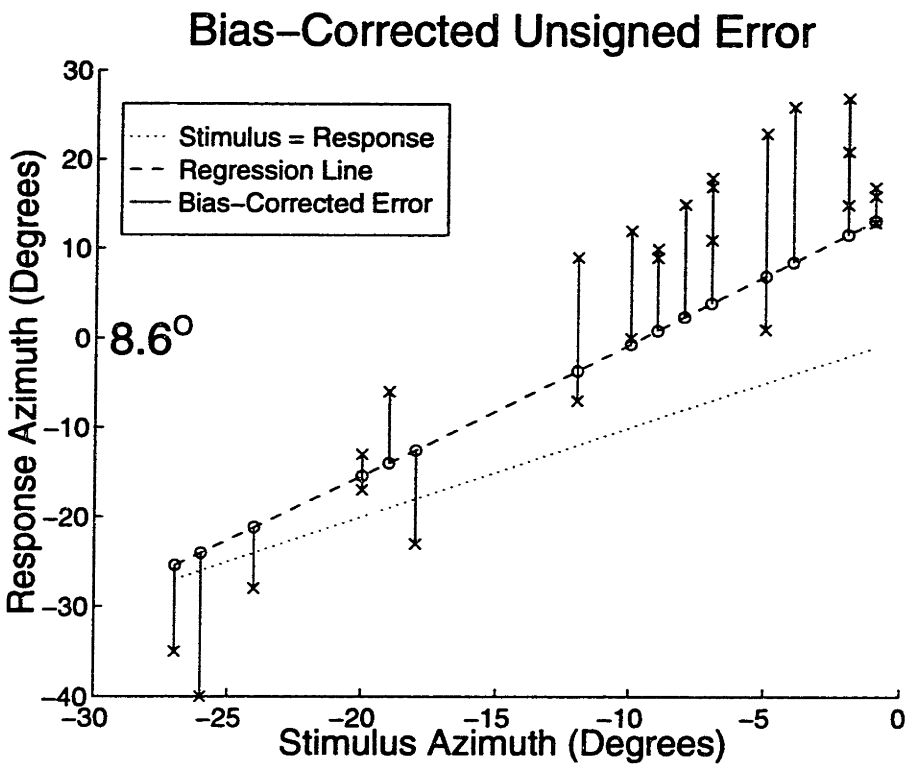


Figure 6-6: Calculation of bias-corrected unsigned error. First the data were sorted into overlapping bins representing a particular region of space. In this example, the data are from subject KMY and represent elevations less than −20° and distances less than 25 cm. Then the quadratic regression line curve fitting the response azimuth to the stimulus azimuth was calculated. Next, the data were sorted into overlapping bins by azimuth: the data shown represent the 18% of the trials closest to 0° in azimuth. Finally, the bias-corrected unsigned error was calculated from the mean distance from each response location to the regression line (8.6°). Note that this figure is limited to a single bin in azimuth, and that in this small region the quadratic regression curve resembles a line.

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 141

6.4.4 Front-Back Reversals

In this experiment, front-back reversals occurred in approximately 10% of all trials. In the raw azimuth plots (Figures 6-2 to 6-5), each 'x' represents a data point where a front-back reversal occurred. Recall that reversals were declared when the azimuth error was reduced at least 10° by reflecting the response across the frontal plane. The top number on the left side of each panel tells the percentage of trials where front-back reversals occurred in each panel. Figure 6-8 summarizes the relationship between the percentage of front-back reversals and the source location. Four important observations can be made from the reversal data.

1. Subject CLL has far more reversals than any other subject, and dominates the mean reversal percentages across subjects. CLL reverses 35% of the trials in the closest distance bin, five times rate of the second highest subject. In certain locations, CLL reverses the majority of trials. In the left middle panel of Figure 6-5, for example, virtually every source placement from $-45°$ to $0°$ in azimuth resulted in a front-back reversal.

2. Only CLL shows a significant distance dependence in the percentage of front-back reversals. CLL reversed a significantly larger percentage of trials at close and medium distances (< 50 cm) than at far distances (> 50 cm) (one-tailed t-test, $\alpha = 0.005$).

3. Relatively few reversals occur at middle elevations. The majority of reversals occur above and below the horizontal plane (although CLL has many reversals at middle elevations in the front hemisphere). One-tailed t-tests ($\alpha = 0.005$) indicate that all four subjects reversed a significantly larger percentage of trials at high elevations ($> 20°$) than at middle elevations, and that three subjects (DSB, KMY, and DTD) reversed a significantly larger percentage of trials at low elevations $< -20°$ than at middle elevations. Only subject CLL reversed a significantly larger percentage of trials at high elevations than at low elevations.

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 142

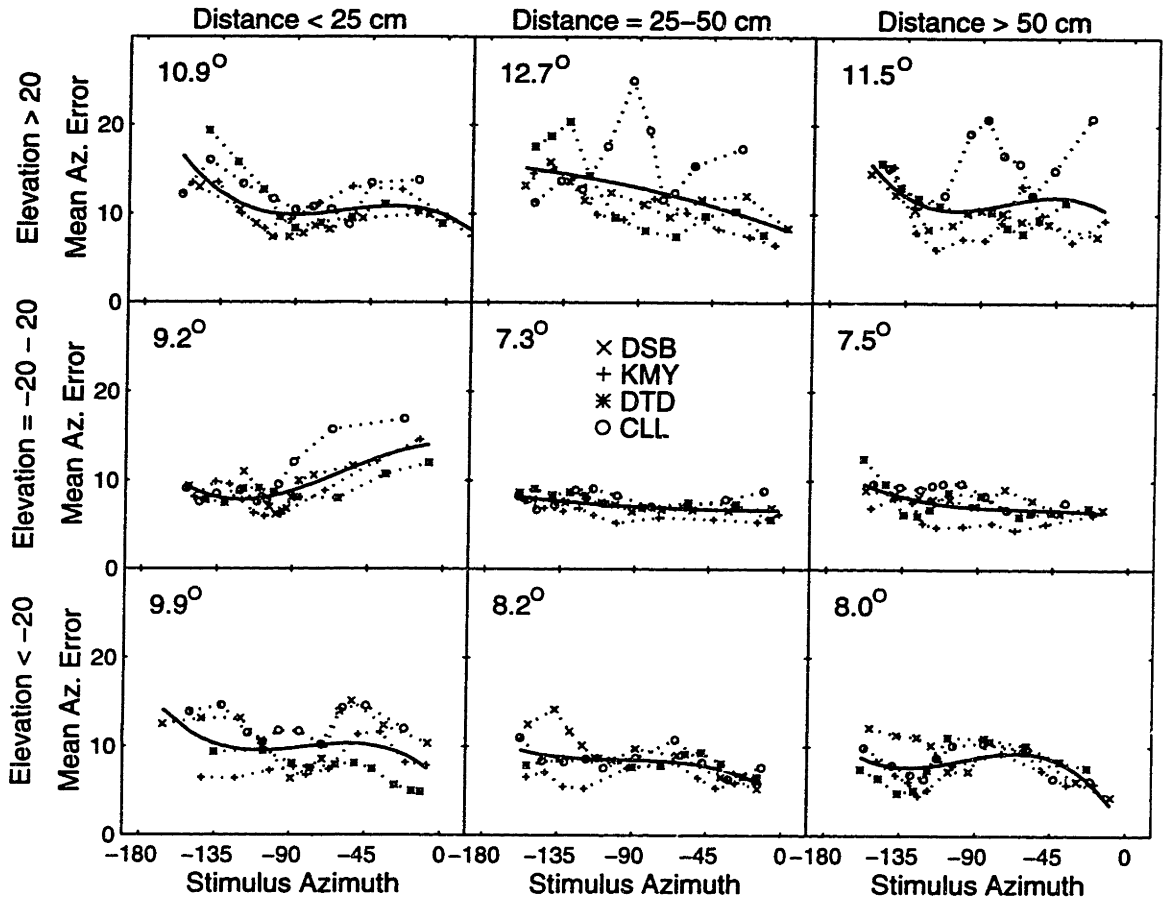


Figure 6-7: Directional dependence of the bias-corrected unsigned error in azimuth. The nine panels show the dependence of the mean unsigned error in azimuth on the azimuth location for different locations in space. The columns represent different distances, with increasing distances from left to right. The rows represent different elevations, with high elevations in the top row and low elevations in the bottom row. In each of the nine elevation and distance regions, the non-reversed azimuth responses were sorted by azimuth into 10 overlapping bins, each containing 18% of the total trials. Each symbol shows the mean azimuth of each bin (abscissa) and the mean bias-corrected unsigned error within each bin (ordinate) for one subject. The solid line is the third-order polynomial best fitting the data of all four subjects. The mean unsigned error across all four subjects is shown in the upper-left corner of each panel.

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 143

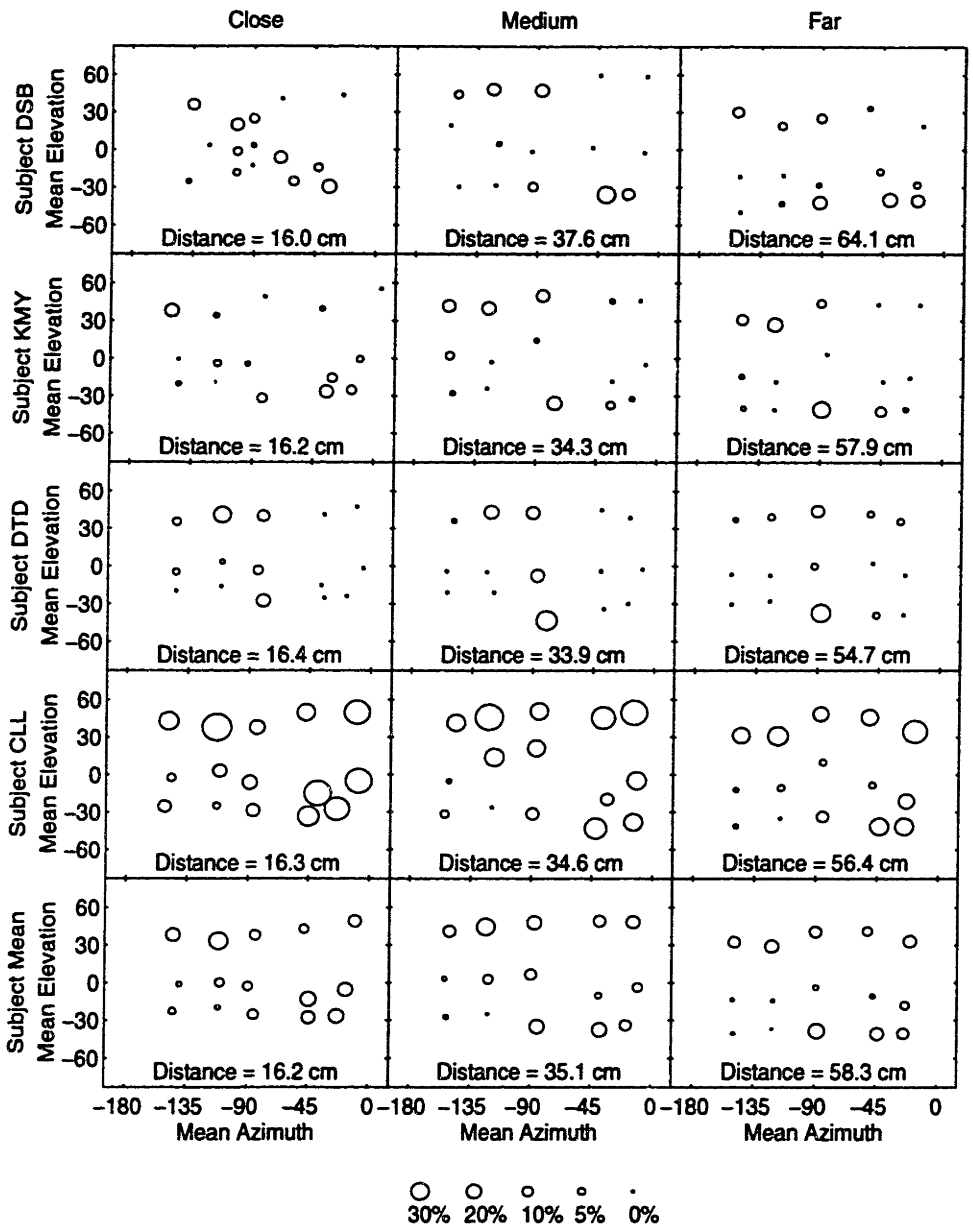


Figure 6-8: Spatial distribution of front-back reversals. First, the data were sorted into three non-overlapping distance bins, then each distance bin was sorted into eight overlapping azimuth bins, and finally each azimuth bin was sorted into three non-overlapping elevation bins. The number of reversals in each bin is shown as a function of mean location for each individual subject and averaged across all four subjects, where a reversal is defined as any trial where the response was at least 10° closer to the mirror image of the source location across the frontal plane than to the actual source location (see text). For clarity, only five bins are shown in azimuth. The three bins to the side and the rear are non-overlapping, while the two bins near 0° are overlapping. The percentage reversals at each location are shown by the size of the circle, according the the code shown at the bottom of the figure.

Exhibit 2013
IPR2023-01024
Page 144

4. In the rear hemisphere, the vast majority of reversals occur at high elevations. Very few sources placed behind and below the subject result in reversals. In contrast, almost all reversals in the front hemisphere for subjects DSB and KMY occur at low elevations. The variations in the placement of reversals across subjects is not surprising, since the subjects must essentially make an arbitrary decision about the true location of the source whenever they are unsure about the actual hemisphere of the source.

6.4.5 Elevation Error

The raw elevation data are shown in Figures 6-9 to 6-12. These results differ from the raw azimuth data both in the lack of any reversals and the relatively greater spread of responses. Figure 6-13 shows the dependence of the bias-corrected unsigned error in elevation on the source location.

1. Elevation localization performance was best to the side, and worst to the rear. A one-tailed t-test on the bias-corrected, unsigned errors reveals that each of the four subjects was significantly more accurate in front than in back, and most accurate to the side ($\alpha = 0.01$ level).

2. The error was particularly large for subject CLL when the source was above, behind, and close. The mean error in this region exceeds 25° and is by far the largest error at any location by any subject.

3. Elevation performance did not depend on distance in a consistent way. Two of the subjects (DSB and KMY) had significantly lower errors at distances less than 25 cm than at distances greater than 25 cm, and the other two subjects had significantly lower errors at distances greater than 25 cm than at distances closer than 25 cm (one-tailed t-tests on bias-corrected, unsigned elevation error, $\alpha = 0.05$).

6.4.6 Comparison of Azimuth and Elevation Errors

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 145

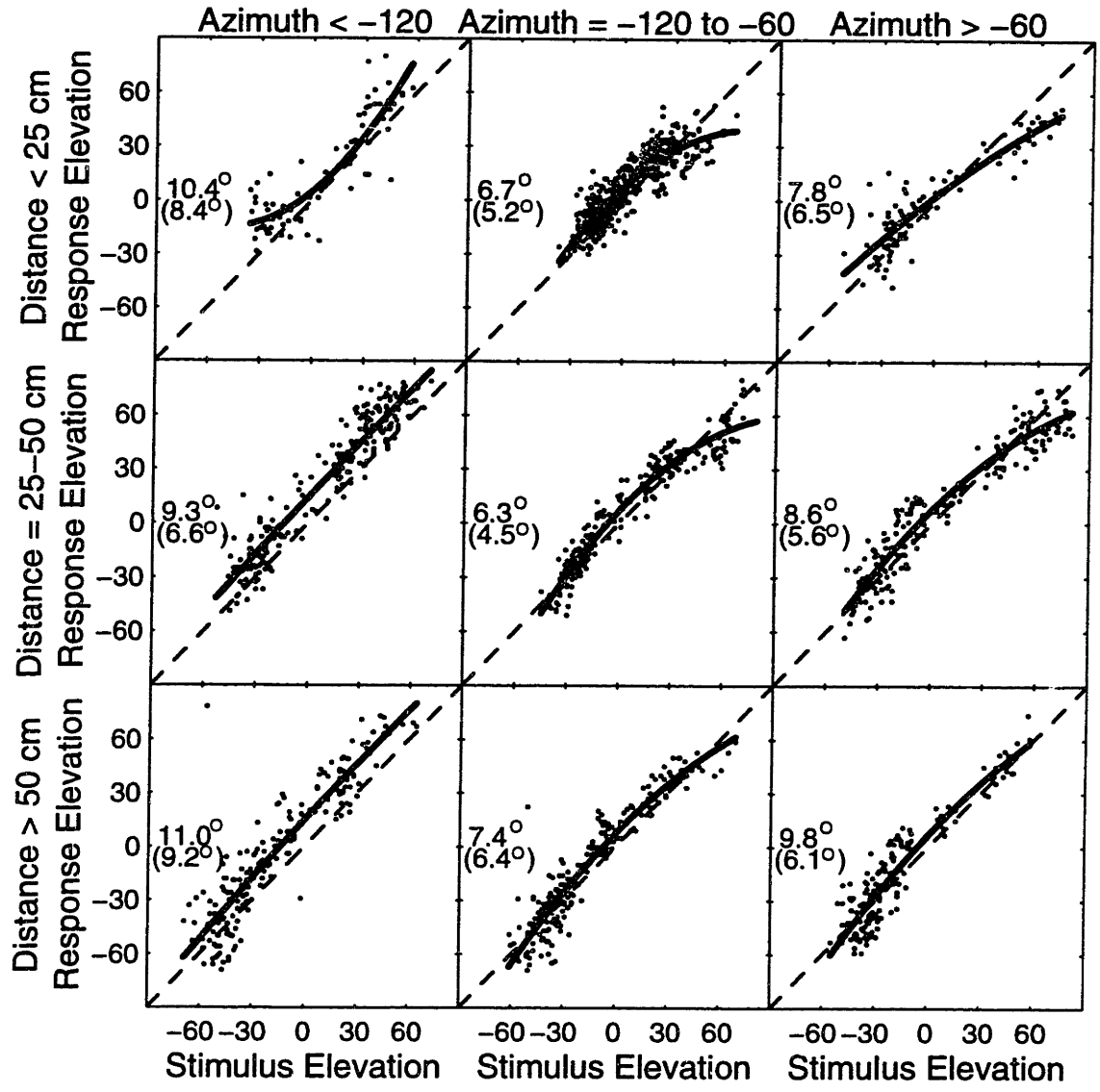


Figure 6-9: Raw Elevation Data for Subject DSB. The nine panels show raw elevation stimulus and response data for subject DSB for nine regions of the right hemisphere. Azimuths were divided into regions less than −120°, from −120° to −60°, and greater than 60°. Distances were divided into regions less than 25 cm, from 25 cm to 50 cm, and greater than 50 cm. A dashed line represents "correct" responses. The bias-corrected unsigned elevation error is shown at the left side of each panel (see Figure 6-7 for details), and the solid line represents the best second-order polynomial fit of the stimulus data to the response locations. Note that the stimuli range from approximately −45° to 80° in elevation. The data are limited at low elevations because the subject's torso and the chin-rest prevented placement at some source locations. For example, low elevations are particularly truncated at close distances behind the subject where their neck and back prevented placement at low elevations.

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 146

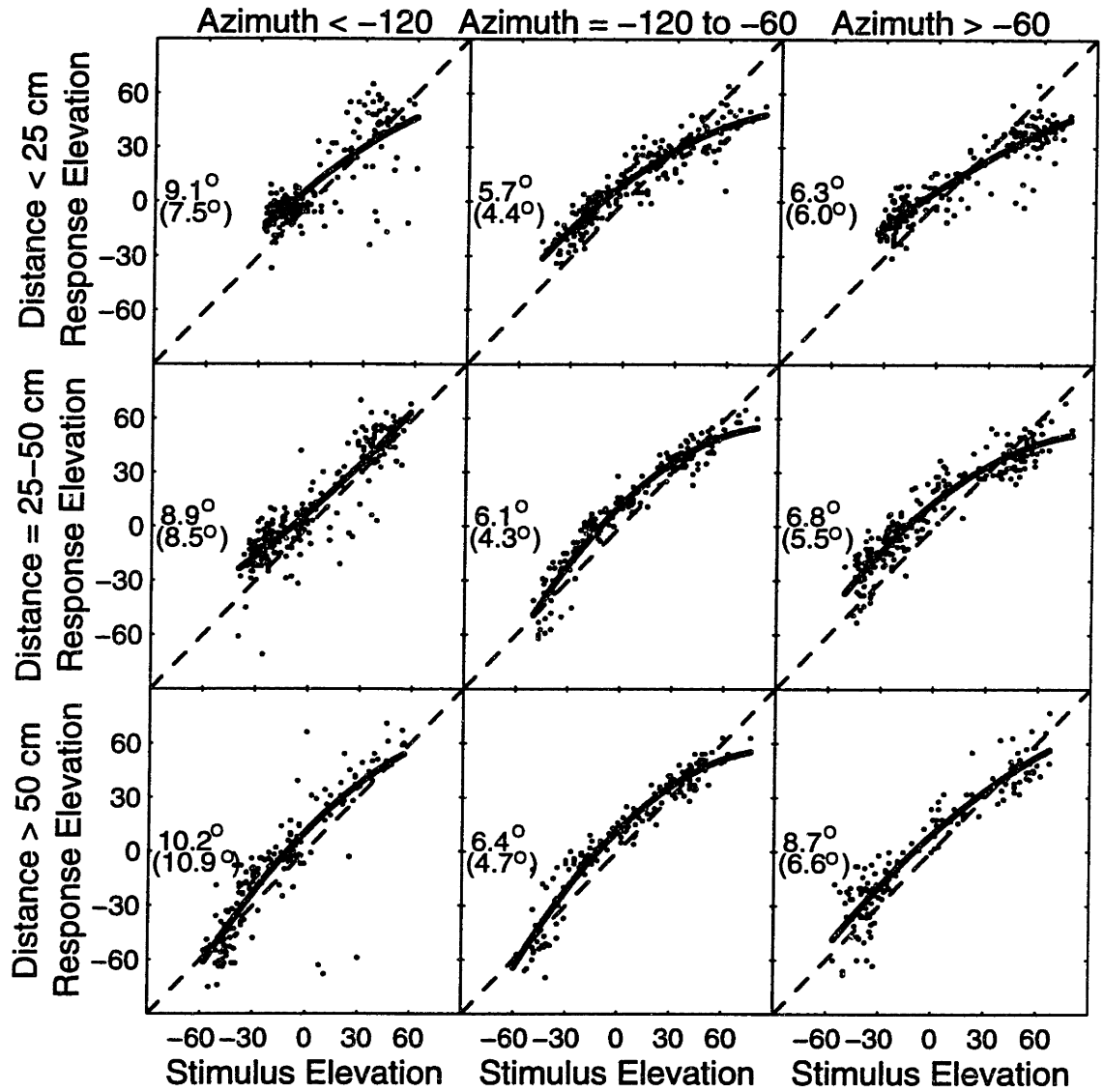


Figure 6-10: Raw elevation data for subject KMY. The format is as described in Figure 6-9.

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 147

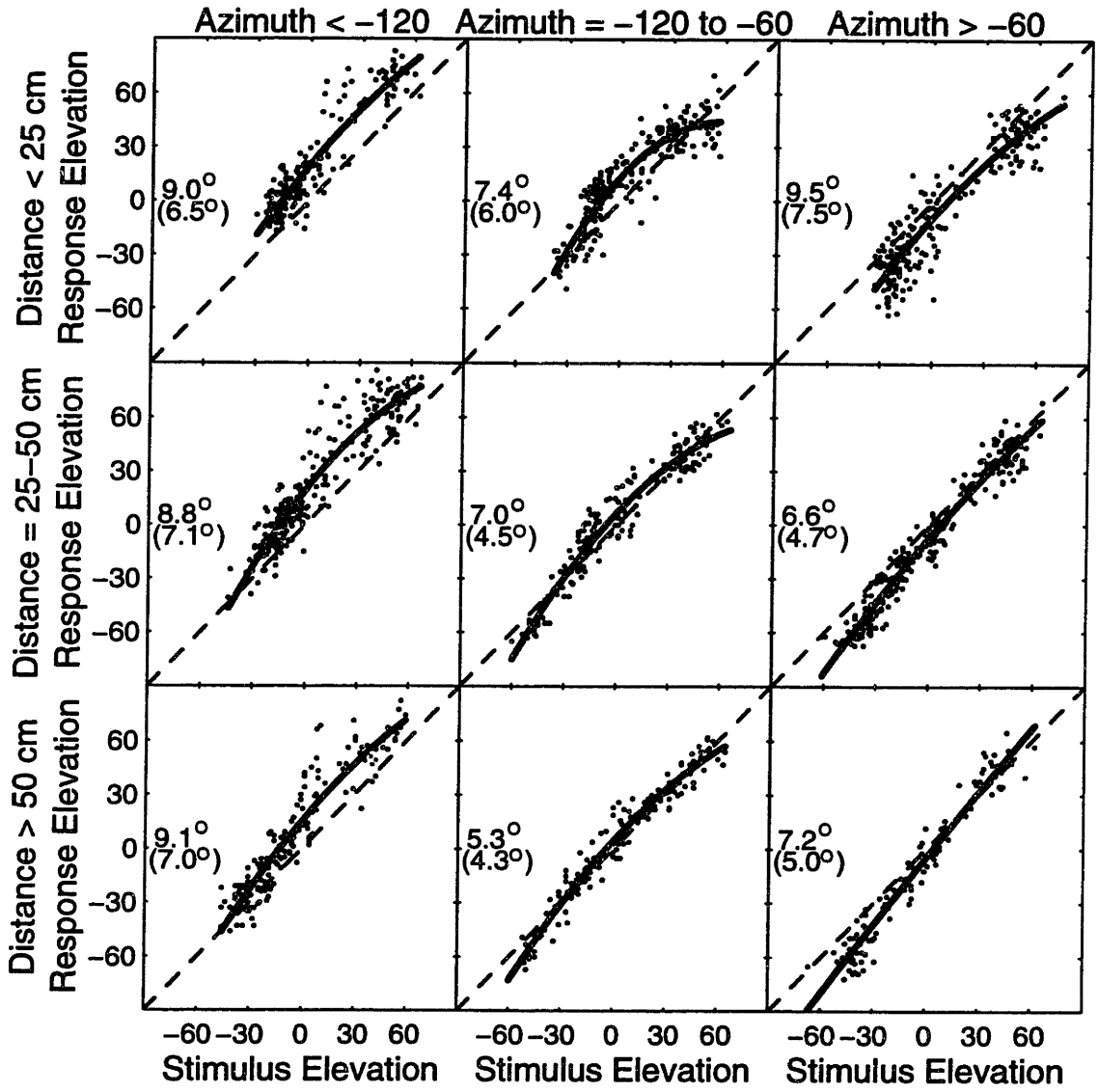


Figure 6-11: Raw elevation data for subject DTD. The format is as described in Figure 6-9.

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 148

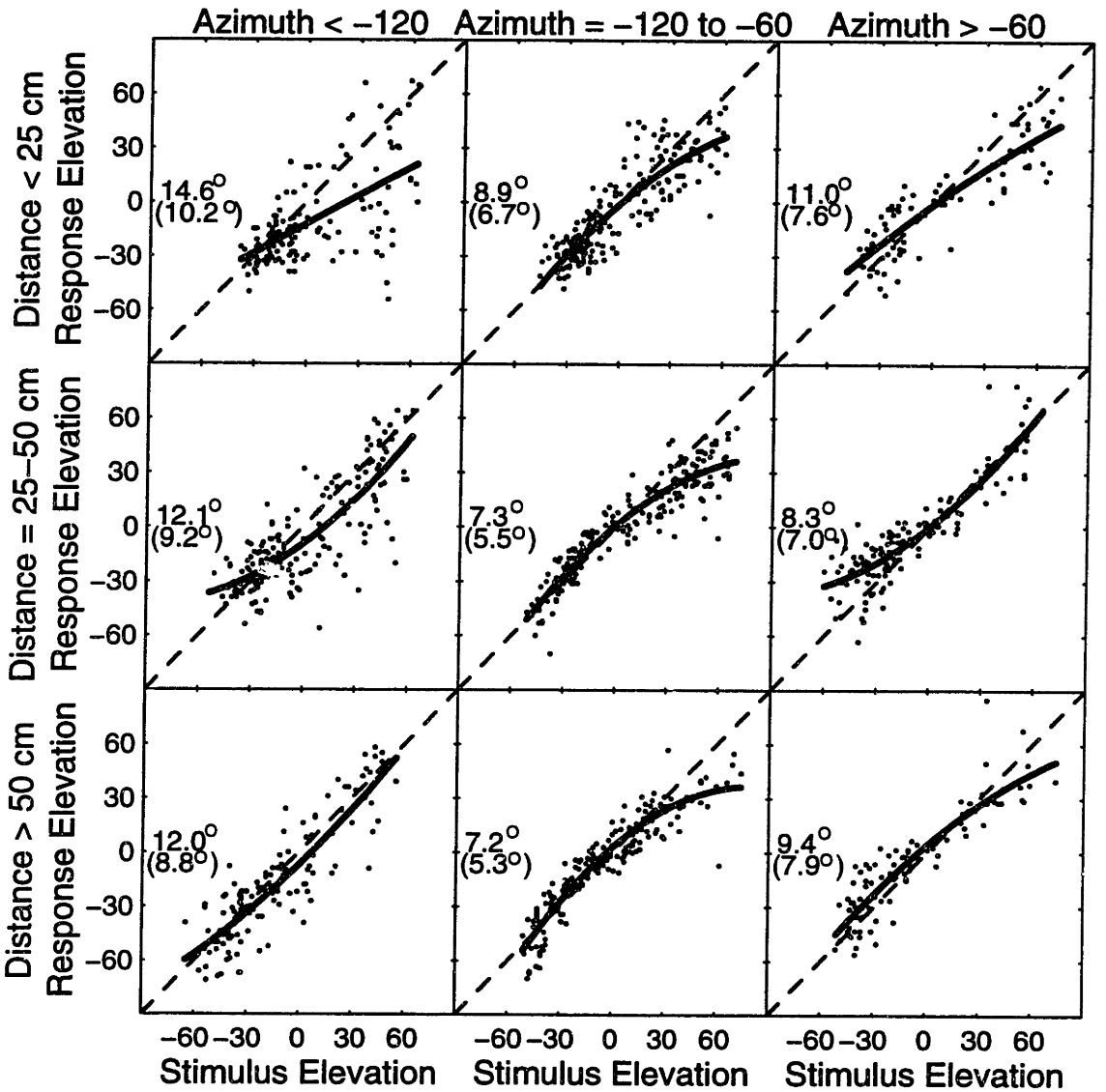


Figure 6-12: Raw elevation data for subject CLL. The format is as described in Figure 6-9.

