MEASUREMENT AND VALIDATION OF BONE-CONDUCTION ADJUSTMENT FUNCTIONS IN VIRTUAL 3D AUDIO DISPLAYS

A Dissertation
Presented to
The Academic Faculty

by

Raymond M. Stanley

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Psychology

Georgia Institute of Technology
August 2009

MEASUREMENT AND VALIDATION OF BONE-CONDUCTION ADJUSTMENT FUNCTIONS IN VIRTUAL 3D AUDIO DISPLAYS

Approved by:

Dr. Bruce N. Walker, Advisor
School of Psychology
Georgia Institute of Technology

Dr. Gregory M. Corso
School of Psychology
Georgia Institute of Technology

Dr. Adrianus J.M. Houtsma
School of Psychology
Georgia Institute of Technology

Dr. Dennis J. Folds
Georgia Tech Research Institute

Dr. Paul M. Corballis
School of Psychology
Georgia Institute of Technology

Date Approved: June 26, 2009

ACKNOWLEDGEMENTS

I would like to thank my academic advisor, Bruce Walker, for his support and encouragement, his creativity in research, and his writing training. I would also like to thank Adrian Houtsma, for his careful feedback and enthusiastic tutoring on psychophysics and electrical engineering concepts. I am also grateful to the other members of my committee, Greg Corso, Paul Corballis, and Dennis Folds, for their efforts throughout all of graduate school to make me a better researcher and provide valuable guidance. I am indebted to Ewan McPherson, who served as an external digital signal processing (DSP) consultant. Without his generous expertise and tutoring, this project could not have been executed. Thanks to Devangi Parikh for her valuable DSP assistance. Many thanks to John Middlebrooks for his sharing of HRTFs and MATLAB code. Thanks to all at Wright Patterson Air Force Base who assisted in the measurement of my HRTFs and other technical matters. Thanks also to Tim Streeter for his assistance in the measurement of HRTFs and DSP, as well as Barb Shinn-Cunningham for allowing use of her lab's facilities and personnel. To my labmates and fellow graduate students at Georgia Tech: I have never been surrounded by so many kind, intelligent, and entertaining people. I want to thank Nestor Matthews for his enthusiastic introduction to psychological science and MATLAB.

I am eternally grateful to my mother for her unconditional love, and my father for setting the example of a relentless work ethic. To my Uncle Mason and Aunt Sally, for their guidance provided at such a crucial point of my life. To my wife, Jenny, who's value I simply cannot put into words. Thank you for your constant unconditional support and patience.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Page

LIST OF SYMBOLS AND ABBREVIATIONS

ANOVA	Analysis of variance
BAF	Bone-conduction adjustment function
BCT	Bone-conduction transducer
CLT	Central Limit Theorem
DTF	Directional transfer function
HRTF	Head-related transfer function
ILD	Interaural level difference
ITD	Interaural time difference
SWAN	System for wearable audio navigation
V3DAD	Virtual 3D audio display

SUMMARY

Virtual three-dimensional auditory displays (V3DADs) use digital signal processing to deliver sounds (typically through headphones) that seem to originate from specific external spatial locations. This set of studies investigates the delivery of V3DADs through bone-conduction transducers (BCTs) in addition to conventional headphones. Although previous research has shown that spatial separation can be induced through BCTs, some additional signal adjustments are required for optimization of V3DADs, because of the difference in hearing pathways. The present studies tested a bone-conduction adjustment function (BAF) derived from equal-loudness judgments on pure tones whose frequencies were spaced one critical band apart. Localization performance was assessed through conventional air-conduction headphones (the "air" condition), BCTs with only transducer correction (the "unadjusted bone" condition), and BCTs with a BAF applied (the "adjusted bone" condition). The results showed that in the elevation plane, the BAF was effective in restoring the spectral cues altered by the bone-conduction pathway. No evidence for increased percept variability or decreased lateralization in the bone-conduction conditions was found. These findings indicate that a V3DAD can be implemented on a BCT and that a BAF will improve performance, but that there is an apparent performance cost that cannot be addressed with BAFs measured using the methodology in the present studies.

CHAPTER 1

INTRODUCTION

1.1. Background and Impetus

Virtual three-dimensional auditory displays (V3DADs) apply digital signal processing to deliver sounds (typically through headphones) that seem to originate from specific external spatial locations. They allow delivery of any sound to any location in the horizontal or median plane, located at a fixed short distance external to the listener's head. V3DADs take advantage of the auditory system's sensitivity to interaural difference cues that lead to localization in the horizontal plane (Blauert, 1997) and spectral cues that lead to localization in the median plane (Blauert, 1969/1970; Burger, 1958; Oldfield & Parker, 1984b; Rayleigh, 1907; Roffler & Butler, 1968; Searle, Braida, Cuddy, & Davis, 1975) as well as externalization (Blauert, 1997; Wenzel, 1992).

The application of virtual 3D audio technology in auditory displays has many important uses, such as increasing the detectability of radio communication signals amidst distracters and noise (e.g., Brungart & Simpson, 2002) and providing orientation cues in cases of vision loss (e.g., Walker & Lindsay, 2006). Because V3DADs are typically delivered via headphones, the detection and localization of environmental sounds can deteriorate because the ears are being covered. This can be a problem when spatialized audio and sounds in the environment are both important for the user's task, such as with audio navigation systems like the System for Wearable Audio Navigation ("SWAN" (Walker & Lindsay, 2006)), or a tactical environment where a soldier needs to hear events occurring around him in combination with radio communications. In addition to performance, the level of comfort or confidence that the individual has when his ears

are covered may decrease – both sighted and visually impaired users of SWAN have reported that they would not use the system outdoors if it covered their ears (Walker, Stanley, & Lindsay, 2005).