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 149

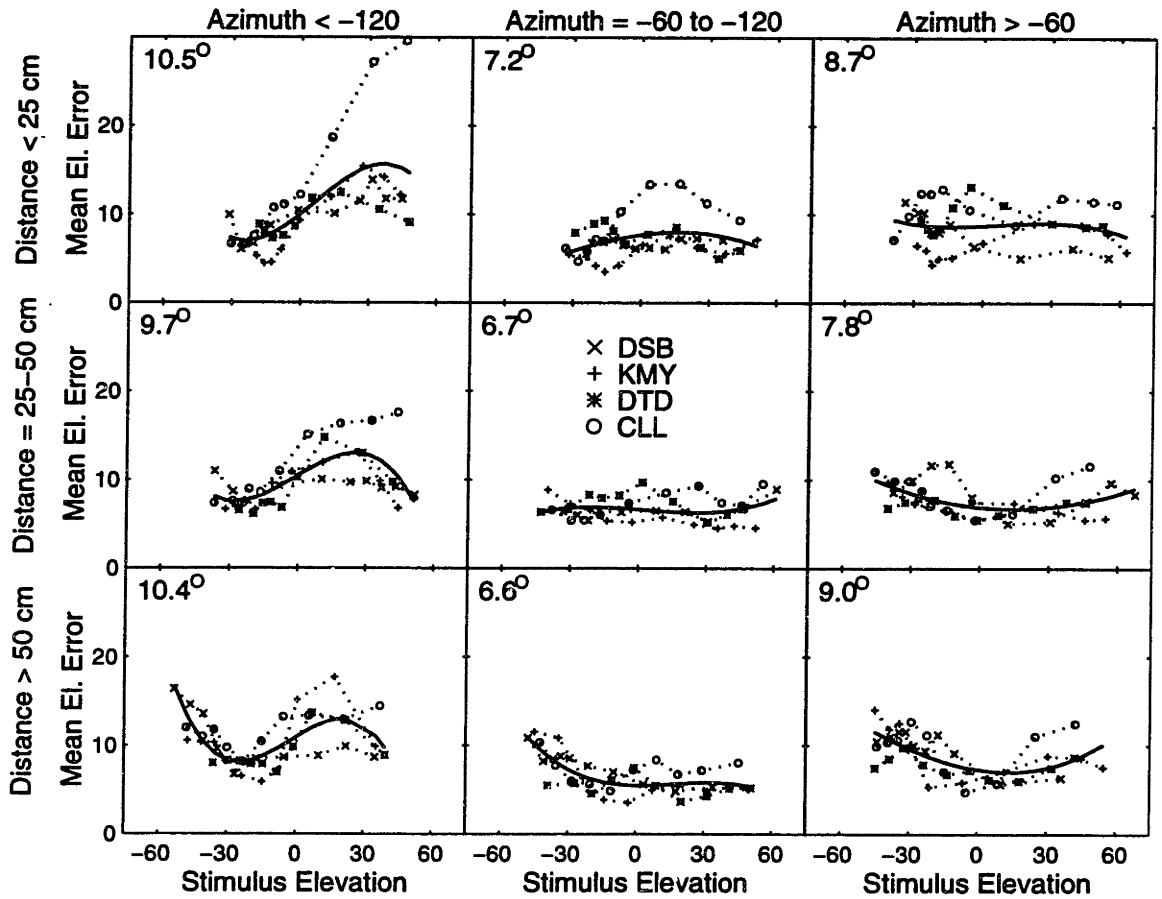


Figure 6-13: Directional dependence of mean unsigned bias-corrected error in elevation. The nine panels show the dependence of the unsigned error in elevation on the elevation location for nine locations in space. As in Figure 6-7, front back reversals have been eliminated, and the data in each panel have been sorted by elevation into 10 overlapping bins, each containing 18% of the total number of trials. Each symbol represents the mean elevation and mean bias-corrected elevation error in a single bin for a single subject. The solid line is the third-order polynomial best fitting the data of the four subjects.

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 150

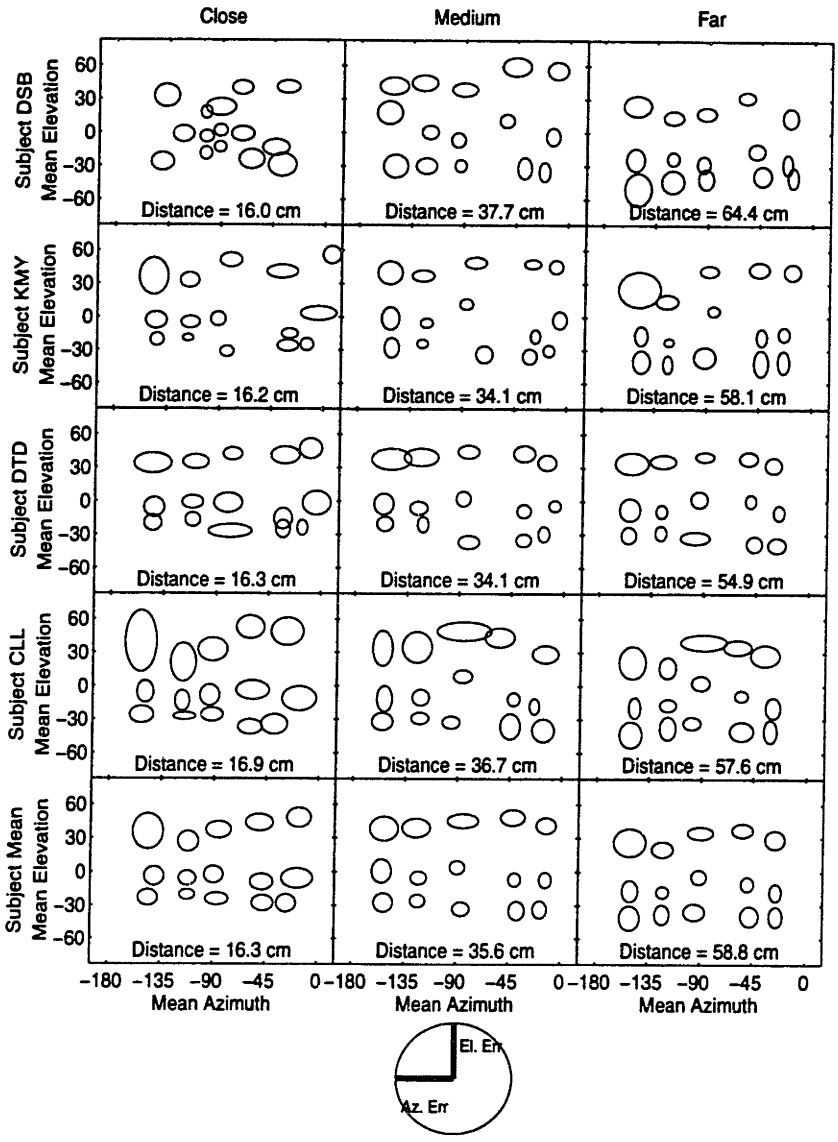


Figure 6-14: Spatial distribution of errors in azimuth and elevation. The non-reversed data for each subject were divided into three non-overlapping bins for distance, three non-overlapping bins for elevation, and eight overlapping bins in azimuth, as in Figure 6-8. The mean unsigned errors in azimuth and elevation (relative to the best linear fit of the response data within the bin) are shown by an oval at the mean azimuth and elevation locations. The width of the oval represents the azimuth error, and the height represents elevation error. The bottom row is the mean across the four subjects. Note that the bias-corrected errors were calculated differently than those in the previous figures. In this figure, the unsigned error was calculated relative to the best linear fit of each variable (azimuth or elevation) in that particular bin, rather than the best quadratic fit across all values of azimuth and elevation. Thus this figure provides more information about the spread of responses in a particular bin and less about the relationship between the stimulus and response locations.

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 151

A general comparison of the response variability in azimuth and elevation is provided in Figure 6-14. This figure is similar to Figure 6-8, but shows the bias-corrected unsigned errors in azimuth and elevation instead of the percentage reversals.

Generally, the worst performance overall is in the region behind and above the subject, both in azimuth and in elevation. The azimuth error is particularly large at high elevations, reflecting the increased sensitivity of azimuth in that region. The increase in azimuth error when the source is close and directly in front of the subject (at middle elevations) is again striking, Note that the errors in azimuth and elevation are generally comparable. The azimuth error is almost universally larger than the elevation error at high elevations, and for relatively distant sources the elevation error is larger than the azimuth error near 0° azimuth, but elsewhere the azimuth and elevation errors are generally comparable.

6.4.7 Response Biases

To this point, the primary focus has been the variability of subject responses in the form of the bias-corrected unsigned error. Systematic directional biases are also of considerable interest. Figure 6-15 shows the response bias (mean uncorrected signed error) in azimuth and elevation as a function of source location. Note that the directional biases are generally invariant to source distance. Although the directional biases differ substantially from subject to subject, the general pattern of biases for each of the subjects is consistent across the three distance bins. Subject DSB, for example, has a bias up and towards the front for sources behind and above the head at all distances, while CLL is generally biased down and towards the front at high elevations. It appears that directional response biases are roughly independent of distance.

Although they are relatively independent of distance, the response biases vary considerably across directions and subjects. When the source was close, high, and in front, however, each of the four subjects exhibited a negative elevation bias and a lateral azimuth bias. Elsewhere the subject biases were inconsistent, and when averaged across all four subjects the biases were relatively small.

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 152

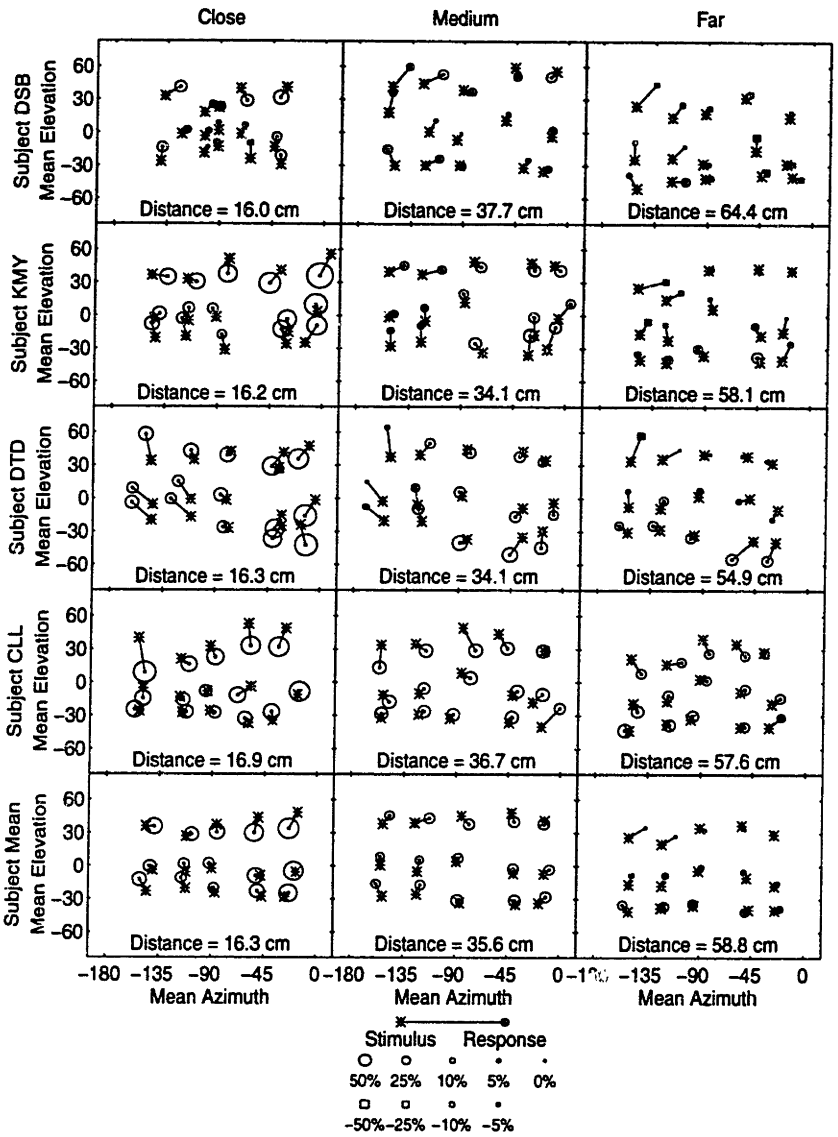


Figure 6-15: Response Biases. Direction of response bias as a function of source location. Details are similar to Figure 6-14. The * is the mean stimulus location, while the . is the mean response location. The circles indicate that the subject overestimated distance (see legend), while the squares indicate an underestimate of distance. Note that there are very few squared because the subjects almost never underestimated distance.

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 153

6.4.8 Discussion

In order to put these results into context, it is useful to compare them to previous estimates of directional localization ability available in the literature. Although no data are available on near-field localization, our results at distances greater than 50 cm can be compared to previous data collected 1 m or further from the subject. Two studies which have evaluated directional localization (position identification) rather than discrimination are Wightman and Kistler (1989a) and Makous and Middlebrooks (1990). There are substantial differences between these two studies and the present study. First, subjects in the present experiment were asked to simultaneously determine the azimuth, elevation, and distance of each stimulus, whereas in the two previous studies subjects only evaluated the azimuth and elevation of the source. The response methods also differ. Wightman and Kistler used verbal reports of location and Makous and Middlebrooks used a head-pointing technique, while in the present a manual pointing method was used. Finally, the stimuli were significantly different. Wightman and Kistler used eight 250-ms bursts of noise with randomized spectral content and a 200 Hz - 14 kHz bandwidth. Makous and Middlebrooks used a single 150-ms burst of noise with a flat spectrum from 1.8 kHz to 16 kHz. This experiment used five 150-ms bursts of noise with a gentle low-pass roll-off from 200 Hz to 15 kHz. Despite these differences it is useful to make comparisons across the three studies, especially in terms of relative performance in different regions of space.

Wightman and Kistler provide information on the angular error at three elevation locations and three azimuth locations. In general, the angular errors in their study are comparable to those in this experiment. In particular, the results of both studies indicate that performance is worst above and behind the subject, and is generally better to the side than in front or in back. The magnitudes of the errors are considerably lower in this study than in the Wightman and Kistler study (15.2° mean error at distances greater than 50 cm across the nine location bins compared to 21.1°). This discrepancy is likely the result of the scrambled amplitude spectrum in the Wightman and Kistler study and their use of verbal report rather than direct location as a

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 154

response method (See Chapter 4 for a comparison of direct location to verbal report).

Makous and Middlebrooks provide a map of the standard deviations and unsigned errors in as a function of source location. Their data indicate much less response variability than was found in the present experiment. For example, their results indicate a standard deviation of only 1.9° in azimuth and 3.3° in elevation for sources directly in front of the listener (0° azimuth and −5° in elevation). In contrast, this experiment indicates standard deviations (calculated from the root mean square of the bias-corrected unsigned error) of approximately 6° in azimuth and 8° in elevation in that region (at distances greater than 50 cm). In terms of relative performance, their data indicates that azimuth errors are much larger to the side than in front. In contrast, the data from this experiment indicate that the response variability in azimuth is roughly comparable in front and to the side. Makous and Middlebrooks also provide information on elevation performance, but they use a two-pole coordinate system that cannot be compared directly with the data from this experiment. Their general finding that azimuth localization is relatively more accurate than elevation localization in front and that the reverse is true to the side is, however, consistent with these results.

One characteristic of localization performance shared by all three studies is that directional judgments are least accurate when the source is located above and behind the head. In both the Makous and Middlebrooks experiment and in our experiment, the most plausible explanations for poor performance in this region are related to response method. The head-pointing response in the Makous and Middlebrooks experiment required the subjects to turn their heads nearly 180° to respond in this region, which is certainly a difficult task. In our experiment, subjects may have had difficulty reaching the region behind and above their heads with the response wand. Although we tested our response method and found no significant differences between response accuracy in the front and rear hemispheres, we did not collect enough data to evaluate specific regions of space such as the region behind and above the head (Chapter 4). The response method explanation is weakened by Wightman and Kistler's reports of poorest performance behind and above the subject using a verbal

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 155

response. Since all three studies report poorest performance in this region, using three different response methods, it is likely that there are some perceptual problems in localizing sound behind and above the head. The reasons for poor perception in this region are not obvious, however.

Both the Makous and Middlebrooks study and the Wightman and Kistler study provide detailed information about front back reversals. Our definition of a front-back reversal is more conservative than theirs (we require a response to be at least 10° closer to the mirror image of a source across the frontal plane to be considered a reversal while they only require the response to be closer to the mirror image), so the figures cannot be compared directly. Both studies report reversals in approximately 6% of all trials, while we found reversals in 7% of our trials at distances greater than 50 cm. Since our definition is considerably more conservative, this implies a larger number of reversals in our study. In the Wightman and Kistler study, almost all of the reversals occurred at high elevations to the side or in front of the listener. In the Makous and Middlebrooks study, most of the reversals were at high elevations behind the listener. In our results, reversals occur at high elevations in front of and behind the listener, and at low elevations in front of the listener. The number and distribution of reversals appears to depend primarily on the individual subject, as they vary significantly across subjects in all three studies.

The comparison of the data at distances greater than 50 cm with previous data is useful for establishing a baseline for comparison with the results at closer distances. Of primary importance in this study, however, is the effect of an extremely close source on directional localization ability. The angular error increases significantly as distance decreases (Table 6.1), but this increase is at least in part a result of larger response biases when the source is close, rather than larger response variability. Response variability, measured by the mean unsigned errors in azimuth and elevation, did not vary consistently with distance. The azimuth error was significantly larger at close distances (< 25 cm) than at far distances (> 50 cm) for only two of the four subjects. The elevation bias-corrected error was significantly larger at close distances than at far distances for two subjects, but significantly larger at far distances than

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 156

at close distances for the other two subjects. There is also some reason to believe that experimental error is slightly greater for very close sources than for more distant sources. At locations very close to the head, direction is very sensitive to small displacement errors. At 12 cm, for example, a 1 cm error in the subject response, or in the measurement of the stimulus and response locations, can cause a directional error of nearly 5°. When the increased error sensitivity of the response method for very near sources is weighed against the relatively minor decrease in performance at close distances, it appears that source distance has, at most, a marginal effect on directional accuracy in the near-field.

Front-back reversals increased slightly at close distances for all four subjects, but only one subject (CLL) reversed a significantly larger percentage of trials at close distances than at far distances. CLL appeared to be a poor localizer in general, as he experienced substantially more front-back confusions than the other subjects even at the greatest distances tested. Although data are only available from one subject, it may be the case that the localization problems of poor localizers are exacerbated when the source is very near the head, but that normal localizers may be unaffected by sources very close to the head.

Azimuthal localization in the region directly in front of the head appears to be unusually sensitive to distance. Azimuth error increases substantially as distance decreases when the source is at middle elevations, and directly in front of the listener (i.e. the region directly in front of the face). The mean unsigned error more than doubles as the distance of a source directly in front of the face decreases below 25 cm (see Figure 6-7), and this trend is evident in the data of all four subjects. Further research is necessary to determine the cause of this degradation in performance.

6.5 Distance Localization Results

One of the primary motivations for this experiment was an examination of the accuracy of auditory depth perception for nearby sources. The raw data for near-field distance perception are provided in Figures 6-16 to 6-19. In comparison to the raw

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 157

azimuth and elevation data, it is apparent that the log distance responses are considerably less accurate than either of the directional responses. The log distance responses are, however, approximately linearly related to the log stimulus locations. The sample correlation coefficient, which measures the degree of linear relationship between two variables, provides a good characterization of the relationship between the stimulus and response locations. The correlation coefficient between the log stimulus location and log response location is given at the left side of each panel in Figures 6-16 to 6-19.

The raw distance data indicate that the stimulus-response correlation coefficient is strongly dependent on the lateral position of the source. The correlation coefficient is much higher to the side than in front or back. A more detailed analysis of the relationship between distance performance and azimuth is provided in Figure 6-20. The correlation coefficient is clearly highest to the side of the listener and lowest in front and in back. Furthermore, the correlation coefficient is lower at high elevations, and decreases more rapidly as the azimuth moves away from $-90°$. These general trends are consistent across all four subjects. Observe that the data are roughly symmetric in the front and rear hemispheres, but that there are no data points behind $-150°$ in azimuth.

The correlation coefficient measures how well the response location can be explained by source location, but it can be misleading if the source-response relationship is non-linear. Therefore it is useful to examine the raw data in front and to the side of the listeners to verify that distance performance is well characterized by the correlation coefficient. Figure 6-21 shows the raw data for the locations directly in front and directly to the side of subject DSB. These data, which are typical of the data for the other subjects, show that the primary reason for the decrease in correlation in the median plane is that the slope of the stimulus-response line is much lower. In fact, almost all of the responses in the front bin are grouped around 60 cm, independent of the actual stimulus location. A similar pattern is found in the data from all four subjects: the slope of the the log-log stimulus response line increases systematically as the source moves towards the side ($-90°$ in azimuth), while the constant offset is

Sight Effect LLC
Exhibit 2013
IPR2023-01024
Page 158

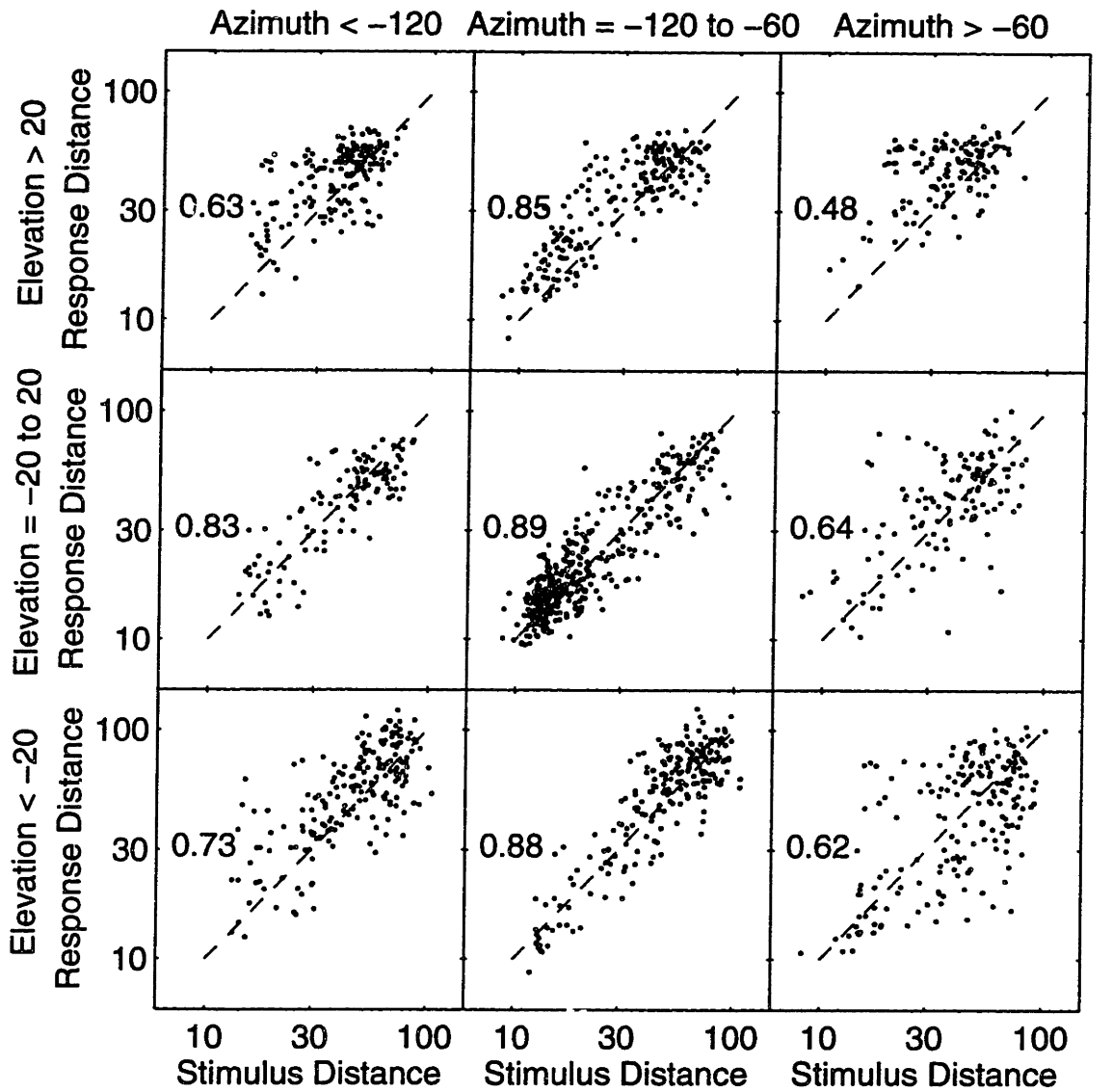


Figure 6-16: Raw Distance Data for Subject DSB. The columns represent different azimuths, and the rows represent different elevations. The data are plotted on a log-log scale, and the correlation coefficient of the log stimulus distance and log response distance is shown at the left of each panel. The dashed line indicates "correct" responses.

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 159

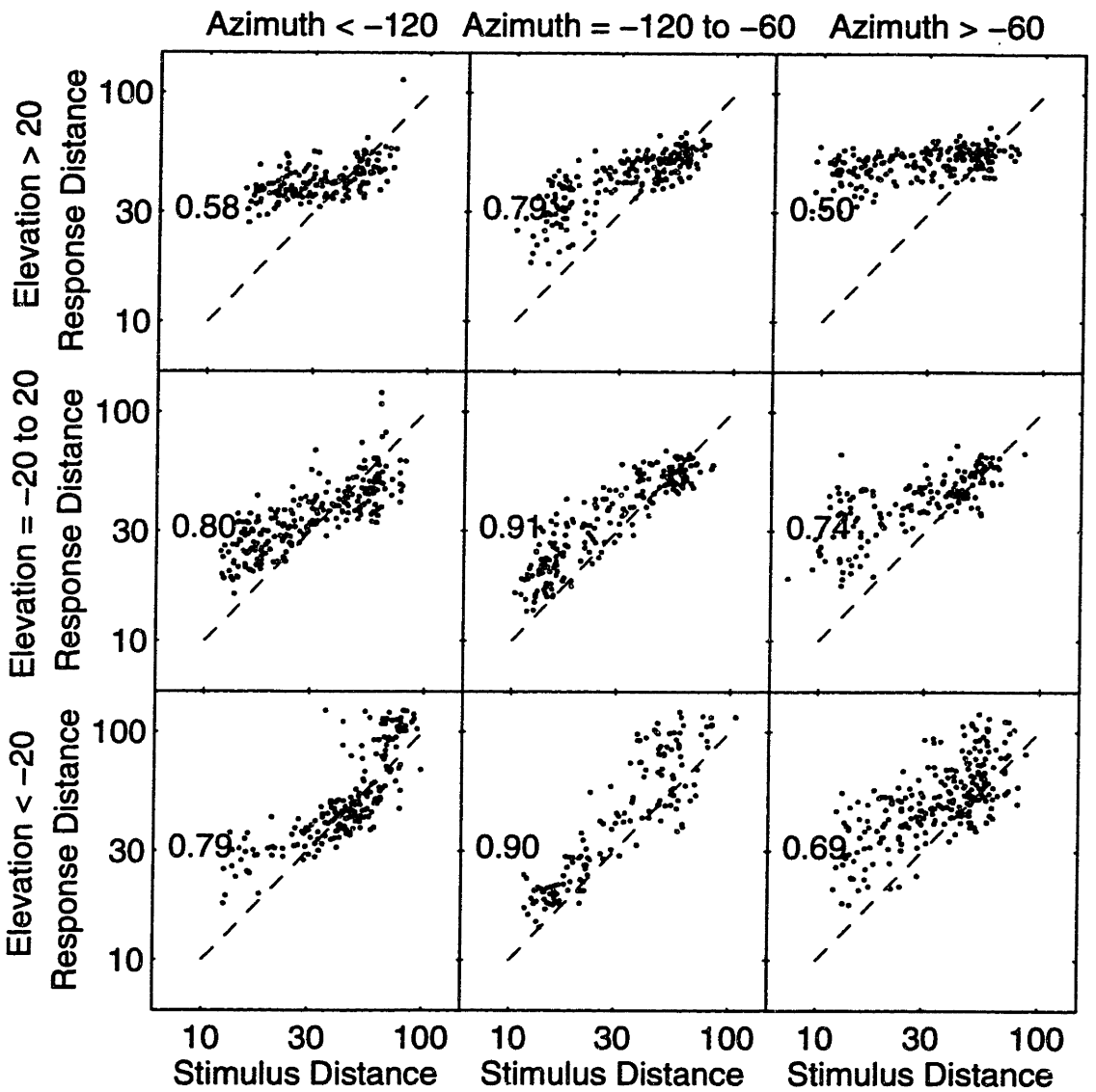


Figure 6-17: Raw Distance Data for Subject KMY. Format is similar to Figure 6-16.

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 160

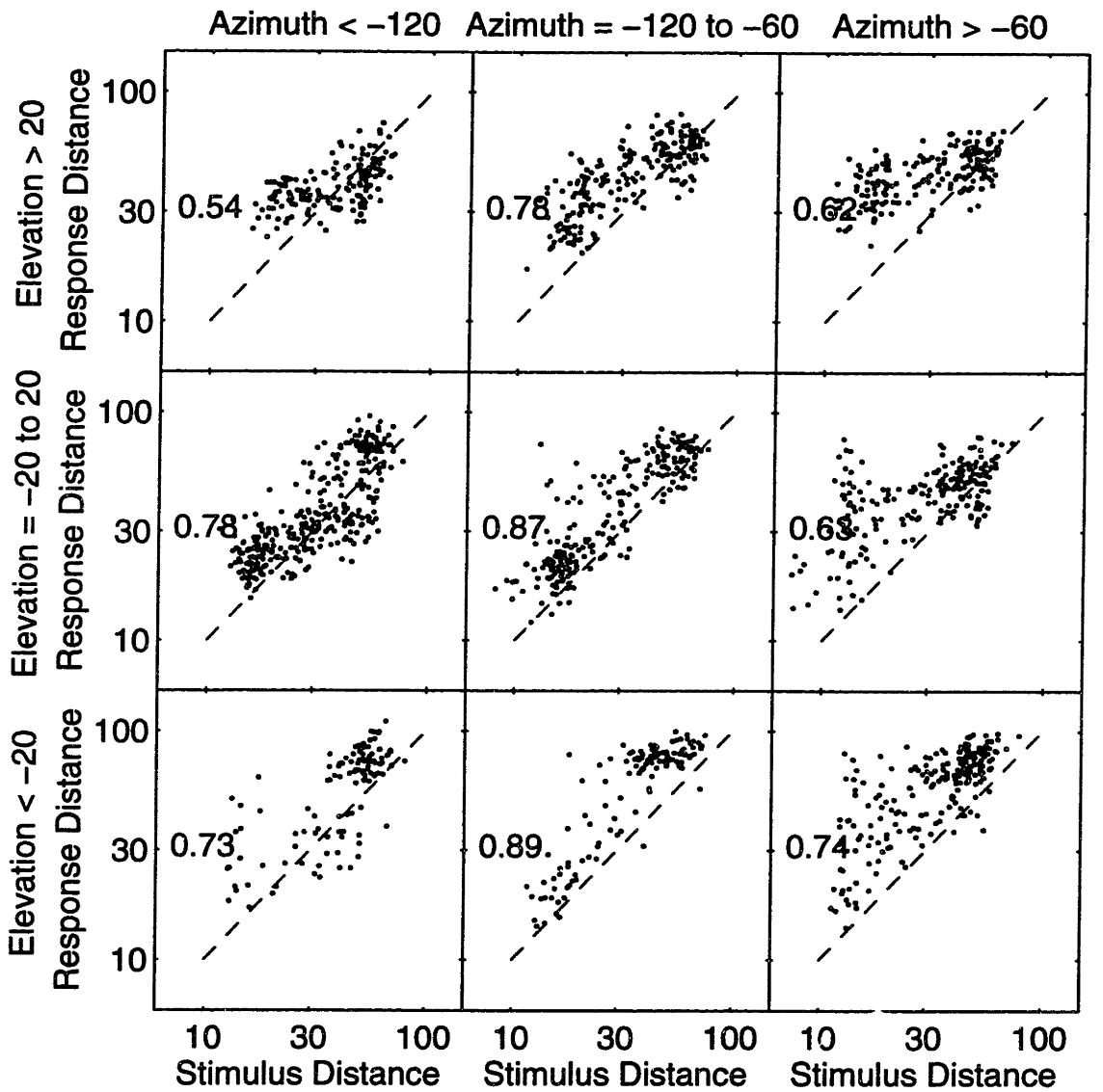


Figure 6-18: Raw Distance Data for Subject DTD. Format is similar to Figure 6-16.

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 161

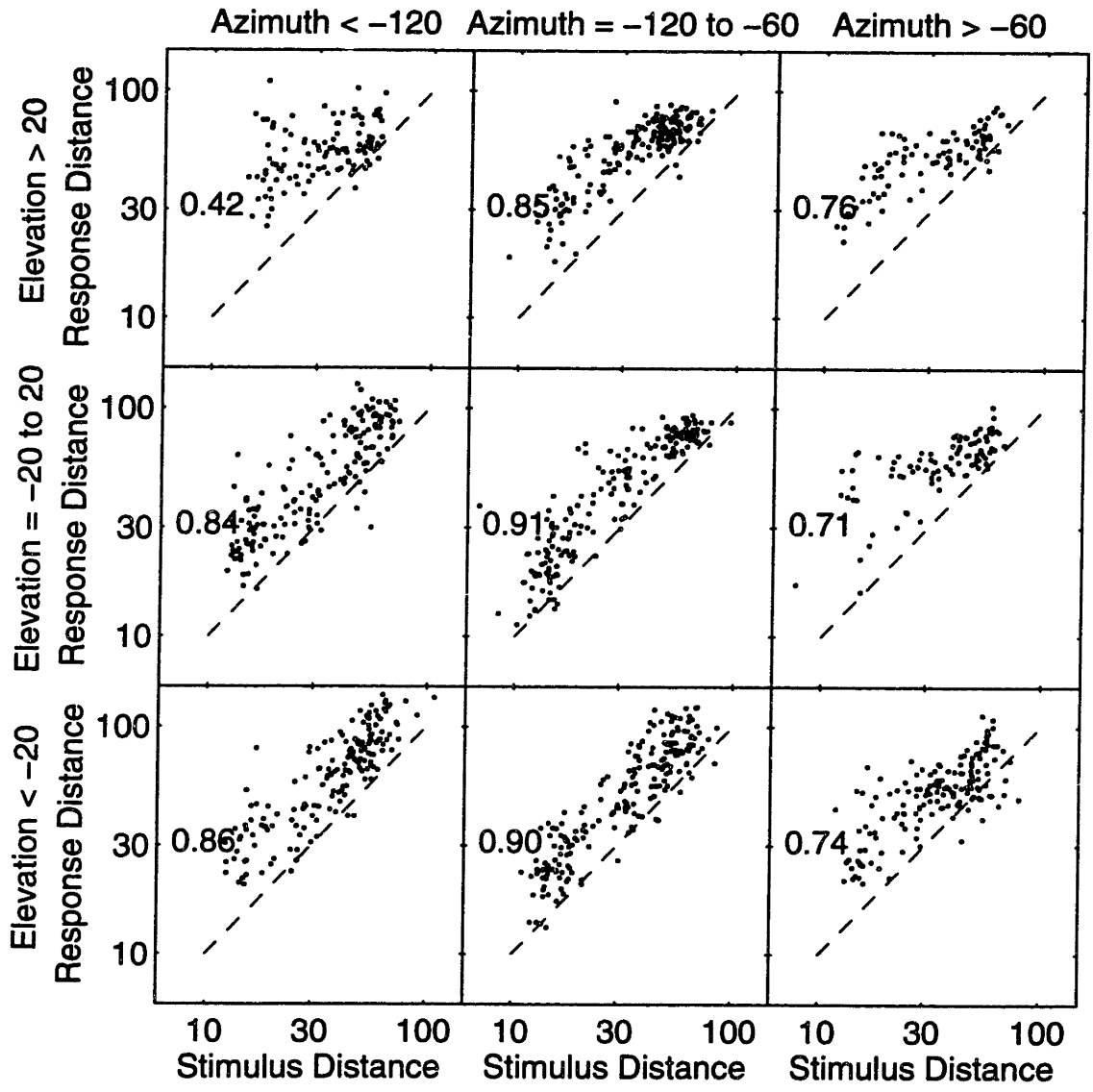


Figure 6-19: Raw distance data for subject CLL. Format is similar to Figure 6-16.

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 162

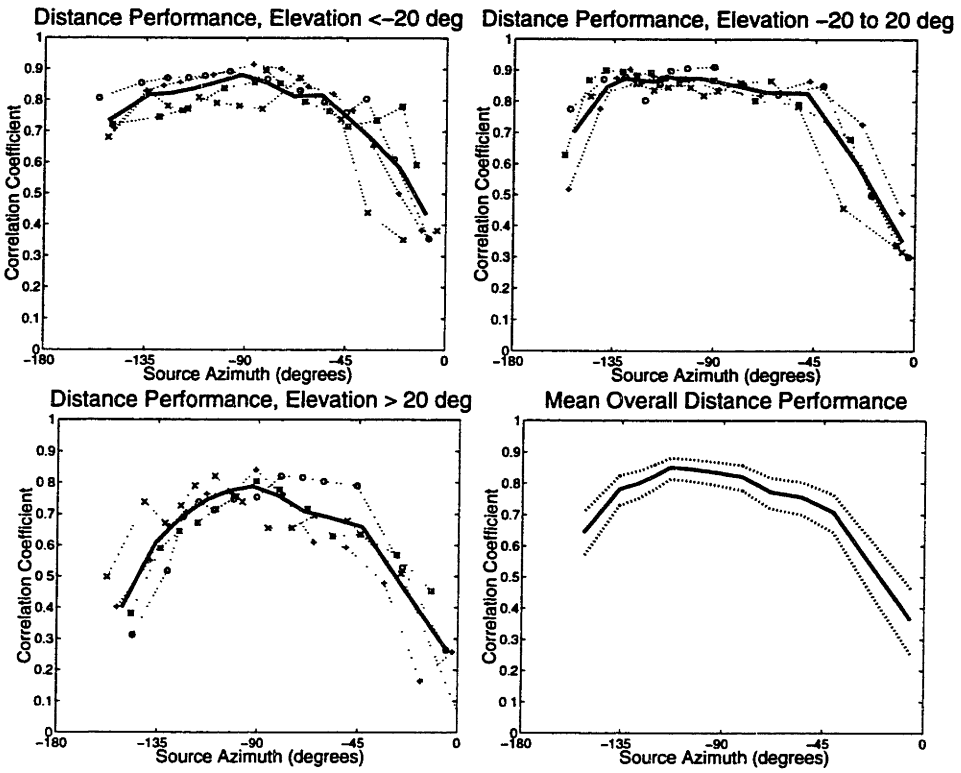


Figure 6-20: Correlation of log stimulus distance and log response distance as a function of source azimuth and elevation. Each subject's responses were divided into three bins according to elevation, and then sorted by azimuth into 13 overlapping bins each containing 14% of the total number of trials. The correlation in each bin for each subject is plotted as a function of the mean azimuth value in that bin. The solid lines represent the mean values across each subject. The first three panels represent trials at high, medium, and low elevations. In the lower-right panel, the coefficients were calculated separately for each subject, and then transformed into approximately Gaussian random variables using the Fisher transformation (Devore, 1991). The transformed data were then averaged together, and the resulting value in each bin was transformed back into the mean correlation coefficients in the last panel. The dotted lines above and below the average represent the 5% confidence interval of the correlation estimate, indicating that the increase in correlation around $-90°$ is highly significant.

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 163

largest near the median plane (Figure 6-22). Note that the slope of the regression line is always less than 1.0, and is less than 0.3 directly in front of the subject, indicating a compression distance responses into a smaller range than the span of actual source distances. This compression of distance responses in the median plane is also evident in the distance response biases (Figure 6-15). The subjects had a strong tendency to overestimate the distance of close sources, especially in front, but were unbiased for more distant sources.

The last panel of Figure 6-22 shows that the azimuthal dependence of distance accuracy with an alternate measure of accuracy — the standard deviation of the signed percent distance error. This measure of accuracy closely mirrors the correlation coefficient, with the best performance to the side and worst performance in the median plane. By this measure of distance performance, the distance judgments were generally less accurate at close distances than at greater distances (Figure 6-23). The standard deviation of the signed percent distance error increased substantially at close distances for subjects KMY, DTD, and CLL. Subject DSB's errors, however, did not increase at close distances. While the decrease in percentage error with distance seems dramatic, it is important to note that the response distance was effectively limited to the range from 0.1 m to 1 m, and that to some extent the reduction in percent error at the furthest distances is an edge effect— it was impossible to overestimate the distance of a source near 1 m. Also note that the absolute distance error, measured in cm rather than percent, increases with distance.

It is difficult to compare distance localization in this experiment to previous results. Few studies have directly examined auditory distance perception as a function of direction, and even fewer have examined distance perception for nearby sources. We know of only two studies which have actively examined near-field distance perception. Ashmead, LeRoy, and Odom (1990) measured the smallest discriminable percent decrease in the distance of a sound source in the median plane at 1 m. When the amplitude of the source was corrected for distance effects (a control condition), they found the threshold change in distance was approximately 16%. Although this figure is artificially low due to ceiling effects in their adaptive procedure, at least

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 164

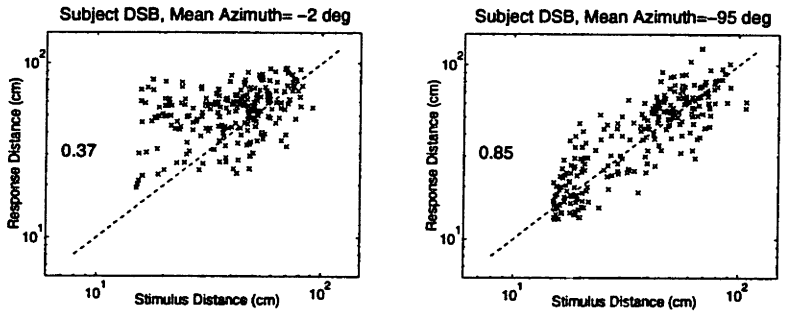


Figure 6-21: Raw distance data for subject DSB in front and to the side. These plots show the data for subject DSB in the 1st and 7th azimuth bins used to calculate the results in Figure 6-20. In front, the responses tend to be clustered around 60 cm independent of the source location, while to the side the response distance varies systematically with stimulus distance.

four subjects did not show any ceiling effects, indicating that subjects were able to determine the distance of the sources based on non-amplitude cues. Although the thresholds found in this experiment were large, the results do indicate that subjects are able to obtain some information about the distance of a near-field source in the median plane, which is consistent with our finding of a non-zero correlation coefficient between stimulus and response distance in the median plane.

Simpson and Stanton (1973) performed an experiment specifically designed to look for binaural distance cues for close sources. In this experiment, subjects were asked to estimate the distance of a sound source directly in front of the listener placed at one of five locations ranging from 30 cm to 2.7 m. Some of the subjects used a fixed head position during the experiment, some were allowed to turn their heads, and some were required to move their heads. Simpson and Stanton found that head motion had no significant effect on distance perception. Since the results of the current experiment indicate that distance accuracy is substantially better for lateral sources than for medial sources, it is surprising that Simpson and Stanton's subjects were not able to judge distance more accurately when they were allowed to turn their heads away from the sound source. Amplitude and reverberation in the Simpson and Stanton study may account for the discrepancy. The amplitude of the source was fixed during their experiment, and their subjects were seated in the corner of a sound-treated listening

Eight KHz, LLC
Exhibit 2013
IPR2023-01024
Page 165

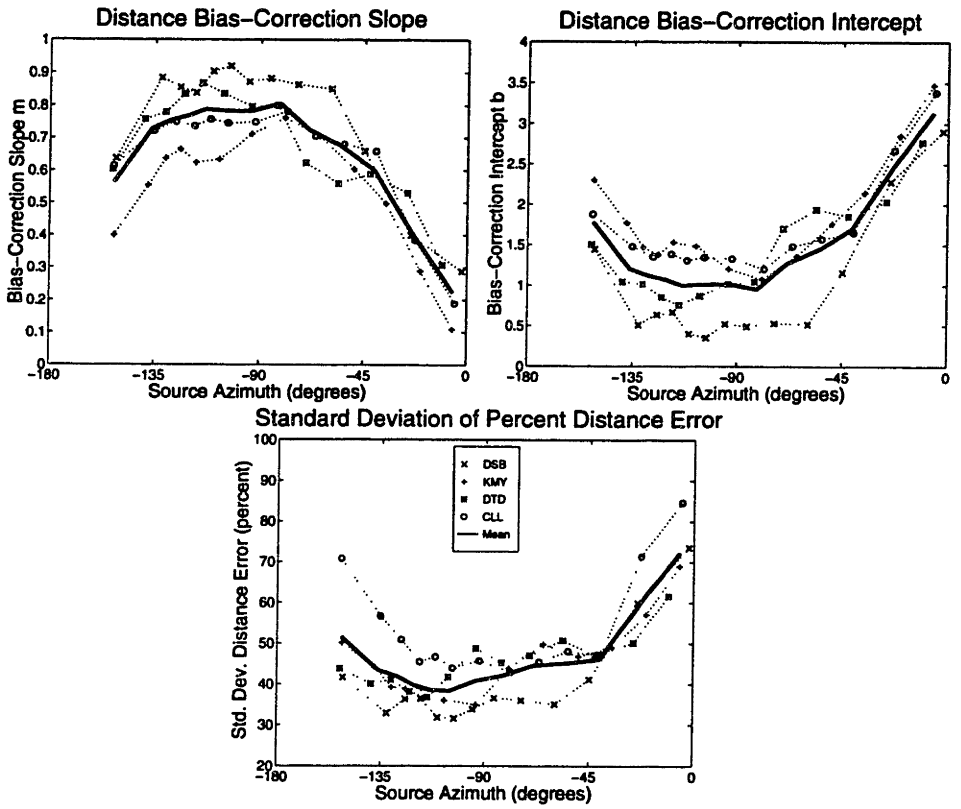


Figure 6-22: Slope and intercepts of regression lines relating the log response distances to the log stimulus distances for each subject. The raw data for each subject were sorted by azimuth into 13 overlapping bins each containing 13% of the total data points. In the first two panels, the ordinate is the mean azimuth location in each bin and the abscissa is the y-intercept c and slope m of the log-log regression line, which estimates response distance from the stimulus distance r by the equation $e^{b+m\ln r} = e^b r^m$. The last panel shows the standard deviation of the percentage distance error in each azimuth bin as an alternative measure of performance.

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 166

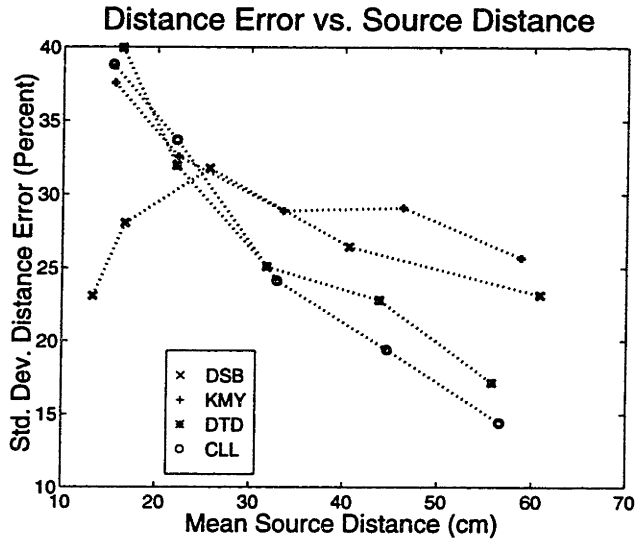


Figure 6-23: Standard deviation in percent distance error as a function of source distance for lateral sources. The raw data at the side of each subject (the middle third of azimuth values) were sorted by distance into 5 overlapping bins each containing 40% of the total trials. The standard deviation of the percent distance error in each bin is plotted versus the mean distance of the trials in each bin. The percent error decreases as distance increases for every subject except DSB.

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 167

booth, with their heads only 25 cm from either wall. Thus it is likely that their subjects were able to use amplitude and reverberation cues to judge distance, and these cues may have dominated the binaural distance cues in their experiments.

In contrast to the Simpson and Stanton paper, two studies of far-field localization have indicated that distance perception is better for sources along the interaural axis than for sources in the median plane when the amplitude of the source is randomized. Holt and Thurlow (1969) found that subjects could accurately determine the relative distances of the sound sources when they were lined up with the interaural axis (rank order correlation of 0.93), but not when the sources were directly in front of the subject. Gardner (1969) informally reported a similar result. The relationship between azimuthal position and distance localization accuracy found in these earlier studies is in agreement with this the results of this experiment, but we cannot explain why subjects were able to perform so well in the far-field where binaural distance cues are largely absent.

Other than the observations of Holt & Thurlow and Gardner, no previous studies have indicated that distance perception is better for lateral sources than medial sources at close distances. Furthermore, the strong correlations found in this study (as large as 0.85 for sources near 90°) indicate that distance perception is reasonably accurate in this region, and in fact may be more accurate than in any previous study where overall level cues were not available, especially considering the additional requirement of simultaneously determining the azimuth and elevation of the source in this experiment.

6.6 Comparison of Results to Near-Field HRTF Measurements

By comparing the results of this psychoacoustic experiment with previously measured head-related transfer functions in the near-field, we can gain valuable insights into the mechanisms of near-field localization. The features of the near-field HRTFs, along

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 168

with previous results from far-field localization experiments, can explain the relatively weak distance dependence of directional localization, as well as the relatively accurate distance judgments for lateral sources.

Although three of the four subjects were slightly less accurate at azimuthal localization when the sound source was close to the head, the decrease in performance was relatively minor except for locations directly in front of the face. Similar horizontal localization performance in the near- and far-fields may indicate that low-frequency ITDs dominate azimuth judgments in the near-field as they have been shown to do in the far-field. Previous work by Wightman and Kistler (1992) has shown that ITDs tend to dominate azimuthal localization when the stimulus contains low-frequency energy. Note, however, that Wightman and Kistler's experiments manipulated the time delay in HRTFs measured in the far-field. The low-frequency time delay clearly dominates perception with the far-field HRTFs, where the IID was significant only at high-frequencies. When the source is in the near-field, however, the IID can be large even at low frequencies, and the Wightman and Kistler data provide no direct evidence that the ITD dominates the influence of low-frequency IID on azimuth perception. The absence of a strong lateral azimuth bias for nearby sources provides some indirect evidence that ITD dominance extends into the near-field. In the near-field, an increase in IID could result either from a source moving closer to the head or from a source moving away from the median plane. If azimuth judgments were based on IID, one might expect listeners to confuse the distance and direction of the source in the near-field, resulting in a lateral bias for nearby sources. There is, however, no indication of such a bias in the data. The lack of lateral directional biases for nearby sources, coupled with comparable directional accuracy in the near- and far-fields, indicate that near-field azimuth perception is most likely based on ITDs which are essentially independent of source distance.

An interesting result which cannot be readily explained by the HRTFs is the relatively poor accuracy in azimuth found at locations directly in front of the face. In far-field localization, accuracy is generally best near 0° in azimuth because of the keen sensitivity of the auditory system to changes in ITD near 0 μs (both in localization

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 169

and discrimination). In contrast, our data indicate that at middle elevations and distances < 25 cm, response variability is greatest near 0°. It is unlikely that the response method can account for this result, and there is no indication that the ITD and IID cues that allow accurate perception of azimuth in front at large distances are significantly reduced when the source is close. (In fact, the IID is more sensitive to changes in azimuth when the source is close.) It is true, however, that human listeners are not generally accustomed to encountering sound sources in this region and, moreover, are generally uneasy about having any objects "in their face". The degraded performance in this region is quite interesting, and requires further investigation.

The psychoacoustic results indicate that elevation perception does not depend on distance in a systematic way. Two subjects performed slightly better in elevation when the source was distant, and two performed better when the source was close. This is consistent with the observation (see Chapter 5) that the high-frequency features of the HRTF which change systematically with elevation are relatively independent of distance.

The distance perception abilities of our subjects, and in particular their ability to make unbiased, accurate distance judgments about lateral sources and their inability to make distance judgments about medial sources, suggest that the variations in the IID with angle and distance provide a useful binaural near-field distance cue. In the far-field, the IID varies only with direction. In the near-field, the IID increases as the source approaches the head. The usefulness of this increase as a distance cue is related to the range over which the IID varies in a particular direction. The span of possible IIDs is largest when the source is to the side, and decreases to zero in the median plane (Figure 6-24) . This span mirrors the distance performance by the subjects, which was also best for lateral sources, and worst in the median plane. In fact, the only major discrepancy between distance localization accuracy and the range of possible IID values is that localization performance appears to plateau in the region from −45° to −135°, while the span of IIDs increases systematically up to −90°. This could be explained by the well-known range effect in stimulus identification experiments, which causes sensitivity to changes in a stimulus to decrease when the

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 170

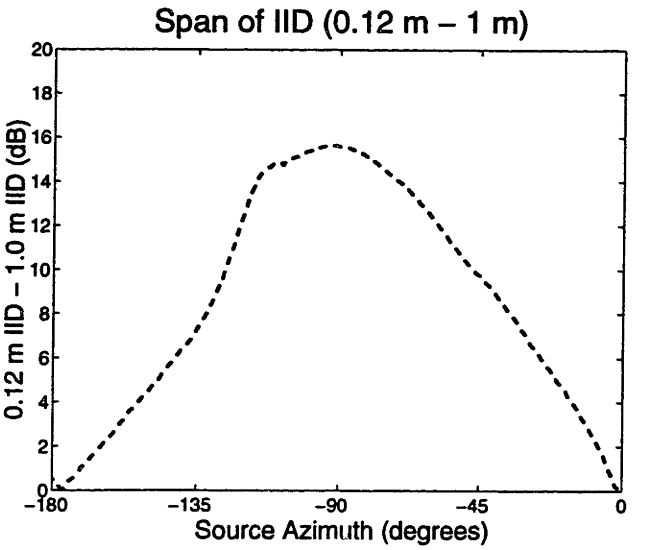


Figure 6-24: Variation in IID with distance. Near-field HRTFs measured with a KEMAR manikin (Chapter 5) in the horizontal plane were used to calculate the mean interaural intensity difference over the frequency range of 200 Hz to 1.5 kHz at distances of 0.12 m and 1.0 m. This figure shows the difference between the IID at 0.12 m and the IID at 1.0 m as a function of azimuth. To the extent that near-field distance perception relies on IID information, this plot roughly indicates the amount of available distance information as a function of the azimuth position.

range of possible values increases (Durlach & Braida, 1969; Koehnke & Durlach, 1989). The range effect, which is based on memory noise rather than sensory noise, could explain the saturation in performance seen in this experiment. Note that IID-based distance cues could also explain the decrease in performance at high elevations (and the more rapid decrease in performance away from $-90°$), since the IID is lower at high elevations than in the horizontal plane.

A surprising result which is inconsistent with the measured HRTFs is the decrease in percentage distance error as distance decreases (for three of the four subjects). The percentage change in distance required to generate a one-JND increment in IID decreases substantially with distance. Thus, a listener judging the distance of a sound source based solely on IID should be sensitive to *smaller* percentage changes when the source is close, and the standard deviation in the percentage distance error should be lower for nearby sources than for more distant sources. There are two possible factors

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 171

which may explain why this effect was not found in the data. First, the percentage distance error near 1 m was artificially low because the range of responses was limited to locations within 'arm's reach' of the listener. As a result, the listeners could never overestimate the distance of a source at 1 m. Furthermore, they could at most underestimate the distance of a source at 1 m by 100%, while they could overestimate the distance of a source 20 cm away by as much as 400%. These restrictions in the range of responses would tend to artificially inflate the percentage distance error for nearby sources. In addition, there is probably some portion of the error that is constant, rather than proportional to source distance. This constant error would contribute proportionally more to the overall percent distance error for closer sources. Direct measurements of distance JNDs are needed to determine whether listeners are actually more sensitive to percentage changes in distance for close sources.

6.7 Conclusions

The general results of these experiments can be summarized as follows:

- The angular error, which includes the effects of response bias and response variability, increases as the source approaches the head, particularly in front and behind the listener.
- The bias-corrected mean absolute azimuth error generally increases slightly at close distances. The increase is dramatic, however, for sources directly in front of the face, where the error approximately doubles as the source moves closer than 25 cm.
- The bias-corrected mean absolute elevation error is lowest for lateral sources and greatest behind the listener. It did not vary consistently with distance.
- Distance perception is most accurate for lateral sources and least accurate near the median plane. For lateral sources, the distance judgments were highly correlated with the actual source position ($r > 0.85$), and were relatively unbiased;

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 172

in the median plane, the correlations were low ($r < 0.4$). The results generally indicate better distance perception in the near-field than in any previously reported studies involving sources of unknown strength in anechoic conditions.

- The results were largely consistent with previously measured HRTFs in the near-field, which indicate that IID varies with distance in the near-field, while ITDs are roughly independent of distance, and in particular with the hypothesis that IIDs are an important binaural distance cue in the near-field.

It appears that directional localization is modestly degraded when sources are close to the head, but that distance perception is significantly improved, at least for sources away from the median plane. Additional experiments are necessary to fully understand the mechanisms of near-field localization. The next chapter will look at the effects of different stimuli (e.g., band-limited, monaural, or fixed amplitude) on near-field localization. Additional experiments will be necessary to determine JNDs in azimuth, elevation and distance in the near-field. Of particular interest in these experiments are the region of poor azimuth resolution at locations directly in front of the face, and the relationship between the threshold change in distance and the distance of the source.

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 173

Chapter 7

The Effects of Stimulus Characteristics on Near-Field Localization

Abstract

A series of experiments have been performed to examine the localization of nearby sound sources under a variety of conditions. A previous experiment (Chapter 6) has shown that in the region within 1 m of a listener's head, which we define as the near-field, directional localization with a broadband source is roughly comparable to localization at distances greater than 1 m. In contrast, distance localization appears to be more accurate in the near-field than in the far-field for lateral sources. In this experiment, near-field localization performance was tested with a fixed-amplitude source, with one ear occluded by an ear-plug and muff, and with a high-pass and low-pass filtered stimulus. Directional performance in each condition conformed with previous results from far-field experiments. In the fixed-amplitude condition, tistance performance was improved slightly at all azimuth locations, with the largest improvement near the median plane. In the monaural condition, distance performance was very poor, with stimulus-response correlations never exceeding 0.4. In the low-pass condition, distance performance was nearly as good as in the broadband condition, but in the high-pass condition distance localization was substantially degraded. The results indicate that low-frequency interaural intensity differences in the near-field facilitate accurate distance perception in that region.

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 174

7.1 Introduction

The problem of identifying the location of a sound source in space from auditory information has been studied extensively, but until recently little attention has been given to the unique aspects of localization in the immediate vicinity of the head. In the psychoacoustic "far-field", which we define as the region greater than 1 m from the head, the head-related transfer function (HRTF) is roughly independent of distance. In contrast, in the "near-field" (within 1 m of the head), the HRTF changes systematically with distance. This study extends a previous study of near-field localization with a broad-band source (Chapter 6) to a variety of conditions, including fixed stimulus amplitude, monaural listening, and highpass and lowpass filtered stimuli. The results are compared to the broadband condition, and to previous results from far-field localization experiments under similar conditions.

7.2 Background

The important role that the frequency content of the source plays in auditory localization has been recognized from the earliest days of auditory localization research. Lord Rayleigh's famous Duplex Theory of localization (1907) is based on the idea that different mechanisms of localization are dominant depending on the frequency spectrum of the source. At low frequencies (below 1500 Hz), where head shadowing is minimal, the interaural phase delay is the primary localization cue. At higher frequencies, where the phase difference between the ears is ambiguous, interaural intensity differences caused by head shadowing dominate horizontal localization. More recently, researchers have noted that the complex pattern of folds in the outer ear provide high-frequency spectral localization cues which are important in determining the elevation of a source (Roffler & Butler, 1968; Gardner & Gardner, 1973; Butler et al., 1980) and resolving front-back confusions (Muscicant & Butler, 1984; Oldfield & Parker, 1984).

The different localization mechanisms function in different frequency ranges, so the

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 175

spectral content of a stimulus has a strong influence on localization accuracy. When the source contains only low-frequency energy, localization relies almost entirely on interaural time delays. Under these conditions, the ITD provides accurate information about the lateral position of the source, but cannot be used to distinguish between sources at the same lateral position in the front and rear hemispheres, or to determine the elevation of a sound source. Musicant and Butler (1984) found that the number of front-back reversals increased substantially when the stimulus was low-pass filtered at 4 kHz, and dramatically when the stimulus was low-pass filtered at 1 kHz. Hebrank and Wright (1974) found that subjects were unable to accurately judge the elevation of a sound source in the median plane (with more than 50% of responses within 45° of the source location) without spectral content in the stimulus above 8 kHz. When the source contains only frequencies below 4 kHz or so, localization performance is poor in elevation and the number of front-back reversals increases substantially.

In contrast, localization accuracy is only slightly degraded when a source contains only high-frequency energy and the source is not narrow-band. Hebrank and Wright (1974) found that elevation judgments were quite accurate when the stimulus was high-pass filtered above 4 kHz (100% of responses were within 45° of the actual source location and 60% were within 15°). Musicant and Butler reported no significant differences between the accuracy of azimuthal localization with a broadband noise and a noise high-pass filtered above 4 kHz. These results indicate that the localization at high frequencies probably does not rely solely on IIDs, as hypothesized in the Duplex Theory. There is some evidence that subjects are able to obtain interaural time delay information from the envelopes of high-frequency sounds (Henning, 1974; McFadden & Pasanen, 1976; Trahoitis & Bernstein, 1986; Yost et al., 1971). In addition, Musicant argued that the directional characteristics of the pinnae play an important role in azimuthal localization, and found that localization accuracy decreased substantially when the folds of pinnae were filled with putty.

When binaural information is removed from the stimulus by occlusion or impairment of one of the ears, localization is seriously degraded in azimuth but only slightly impaired in elevation. Azimuthal localization judgments are extremely inaccurate, on

Eight Bit, LLC
Exhibit 2013
IPR2023-01024
Page 176

the side of the occluded ear (Wightman & Kistler, 1987; Slattery & Middlebrooks, 1994) and exhibit a strong lateral bias on the side of the normal ear (Slattery & Middlebrooks, 1994; Butler, 1987). In contrast, elevation accuracy is only slightly degraded under monaural listening conditions (Oldfield & Parker, 1986; Slattery & Middlebrooks, 1994).

Source characteristics also have an influence on distance perception. In particular, the perceived distance of an unfamiliar source can be increased by lowpass filtering the stimulus (Little et al., 1992). While this effect is often explained by the greater absorption of high-frequency sound energy by the atmosphere, a more realistic explanation is that low-frequency sound propagates around obstacles in a room more efficiently than high-frequency sound and consequently relatively more low-frequency energy reaches a listener from a distant source in an enclosed space.

Examining auditory localization under a variety of source conditions has significantly advanced our understanding about the mechanisms of spatial hearing. To date, however, no studies have examined the influence of source spectrum or monaural listening conditions on the localization of sound sources near the head. Although researchers have recognized that interaural intensity differences are substantially larger in the immediate vicinity of the head than in the far-field (Hartley & Frey, 1921), and have noted the possible importance of near-field IIDs as distance cues (Coleman, 1963), only in the past few years has near-field localization received serious attention. Recent studies have calculated the near-field head-related transfer function for a rigid sphere model of the head (Brungart & Rabinowitz, 1996) and verified the results with measurements on the surface of a bowling ball (Duda & Martens, 1997). Another study has measured near-field HRTFs on a KEMAR manikin with a compact acoustic source (Chapter 5). The changes that occur in the HRTF in the near-field can be summarized as follows:

- The interaural intensity difference increases dramatically for all frequencies as the source approaches the head.
- The interaural time delay is roughly independent of distance.

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 177

- Diffraction by the head slightly low-pass filters stimuli located close to the ear.
- The high frequency features of the HRTF which are dependent on elevation are relatively insensitive to source distance.
- The high frequency features of the HRTF are governed by the orientation of the source relative to the ear, and are effectively compressed around the interaural axis at close distances.

The current study is based on a previous psychoacoustic experiment which examined near-field localization for a broadband source (Chapter 6). In that study, listeners were asked to place a response sensor at the perceived location of a noise burst stimulus generated by a compact, non-directional source at a random location in the right hemisphere. The major conclusions of that study are summarized below:

- Azimuthal localization accuracy decreased only slightly as the source approached the head. This relative independence of directional localization on distance may indicate that horizontal localization in the near-field is dominated by ITDs rather than IIDs. Azimuthal localization was, however, inaccurate at close distances directly in front of the subject's face.
- Elevation localization accuracy did not change consistently with source distance. Two subjects performed better at short range than at long range, and two subjects performed better at long range than at short range.
- Distance localization was accurate and relatively unbiased for lateral sources but inaccurate in the median plane. The results indicated the importance of IID in near-field distance perception.

The previous study focused on a general overview of near-field auditory localization, and provided a map of localization accuracy as a function of the azimuth, elevation, and distance of the source. This study will extend the previous results by measuring relative localization performance under four different listening conditions exploring the consequences of fixing the source amplitude, occluding one ear with

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 178

an ear-plug and muff, and high-pass and low-pass filtering the stimulus. The results in each of these conditions are compared to results from the broadband condition, and to previous far-field experiments under similar conditions. The results are also discussed in relation to the near-field characteristics of the HRTF.

7.3 Experiment 1

7.3.1 Method

The procedure used in Experiment 1 was identical to the one used in the earlier study (Chapter 6), but it is summarized here for convenience. Four right-handed male subjects with normal hearing participated in the study, all of whom had previously participated in the earlier study. The subject was seated on a stool in an anechoic chamber, and was provided with a chin-rest to immobilize the head during the experiment. The experimenter stood in a fixed location approximately 1.5 m to the right of the subject, and manually placed the sound source prior to each trial. The source was specifically designed to approximate the compact, non-directional properties of an acoustic point source, and was shaped to allow the experimenter to place the source anywhere in the right hemisphere of the subject without moving from a stationary position. Random numbers, read to the experimenter through an earphone prior to each trial, guided the placement of the source over a range of positions from 0° to −180° in azimuth, −60° to 90° in elevation, and 10 cm to 100 cm in distance.

A Polhemus Three-Space Tracker recorded the stimulus and response positions. The Polhemus source was rigidly mounted on the chin-rest to maintain a constant position relative to the subject's head. One of the two position sensors, mounted near the tip of the point source, measured the stimulus location during each trial. The second position sensor, which was mounted on the end of a 20 cm wooden wand, was moved to the perceived location of the stimulus by the subject and measured the response location. This *direct-location* response method was described fully in an earlier assessment of near-field response paradigms (Chapter 4).

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 179

Under most conditions, the stimulus amplitude was normalized to eliminate amplitude-based distance cues. The normalization factor was calculated from the source position prior to each trial according to the formula:

$$\frac{1}{\frac{50}{\textit{Distance to left ear (cm)}} + \frac{50}{\textit{Distance to right ear (cm)}}}.$$

This normalization factor emphasizes the distance from the source to the right ear when the source is close to the head, but includes the contribution of both ears when the source is further away. In addition to this normalization, the amplitude of the stimulus was roved randomly over a 15 dB range (from 0 dB to 15 dB of attenuation in 1 dB steps).

The overall procedure for each trial can be summarized as follows:

1. The subject closed his eyes while the experimenter moved the source to a random location according to three random numbers generated by the control computer.

2. The control computer recorded the location of the point source, determined the proper amplitude normalization factor, and generated the stimulus.

3. The experimenter moved the source to a neutral position, and told the subject to open his eyes and move the response sensor to the perceived location of the sound.

4. The subject finished the response, the experimenter pressed the response switch to tell the computer to record the response location, and the subject closed his eyes in preparation for the next trial. Note that the subject was not provided with any feedback.

Stimulus

The stimulus consisted of five short (150 ms) bursts of noise, separated by 30 ms intervals. The waveforms were constructed from white Gaussian noise, which was first low-pass filtered with a gentle 6 dB/octave roll-off. The Gaussian stimulus was

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 180

Table 7.1: Stimulus conditions. In the monaural stimulus condition, binaural cues were eliminated by occluding the ear subject's left ear with an ear-plug and an ear-muff modified to cover only the one ear. The Max SPL is the maximum sound pressure level generated by the stimulus when the roving attenuation level was at 0 dB, measured when the distance to the head was 1 m.

Stimulus		Passband	Binaural	Roved	Gating	Max SPL
Broadband	(BB)	0.2-15 kHz	y	y	Rect.	59 dBA
Fixed Amplitude	(FA)	0.2-15 kHz	y	n	Rect.	51 dBA*
Monaural	(MO)	0.2-15 kHz	n	y	Rect.	59 dBA
Highpass	(HP)	3-15 kHz	y	y	$\cos^2$	51 dBA
Lowpass	(LP)	0.2-3 kHz	y	y	$\cos^2$	66 dBA

* In the fixed-amplitude condition roving was not used, but the signal was attenuated 8 dB to place its amplitude in the middle of the roved range. The value given was measured at 1 m and, since the amplitude was not normalized for distance, the effective amplitude was greater at locations closer than 1 m.

then band-pass filtered with sharp (120 dB/decade) roll-offs in the stop-bands and a pass-band determined by the stimulus condition. The waveforms were also filtered to flatten the irregular spectral response of the point source. In the baseline stimulus conditions, rectangular gating was used, while in the highpass and lowpass conditions, 50 ms $\cos^2$ ramps were used to reduce transients. In each stimulus condition, the maximum amplitude (prior to normalization) was chosen to generate the greatest possible non-distorted output from the point-source. The source produced low-frequency sound more efficiently than high-frequency sound, resulting in relatively lower output in the highpass condition. The details of all five stimulus conditions are provided in Table 7.1.

Data Collection

The data were collected in blocks of 100 trials, with each block requiring approximately 20 minutes. Four or five blocks of trials were recorded in each 2-hour session. The data was collected consecutively for each of the five conditions. First, each subject participated in a few trial blocks to familiarize them with the experimental procedure. Next, 2000 trials per subject were collected in the broadband condition of the previ-

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 181

ous study. Then 500 trials per subject were collected consecutively in the broadband, monaural, high-pass, low-pass, and fixed-amplitude conditions. The subjects were informed that the source level was constant in the fixed-amplitude condition, and of course understood the nature of the monaural condition, but only subject DSB was aware of the composition of the high-pass and low-pass stimuli.

Coordinate system

When dealing with locations close to the head the coordinate system, and in particular the location of the origin within the head, must be chosen carefully. In this experiment, the coordinate system was based on the geometry of the subject's head. Prior to each block of trials, one of the position sensors was used to record the locations of the subject's nose, left ear canal opening, and right ear canal opening. These locations were used to define a coordinate system with the y axis along the interaural axis, the x axis as the perpendicular bisector of the interaural axis passing closest to the nose, and the z axis as the cross-product of the y and x axes. In this coordinate system, positive azimuth locations are in the left hemisphere, and positive elevations values are above the horizontal plane. The location directly in front of the head is at 0° azimuth and 0° elevation.

7.3.2 Directional Localization

Data Analysis

Four metrics were used to evaluate directional localization accuracy in each of the five stimulus conditions: the overall angular error, and, in azimuth and elevation, the signed errors, bias-corrected unsigned errors, and stimulus-response correlation coefficients. The overall angular error is simply the angle between the vector from the origin to the stimulus location and the vector from the origin to the response location. This error is always non-negative, and its mean value is an indication of overall accuracy (including the effects of both systematic directional response bias and response variability). The inability to distinguish between response bias and response

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 182

variability limits its usefulness, but the angular error does allow direct comparison of the results with previously published localization studies.

The signed errors in azimuth and elevation are capable of distinguishing between bias and variability. The mean signed error is a direct measure of response bias, while the standard deviation of the signed error is a measure of response variability. Unfortunately, the mean signed error across all locations will include both response variability and bias in an experiment with continuously distributed source locations and an underlying subject response bias that varies with source location. Most previous studies have repeatedly measured localization accuracy at a discrete set of response locations, allowing independent calculation of the mean and standard deviation at each location. In this experiment, however, the source locations were continuously distributed in space and no repeated measures were available. Although it is possible to collect trials in a particular region together into a bin and perform statistics on the signed error within each bin, this will result in artificially high values of standard deviation to the extent that the response bias varies over the region spanned by the bin. A metric of response error was developed to isolate response variability from the effects of response bias. This error value, the bias-corrected error, was defined as the mean absolute (unsigned) error between the response location on each trial and the second-order polynomial best fitting the stimulus data to the response data. The bias-corrected error is similar to the standard deviation in traditional, repeated measures localization experiments in that it measures the spread of responses around the mean, but it also accounts for mean response biases that change with source location.

The bias-corrected error measures the spread of subject responses around the mean, but it does not provide much information about overall performance in the localization task. For example, a subject who responds at a fixed location independent of the stimulus position will have zero bias-corrected error. The degree to which the response location depends on the stimulus location can be measured by the linear correlation coefficient between the stimulus and response locations. Recall that the square of the correlation coefficient measures represents the percentage of variance in the response location which can be explained by a simple linear relationship between

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 183

the stimulus and response locations. A weakness of the correlation coefficient is that it is inherently non-linear, and cannot simply be averaged across subjects. The Fisher transformation (Devore, 1991) converts the correlation coefficients into approximately Gaussian random variables, which can be averaged or analyzed statistically.

Front-back confusions require special attention in the analysis of directional localization accuracy. In order to reduce the number of reversals in the vicinity of $-90°$ in azimuth, a relatively conservative definition of front-back confusions was used. Trials were considered reversals only if the response location was at least 10° closer to the mirror image of the source across the frontal plane than to the actual source. The number of front-back reversals was calculated in each condition, and the trials where front-back confusions occurred were eliminated from the analysis of directional localization.

Azimuth Performance

Overall, localization in azimuth was most accurate in the broadband and fixed-amplitude conditions, slightly less accurate in the lowpass and highpass conditions, and least accurate in the monaural condition. This trend is seen most clearly from the linear correlation coefficient between the stimulus and response azimuths averaged across all subjects in Table 7.2. Although the differences in the correlation across conditions are small, because of the extreme sensitivity of the correlation coefficient near 1, they are sufficiently large to divide the conditions into three groups. The correlation coefficients in the broadband and fixed amplitude conditions were significantly larger than in the highpass and lowpass conditions, while the correlation in the monaural condition was significantly lower than in any other condition (from Fisher transformation of r, α=0.005).

The same general ordering is seen in the mean absolute bias-corrected error, which measures the spread of response locations around the mean (Table 7.2 and the lower-left value in each panel of Figure 7-1). Again the fixed amplitude and broadband conditions had a lower mean error than the highpass condition. The monaural error was not significantly greater than the error in the lowpass condition, but it was

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 184

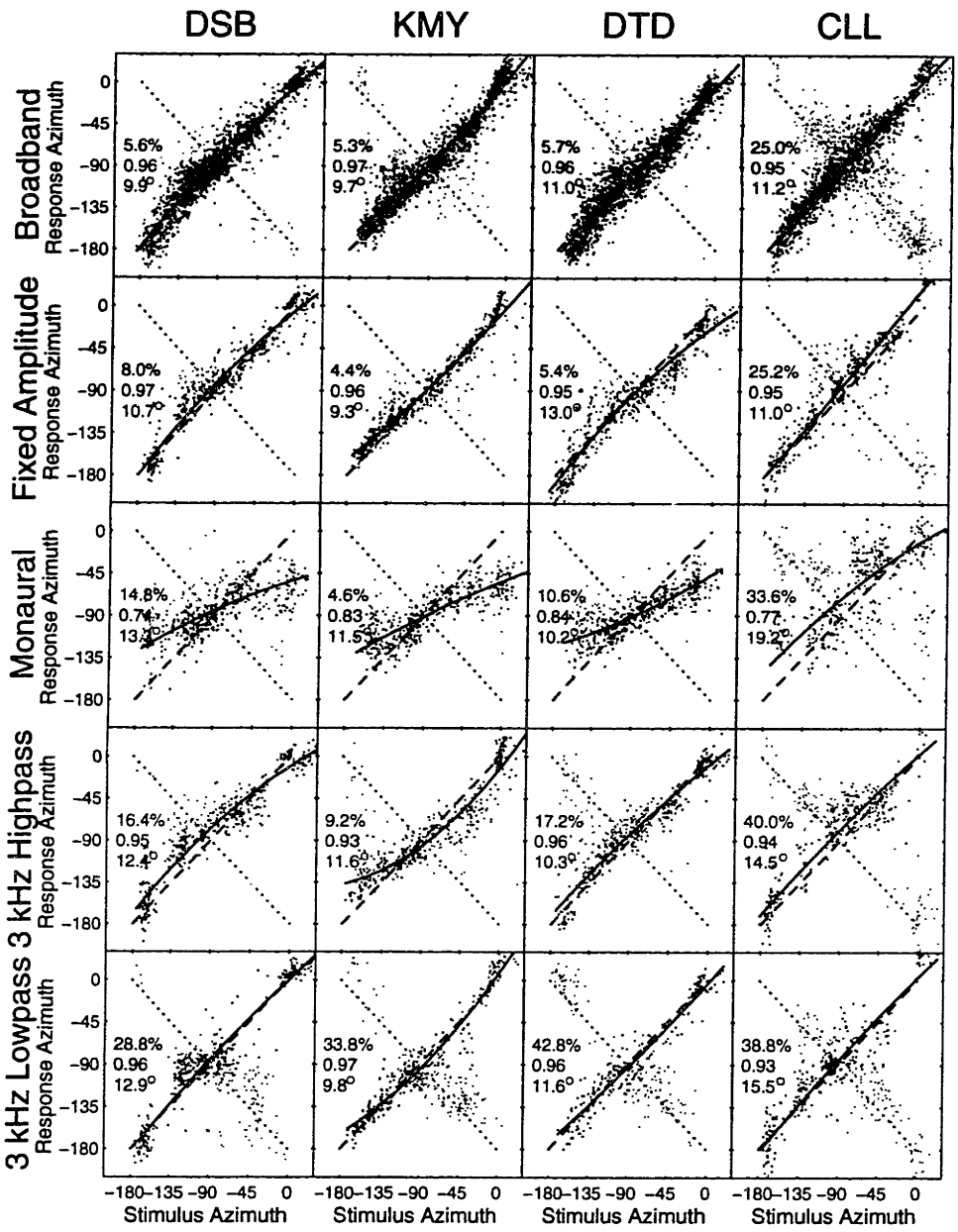


Figure 7-1: Raw stimulus-response data in azimuth from Experiment 1. The data are divided into normal trials (large dots) and trials where a front-back reversal occurred (small dots), and the percentage of reversals is the top number at the left of each panel. In each panel, a dashed line shows the location of "ideal" responses and the dotted line shows the location of perfect reversals. The correlation coefficient for each condition is at the middle left of the panel, while the mean bias-corrected error is the bottom number at the left of each panel. The solid black line is the second order polynomial fit of the response location to the stimulus location used to calculated the bias corrected errors.

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 185

Table 7.2: Summary of statistics measuring directional localization accuracy in each condition. The standard deviations were calculated separately for each subject and averaged across subjects to give the value in the table. Note that the unsigned errors are the bias-corrected errors calculated from the absolute value of the difference between the response value and the best second order polynomial fit of the stimulus locations to the response locations. The data were divided into three bins in azimuth ($< -120°$, $-120°$ to $-60°$ and $> -60°$), three bins in elevation ($< -20°$, $-20°$ to $20°$, and $> 20°$), and three bins in distance (< 25 cm, 25 cm to 50 cm, and > 50 cm). The mean bias-corrected errors were calculated separately in each of these nine bins for each of the four subjects and averaged together. The overall correlation coefficients were calculated separately for each of the three elevation bins in azimuth, and each of the three azimuth bins in elevation, and then averaged together using the Fisher transform. Note that the correlation values in this table are more informative than the correlation values in Figure 7-1 because they represent the average correlation value at three elevation locations, and the biases change significantly enough in the three elevation regions to produce a higher average correlation than the overall correlations shown in Figure 7-1. BB=Broadband; FA=Fixed Amplitude; MN=Monaural; HP=Highpass; LP=Lowpass.

	BB		FA		MN		HP		LP	
Angular:	16.6°	(10.0°)	17.0°	(10.0°)	29.7°	(17.7°)	22.6°	(13.6°)	30.2°	(20.0°)
Azimuth:										
Signed	1.3°	(14.2°)	1.7°	(17.0°)	−1.0°	(31.2°)	3.8°	(18.6°)	0.6°	(16.6°)
Unsigned	9.4°	(7.8°)	8.9°	(7.2°)	11.1°	(8.7°)	9.8°	(7.9°)	10.6°	(8.7°)
Correlation	0.964		0.967		0.812		0.953		0.956	
Reversals	10%		11%		16%		21%		36%	
Elevation:										
Signed	3.2°	(14.7°)	2.4°	(13.2°)	−0.4°	(21.5°)	2.1°	(21.1°)	−8.1°	(31.3°)
Unsigned	8.5°	(7.3°)	7.3°	(6.0°)	9.4°	(8.1°)	9.7°	(8.2°)	8.6°	(7.1°)
Correlation	0.918		0.942		0.803		0.858		0.427	

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 186

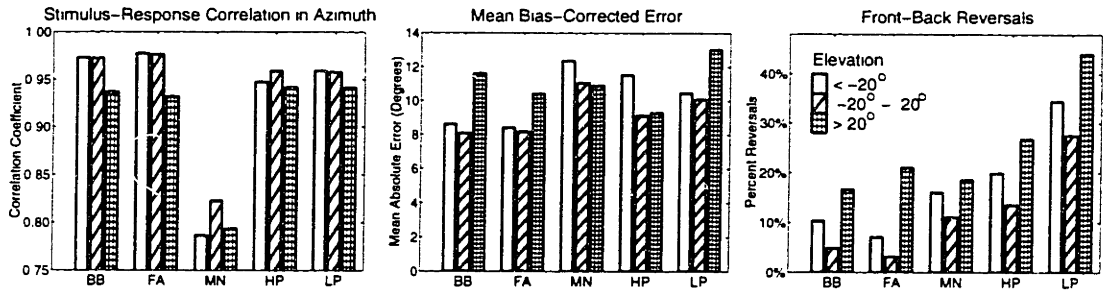


Figure 7-2: Azimuth performance as a function of source elevation. The bias-corrected errors were calculated separately for each subject in each of three distance bins (<25 cm, 25-50 cm, >50 cm) and averaged together. BB=Broadband; FA=Fixed Amplitude; MN=Monaural; HP=Highpass; LP=Lowpass.

significantly higher than in the other three conditions. Also note that the standard deviation of the signed azimuth error (Table 7-1) is greatest in the monaural condition, and smallest in the broadband condition.

Another important measure of azimuth localization is the frequency of front-back confusions (top left of each panel in Figure 7-1 and Table 7.2). The overall percentage of reversals varied considerably across the four conditions. The broadband and fixed amplitude conditions produced the smallest number of reversals, and were not significantly different from one another. The monaural condition produced fewer reversals than the highpass condition, and the lowpass condition produced the largest number of reversals by a wide margin (ordering significant at the α=0.005 level from one-tailed t-tests). Note that subject CLL experiences front-back confusions more frequently than any other subject. In every condition except lowpass, CLL reverses at least twice as many responses than any other subject. In the lowpass condition, CLL's reversal rate did not increase as much as the other subjects, but may have been limited by a ceiling effect at the level of pure guessing (50% reversals).

The previous study describing the broadband condition indicated that azimuth performance was strongly dependent on elevation, and that performance was best at middle elevations and worst at high elevations. This was also generally the case in the other four stimulus conditions (Figure 7-2). In every condition except monaural listening, the correlation coefficient was significantly lower at high elevations (> 20°)

Eight Kilohertz
Exhibit 2013
IPR2023-01024
Page 187

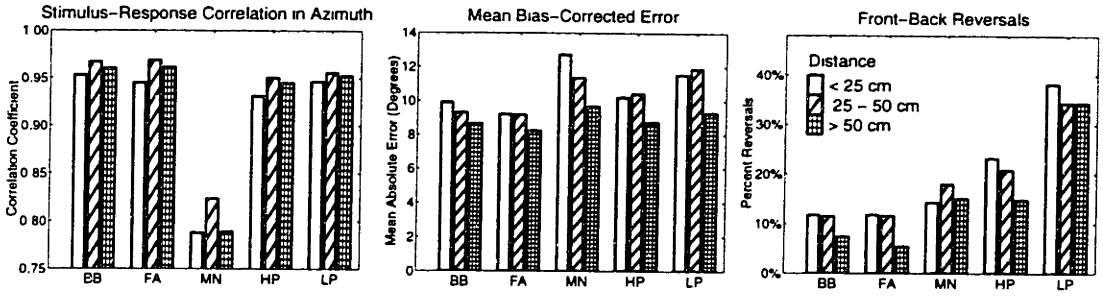


Figure 7-3: Azimuth performance as a function of source distance. The bias-corrected errors were calculated separately for each subject in each of three elevation bins ($< -20°$, $-20°$ to $20°$, $> 20°$) and averaged together. BB=Broadband; FA=Fixed Amplitude; MN=Monaural; HP=Highpass; LP=Lowpass.

than at middle elevations ($-20°$ to $20°$) (from Fisher transform, α=0.025). Front-back reversals were also most frequent at high elevations and least frequent at middle elevations in each of the five conditions (one-tailed t-tests, α=0.005). The only deviations from the general pattern are the unusually large bias-corrected errors at low elevations in the monaural and highpass conditions. It is not surprising that performance was generally worse at high elevations because azimuth is an increasingly sensitive measure of location at high and low elevations. Asymmetrical performance at high and low elevations occurs at least in part because interference by the torso and chin-rest prevented placement of the source much below $-60°$ in elevation, while some stimuli at high elevations were near $+90°$.

Another finding of the previous study of near-field localization was that the relationship between azimuthal localization and distance varied across subjects, but that performance was generally worse for close sources (< 25 cm) than for far sources (> 50 cm). Again, this finding was consistent across all of the binaural stimulus conditions (Figure 7-3). The correlation coefficient was consistently lower at close distances than at far distances, while the bias-corrected error and the percentage of reversals were higher at close distances than far distances. The correlation coefficient was always highest (slightly) at intermediate distances (25-50 cm). In the monaural condition, the bias-corrected error was largest at close distances, and the number of reversals was greatest at intermediate distances, but the differences in the correlation coefficient

Apple Inc.
Exhibit 2013
IPR2023-01024

cient were not significant. Note that, while localization in azimuth does depend on distance and is generally less accurate at close distances, the decrease in performance for nearby sources is relatively modest.

7.3.3 Elevation Performance

Elevation performance, measured by the correlation coefficient, was best in the fixed amplitude and broadband conditions, worse in the highpass and monaural conditions, and very poor in the lowpass condition (Table 7.2). This ordering was highly significant (one-tailed t-tests, $\alpha = 0.001$). Response variability, measured by the bias-corrected unsigned error, was lowest in the fixed amplitude condition, slightly higher in the broadband and lowpass conditions, and highest in the monaural and highpass conditions. Note that the lowpass condition is an unusual case where the correlation coefficient was extremely low but the response variability was modest. This discrepancy occurs primarily because subject DSB clustered all of his responses around 0° in elevation in the lowpass condition (Figure 7-4).

The raw data (Figure 7-4) show that the subjects made less accurate responses at high elevations than at lower elevations. The spread of responses is generally larger in the upper hemisphere, and the second-order polynomial best fitting the response data to the stimulus locations decreases in slope at high elevations, indicating lower sensitivity.

The mean bias-corrected error in elevation was strongly dependent on source azimuth, but the correlation coefficient generally was not (Figure 7-5). The correlation coefficient was slightly lower for sound sources behind the listener (azimuth $< -120°$) than for sources in front or to the side, especially in the monaural and highpass conditions. The mean bias-corrected error was always lowest for lateral sources ($-120°$ to $-60°$) and was generally highest for sources in the rear. Elevation performance is generally better for lateral sources than for medial sources because changes in elevation near $\pm 90°$ produce changes in the IID and ITD in addition to the pinnae-based elevation cues.

As was the case for azimuth, localization accuracy in elevation was only slightly

Apple Inc.
Exhibit 2013
IPR2023-01024
Page 189

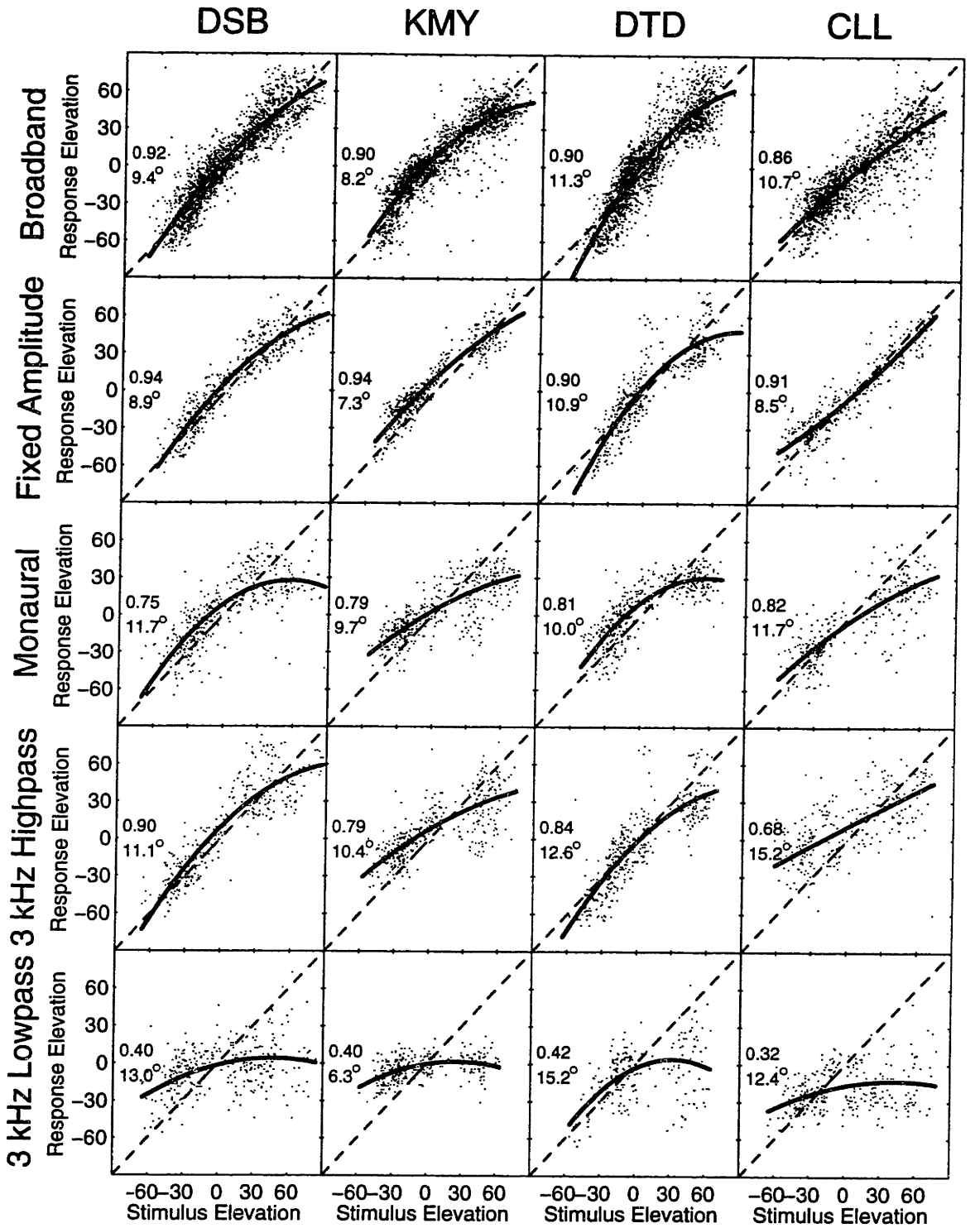


Figure 7-4: Raw stimulus-response data in elevation for Experiment 1. As in Figure 7-1, the dotted line indicates "correct" responses, while the solid black line shows the best second order polynomial fit of the response data to the stimulus data. The number at the top left of each panel is the linear correlation coefficient, while the number at the bottom left of each panel is the mean bias-corrected error.

Eight KHz, LLC Exhibit 2013

IPR2023-01024
Page 190

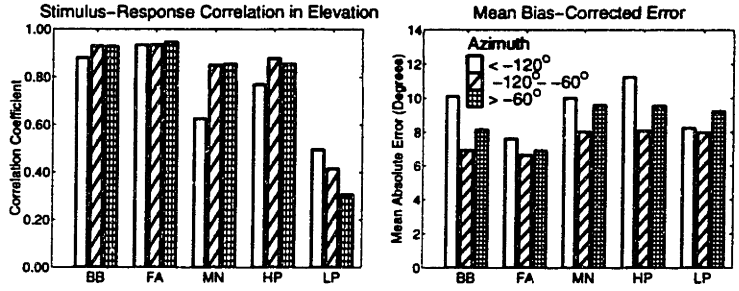


Figure 7-5: Elevation performance as a function of source azimuth. The bias-corrected errors were calculated separately for each subject in each of three distance bins (<25 cm, 25-50 cm, >50 cm) and averaged together.

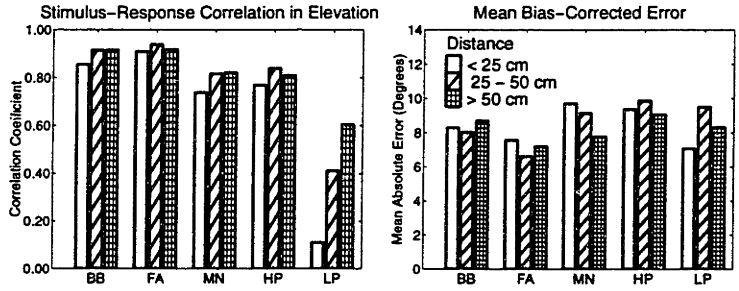


Figure 7-6: Elevation performance as a function of source distance. The bias-corrected errors were calculated separately for each subject in each of three azimuth bins ($< -120°$, $-120°$ to $-60°$, and $> -60°$) and averaged together.

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 191

dependent on distance (Figure 7-6). The correlation coefficients in elevation were always significantly lower at close distances (< 25 cm) than at middle distances (25-50 cm) (Fisher transformation, α=0.005). Note that the decrease in correlation is deceptively large in the lowpass condition because the correlation coefficient is relatively insensitive at low correlation values. The mean bias corrected errors was not dependent on distance in any systematic way.

7.3.4 Response Bias

Directional biases are displayed graphically in Figure 7-7. In the broadband conditions (broadband and fixed-amplitude), the response biases are relatively modest, and the general patterns of bias are similar across each of the three distance bins. Note, however, that there is a lateral bias for the closest sources above and in front of the listener in these conditions. In the highpass condition, the response biases tend to be larger than in the broadband conditions, but they continue to be roughly constant across distance. The most interesting response biases are in the monaural and lowpass conditions. In the monaural condition, the responses are strongly biased in the direction of the unoccluded ear. Such a pattern of bias has been reported in previous studies of monaural localization. In the lowpass condition, virtually all of the responses are clustered around the horizontal plane, independent of the actual source location. This pattern of response bias provides further evidence that the subjects were receiving little or no information about the elevation of the source when the stimulus is limited to frequencies below 3 kHz.

7.4 Discussion

The results of the broadband condition were discussed in great detail and compared to previous results from far-field localization experiments in a previous paper (Chapter 6), and those analyses will not be repeated here. The focus of this experiment was the relative accuracy of near-field localization accuracy in the five stimulus conditions. The results in each of the four new stimulus conditions are discussed below.

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 192

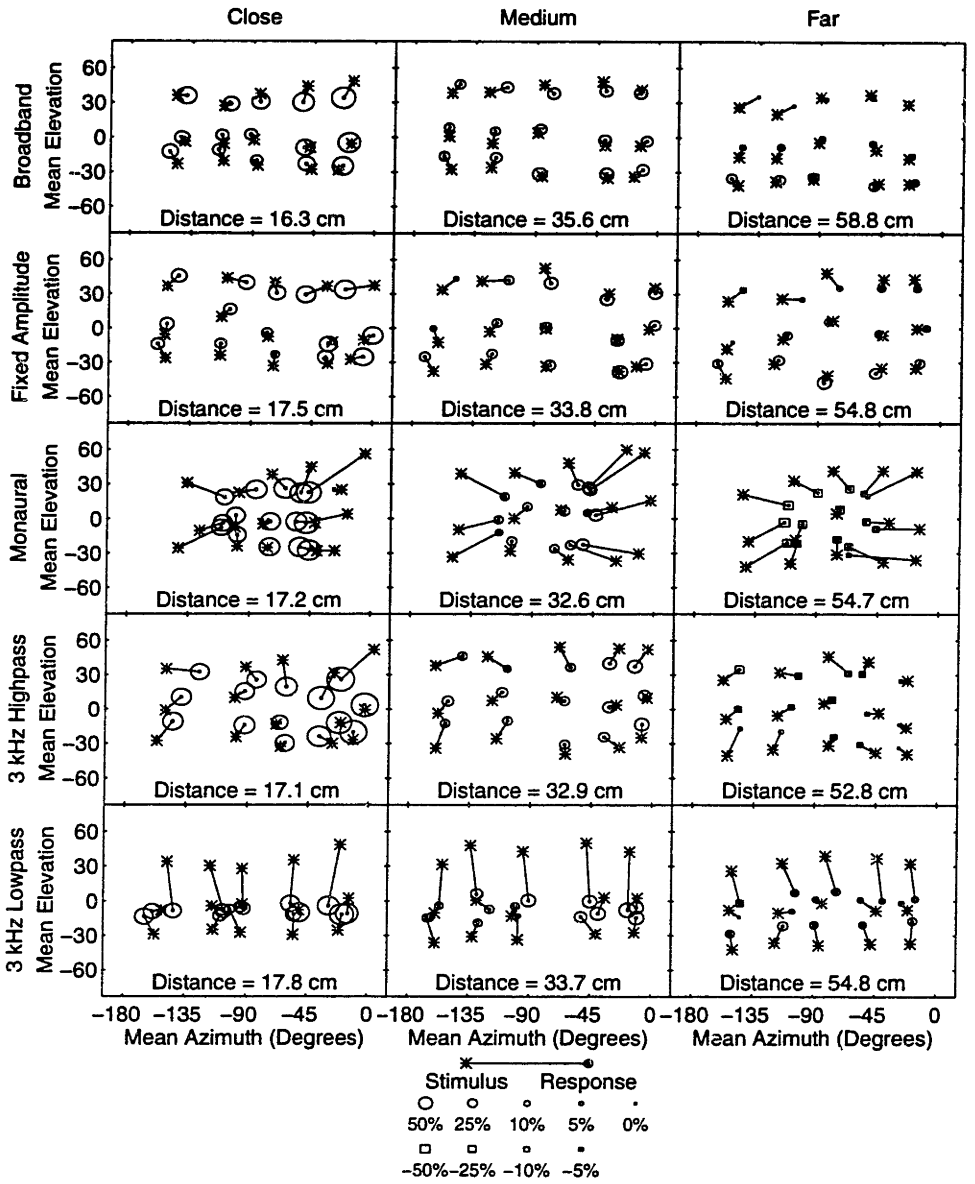


Figure 7-7: Response Biases. Direction of response bias as a function of source location. In this figure, the trials have been sorted into three equal-sized non-overlapping bins in elevation, eight equal-sized overlapping bins in azimuth, and three equal-sized bins in elevation. Only five of the eight bins in azimuth are shown, for clarity. The * is the mean stimulus location within each bin, while the . is the mean response location. The circles indicate that the subject overestimated distance (see legend), while the squares indicate the subject underestimated distance.

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 193

- **Fixed Amplitude:** Directional localization performance in the fixed amplitude condition was virtually identical to the random-amplitude (broadband) condition in every way. The angular error of 16.6° is nearly the same as the 17° error in the broadband condition. The correlation coefficient, mean bias-corrected error, and number of reversals in azimuth were also nearly identical. The direction and magnitude of the biases were also similar in the two conditions. In fact, the only significant difference between the two conditions was that elevation performance, measured both by correlation and by the bias-corrected error, was slightly better in the fixed amplitude condition.

  The similarity between the broadband and fixed amplitude conditions is not surprising. The only difference between the two conditions was the amplitude of the source, and under normal conditions directional localization judgments are based on interaural differences and spectral cues, and should be effectively independent of amplitude.[1] The similarity between the two conditions, therefore, should be viewed simply as a measure of the repeatability of the experimental procedure.

- **Monaural:** In contrast to the fixed amplitude condition, performance in the monaural condition was drastically different from the broadband condition. The decrease in performance was most pronounced in azimuth, where response variability was significantly larger than in the broadband condition. The response biases, which were characterized by a strong pull in the direction of the ear, were also dramatically different from the other conditions. This type of response behavior has been documented in previous studies of monaural localization (Wightman & Kistler, 1987). Perhaps more interesting is the relatively modest degradation in elevation accuracy in the monaural condition. Elevation performance was comparable to the binaural highpass conditions and better than the binaural lowpass condition. The increase in front-back confusions

[1]Note that under unusual listening conditions, e.g. a unilateral hearing impairment, directional judgments may be based, in part, on the overall amplitude of the source.

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 194

(from 10% to 16%) was also relatively small, but note that the bias towards $-90°$ in azimuth tended to move reversed responses into the region where reversals were not 20° closer to the mirror image of the source location than to the source location and were not counted as reversals. The elevation and reversal results are consistent with the widely held belief that the directional properties of the pinna at high frequencies are largely responsible for resolving the locations of sources within the "cone of confusion" where interaural differences are ambiguous. The available cues from one ear are apparently sufficient to allow subjects to make reasonably accurate judgments about elevation and correctly resolve front-back confusions, even though they are unable to accurately judge azimuth without binaural difference cues.

- **Highpass:** In the highpass condition, localization accuracy was slightly lower than in the broadband condition both in azimuth and in elevation. Although no data are available directly comparing the localization of broadband noise to the localization of highpass noise, these results were roughly consistent with those of Musicant and Butler (1984) who measured localization performance for noise high-pass filtered at 4 kHz. Musicant and Butler did, however, report a much smaller number of front-back confusions (approximately 0.5% vs. 21% in this experiment), possibly because their 5 s stimulus may have allowed some small head motions. The results of the highpass localization condition are particularly interesting in the near-field because the interaural intensity difference, which according to the widely held duplex theory of localization dominates azimuth localization at high frequencies, varies substantially with distance across all frequencies. Since the IID does not correspond to a unique lateral position in the near-field, one might expect listeners to confuse the distance and direction of a near-field source containing only high frequencies. In particular, it seems likely that subjects would be unable to distinguish between a close source near the median plane and a more distant source at a more lateral position with the same average IID. At the very least, one would expect this ambiguity to

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 195

result in a very strong lateral bias for sources near the head. The results do not, however, conform to these expectations. Although there is a relatively strong lateral bias in the highpass condition (Figure 7-7), the bias is only slightly larger at close distances than at the furthest measured distances. Furthermore, an examination of the raw azimuth data (Figure 7-1) reveals that only one subject (KMY) exhibited a strong tendency to compress his responses around $-90°$ in the highpass condition (note that almost all of his responses fall between the dotted and dashed lines).

The most probable explanation for the failure of the subjects to exhibit significant lateral bias in the highpass experiment is that the subjects were able to obtain salient interaural timing information either from the onset of the stimulus or the envelopes of the high-frequency stimulus. The 50 ms ramps used to gate the high-pass stimulus may not have been sufficient to eliminate all onset cues. Furthermore, researchers have demonstrated that subjects are sensitive to envelope delays in bandpass noise (Trahoitis & Bernstein, 1986). The availability of accurate interaural timing information from either of these sources would explain the reasonably accurate localization judgments in the high-pass condition, and the absence of a strong lateral bias for nearby sources.

- **Lowpass:** In the lowpass condition, subjects were able to make reasonably accurate judgments about the lateral position of the source, but they were almost completely unable to estimate source elevation or resolve front-back confusions. The subjects tended to respond near the horizontal plane, independent of the true source elevation, and the correlation coefficient in elevation was also only about half as large as in the second worst condition. Three of the subjects experienced more than twice as many front-back reversals in the lowpass condition than in any other condition. Again, these results illustrate the importance of the high-frequency directionality of the pinna in determining the elevation of a source and distinguishing between sources in the front and rear hemisphere. Since these high-frequency directional cues only occur above approximately 4

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 196

kHz, they were not available in the low-pass condition. Consequently, the subjects exhibited large numbers of reversals and very poor elevation accuracy.

The overall directional accuracy in each of the five stimulus conditions, and in particular the relative performance in each of the five conditions in azimuth, elevation, and front-back confusions, was consistent with the results of previous localization studies conducted in the far-field. This consistency both confirms the procedure used in the experiment, and tests the conjecture, based on measured head-related transfer functions in the near-field, that directional localization performance is similar in the near- and far-fields. This experiment also provided the opportunity to evaluate exactly how directional accuracy changes with distance in the near-field. The results were mixed, but in general performance was slightly lower both in azimuth and elevation for very near sources (< 25 cm) than for more distant sources (> 50 cm). In all cases, however, the distance dependence was relatively weak. Again, this confirms the observation that directional localization is relatively similar in the near- and far-fields.

7.5 Distance Localization

The directional localization performance measured in Experiment 1 generally conformed to the results of earlier studies which have examined far-field auditory localization under a variety of stimulus conditions. In directional localization, there does not appear to be a large change in performance for sources in the near-field. The accuracy of auditory distance perception, however, was dramatically different in the near-field. As reported previously (Chapter 6), near-field distance perception with a broadband, random-amplitude source is significantly more accurate for lateral sources than for sources near the median plane. Furthermore, the distance judgments for lateral sources were highly correlated with the actual source location (log-stimulus log-response correlation coefficients were in excess of 0.85) and were relatively unbiased. Although it is difficult to make direct comparisons between these results and those of previous far-field distance localization experiments, it appears that near-field

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 197

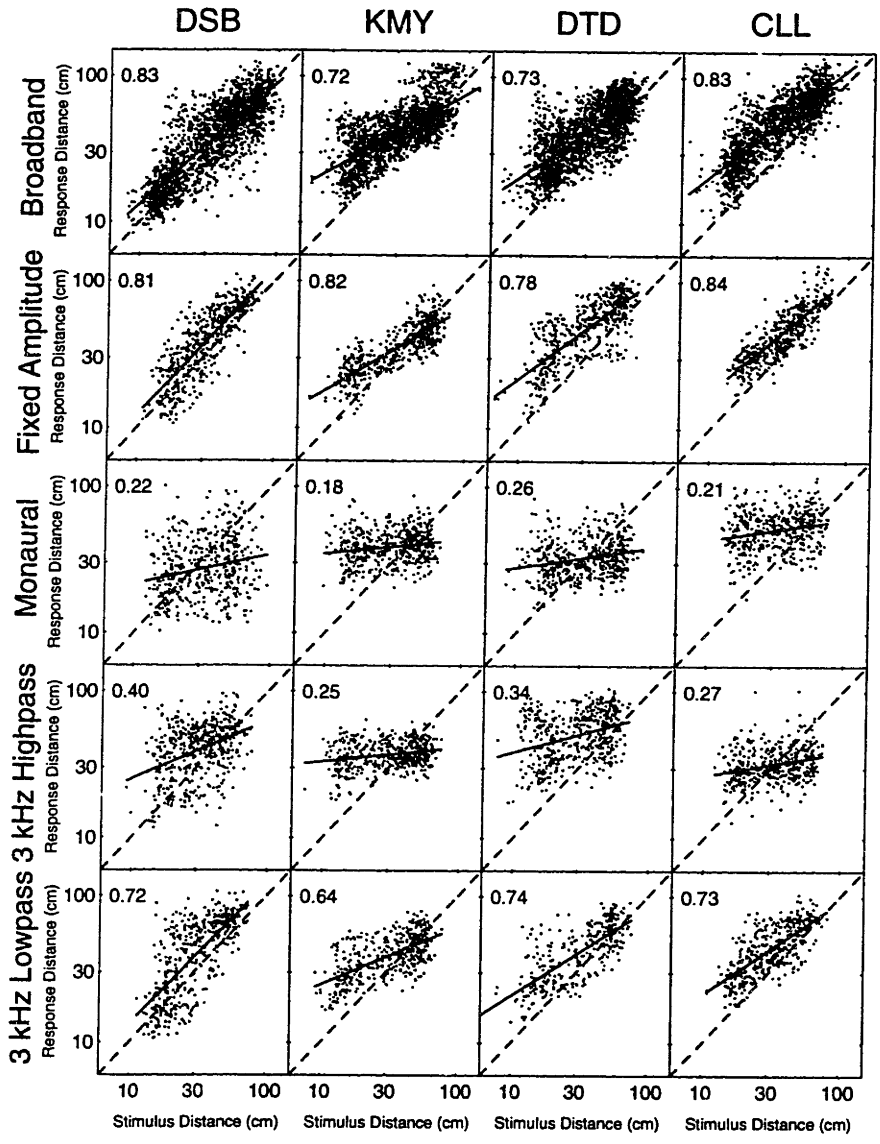


Figure 7-8: Raw Distance Stimulus-Response Data. These data include all azimuth and elevation locations. The errors in distance increased proportionately with stimulus distance, so the data are shown on a log-log scale. The solid line in each panel of the figure represents the best linear fit of the stimulus location to the response location, and the number at the top left of each panel is the correlation coefficient between the log stimulus and log response distances.

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 198

distance perception for lateral sources is significantly better than has been reported in far-field experiments in the absence of room reverberation and amplitude cues.

The results of the current experiments indicate that near-field distance perception is facilitated by binaural information, and imply that low-frequency interaural intensity cues provide the primary distance cue for sources near the head. Overall distance performance was best in the fixed amplitude condition, and only slightly less accurate in the broadband and lowpass conditions (Figure 7-8). Performance in the highpass condition was substantially worse than in the lowpass condition, and performance in the monaural condition was extremely poor. As has been reported in previous localization experiments, the subjects (with the exception of DSB) tended to compress their responses over a narrower range than the range of stimulus locations (generating a linear fit with a slope less than one). In most cases, this caused the subjects to significantly overestimate close distances.

Distance performance was highly dependent on source azimuth, and subjects were much more accurate at judging the distance of lateral sources than sources near the median plane. Figure 7-9 and Figure 7-10 show the raw stimulus response data for sources in front ($-20° <$ to $< 20°$ azimuth) and to the side ($-110° <$ to $< -70°$ azimuth) of the listener. When the source was to the side of the listener, the responses were highly correlated with the stimulus distance in the broadband, fixed amplitude, and lowpass conditions. The responses in the broadband and fixed amplitude conditions were also relatively unbiased except for slight tendency to overestimate distance (note the slope of the best linear fit in this case is near 1). The responses in the monaural and highpass conditions were much less dependent on the stimulus distance, although they are still positively correlated with distance.

When the source was in front (Figure 7-9), performance is substantially degraded in every condition. The correlation was clearly highest in the fixed amplitude condition, but the responses were compressed and the subjects overestimated the distance of nearby sources. In the broadband and lowpass conditions, the responses are only weakly correlated with stimulus distance, while in the monaural and highpass conditions the responses are not significantly correlated with stimulus location at all

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 199

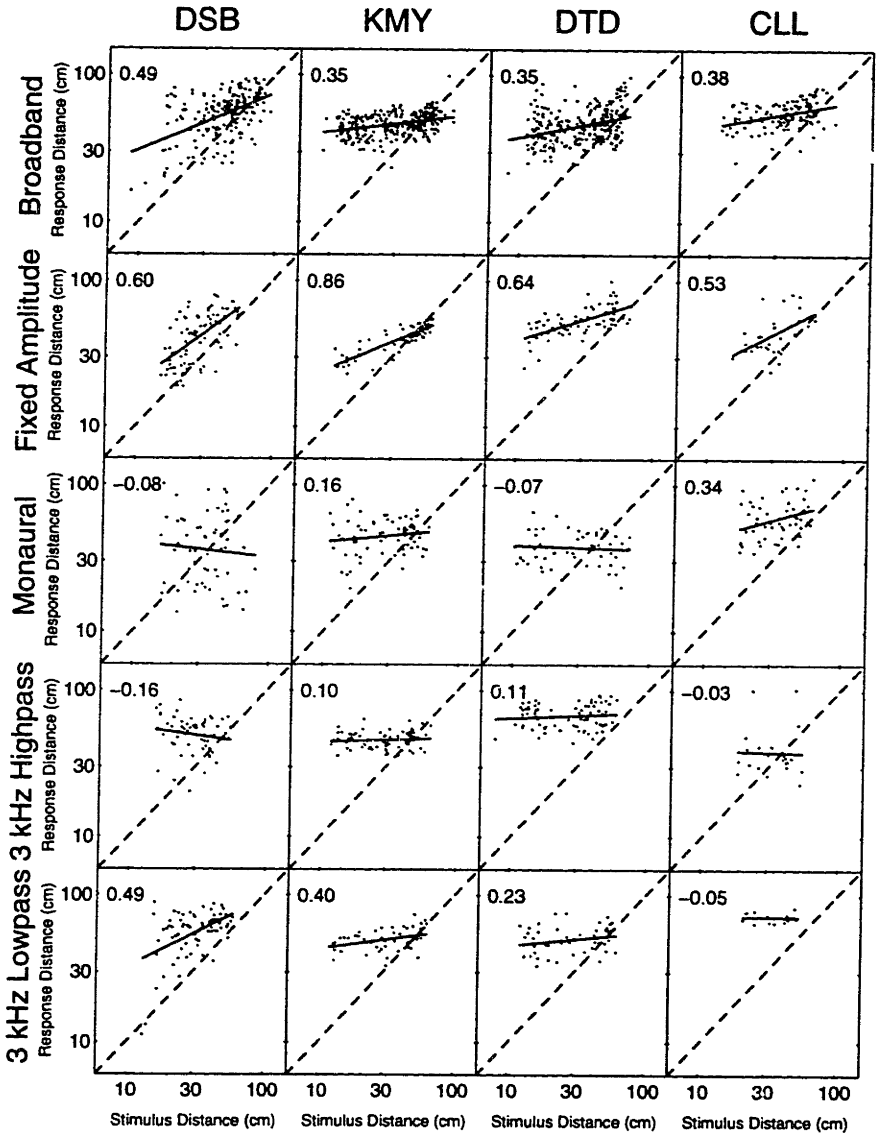


Figure 7-9: Raw Distance Stimulus-Response Data for sources in front ($-20° <$ to $< 20°$). Otherwise similar to Figure 7-8.

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 200

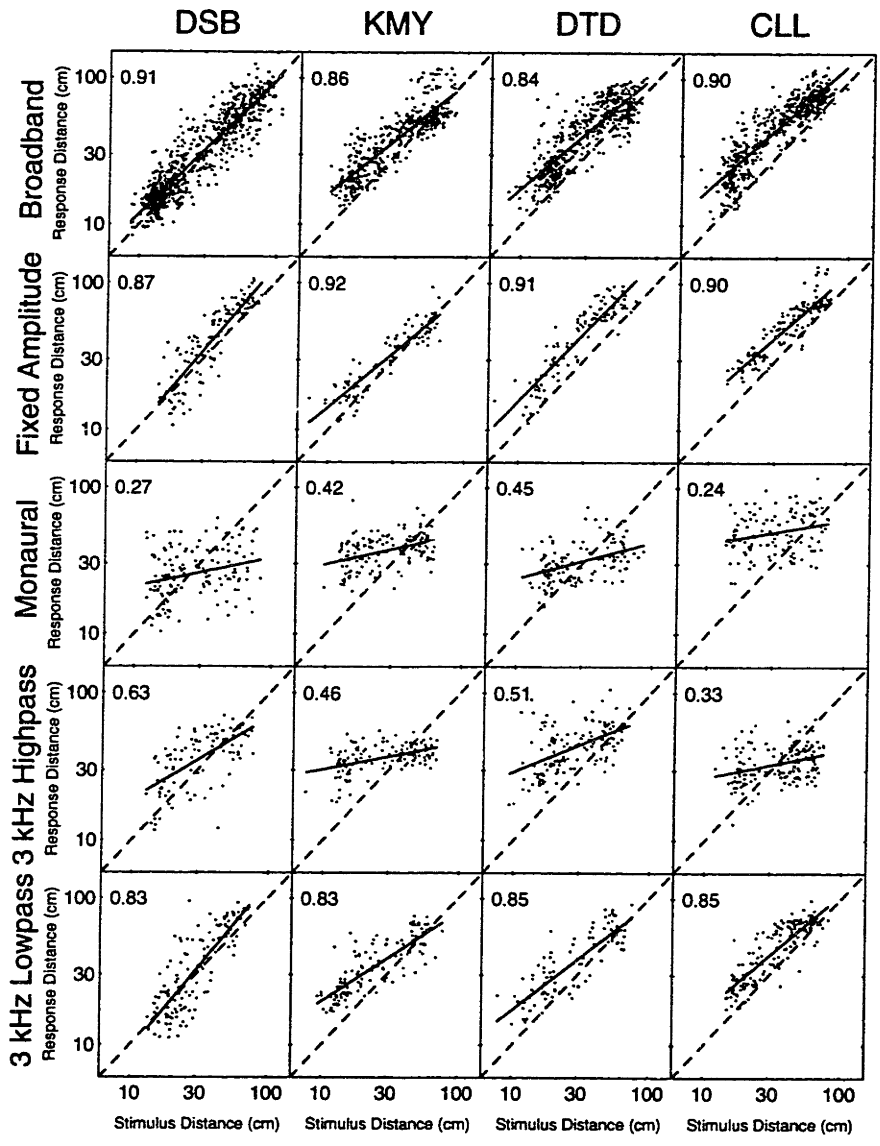


Figure 7-10: Raw Distance Stimulus-Response Data for sources to the side ($-110° <$ to $< -70°$). Otherwise similar to Figure 7-8.

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 201

(except CLL in the monaural case).

A more comprehensive measure of distance performance as a function of azimuth is given in Figures 7-11 and 7-12, in which the data are sorted by azimuth into 13 overlapping bins and the correlation coefficient is shown as a function of the mean azimuth in each bin. In each of the five conditions, performance is lowest near 0° azimuth, increases systematically as the source moves towards −90°. Performance in the fixed amplitude condition, in which all subjects performed best at all azimuth locations, generally decreases less as the source moves toward the median plane than the broadband or lowpass conditions. Consequently, fixed amplitude distance performance was only slightly better than broadband or lowpass near −90°, but was considerably better near 0° and −180°. Performance in the broadband and lowpass conditions was virtually identical everywhere except behind the listener, where the correlation coefficients were slightly higher in the broadband condition. Performance in the highpass condition was substantially worse than in the broadband, monaural, or lowpass conditions, and performance in the monaural condition was extremely poor, with mean correlations never exceeding 0.4.

This trend is consistent for each of the four subjects. Although the data for the individual subjects are considerably noisier than the mean data, the increase in performance for lateral source locations and the relative ordering of the five stimulus conditions was essentially identical for each of the four subjects. Note that subject KMY performed particularly well in the fixed amplitude condition near the median plane, while subject DSB generally performed better in the highpass condition than the other subjects.

These results provide strong evidence that binaural cues are critical for near-field distance perception. In all four binaural conditions, distance judgments were considerably less accurate near the median plane than near −90°. Since interaural time and intensity differences are smallest in the median plane, this result is consistent with a distance perception mechanism based on binaural differences. Further evidence of the importance of binaural differences to auditory distance perception is provided by the extreme inaccuracy of distance judgments when one ear was occluded by an

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 202

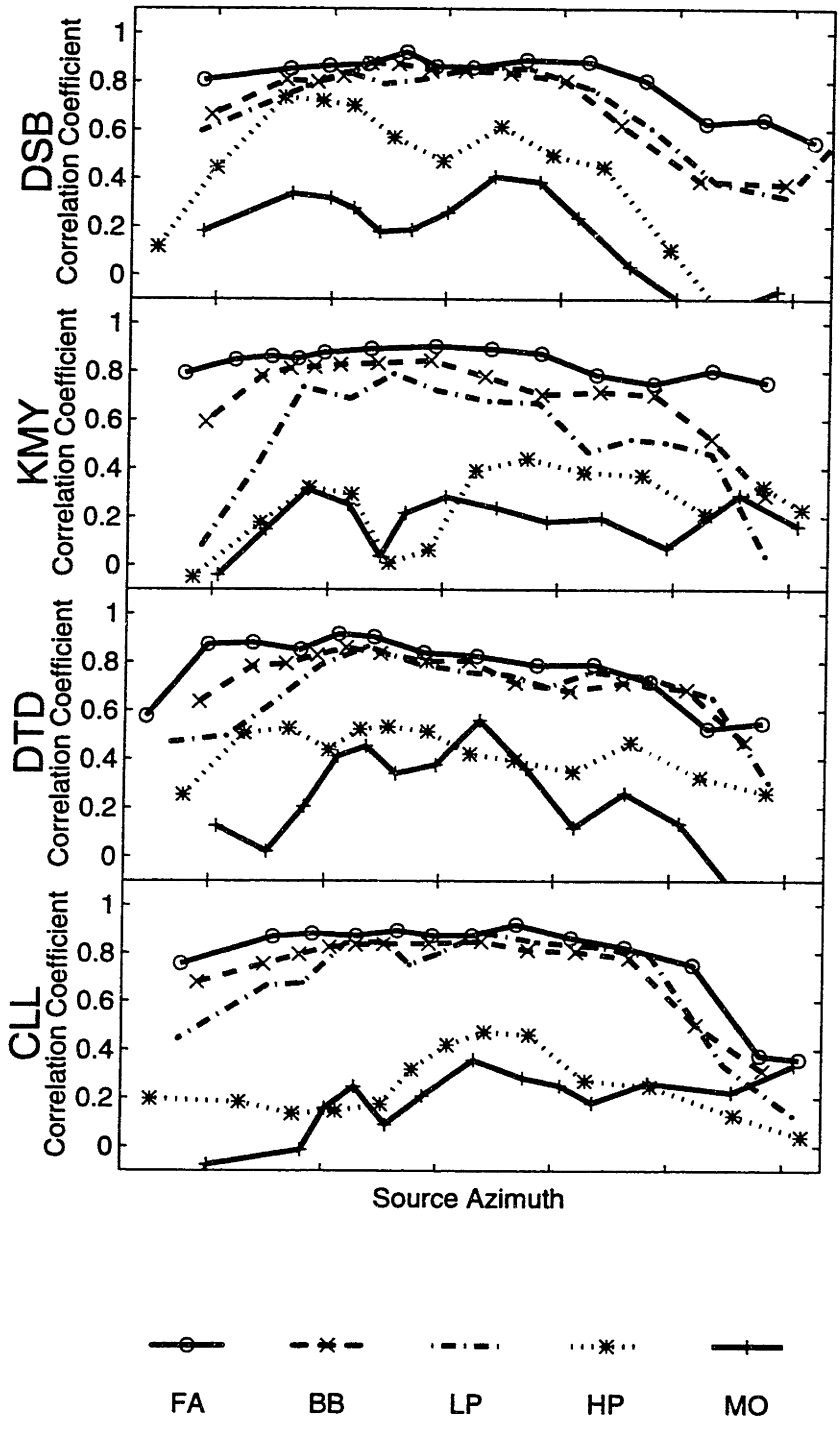


Figure 7-11: Stimulus-response correlation in distance as a function of source azimuth for each subject. The data for each subject in each condition were sorted by azimuth into 13 overlapping bins each containing 14% of the total trials. The correlation coefficient between the log of the stimulus location and the log of the response location in each bin is plotted here as a function of the mean azimuth location of each bin.

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 203

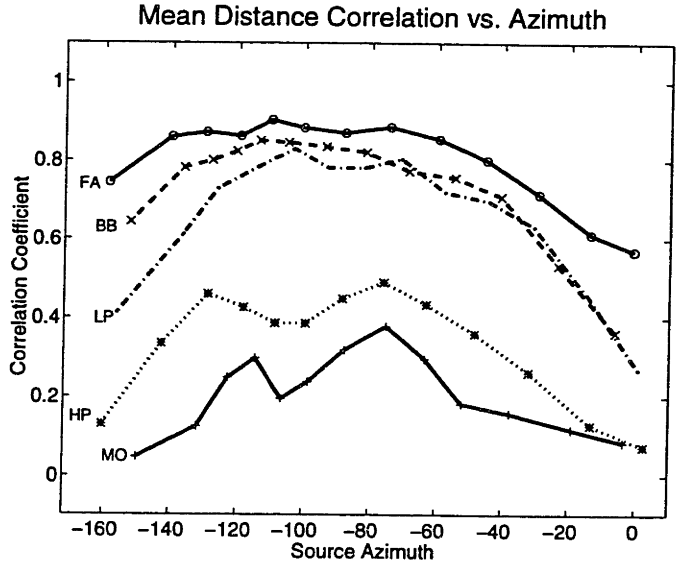


Figure 7-12: Mean overall stimulus-response correlation in distance. The data for each subject in Figure 7-11 has been averaged together by plotting the mean correlation coefficient (averaged using the Fisher transform) in each of the 13 bins versus the mean azimuth value of each bin.

ear plug and ear muff. A distance perception mechanism based on monaural spectral information, rather than interaural differences, would not be degraded as severely in the monaural listening condition.

The results in the fixed-amplitude condition also support the importance of binaural cues to near-field localization. The subjects were able to use the amplitude cue to determine distance in the near-field source, so it is not surprising that they were better able to judge distance in fixed-amplitude condition than in the roved-amplitude conditions. Note, however, that the improvement in distance perception generated by the amplitude cues is relatively small for lateral sources, and relatively large near the median plane. When the source is to the listener's side, binaural information provides most of the distance cue and the additional information provided by the amplitude cue only slightly improves performance. In contrast, the binaural cue is weak near the median plane and the improvement in performance provided by the amplitude cue is more dramatic (the correlation coefficient is more than twice as large in the fixed amplitude condition than in the broadband condition at 0°).

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 204

Based on the results in the conditions tested, and previously measured near-field head-related transfer functions on a KEMAR manikin (Chapter 5), low-frequency interaural intensity differences appear to be the primary cue for auditory distance perception in the near field. The results in Figure 7-12 show that distance localization performance is comparable when the source is broadband and when the source is low-pass filtered below 3 kHz. When the source is high-pass filtered above 3 kHz, however, performance degrades substantially. These results indicate that low-frequency components, below 3 kHz, are essential to near-field distance perception. Performance is also poor in the monaural case, suggesting that the low-frequency distance information is due to interaural differences. Since the near-field HRTFs show that the ITD is roughly independent of distance in the near-field, while the IID increases dramatically as a lateral source approaches the head, the low-frequency IID must be the primary near-field distance cue.

This explanation of near-field distance perception is appealing because it indicates that near-field distance judgments are based on a feature of the HRTF which is essentially orthogonal to the features used to identify the direction of a far-field source, and because it provides an application for human sensitivity to changes in the IID at low frequencies where large IIDs never occur for far-field sources. In the far-field, the interaural intensity difference at low frequencies is never more than a few decibels, and the classic duplex theory of localization is based on the idea that IIDs do not change sufficiently with direction to provide a useful localization cue. In fact, previous research (Wightman & Kistler, 1992) has shown that in directional localization the low-frequency time delay dominates the interaural intensity difference at all frequencies whenever the stimulus contains low-frequency energy. If this results also applies to the near-field, the directional localization of a stimulus containing low-frequency energy will be based on the ITD and will ignore the low-frequency IID. Although the low-frequency IID is irrelevant to directional localization, it appears to be vital to distance localization. In the near-field, low frequency IIDs increase rapidly as the source approaches the head, and can be as large as 20 dB or more if the source is very close to one ear. In fact, a low frequency IID in excess of a few dB can only

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 205

occur when a source is very near the head. Although the near-field IID varies with direction of the source, the direction can be determined accurately from the ITD, and when the direction is known the IID provides a powerful distance cue everywhere except the median plane. Thus a distance perception mechanism based on the ITD and the low-frequency IID could explain the localization performance found in this experiment.

Such a mechanism also indicates an important application for human sensitivity to small changes in the low-frequency IID over a wide range of IIDs. Previous research (Hershkowitz & Durlach, 1969; Hafter, Dye, Nuetzel, & Aronow, 1977) has shown that listeners are sensitive to changes in IID of less than 1 dB over a 30 dB range. At distances of 1 m or more, however, the IID generated by a source in a free field would never exceed 2-3 dB. If only far-field listening conditions are considered, the ability to detect small changes in the low-frequency IID over a wide range of IIDs does not appear to serve any useful purpose. But this ability would be vital if low-frequency IIDs are used to determine distance in the near-field.

Still, it is surprising that the high-frequency IIDs provide so little information about the distance of a sound source. As discussed previously in Chapter 5, the most basic model for near-field auditory distance perception assumes that the ITD, which is roughly independent of distance, is used to determine the lateral position of the source, and then the IID, which varies systematically as a function of distance for a fixed source direction, is used to determine the distance of the source. Since the IID increases across all frequencies as the distance of a near-field source decreases, and actually increases more at high frequencies than at low frequencies, it is not immediately obvious why listeners cannot judge the distance of a source from the high-frequency IID. One possibility is that listeners cannot determine the direction of a sound which is limited to high-frequencies because no low-frequency ITD information is available and the IIDs at high frequencies vary with both the direction and distance of the source. This argument fails to fully explain the results, however, because the subjects were reasonably accurate at judging the direction of the high-pass filtered source. The directional localization data shows that the subjects were able to

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 206

determine the direction of a high-pass source, most likely from interaural amplitude delays or interaural onset information. It is puzzling that the subjects were unable to determine the distance of the high-pass stimulus despite the fact that had relatively accurate knowledge about the direction of the source. Two possible factors may explain these results:

- The near-field IID at high frequencies may be so large that the auditory system is no longer able to detect small changes in the IID. At low frequencies, the IID is relatively small when the source is in the far-field, and it doesn't increase above 20-30 dB as the source approaches the head. At 500 Hz, for example, the IID increases from 4 dB to 20 dB as a source at 90° moves from 1 m to 12 cm (Chapter 5). At high-frequencies, however, the IID can be large in the far-field and becomes extremely large in the near-field. At 6 kHz, the IID increases from 30 dB to 45 dB as a source at 90° moves from 1 m to 12 cm. Most studies examining the IID have reported that sensitivity to increases in the JND does not change significantly as the initial IID varies over a wide range (Hershkowitz & Durlach, 1969; Hafter et al., 1977). Nevertheless, it is likely that if the IID becomes sufficiently large, sensitivity to changes in the IID will decrease; in the limit, a sufficiently large IID will cause the signal at the contralateral ear to fall below the threshold of hearing. Since the IIDs at high frequencies are large even in the far-field, they may become so large when the source approaches the head that the auditory system saturates and can no longer detect changes in the IID. This saturation would explain the inability of listeners to determine the distance of a high-frequency near-field source.

- A second possible explanation for poor high-frequency distance localization follows from the assumption of the Duplex Theory that interaural timing information is salient primarily at low-frequencies. If one assumes that the mechanisms of the auditory system that determine the ITD of a low-frequency signal (which are known to dominate the perception of direction when the stimulus contains low-frequency energy (Wightman & Kistler, 1992)) are fundamentally differ-

Eight Horiz LLC
Exhibit 2013
IPR2023-01024
Page 207

ent than those that determine the ITD of a high-frequency signal (from the envelopes of the signal, for example), then it is possible that our ability to determine the distance of a source from its direction (determined from the ITD) and its IID requires low-frequency interaural timing information.

The inability of listeners to determine the distance of high-pass filtered signals is one of the more interesting results of these near-field localization experiments. While it is clear that low-frequency IIDs play an important role in near-field localization, further investigation is required before the mechanisms of auditory distance perception are fully understood.

7.6 Experiment 2

The results of Experiment 1 provide strong evidence that near-field distance perception is based on the low-frequency interaural intensity difference. The relatively poor performance in the monaural condition and near the median plane in the binaural conditions indicate the importance of binaural information, and the poor performance in the highpass condition implies the importance of low-frequency energy. Further tests are needed, however, to eliminate the possibility that variations in overall amplitude across stimulus conditions significantly influenced the results, and to ensure that the subjects were not using information other than the stimulus to make distance judgments.

In the first experiment overall, amplitude differed across stimuli, from a maximum of 64 dBA SPL for the lowpass condition to a maximum of 49 dBA SPL in the highpass conditions. Although the subjects performed adequately in the directional localization task in the highpass condition, indicating that the stimulus was clearly audible, they performed much worse in the distance localization task. A further experiment was necessary to verify that overall amplitude was not responsible for the poor highpass performance in the first experiment.

Another possible confounding factor in the first experiment was the use of extraneous information to make distance judgments. In the first experiment, 100 trials

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 208

were collected from each subject with the auditory stimulus eliminated in order to verify that extraneous information was not significantly contributing to the subject's responses. This experiment showed that the subjects were receiving almost no extraneous information about the direction of the source (overall angular errors exceeded 50°), but that they were extracting some information about the source distance. The average log-distance correlation coefficient was 0.16, and it was as high as 0.33 for one subject. This small amount of distance information is largely irrelevant in the conditions where performance was relatively good (lowpass, broadband) and the mean overall correlation was in excess of 0.75, but it complicates the accurate assessment of distance performance in conditions where measured performance was poor (monaural, highpass). Further investigation showed that the extraneous information was a combination of moving shadows which were visible through closed eyelids, a small amount of amplifier noise from the source, and the perception of air currents on the face from the moving source.

In order to ensure that overall stimulus amplitude and extraneous distance information did not significantly influence the distance results in the first experiment, a second experiment was performed to test distance accuracy for sources at $-90°$ azimuth and $0°$ elevation.

7.6.1 Experimental Procedure

The procedure for Experiment 2 was similar to the procedure in the first experiment described previously, except for three important changes:

1. Source locations were limited to the interaural axis ($-90°$ in azimuth, $0°$ in elevation). During each trial, the computer produced only a single number, ranging from 1 to 6, which the experimenter used to determine the approximate distance of the source from the head. In order to ease the placement of the source lateral to the listener, the experimenter stood behind and to the right of the listener, rather than directly to the right as in the first experiment.

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 209

2. The same signals from the first experiment (broadband, lowpass at 3 kHz, highpass at 3 kHz) were used in Experiment 2, but amplitude of the stimuli were adjusted to produce the same maximum sound pressure level (48 dB SPL at 1 m) in each condition. In addition, two more highpass conditions were added, one with a cutoff frequency of 1.5 kHz, and one with a cutoff frequency of 0.75 kHz. The randomization of the stimulus amplitude on each trial was also modified slightly. The range over which the normalized stimulus amplitude was roved was increased 5 dB, from 15 dB to 20 dB, by limiting the maximum amplitude at distances greater than 0.5 m. Thus the maximum amplitude was slightly lower at 1 m than at 0.5 m in this condition (but the difference was small relative to the range of the amplitude rove).

3. Several steps were taken to eliminate extraneous distance information. A blindfold eliminated any possible visual cues. A fan blew a stream of air across the subject's face during the experiment in order to eliminate any air currents caused by the moving source. And a masking signal, generated by a loudspeaker located on the floor of the anechoic chamber 1 m in front of the subject, obscured any audible amplifier noise. The masking signal consisted of USASI noise, lowpass filtered at 10 kHz, and its intensity was 49 dBA SPL at the location of the subject's head. The masker signal was continuously generated throughout the experiment. Note that, although the masker was louder than the stimulus at the location of the head, the spatial separation of the signal and masker ensured that the stimulus was audible.

7.6.2 Results

The results of Experiment 2 support the important results of Experiment 1. Performance is best in the lowpass and broadband conditions and is degraded significantly in the highpass and monaural conditions (Figure 7-13). Thus it does not appear that the lower overall amplitude of the stimulus in the highpass condition of Experiment 1 was responsible for poor distance performance. It is interesting, however, that the

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 210

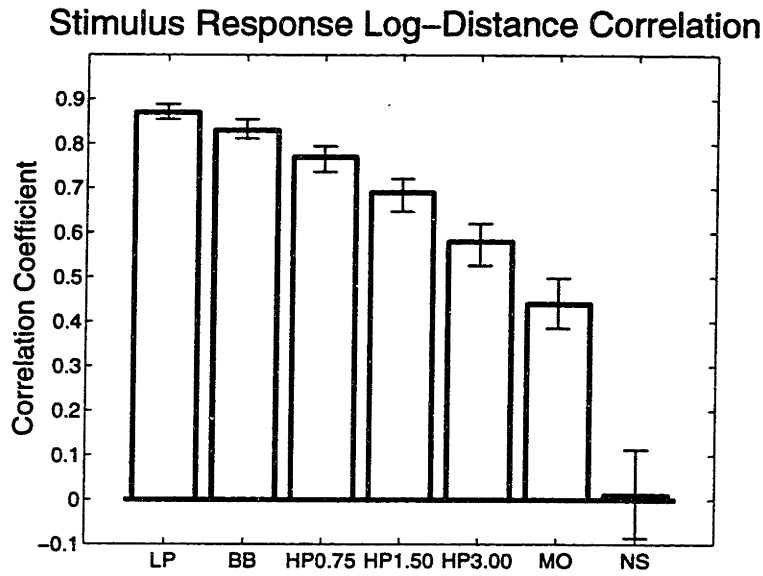


Figure 7-13: Correlation coefficients between log-stimulus distance and log-response distance in Experiment 2. The results have been averaged across the four subjects using the Fisher transform. The bars represent the 95% confidence intervals in each condition. LP=Lowpass; BB=Broad Band; HP0.75=High-pass, 750Hz cutoff; HP1.5=Highpass, 1.5 kHz cutoff; HP3.00=Highpass, 3 kHz cutoff; MO=Monaural; NS=No Sound.

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 211

stimulus-response correlation is higher in the lowpass condition than in the broadband condition. Since the stimuli all had equal overall amplitude, the lowpass stimuli contained more low-frequency energy than the broadband stimuli, and thus was not masked as effectively as the broadband condition. It is further testimony to the importance of low frequency energy to near-field distance perception that this increase in low-frequency energy allowed subjects to perform significantly better in the lowpass condition than the bandpass condition in the presence of masking noise.

The results of Experiment 2 also confirm the existence of monaural near-field distance cues. The mean stimulus-response distance correlation was similar in the monaural conditions of Experiments 1 and 2 (approximately 0.45 at $-90°$), but in the first experiment it was not clear how much extraneous information contributed to this correlation (since subject KMY was able to achieve a distance correlation of 0.33 with no stimulus). In Experiment 2, where the blindfold, fan, and masking noise rendered the subjects unable to make accurate judgments about source location in the no-stimulus condition, they clearly were still able to obtain some distance information in the monaural condition. A surprising result from the raw data (Figure 7-14) is that the monaural distance correlation does not seem to be purely a result of correctly identifying very near sound sources. Subject DSB's responses were substantially more accurate when the source was within 15 cm of the head, but the other subjects' responses were not. All four subjects' mean responses increase slightly with source distance up to 1 m. This is surprising, because previous analysis of near-field head-related transfer functions (Chapter 5) indicated that spectral localization cues, primarily in the form of a slight emphasis of low frequencies for nearby sources, were most pronounced in the region within 25 cm of the head. Further research, including the measurement of monaural distance JNDs and the evaluation of monaural distance performance with a variety of stimulus spectra, is necessary to fully understand the mechanisms that provide some distance information under monaural listening conditions. It is evident, however, that monaural distance cues are weaker than binaural cues and that binaural distance cues dominate when sources are located outside the median plane.

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 212

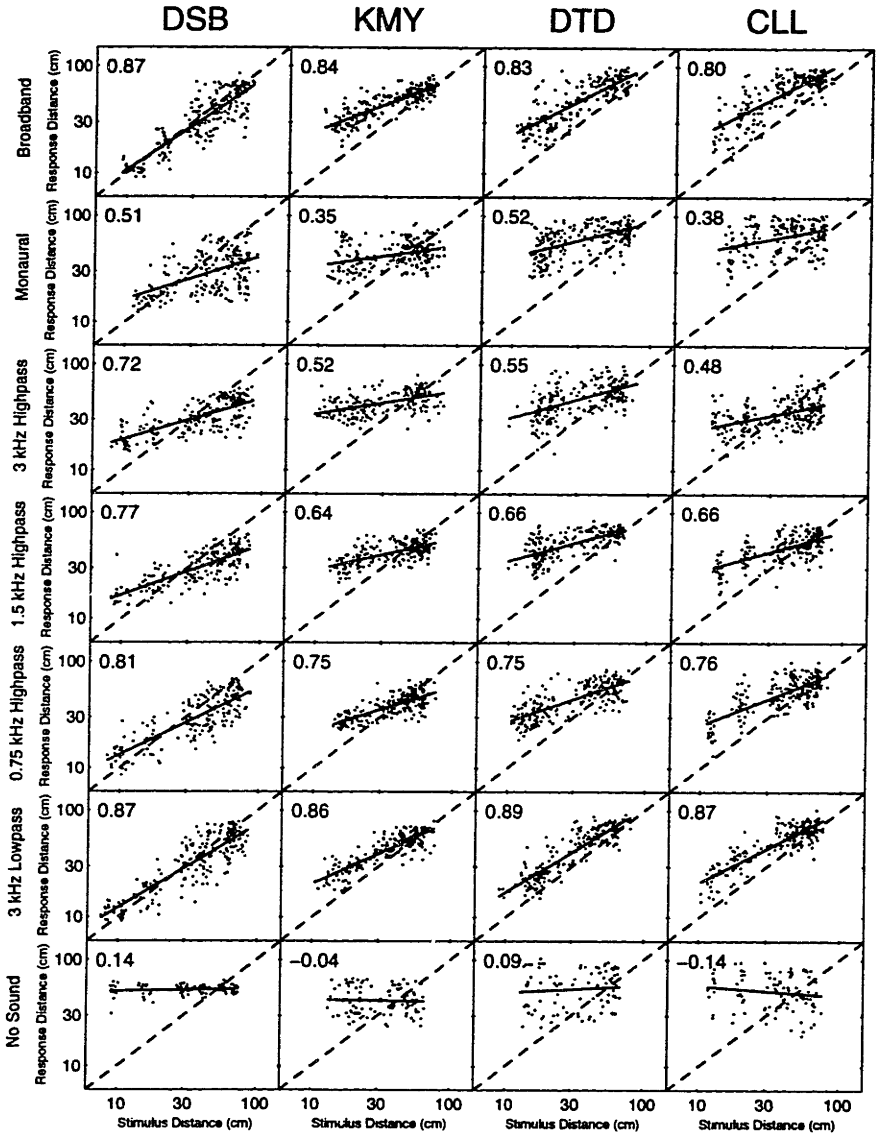


Figure 7-14: Raw Distance Stimulus-Response Data for Experiment 2. The data are for sources positioned along the interaural axis. The dashed lines represent "correct" responses, while the solid line is the best linear fit of the stimulus data to the response data. The number at the top left of each panel is the linear correlation coefficient.

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 213

A final important result of Experiment 2 is the relative performance in the three highpass conditions, with cutoff frequencies of 3.0 kHz, 1.5 kHz, and 0.75 kHz. Although overall amplitude was the same in each of the three conditions, the distance correlation coefficient increased systematically as the cutoff frequency decreased for each of the four subjects, and the performance was better in the broadband condition than in the 0.75 kHz highpass condition. This result provides further evidence that low-frequency energy is required for accurate near-field distance perception. It also indicates that some useful distance information occurs in the frequency range from 1.5 kHz and 3 kHz, and that a significant portion of the useful distance information in a broadband stimulus occurs below 750 Hz. Thus it appears that even very low-frequency interaural intensity differences are important to auditory depth perception.

7.7 Conclusions

The results from this study can be summarized as follows:

- Directional localization performance is similar in the near- and far-fields. Although there was a slight decrease in the accuracy of directional localization for nearby sources (in terms of the correlation coefficients in azimuth and elevation and the number of front-back confusions), the difference was not dramatic. It appears that directional localization accuracy does not depend strongly on source distance, even when the source is very close to the head.

- Directional accuracy in each of the five stimulus conditions was consistent with the auditory localization literature. Performance in the fixed amplitude and broadband conditions was almost identical. In the monaural condition, subjects were able to determine whether sources were in the front and rear hemisphere, and make reasonable guesses about the elevation of the source, but were unable to accurately determine source azimuth. In the highpass condition, localization performance was modestly degraded both in azimuth and elevation. In the low-

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 214

pass condition, accuracy was only slightly degraded in azimuth, but the number of front-back confusions increased dramatically, and elevation localization was severely degraded. These results conform to the widely held theory that interaural time and intensity differences allow subjects to determine the lateral position of a sound source, while high-frequency monaural pinna cues allow listeners to determine source elevation and to resolve front-back confusions.

- Some interaural timing information was apparently available to the subjects in the high-pass condition. Although the interaural intensity difference for a fixed source direction increases dramatically in the near-field, the subjects did not appear to confuse source distance and source direction in the high-pass condition. They were not able to judge distance accurately, but they were able to judge direction reasonably well. Furthermore, they did not show a systematically larger lateral bias in direction at close distances, as would be expected if the subjects were relying solely on interaural intensity differences to determine source direction. This result suggests that the subjects were able to obtain interaural timing information, either from envelope delays or onset cues, that allowed them to accurately judge source direction.

- Distance perception is reasonably accurate for near-field sources outside the median plane, and the distance perception in this region appears to depend primarily on low-frequency binaural cues. In the broadband condition, the correlation between the log of the stimulus distance and the log of the response distance was as high as 0.85. This indicates distance perception superior to that found in far-field experiments under anechoic conditions with a random-amplitude source. Distance performance was poor in the monaural condition, and was worse near the median plane than near the interaural axis. This pattern of performance indicates the importance of binaural cues to distance perception. Distance perception was also significantly worse in the highpass condition than in the lowpass condition, which implies the importance of low-frequencies to distance perception. Measurements of the HRTF in the near-field have in-

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 215

dicated that interaural time delay is relatively independent of distance in the near-field, so by elimination it appears that low-frequency interaural intensity differences are responsible for distance perception in the near-field. The results of Experiment 2 indicate that distance performance is enhanced by the addition of frequency content from 750 Hz to 3 kHz, but that additional distance information is provided by lower frequencies (below 750 Hz).

- Amplitude cues are more important to auditory distance perception at medial locations than at lateral locations. In the median plane, binaural cues are relatively weak, and amplitude cues dominate distance perception. At lateral locations, however, the binaural cues provide most of the distance information. Note that distance perception is better near the interaural axis when the amplitude of the source is randomized than in the median plane when the amplitude of the source is fixed. This indicates that near-field binaural distance cues are more salient than the amplitude-based distance cue which has been found to dominate distance judgments in the far-field.

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 216

Chapter 8

Conclusions

The experiments described in this dissertation provide a basic understanding of near-field localization. As far as we know, this research has been the first comprehensive effort to examine both the physical acoustic and psychoacoustic issues involved in localizing nearby sources.

8.1 Summary of Results

Because of the novelty of the research, two preliminary efforts were required before proceeding to the primary experiments. First, a compact, non-directional, mobile sound source was needed both for measuring near-field HRTFs and for the psychoacoustic localization studies. To this end, a novel acoustic point source was developed (see Appendix C). Second, a response method was required which would allow subjects to reliably indicate the location of a nearby source in three dimensions. Again, no methods described in the literature were directly applicable, so an experiment was conducted to evaluate four possible near-field response methods. The results, presented in Chapter 4, indicate that the "direct-location" response, in which subjects moved a response sensor directly to the perceived location of the sound, was considerably more accurate than any of the other methods tested. The results also provided some insights into our ability to transform locations from our own reference frame to the reference frame of a manikin head which, though unrelated to the main topic of

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 217

this research, are nonetheless interesting.

Once these preliminary tasks were done, the focus of the research turned to the physical acoustic aspects of near-field localization (Chapter 5). Initially, the near-field head-related transfer function was calculated for a head modeled by a rigid sphere. These calculations updated and extended similar ones performed by Hartley and Frey in 1921 (Hartley & Frey, 1921), and confirmed the earlier results: the interaural intensity difference increases dramatically in the near-field, while the interaural time delay is roughly independent of distance. The monaural transfer functions from the sphere model indicated that diffraction by the sphere causes a relative boost to low frequency energy when a source is close to the ear which could potentially serve as a monaural distance cue. Later, near-field HRTFs were measured with the KEMAR manikin and the acoustic point source. These measurements confirmed all three of these aspects of the sphere-model HRTFs, and provided additional information about the high-frequency behavior of the HRTF. In particular, the elevation-dependent high-frequency features of the HRTFs appeared to be roughly independent of distance, while the azimuth-dependent high-frequency features appeared to be compressed around the interaural axis. This last feature was believed to be caused by the discrepancy between the location of the source relative to the pinna and the location of the source relative to the head: the high-frequency features are generally governed by the pinna, and the direction of a nearby source relative to the pinna changes more rapidly near the interaural axis than near the median plane.

The final thrust of the research was a series of psychoacoustic experiments designed to measure spatial perception in the near-field. Initially, extensive data were collected with a broadband stimulus in order to evaluate localization accuracy as a function of source location in the near-field (Chapter 6). These results indicated that directional localization accuracy is at least comparable in the near- and far-fields. Directional localization degraded slightly as the source moved closer than 25 cm from the head. The distribution of errors was similar to those previously reported in far-field localization experiments, and the general direction of the response biases were roughly independent of distance. Curiously, the one region where localization accu-

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 218

racy degraded most severely for very near sources was the region directly in front of the face. This result is not easily explained by the physical acoustics of near-field localization, and requires further research.

In contrast to directional localization, distance localization was considerably different in the near-field than in previous far-field experiments. The subjects were able to make accurate, relatively unbiased estimates of source distance for lateral sources. In fact, audio distance perception for lateral sources in the near-field (with stimulus-response correlations as high as 0.85) appears to be more accurate than in any previous study of audio distance perception under similar (random amplitude, anechoic environment) conditions. In the median plane, however, distance localization was considerably less accurate. This result strongly suggests that binaural cues are important for distance perception in the near-field.

The broadband localization experiment was followed by a series of experiments involving different source conditions, which were described in Chapter 7. Fewer trials were collected in each of these conditions, so it was impossible to construct a detailed map of performance as a function of source location. It was possible, however, to compare the relative performance in each of the four conditions. The conditions were:

1. **Fixed-Amplitude:** The amplitude of the source was not randomized, so the amplitude distance cue was available. Directional localization was, not surprisingly, nearly identical to the broadband condition. Distance performance was much better than in the random amplitude condition near the median plane, but only slightly better for lateral sources. Thus it appears that amplitude cues dominate in the median plane, but they provide only a minor benefit in lateral regions where binaural cues are available. This contrasts with previous studies which have shown that distance judgments in the far-field can be explained almost entirely by the amplitude cue (Strybel & Perrott, 1984).

2. **Monaural:** One ear was occluded by an ear-plug and muff. As in previous far-field localization studies, azimuth localization was severely degraded and

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 219

characterized by a strong bias in the direction of the unoccluded ear, while elevation localization accuracy was only slightly reduced. The subjects were unable to accurately judge distance in the monaural condition, providing further evidence that binaural cues dominate near-field distance perception.

3. **High-pass Filtered:** The stimulus was high-pass filtered above 3 kHz. Directional localization both in azimuth and elevation was slightly degraded. Azimuthal localization was not reduced as severely as would be expected if the subjects were relying solely on interaural intensity differences (as predicted by the duplex theory of localization), as they showed no indications of being confused by the interactions between lateral position and direction in the near-field IID. This indicates they may have been able to obtain interaural timing information from onsets or high-frequency envelope information. Despite their reasonably accurate judgments of source direction, the subjects were unable to make accurate distance judgments in the high-pass condition.

4. **Low-pass Filtered:** The stimulus was low-pass filtered below 3 kHz. The number of front-back reversals increased dramatically, and elevation judgments were exceptionally poor, as would be expected from the elimination of high-frequency pinna cues. Distance perception, however, was nearly as good as in the broadband condition.

The results from these stimulus conditions provide very strong evidence that near-field auditory distance perception is primarily binaural, and that it depends on low-frequency information. Since the ITD is roughly independent of distance at low frequencies, this implies that near-field distance perception must be based on low-frequency IIDs. Low-frequency IIDs are an appealing explanation for near-field distance perception, because such an explanation implies that near-field distance perception is based on a component of the HRTF which is essentially orthogonal to directional localization. Low-frequency IIDs exist only in the near-field, and previous work has shown the ITD dominates directional localization when a far-field stimulus contains low-frequency components. If this dominance also occurs in the near-field,

Eight Bit, LLC
Exhibit 2013
IPR2023-01024
Page 220

then the low-frequency IIDs which govern near-field distance perception are ignored by the directional localization mechanisms. Low-frequency IID-based depth perception also provides an important application for human sensitivity to small changes in the low-frequency IID over a wide range of values which would never naturally occur in the far-field. Despite the theoretical appeal of a near-field distance perception mechanism based on low-frequency IIDs, it is surprising that listeners are not able to use the IID at high frequencies to determine the distance of a near-field source. This result is especially puzzling because the high-frequency IIDs increase as much as or more than the low-frequency IIDs as the source approaches the head. The inability to judge the distance of high-frequency sounds may occur because the IIDs at high-frequencies for a near-field source are so large they saturate the auditory system. Further investigation is necessary to fully understand the mechanisms of near-field auditory depth perception.

A final experiment was designed to verify that extraneous (non-audio) cues did not significantly distort the distance-perception results in the previous experiments. The addition of a blindfold, a fan, and a masking speaker completely eliminated the ability of the subjects to judge distance without the auditory signal. Under these conditions, the subjects' responses in the monaural condition were still positively correlated with source distance, verifying the existence of a monaural spectral cue. Additional high-pass conditions were also tested, and they indicated that frequencies below 750 Hz make a significant contribution to near-field auditory distance perception.

While many additional experiments are necessary to provide a comprehensive understanding of near-field localization, these experiments have conclusively indicated that subjects are able to make relatively accurate distance near-field judgments based on changes in the low-frequency IID with source distance. The ability to make accurate distance judgments about an unfamiliar source is apparently unique to the near-field, and we hope that our findings will lead to renewed interest in this previously ignored area of spatial hearing. The next section outlines some unanswered questions about near-field localization which should be the focus of future research.

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 221

8.2 Other Possible Issues in Near-Field Localization

8.2.1 Psychoacoustic Issues

- Discrimination thresholds for near-field localization: What are the JNDs for azimuth, elevation, and distance as a function of source position in the near field? The present experiments measured only absolute localization ability, not the ability to discriminate changes in the location of a source. Of particular interest is the region directly in front of the face, where azimuthal localization was found to be relatively poor in these identification experiments, and the relationship between the percent JND in distance and source distance, which did not seem to agree with the predictions of the near-field HRTFs. Free-field measurement of these JNDs would require an apparatus capable of incremental manipulation of source positions in three dimensions, so it may be more feasible to use virtual sound sources for JND experiments.

- Are their any conditions where the increased high-frequency IIDs in the near-field cause systematic lateral directional biases? In these experiments, the subjects were evidently able to obtain some interaural timing information and make accurate directional judgments. Would this ability disappear if slower onsets or a different type of high-frequency stimulus were used? Are the expected biases seen with sinusoidal sources?

- Does the compression of the high-frequency features of the HRTF around the interaural axis ever significantly effect near-field directional localization? In our monaural experiment, any effects due to this compression (which should cause a medial bias) were masked by the strong lateral bias in the direction of the unoccluded ear. In order to detect this effect, it will be necessary to carefully compare monaural localization in the near- and far-field for the same subjects.

- How do head-motions affect near-field localization accuracy? In the near-field,

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 222

small translational movements could be as important as the rotational movements known to be important to far-field localization. In order to exploit information from such movements, the brain must have accurate knowledge of the precise location of the head at all times. Is near-field localization improved dramatically by head-motions?

- How accurate is near-field localization in situations where the possible source locations are constrained? For instance, if the source were known to be somewhere on a vertical wall in the near field, how accurately could a listener pinpoint the location of that source?

- Can near-field audio streams be segregated from far-field audio streams? Can listeners selectively tune their attention to sources 0.5 m away and 3 m away in the same direction?

- Do prejudices about the stimulus dominate near-field localization cues: Will a shouted voice attenuated and generated 0.25 m from the head still sound far away?

- Are physical interactions with the source important? We can move our hands about in the near-field and potentially interact with a sound source. Does this affect our near-field localization ability?

8.2.2 Virtual Display Issues

- How can HRTFs collected in three dimensions be interpolated to create realistic localization information at all near-field locations from a finite number of measurements?

- How well can listeners be trained to associate near-field auditory cues with far-field events? For example, could pilots be trained to associate audio images half a meter away with objects five kilometers away?

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 223

- How should a virtual display handle situations where the operator moves a virtual sound source to a position inside the head? Should unnatural cues which produce lateralized rather than localized sound be used?

These are some of the questions for future research in near-field localization. We believe that this area has the potential to become an important topic in psychoacoustic research, and an important resource for the design of improved virtual audio systems.

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 224

Appendix A

Coordinate System Definitions

One of the key goals of this research is a comparison between judgments of location relative to the subject's own head and judgments of location relative to a manikin head. This requires a coordinate system which is consistent for both the subject's frame of reference and the manikin head's frame of reference. Three features common to both the human and manikin head were used as the basis of this coordinate system: the openings of the left and right ear canals, and the tip of the nose. The locations of these three features generate a coordinate system for each head centered at the midpoint of the interaural axis and with a horizontal plane approximately parallel with the floor. Specifically, head-referenced Cartesian axes were defined as follows: The Y axis of the coordinate system is the interaural axis of the head, and is positive on the left side. The X axis is the perpendicular bisector of the interaural axis passing closest to the tip of the nose, and is positive for locations in front of the head. The Z axis is perpendicular to both the X axis and Y axis and is positive above the head.

The reference points for these coordinate systems were measured with the 3-Space tracker before each block of trials. The tracker provided the X, Y, and Z locations of each position relative to the 3-Space source, which was mounted on the chin-rest. The locations of the left and right ears and the tip of the nose, represented as column vectors $\vec{E}_l, \vec{E}_r$, and $\vec{N}$, were measured for the subject, the large manikin head, and the small manikin head with a 3-Space Tracker sensor. These locations were used to determine the origin and the directional cosines of the X, Y, and Z axes for each of

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 225

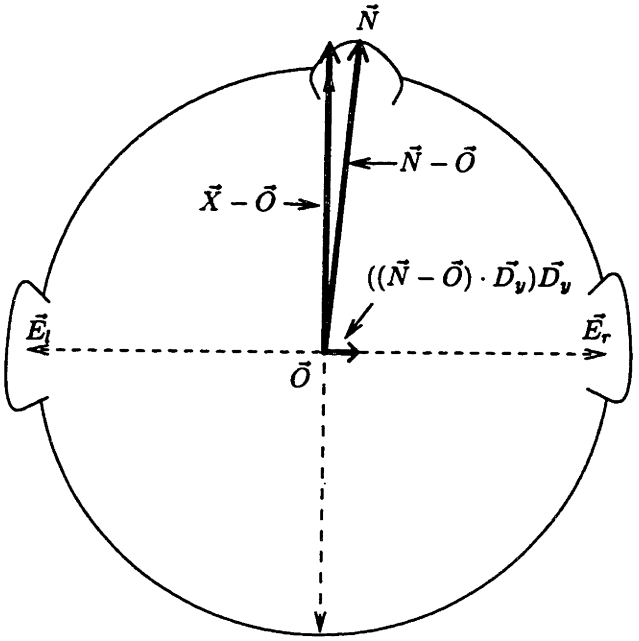


Figure A-1: Definition of the X-axis

An example where the vector $\vec{N}-\vec{O}$ is not perpendicular to the interaural axis. The projection of $\vec{N}-\vec{O}$ onto the Y axis, $(((\vec{N}-\vec{O})\cdot\vec{D_Y})\vec{D_Y})$, is subtracted from $\vec{N}-\vec{O}$ to yield the vector $\vec{X}-\vec{O}$, which is a perpendicular bisector of the interaural axis. Normalization of $\vec{X}-\vec{O}$ gives the directional cosines of the X axis.

the three heads.

The origin of the coordinate system $\vec{O}$ is defined as the midpoint of the interaural axis, $\frac{\vec{E_l}+\vec{E_r}}{2}$. The interaural axis also defines the directional cosines of the Y axis, $\vec{D_Y} = \frac{\vec{E_l}-\vec{E_r}}{|\vec{E_l}-\vec{E_r}|}$. The X axis is defined by the tip of the nose, the origin $\vec{O}$, and the directional cosines of the Y axis $\vec{D_Y}$. Ideally, the X axis should pass through the tip of the nose, but the vector $\vec{N}-\vec{O}$ is not necessarily perpendicular to the Y axis, as seen in Figure A-1. When this is true, the projection of $\vec{N}-\vec{O}$ onto the Y axis is subtracted from $\vec{N}-\vec{O}$ to determine the perpendicular bisector of the interaural axis which is closest to the tip of the nose. If we call this vector $\vec{X}-\vec{O}$,

$$\vec{X}-\vec{O} = (\vec{N}-\vec{O}) - ((\vec{N}-\vec{O})\cdot\vec{D_Y})\vec{D_Y},$$

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 226

and the directional cosines of the X axis are found by normalization:

$$\vec{D_X} = \frac{\vec{X} - \vec{O}}{|\vec{X} - \vec{O}|}.$$

Finally, the directional cosines of the Z axis are obtained by taking the cross product of the $\vec{D_X}$ and $\vec{D_Y}$ vectors, so $\vec{D_Z} = \vec{D_X} \times \vec{D_Y}$.

The directional cosines can be used to convert the Cartesian coordinates of any location, measured relative to the head tracker, into the transformed coordinate system relative to the head. Let $\vec{S}$ be an XYZ location relative to the source of the 3-Space tracker. First the XYZ coordinates are moved relative to the center of the head by subtracting the origin of the transformed coordinate system $\vec{O}$ from $\vec{S}$. Then the XYZ coordinates of the vector are projected onto the X, Y, and Z axes of the transformed coordinate system by matrix multiplication

$$\vec{S_T} = (\vec{S} - \vec{O})' \cdot [\ \vec{D_X} \quad \vec{D_Y} \quad \vec{D_Z}\ ].$$

where $'$ denotes matrix transposition. The resulting vector $\vec{S_T}$ can be used to determine the azimuth, elevation, and distance of the target relative to the coordinate system defined by one of the three heads, thereby allowing a direct comparison of all four response methods.

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 227

Appendix B

Derivation of the Sphere Model

The interaural differences associated with a nearby sound source can be approximated by modeling the head as a rigid sphere. This model is advantageous because many mathematical descriptions of the acoustic properties of rigid spheres are available. This model was developed based on the work of Rabinowitz et al. (Rabinowitz et al., 1993) and was originally intended to examine the frequency scalability of head-related transfer functions for an enlarged head. The model is equally applicable to examining near-field localization.

In this model the head is represented by a rigid sphere, with radius a. Two pressure sensitive "ears" are located at diametrically opposed points on the surface of the sphere. The sound source is a point source radiating spherical acoustic waves, and it is located at distance r from the center of the head, and at angle α from the perpendicular bisector of the interaural axis.

The equations relating the sound pressure at each ear to the velocity of the source are based on the (reciprocal) equations in Morse and Ingard (Morse & Ingard, 1968) deriving the expression for the sound pressure radiated by a sphere whose surface vibrates with velocity $U(\alpha)e^{-j\omega t}$. The velocity function $U(\alpha)$ can be expressed in terms of the Legendre basis functions $P_m(\cos\alpha)(m = 0, 1, 2, ...)$.

$$U(\alpha) = \sum_{m=0}^{\infty} U_m P_m(\cos\alpha). \tag{B.1}$$

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 228

The coefficients U_m are related to $U(\alpha)$ by

$$U_m = (m + \frac{1}{2}) \int_0^\infty U(\alpha) P_m(\cos\alpha) \sin\alpha d\alpha. \tag{B.2}$$

The harmonic form of the pressure wave generated by this source at distance r and angle α is given by

$$\mathbf{p} = \rho_0 c \sum_{m=0}^{\infty} U_m P_m(\cos\alpha) \frac{h_m(kr)}{h'_m(ka)}. \tag{B.3}$$

Here $h_m(kr)$ is the spherical Hankel function and $h'_m(ka)$ is the derivative of the spherical Hankel function. The ratio of the Hankel function to $h'_m(ka)$ represents the radial dependence of the pressure. The Legendre polynomial terms represent the angular dependence.

We want to find the pressure wave radiated by a point source on the surface of the sphere with velocity u_0. This point source can be approximated by a small disk on the surface of the sphere extending over an arc of $\frac{2\Delta}{a}$ radians. In this case, the velocity on the surface of the sphere can be defined as

$$x = \begin{cases} u_0 & 0 \leq \alpha \leq \frac{\Delta}{a} \\ 0 & \frac{\Delta}{a} < \alpha \leq \pi. \end{cases} \tag{B.4}$$

Now the coefficients U_m become

$$U_m = (m + \frac{1}{2}) u_0 \int_0^{\frac{\Delta}{a}} P_m(\cos\alpha) \sin\alpha d\alpha, \tag{B.5}$$

which converges as $\Delta \to 0$ to

$$U_m = \frac{1}{2}(m + \frac{1}{2}) u_0 (\frac{\Delta}{a})^2. \tag{B.6}$$

The total expression for the pressure at a point (r, α) due to a point source on the surface of the sphere with velocity u_0 then becomes

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 229

$$\mathbf{p} = \frac{\rho_0 c u_0}{2}(\frac{\Delta}{a})^2 \sum_{m=0}^{\infty}(m+\frac{1}{2})P_m(\cos\alpha)\frac{h_m(kr)}{h'_m(ka)}. \tag{B.7}$$

From the acoustic reciprocity theorem (Cook, 1996), it follows that this expression is also the pressure at a point on a surface of the sphere resulting from a point source at location (r, α). Specifically, the reciprocity theorem states that the pressure at point B due to a velocity source at point A is exactly the same (in magnitude and phase) as the pressure at point A generated by the same velocity source at point B. This implies directly that the above equation is, in fact, the pressure on the surface of a sphere caused by a point source at location (r, α). Also, notice that the term $u_0\Delta^2\pi$ can be recognized as a single term representing the volume-velocity strength of the unit source. [1]

In order to examine the near-field localization cues predicted by this model, it is necessary to find the magnitude and phase of the pressure on the surface of the sphere as a function of the distance to the source r, the radius of the sphere a, the angle between the source and the point on the surface of the sphere α, and the frequency of the sound f. This complex function, denoted by $\mathbf{P}_s$, is given by

$$\mathbf{P}_s(r, a, \alpha, f) = \frac{q}{a^2}\sum_{m=0}^{\infty}(m+\frac{1}{2})P_m(\cos\alpha)\frac{h_m(\frac{2\pi f r}{c})}{h'_m(\frac{2\pi f a}{c})}, \tag{B.8}$$

where q is the constant $\frac{c\rho_0 u_0 \Delta^2}{2}$.

[1] The factor of π is canceled by a value of π in the denominator in the integration of Equation B.6.

Exhibit 2013
IPR2023-01024
Page 230

Appendix C

Acoustic Point Source

Abstract

An approximation to an acoustic point source has been developed which produces relatively non-directional acoustic signals over a wide frequency range (200 Hz-15 kHz). This source differs from previous point-source systems in several important ways: 1) The use of a high-output electro-dynamic horn driver in place of a conventional cone loudspeaker to power the unit; 2) The use of a relatively long, flexible tube to carry the signal away from the driver, allowing the driver unit to be acoustically isolated from the point source and also allowing easy placement of the point source; 3) The use of a rigid sleeve around the distal end of the tube to allow more convenient placement of the source; 4) The use of an electromagnetic tracking system to accurately measure the effective location of the point source without interfering with its output. The resulting system has a variety of potential applications where a compact, non-directional, high-output source is required, include acoustic and psycho-acoustic measurements in the near-field.

C.1 Purpose

Under certain circumstances, it is desirable to make acoustic measurements with an acoustic "point source". A point source is defined as an infinitesimally small sound source which produces a finite quantity of acoustic power. Usually it is modeled as a pulsating sphere of negligible dimensions producing a finite volume velocity at its surface.

Point sources have two important characteristics which cannot be duplicated in any physically realizable acoustic transducers. They radiate sound from a single

Eight Sleep, LLC
Exhibit 2013
IPR2023-01024
Page 231

location in space, and they radiate sound omnidirectionally. Unfortunately, it is impossible to build an infinitesimally small sound transducer with these characteristics. While it is, of course, possible to build a small loudspeaker, there is a clear trade-off in loudspeaker design between small size and low-frequency output. The system described here is unique in that it is able to generate sound from a compact region of space which is both largely non-directional at relatively high frequencies and relatively powerful at low frequencies. The current rendition of this system generates sound from a location only 1.3 cm in diameter, is capable of generating reasonably strong output from 200 Hz to 15 kHz, and has a 3 dB beamwidth of approximately 120° at 15 kHz. The system is also equipped with an electromagnetic position sensing system that allows accurate measurement of the effective position of the source.

The purpose of this source is to enhance the accuracy of acoustic measurements in situations where conventional loudspeakers capable of producing enough low-frequency output are not sufficiently small or non-directional. An example of such an application is the measurement of Head-Related Transfer Functions (HRTFs) in the near-field. The HRTF is the transfer function from the pressure at a sound source at some location in space to the pressure that actually reaches the eardrums of a human listener. This transfer function includes the propagation of sound from the source to the head, the diffraction of the head and torso, the spectral shaping of the outer ear or pinna, and the ear-canal resonance. Historically, most HRTF measurements have been made at distances of 1 m or more, where the dimensions of a loudspeaker are essentially negligible (they extend only over a few degrees in azimuth and elevation) and the location of the source relative to the head is easily determined. Current research efforts are underway to measure HRTFs at distances less than 1 m. At locations close to the head even a relatively small loudspeaker can extend over a region 25° or more across, and the exact orientation of the source relative to the head is more difficult to determine. In order to measure the near-field HRTF at a well defined location, a point source with some mechanism for accurate positioning is required.

The device is also useful in applications other than measurements when a compact, wide-bandwidth, non-directional source is useful. For example, the point-source can

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 232

be used to conduct psycho-acoustic localization experiments in the near-field.

C.2 Background

Clearly the most conventional transducer from generating an acoustic signal from an electrical input is a loudspeaker. HRTF measurements, for example, have traditionally used conventional loudspeakers, 7 cm or larger in diameter, to generate the acoustic stimulus. At distances of 1 m or more, such loudspeakers are perfectly adequate. At close distances, however, there are serious problems associated with loudspeaker measurements:

- The precise location of a loudspeaker is not well defined in the near-field. The stimulus is generated by the entire diaphragm of the loudspeaker, and at close distances this may extend over a large region of space: at 12 cm, for example, a 7 cm loudspeaker covers an arc in excess of 30°. The HRTF measured will be, in effect, the average HRTF over the entire region covered by the loudspeaker.

- The directional properties of the loudspeaker may taint the HRTF. When the speaker is near the listener, the high-frequency directionality of the speaker will cause the sound pressure reaching the head and torso to vary according to the orientation of that region relative to the speaker. This may significantly effect the measured HRTF.

- The axial response of a loudspeaker is complicated by its distributed geometry at very close distances. At distances less than $2\frac{a^2}{\lambda}$, where a is the radius of the loudspeaker and λ is the wavelength of the sound, the intensity along the axis of the loudspeaker does not decrease monotonically with distance, but rather passes through a series of maxima of constant amplitude with intervening nulls (Kinsler & Frey, 62). For a 15 kHz sound generated by a 7 cm loudspeaker, this effect complicates measurements at distances less than 10 cm from the surface of the head (approximately 20 cm from the center of the head).

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 233

- A loudspeaker is generally large enough to provide a reflective surface when sufficiently close to the head. Sound generated by the speaker might be reflected off the head, the be reflected again off the source and back toward the head. These second-order reflections could corrupt a near-field HRTF measurement.

For these reasons, an ordinary loudspeaker cannot be used effectively to make near-field HRTF measurements. The key to eliminating the problems associated with loudspeaker measurements is reducing the effective area of the source. Every realizable transducer has finite dimensions, and therefore generates a positive particle velocity over some finite region of space. The sound pressure generated by such a source at particular location in space is found by dividing the moving surface of the source into infinitesimal regions. The contribution of each region is determined by assuming that region is a point source with a certain volume velocity. The surface integral of these contributions over the area of the transducer determines the total signal. In acoustic measurements of the transfer function from a sound source at a particular location in space to a receiver at some other location in space, any measurement with a conventional transducer will in fact be the average transfer function over the region covered by the transducer. In order to control the exact location of a sound source, it is necessary to make the area of the transducer as small as possible.

One possible approach to this problem is the use of extremely small loudspeakers. This would certainly reduce the problems of location, directionality, axial response, and reflections described above. However, due to radiation impedance, there is an inverse relation between the efficiency of a loudspeaker at low-frequencies and the size of the loudspeaker. Thus extremely small loudspeakers cannot effectively reproduce wide-band stimuli.

A second approach to the problem is to generate a wide-band stimulus with a relatively large conventional cone loudspeaker and connect this speaker, through an enclosed cavity, to a small diameter metal tube. The sound then propagates down the tube and radiates from the small orifice at the opening of the tube. This is exactly the approach used by Shaw and Teranishi (1968). They connected a loudspeaker to small enclosure, which then opened into a rigid tube, 30 cm long and 1 cm in diameter.

Eight kHz LLC
Exhibit 2013
IPR2023-01024
Page 234

Sound propagated down the tube, and approximated a point source at its opening. A pressure microphone at the opening of the tube was used to actively control the output of the source and maintain a flat frequency response. This approach was effective, in that it produced output from 100 Hz to 15 kHz and had a 2 dB beamwidth of 90° at 15 kHz, but apparently was incapable of producing a stimulus below 1 kHz. There are two reasons why this type of system cannot generate low frequency sounds. First, the radiation impedance of the small tube is very high, especially at low frequencies, and a conventional cone loudspeaker is simply not powerful enough produce much output below 1 kHz. Second, it is extremely difficult to prevent low-frequency energy from leaking out of the loudspeaker enclosure. It generally takes extremely massive barriers to prevent the propagation of sound at low-frequencies, and a point-source enclosed with such massive baffling material would be unwieldy at best.

This point source significantly improves the method used by Shaw and Teranishi in two ways. First, it uses a high-output electro-dynamic horn driver in place of the loudspeaker. This driver is sufficiently powerful to drive even the high impedance of a small-diameter tube at low frequencies, and is more easily adapted down to the small diameter of the tube than a loudspeaker enclosure. Second, a long (3.5 m) section of flexible, thick-walled nylon tubing in place of the rigid metal tube used by Shaw and Teranishi. The use of flexible tubing has two important advantages. First, the driver can be located 1-2 m away from the opening of the tube. This allows the driver unit to be baffled with any amount of material to reduce leakage at low-frequencies, and reduces the effect of any such leakage because the opening of the source is much closer to the receiver than the interfering leakage from the driver. Second, the flexible tube makes the actual placement of the source very convenient, and in fact the actual source can easily be manipulated by hand without moving the massive driver unit. In fact, a specially shaped wand has been developed to allow a stationary operator to move the point source anywhere in the right hemisphere of a listener within 1 m of the head, which is particularly useful in near-field psychoacoustic measurements.

A patent search reveals a third, relatively novel way to simulate an acoustic point source (Burton, 1990). Burton's system consists of a stretched, round membrane

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 235

which is driven only at its center. If the membrane material is chosen carefully, vibrations propagate down the membrane at the same speed the sound waves propagate in air. This results in a hemispherically symmetrical sound radiation pattern. While this system approximates an acoustic point source, it still apparently requires a round membrane which may reflect scattered sound waves. Also, this system must be built from scratch and cannot be adapted from commercially available components.

An additional feature of the device is the addition of an electromagnetic position sensor at the opening of the source. Electromagnetic sensors have been in use for a variety of applications for about 20 years, including sensing head and hand positions in virtual reality systems. The use of these systems for locating a source during an acoustic measurement has not been described previously. In HRTF measurements, source position has traditionally been controlled through the use of an automated source placement system, such as a revolving hoop with several speakers placed at regular intervals on the hoop (Wightman & Kistler, 1989b). These systems are only able to control source location in azimuth and elevation.

In the near-field, HRTFs are dependent on distance as well as direction, and some mechanism is required to allow accurate placement of the sound source in three dimensions. An electromagnetic tracking system is particularly well suited to this application, because it can measure both the orientation and location of the sensor. Obviously, the sensor cannot be at the exact location of the opening of the point source. It must be on the tube slightly back from the opening. However, if the end of the tube is mounted in a rigid sleeve, the combination of information about the orientation and XYZ coordinates of the source allow a very accurate measurement of the effective location of the opening of the point source. Note that the sensitivity of electromagnetic position sensors to metal and to magnetic fields preclude their use to measure the location of a loudspeaker.

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 236

C.3 Description, Manner, and Process of Making and Using Device

C.3.1 Description of Device

The best mode of device is shown in Figure C-1. Basically, the system consists of four major components: a high-output acoustic driver, a long flexible Tygon tube, a rigid plastic sleeve around the termination of the tube, and an electromagnetic position sensing system.

The device is based on a high-performance, high-frequency acoustic driver, in this case an Electro-Voice DH1506 (1). This driver is designed for use in conjunction with a large exponential horn in high-output public address systems. When connected to such a horn, the driver is capable of generating extremely loud signals (104 dB SPL at 10 ft), and has a frequency range flat from 500 Hz to 3 kHz and with a controlled roll-off to 20 kHz. In this application, the driver is not connected to a horn but rather is connected directly to a length of 1.3 cm i.d. tubing. The opening of the driver is 3.5 cm in diameter, so a series of fittings are required to mate the tubing to the driver. First, a 1.5" to 1" copper fitting (2) is mounted over the threaded opening of the driver. Teflon pipe-fitting tape was used to fill the gap between the threaded driver opening and the smooth-walled copper fitting. A 1" to 0.75" brass brushing (3) fits into the copper fitting, and is connected directly to a 0.75" to 0.5" hose fitting (4), which acts as a right angle adapter. The driver assembly generates some sound due to leakage from the back of the driver, especially at 8-9 kHz. Therefore the entire assembly is wrapped in sound absorbing material (foam, blankets, etc.) to prevent this sound from propagating. The entire driver unit is quite heavy (> 5 kg), primarily because of the large magnet used in the driver unit.

The tubing used in the device is Tygon transparent tubing, with an internal diameter of 1.3 cm (0.5") and a wall thickness of 0.3 cm (0.125") (5). The tube is 3.5 m in length. The first 2.3 m of the tube are exposed openly. The next 1.16 m of the tube are encased in a sleeve constructed of PVC pipe with an internal diameter

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 237

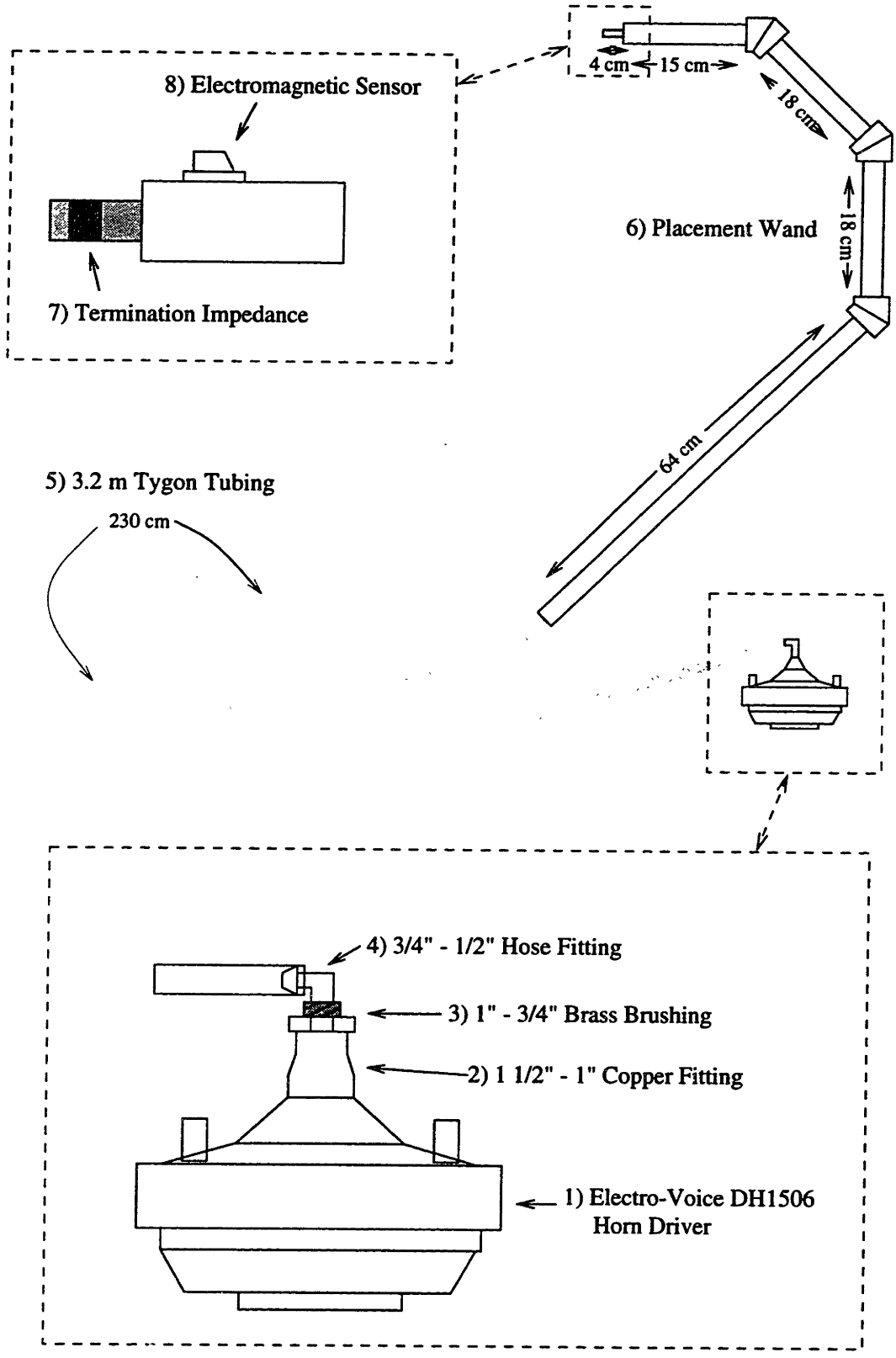


Figure C-1: Description of best mode of device

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 238

of 2.5 cm (1") (6). This sleeve, which acts as a placement wand, includes four strait lengths of pipe, 64 cm, 18 cm, 18 cm, and 15 cm long, and three 45° elbow joints.

The Tygon tubing extends 4 cm beyond the end of the PVC sleeve. At the opening, foam material fills the gap between the tubing and the interior of the sleeve. At the end of the tube, a small amount of acoustic foam has been forced into the opening to act as a terminating impedance. The amount of material used was adjusted to minimize resonances inside the tube (the quarter-wavelength resonance is at approximately 25 Hz).

Just before the end of the PVC sleeve, an electromagnetic sensor is attached by plastic cable ties. In this case, the sensor is from a Polhemus Electronics 3-Space Tracker, which is capable of determining the location of the sensor (relative to a separate electromagnetic source) within 0.25 cm in X,Y,Z coordinates, and the orientation of the sensor within 0.1° in roll, pitch, and yaw. The sensor is positioned so it remains in a fixed location relative to the opening of the tube (which is the effective location of the point source). Specifically, the center of the sensor is located 4 cm above and 6 cm behind the opening of the tube. The position of the opening of the tube can be found from the XYZ and roll, pitch, and yaw coordinates produced by the 3-Space Tracker with the following equations:

$$x_{opening} = x_{sensor} + 6\cos(az)\cos(el) + 4(\cos(az)\sin(el)\cos(rl) + \sin(az) * \sin(rl)), \quad \text{(C.1)}$$

$$y_{opening} = y_{sensor} + 6\sin(az)\cos(el) + 4(\sin(az)\sin(el)\cos(rl) - \cos(az) * \sin(rl)), \quad \text{(C.2)}$$

$$z_{opening} = z_{sensor} - 6\sin(el) + 4\cos(el)\cos(rl). \quad \text{(C.3)}$$

C.3.2 Characteristics of Device

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 239

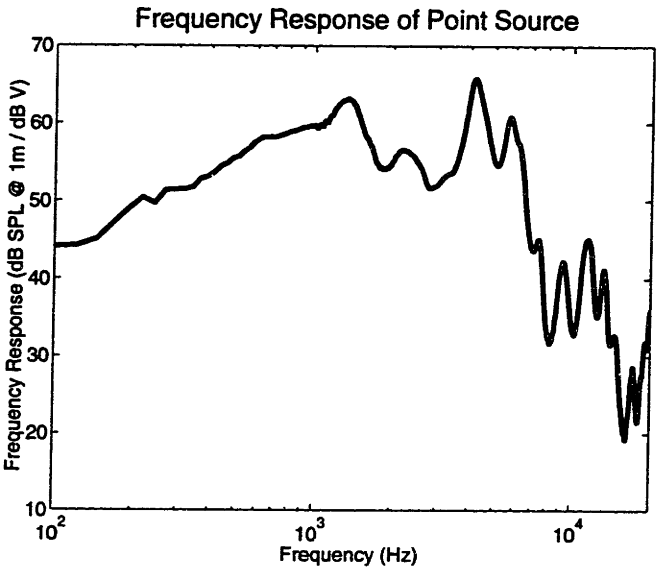


Figure C-2: Source Transfer Function. This figure shows the transfer function of the point source. The measurements were made in an anechoic chamber using a periodic chirp stimulus and a 1024-point FFT. One-third octave smoothing has been applied.

Due to the unconventional transmission path from the driver to the opening of the tube, the frequency response of the system is quite erratic (Figure C-2). The response slopes gently upward from 200 Hz to 1 kHz, then goes through four local maxima up to 6 kHz. Above 6 kHz, the frequency response drops suddenly by 30 dB, and stays at this lower level (with several more local maxima) until dropping dramatically again at 15 kHz. The transfer function is, however, stable to changes in the configuration of the Tygon tubing, so the source can be moved without changing the response characteristics.

The non-directional response of the point source is shown in Figure C-3. As would be expected, the high-frequency sound radiated by the source drops off as the source is rotated away from normal incidence. The 3 dB beam-width of the source is about 120° at frequencies up to 15 kHz. At angles greater than 60°, the high frequency response degrades rapidly, but the source appears to be completely omnidirectional at frequencies up to 2 kHz. In most practical applications, it should be possible to orient the source in such a way that the beamwidth of 120° covers the entire region under investigation.

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 240

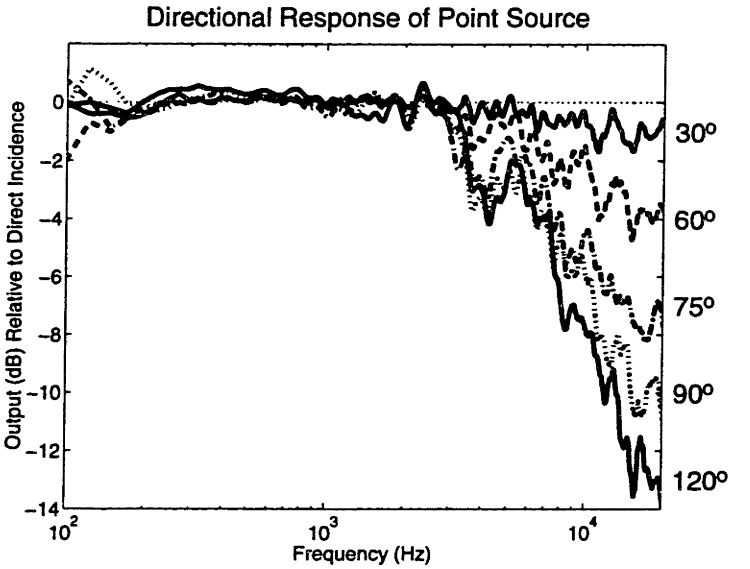


Figure C-3: Source Directionality. The plots show the frequency response of the source at five source directions relative to normal incidence. The measurements were made in an anechoic chamber using a periodic chirp stimulus and a 1024-point FFT. One-third octave smoothing has been used.

In addition to being non-directional, a point source should generate a pressure wave that is inversely proportional to distance at all distances and all frequencies. Figure C-4 shows that the device exhibits this behavior, except for a slight boost at low frequencies (less than 2 dB) when the source is within 2 cm of the receiver. This behavior indicates that the effective area of the source is quite small and that there is no significant leakage from the driver unit interfering with the measurements.

Another concern about the source is the possibility of non-linear operation because of the large radiation impedance at the opening of the small tube. Figure C-5 shows the change in the measured output of the source at 20 cm when the input is attenuated. This device appears to be quite linear.

C.3.3 Use of Device

In ordinary use, the point source driver would be connected directly to a conventional high-power audio amplifier, and driven by any reasonable signal generator. When

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 241

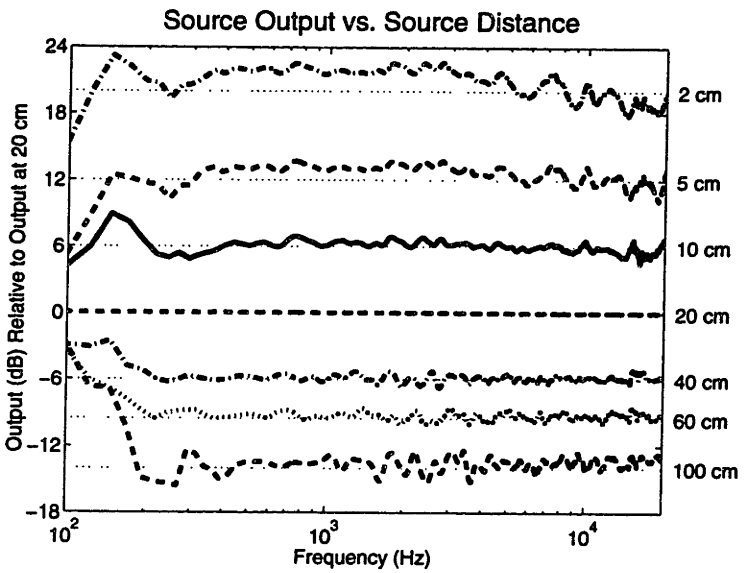


Figure C-4: Source response vs. distance in anechoic chamber. This figure shows the changes in the response of the point source at 6 different distances from the source tip, normalized to the response at 20 cm. The dotted lines represent the predicted source level for an acoustic point source. The measurements were made in an anechoic chamber using a periodic chirp stimulus and a 1024-point FFT. One-twelfth octave smoothing has been used. Note that the discrepancies below 200 Hz are a result of the noise floor of the measurement.

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 242

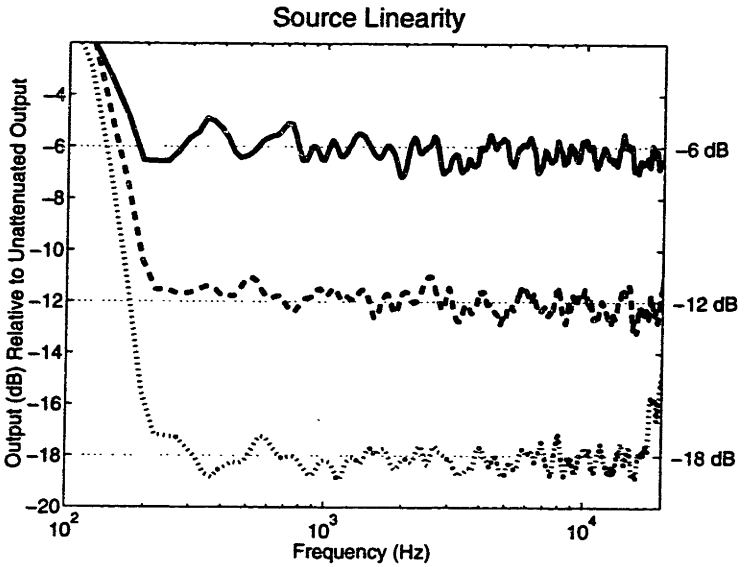


Figure C-5: Source linearity. This figure shows the change in the output of the source when the input is attenuated. The measurements were made in an anechoic chamber using a noise stimulus and a 1024-point FFT (hanning window). One-twelfth octave smoothing has been used. Errors below 200 Hz are a result of the noise floor.

making acoustic measurements, the source of the electromagnetic tracker would be placed in some fixed location relative to the system under test, and the XYZ location of the source relative to the system could be calculated directly with the previously described equations. In general, the placement wand would be clamped in place with a stand, and the sensor measurements would be used to move the source to the desired location relative to the system.

When rapid source placement is required (such as in a psychoacoustic experiment) the curvature of the placement wand is designed to allow a human operator to stand in a fixed location approximately 1 m away from a receiver (a human subject, for example), and be able to move the source to any location within 1 m of the receiver in the hemisphere closest to the operator. The 135° bend in the placement wand enables the operator to keep the source oriented in the direction of the receiver throughout this area, eliminating any undesirable effects due to source directionality at high frequencies. The electromagnetic sensor allows a control computer to determine the exact location of the sound source rapidly even when manually placed by a human

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 243

operator.

The irregular frequency response of the device (Figure C-2) limits the use of the point source directly in many applications, but it is generally possible to compensate for the irregular response. In acoustic measurements, the stability of this transfer function allows it to be completely removed from a measurement. For example, in measuring Head-Related Transfer Functions, the desired quantity under test is the ratio of sound-pressure at the eardrum to the free-field sound pressure at the center of the head. In this application, the point source can be used without spectral compensation because the point-source transfer function is present in both measurements and is eliminated when the ratio between the measurements is calculated.

In other applications, an audio signal with a relatively flat frequency spectrum is required. In this case, the input signal to the point source can be electronically filtered by the inverse of its frequency response. This technique can be used to flatten the spectrum of the audio signal generated by the source. Figure C-6 shows the output of the point source when this technique was used to generate a low-pass filtered noise signal (6dB/octave rolloff above 200 Hz). Despite the large peaks and notches in the frequency response of the system (Figure C-2), many of which are 20 dB or larger in magnitude, the flattened response is within 1-2 dB of the desired response at all frequencies. Here a low-pass filtered signal was chosen because the response characteristics of the system allow greater non-distorted output with a low-pass filtered signal than with a flat signal. A white-noise spectrum could also be generated, but a only a lower total output could be achieved.

A final note is in order about the use of the system. The length and complexity of the propagation path from the source to the opening of the tube and consequent reflections result in a long and complicated impulse response, as well as a tendency towards intermodal distortion in high-output narrow-band signals. For these reasons, the source is better suited to broadband stimuli, and especially to noise signals, than to narrow-band or speech signals. Also, it is better suited to measurements using spectral averaging than to measurements which attempt to evaluate the impulse system directly.

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 244

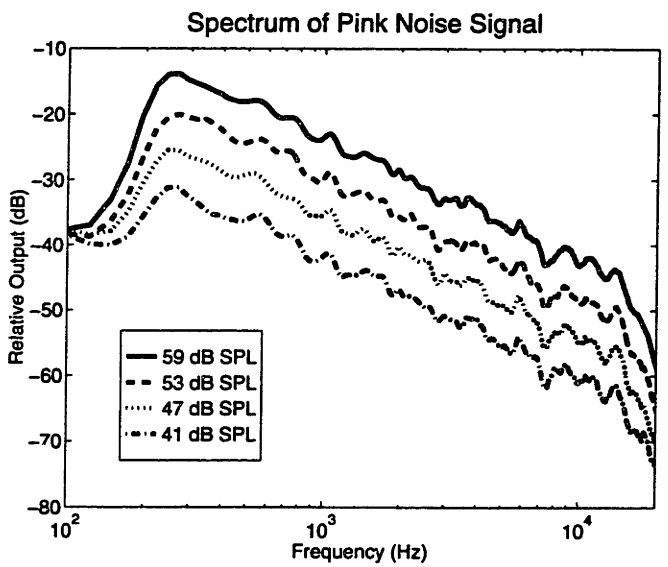


Figure C-6: Flattened Pink Noise Output. This figure shows the electronically flattened pink noise stimulus to be used in the proposed experiment. Four output levels (at 1 m) are shown. The measurements were made in an anechoic chamber using a noise stimulus and a 1024-point FFT (hanning window). One-twelfth octive smoothing has been used.

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 245

Bibliography

Ashmead, D. H., Davis, D. L., & Northington, A. (1995). Contribution of listener's approaching motion to auditory depth perception. *J. Exp. Psychol.: Hum. Percept. Perform.*, *21*, 239–256.

Ashmead, D., Davis, D. L., & Odom, R. D. (1990). Perception of the relative distances of nearby sound sources. *Percept. Psych.*, *47*, 326–331.

Begault, D. R. (1987). *Control of Auditory Distance.* Ph.D. thesis, University of California, San Diego.

Belendiuk, K., & Butler, R. (1977). Spectral cues which influence monaural localization in the horizontal plane. *Percept. Psycho.*, *22*, 353–358.

Blauert, J. (1983). *Spatial Hearing.* Cambridge, MA: MIT Press.

Brungart, D., & Rabinowitz, W. (1996). Auditory localization in the near-field.. Third International Conference on Auditory Display.

Burton, P. (1990). Audio speaker. United States Patent 4,924,506.

Butler, R. (1987). An analysis of the monaural displacement of sound in space. *Percept. Psychophys.*, *41*, 1–7.

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

Butler, R., Levy, E., & Neff, W. (1980). Apparent distance of sounds recorded in echoic and anechoic chambers. *J. Exp. Psychol.*, *6*, 745–750.

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 246

Carlile, S., & Pralong, D. (1994). The location dependent nature of perceptually salient features of the head-related transfer function. *J. Acoust. Soc. Am.*, *95*, 3445–3459.

Chan, J., & Geisler, C. (1990). Estimation of eardrum acoustic pressure and of ear canal length from remote points in the ear canal. *J. Acoust. Soc. Am.*, *87*, 1237–1247.

Cochran, P., Throop, J., & Simpson, W. (1968). Estimation of distance of a source of sound. *Am. J. of Psychol.*, *36*, 1171–1184.

Coleman, P. (1963). An analysis of cues to auditory depth perception in free space. *Psychol. Bull.*, *60*, 302–315.

Coleman, P. (1968). Dual role of frequency spectrum in determination of auditory distance. *J. Acoust. Soc. Am.*, *44*, 631–632.

Cook, R. (1996). Lord rayleigh and reciprocity in physics. *J. Acoust. Soc. Am.*, *99*, 24–29.

Devore, J. (1991). *Probability and Statistics for Engineering and the Sciences.* Brooks/Cole.

Duda, R. O., & Martens, W. L. (1997). Range dependence of the hrtf for a spherical head.. IEEE ASSP Workshop.

Durlach, N., & Braida, L. (1969). Intensity perception. i. preliminary theory of intensity perception. *J. Acoust. Soc. Am.*, *46*(2), 372–383.

Durlach, N., Rigoulos, A., Pang, X., Woods, W., Kulkarni, A., Colburn, H., & Wenzel, E. (1992). On the externalization of auditory images. *Presence*, *1*, 251–257.

Firestone, F. (1930). The phase difference and amplitude ratio at the ears due to a source of pure tone. *J. Acoust. Soc. Am.*, *2*, 260–270.

Foster, S. (1986). Impulse response measurements using golay codes. *IEEE Acoust. Speech Sig. Proc.*, *2*, 929–932.

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 247

Gardner, M. B. (1968). Proximity image effect in sound localization. *J. Acoust. Soc. Am.*, *43*, 163.

Gardner, M. B. (1969). Distance estimation of speech signals. *J. Acoust. Soc. Am.*, *48*, 47–53.

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

Gardner, W., & Martin, K. (1994). Kemar transfer functions. Available on Internet.

Gilkey, R., Good, M., Ericson, M., Brinkman, J., & Stewart, J. M. (1995). A pointing technique for rapidly collecting localization responses in auditory research. *Beh. Res. Meth., Inst. & Comp.*, *27*, 1–11.

Green, D. (1988). *Profile Analysis.* Oxford University Press.

Greene, D. (1968). Comments on 'perception of the range of a sound source of unknown strength'. *J. Acoust. Soc. Am.*, *44*, 634.

Hafter, E., Dye, R., Nuetzel, J., & Aronow, H. (1977). Difference threshold for interaural intensity. *J. Acoust. Soc. Am.*, *61*, 829–833.

Hartley, R., & Frey, T. (1921). The binaural location of pure tones. *Phys. Rev.*, *18*, 431–442.

Hebrank, J., & Wright, D. (1974). Spectral cues used in the localization of sound sources on the median plane. *J. Acoust. Soc. Am.*, *56*, 1829–34.

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

Henning, G. (1968). Some observations on the lateralization of complex waveforms. *J. Acoust. Soc. Am.*, *79*, 446–454.

Hershkowitz, R., & Durlach, N. (1969). Interaural time and amplitude jnds for a 500-hz tone. *J. Acoust. Soc. Am.*, *46*, 1464–1467.

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 248

Hirsch, H. (1968). Perception of the range of a sound source of unknown strength. *J. Acoust. Soc. Am.*, *43*, 373–374.

Holt, R., & Thurlow, W. (1969). Subject orientation and judgment of distance of a sound source. *J. Acoust. Soc. Am.*, *46*, 1584–1585.

Kinsler, L., & Frey, A. (62). *Fundamentals of Acoustics.* New York: Wiley & Sons.

Koehnke, J., & Durlach, N. (1989). Range effects in the identification of lateral position. *J. Acoust. Soc. Am.*, *86*, 1176–1178.

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

Litovsky, R., & Clifton, R. (1992). Use of sound-pressure level in auditory distance discrimination by 6-month-old infants and adults. *J. Acoust. Soc. Am.*, *92*, 794–802.

Little, A., Mershon, D., & Cox, P. (1992). Spectral content as a cue to perceived auditory distance. *Perception*, *21*, 405–416.

Lounsbury, B., & Butler, R. (1979). Estimation of distances of recorded sounds presented through headphones. *Scand Audiol* , *8*, 145–149.

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

Malme, C. (1959). Detectability of small irregularities in a broadband noise spectrum. Tech. rep., MIT Research Laboratory of Electronics.

McFadden, D., & Pasanen, E. (1976). Lateralization at high frequencies based on interaural time differences. *J. Acoust. Soc. Am.*, *59*, 634–1200.

McKinley, R., Ericson, M., & D'Angelo, W. (1994). Three dimensional audio displays: Development, applications, and performance. *Av., Space, and Env. Med.*, *65*, A31–38.

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 249

Mehrgardt, S., & Mellert, V. (1977). Transformation characteristics of the external human ear. *J. Acoust. Soc. Am.*, *61*(6), 1567–1576.

Mershon, D., Ballenger, W., A.D. Little, P. M., & Buchanon, J. L. (1989). Effects of room reflectance and background noise on perceived auditory distance. *Perception*, *18*, 403–416.

Mershon, D., & Bowers, J. (1979). Absolute and relative cues for the auditory perception of egocentric distance. *Perception*, *8*, 311–322.

Mershon, D., & King, L. (1975). Intensity and reverberation as factors in the auditory perception of distance. *Percept. Psycho.*, *18*, 400–415.

Mershon, D., & Philbeck, J. (1991). Auditory perceived distance of familiar speech sounds. *Proceedings of the Psychonomic Society 32nd Annual Meeting.*

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

Middlebrooks, J., & Green, D. (1991). Sound localization by human listeners. *Annu. Rev. Psychol.*, *42*, 135–139.

Middlebrooks, J., Makous, J. C., & Green, D. M. (1989). Directional sensitivity of sound-pressure levels in the human ear canal. *J. Acoust. Soc. Am.*, *86*, 89–107.

Miller, G. (1947). Sensitivity to changes in the intensity of white gaussian noise and its relation to masking and loudness. *J. Acoust. Soc. Am.*, *191*, 609–619.

Mills, A. (1958). On the minimum audible angle. *J. Acoust. Soc. Am.*, *30*, 237–246.

Molino, J. (1970). Perceiving the range of a sound source when the direction is known. *J. Acoust. Soc. Am.*, *53*, 1301–1304.

Moller, H., Sorensen, M., Hammershoi, D., & Jensen, C. B. (1995). Head-related transfer functions of human subjects. *J. Audio Eng. Soc.* , *43*, 300–320.

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 250

Moore, C., Oldfield, S., & Dooley, G. (1989). Detection and discrimination of spectral peaks and notches at 1 khz and 8 khz. *J. Acoust. Soc. Am.*, *85*, 820–836.

Morse, P., & Ingard, K. U. (1968). *Theoretical Acoustics.* McGraw-Hill.

Muscicant, A., & Butler, A. (1984). The influence of pinnae-based spectral cues on sound localization. *J. Acoust. Soc. Am.*, *75*, 1195–1200.

Musicant, A. D., & Butler, R. A. (1984). The psychophysical basis of monaural localization. *Hearing Research*, *14*, 185–190.

Oldfield, S., & Parker, S. (1984). Acuity of sound localization: a topography of auditory space. ii: Pinna cues absent. *Perception*, *13*, 601–617.

Oldfield, S., & Parker, S. (1986). Acuity of sound localization: a topography of auditory space. iii: Monaural hearing conditions. *Perception*, *15*, 67–81.

P. McGregor, A. H., & Todd, M. (1985). Are familiar sounds ranged more accurately. *Percept. and Motor Skills*, *61*, 1082.

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. *J. Acoust. Soc. Am.*, *95*, 3435–3444.

Rabinowitz, W., Maxwell, J., Shao, Y., & Wei, M. (1993). Sound localization cues for a magnified head: implications from sound diffraction about a rigid sphere. *Presence*, *2*, 125–129.

Rayleigh, L. (1907). On our perception of sound direction. *Philos. Mag.*, *13*, 214–232.

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

Rogers, M., & Butler, R. (1992). The linkage between stimulus frequency and covert peak areas as it relates to monaural localization. *Percept. Psycho.*, *52*, 536–546.

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 251

Shaw, E. A. G., & Teranishi, R. (1968). Sound pressure generated in an external-ear replica and real human ears by a nearby point source. *J. Acoust. Soc. Am.*, *44*, 240–249.

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

Simpson, W., & Stanton, L. D. (1973). Head movement does not facilitate perception of the distance of a sound. *Am. J. Psychol.*, *86*, 151–159.

Slattery, W., & Middlebrooks, J. (1994). Monaural sound localization: Acute versus chronic unilateral impairment. *Hearing Research* , *75*, 38–46.

Soechting, J., & Flanders, M. (1989). Sensorimotor representations for pointing to targets in three-dimensional space. *J. Neurophysiology*, *63*, 582–592.

Speigle, J., & Loomis, J. (1993). Auditory distance perception by translating observers. *IEEE 1993 Symposium on Research Frontiers in Virtual Reality.*

Stevens, S., & Newman, E. (1936). The localization of actual sources of sound. *Am. J. Psychol.*, *48*, 151–160.

Stewart, G. (1911a). The acoustic shadow of a rigid sphere with certain applications in architectural acoustics and audition. *Phys. Rev.*, *33*, 467–479.

Stewart, G. (1911b). Phase relations in the acoustic shadow of a rigid sphere; phase difference at the ears. *Phys. Rev.*, *34*, 252–258.

Strybel, T., & Perrott, D. (1984). Discrimination of relative distance in the auditory modality: The success and failure of the loudness discrimination hypothesis. *J. Acoust. Soc. Am.*, *76*, 318–320.

Toole, F., & Sayers, B. M. (1965). Lateralization judgments and the nature of binaural acoustic images. *J. Acoust. Soc. Am.*, *37*, 319–324.

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 252

Trahoitis, C., & Bernstein, L. (1986). Lateralization of bands of noise and sinusoidally amplitude-modulated tones: Effects of spectral locus and bandwidth. *J. Acoust. Soc. Am.*, *79*, 1950–57.

v. Bekesy, G. (1960). *Experiments in Hearing.* New York: Wiley.

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

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

Wenzel, E. (1991). Localization in virtual acoustic displays. *Presence*, *1*, 80–107.

Wightman, F. L., & Kistler, D. J. (1987). Monaural sound localization revisited. *J. Acoust. Soc. Am.*, *101*, 1050–1063.

Wightman, F., & Kistler, D. (1989a). Headphone simulation of free-field listening. i: Stimulus synthesis. *J. Acoust. Soc. Am.*, *85*, 858–877.

Wightman, F., & Kistler, D. (1989b). Headphone simulation of free-field listening. ii: Psychological validation. *J. Acoust. Soc. Am.*, *85*, 868–878.

Wightman, F., & Kistler, D. (1992). The dominant role of low-frequency interaural time differences in sound localization. *J. Acoust. Soc. Am.*, *91*, 1648–1660.

Woodworth, R. (1938). *Experimental Psychology.* New York: Holt.

Yost, W. (1976). Lateralization of repeated filtered transients. *J. Acoust. Soc. Am.*, *60*, 1526–2531.

Yost, W. (1981). Lateral position of sinusoids presented with interaural intensive and temporal differences. *J. Acoust. Soc. Am.*, *70*, 397–409.

Yost, W., Wightman, F., & Green, D. (1971). Lateralization of filtered clicks. *J. Acoust. Soc. Am.*, *50*, 178–181.

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 253

THESIS PROCESSING SLIP

FIXED FIELD: ill. _____ name _____

index _____ biblio _____

► COPIES (Archives) Aero Dewey (Eng) Hum

Lindgren Music Rotch Science

TITLE VARIES. ► ☐ _____

_____

_____

NAME VARIES: ► ☑ Scott _____

_____

IMPRINT (COPYRIGHT) _____

► COLLATION: 253 p _____

_____

► ADD. DEGREE: _____ ► DEPT.: _____

SUPERVISORS: _____

_____

_____ _____

_____ _____

_____ _____

_____ _____

_____ _____

_____ _____

_____ _____

NOTES:

cat'r: date:

page: F47

► DEPT: E.E.

► YEAR: 1998 ► DEGREE: Ph.D.

► NAME: BRUNGART, Douglas S.

Eight kHz, LLC
Exhibit 2013
IPR2023-01024
Page 254