These situations would benefit from an alternative to headphones. Because the auditory system is also sensitive to pressure waves transmitted through the bones in the skull (Békésy, 1960; Kelley, 1937; Tonndorf, 1972), bone-conduction transducers may provide an acceptable solution. Bone-conduction transducers (BCTs) leave the ear canal and pinna uncovered, which facilitates the detection and localization of environmental sounds. Furthermore, BCTs allow the display of auditory information even when earplugs are inserted into the ear canal, which cannot be done with headphones. The ability to protect hearing while simultaneously delivering information via V3DADs may be useful in aviation or tactical environments where loud ambient noise and spatialized radio communications could co-occur.

Although still not widespread, the use of BCTs to deliver sound is not new. Special-purpose BCTs have typically been used in clinical audiology settings to assess the locus of hearing damage in patients (e.g., Robinson & Shipton, 1982; Small & Stapells, 2003). There has also been research seeking to understand the fundamental mechanisms of bone-conduction hearing (Békésy, 1960; Freeman, Jean-Yves, & Sohmer, 2000; Stenfelt, 2006; Stenfelt, Hato, & Good, 2004; Tonndorf, 1966). This research on mechanisms has shown that there is a considerable difference between the bone-conduction pathway and air-conduction pathway, although they share the same end organ of the cochlea.

Sensitivity to bone-conducted lateralization cues (interaural level differences (ILDs) and interaural time differences (ITDs)) is a minimum requirement for implementation of virtual 3D audio on BCTs – without this, no changes in localization on the horizontal plane can be simulated. Research has shown that bone-conducted interaural differences can be detected. Jahn and Tonndorf (1982) were perhaps the first to consider lateralization through a pair of BCTs, reporting that the subjective impression of lateralization occurred for the authors when ILDs and ITDs were invoked. Kaga, Setou, and Nakamura (2001) later measured the ILDs and ITDs at which subjective lateralization first occurred for sounds delivered through bone-conduction and air-conduction transducers. They found that there was not a statistically significant difference in the sensitivities to these cues between the bone-conducted sound and air-conducted sound. Researchers have also measured ILD-lateralization functions for bone-conduction and air-conduction, finding similar functions for bone and air (Stanley & Walker, 2006). In another study, Walker, Stanley, Iyer, Simpson, and Brungart (2005) showed that ITDs and ILDs implemented through BCTs to separate target speech signal from masker speech signals can enhance speech intelligibility with BCTs. Furthermore, people have successfully used the SWAN auditory navigation system with BCTs (Walker & Lindsay, 2005), which guides users to their destination using spatial cues. MacDonald, Henry, and Letowski (2006) implemented a virtual 3D audio display on BCTs and regular headphones and found similar localization performance in the horizontal plane. In addition, others have shown sensitivity to binaural aspects of Bone-Anchored Hearing Aids (e.g., Snik, Beyon, van der Pouw, Mylanus, & Cremers, 1998; Snik, Bosman, Mylanus, & Cremers, 2004).

Together, these studies suggest that sufficient spatial separation can be induced to manipulate an auditory object's location on the horizontal plane for V3DADs implemented on BCTs. However, sufficient spatial separation does not mean that they are just as effective. In general, differences in the spatial audio capabilities between air conduction and bone conduction would not be surprising, given the difference in pathways that sound travels for air-conduction and bone-conduction hearing. Schonstein, Ferré, and Katz (2008) considered localization performance on V3DADs with equalized BCTs relative to equalized headphones for one participant in the horizontal and median planes, and found a decrease all performance metrics with the BCTs. Furthermore, Walker and colleagues (2005) found that although spatial separation improved speech intelligibility, the intelligibility was consistently lower for bone than air. In addition, Stanley and Walker's (2006) lateralization study did not produce perceived lateralization greater than 75 percent. Thus, there could be differences in the perceived lateralization between the pathways when interaural differences corresponding to greater lateralization are tested. MacDonald, Henry, and Letowski (2006) used a restricted number of locations, ambient masking noise, and a band-pass filter that cut out frequencies above 5000 Hz, all of which may have prevented revealing differences in auditory localization between the headphones and BCTs. Limitations in the maximum eccentricity on the horizontal plane for BCTs are likely, given the lower interaural attenuation between the ears for bone conduction than for air conduction (Dean, 1930; Hallmo, Sundby, & Mair, 1992; Hurley & Berger, 1970; Nolan & Lyon, 1981; Snyder, 1973).

In addition to remaining questions about localization in the horizontal plane, little information is available on how localization in the elevation plane and externalization is

altered when delivering V3DADs implemented on BCTs. Given the differences in hearing pathways, there are also likely to be large inaccuracies in the spectral cues when using BCTs to implement V3DADs. Because of the role of spectral cues in median plane localization and externalization, these inaccuracies are likely to result in limited or inaccurate elevation, front/back discrimination, and externalization of auditory objects in V3DADs. The study by Schonstein and colleagues (2008) showed data consistent with this prediction. Given these differences in spatial audio between bone conduction and air conduction, some adjustments to the signal for the bone-conduction pathway are required to test the maximum effectiveness of BCTs in V3DADs.

Before discussing adjustments for V3DADs, the methodology for creating V3DADs will first be briefly reviewed. Following seminal work showing the ability to simulate externalized sound sources over headphones (Doll, Gerth, Engleman, & Folds, 1986), researchers showed that signal processing filters could be used to produce V3DADs through air conduction (e.g., Wightman & Kistler, 1989a, 1989b). Head-related transfer functions (HRTFs) produce V3DADs by providing a filter that any sound can be sent through so that the sound seems to come from a specific location outside of the head. Creating an HRTF consists of several steps. First, a tiny microphone is inserted into a person's ear canal to record the response to a broadband sound emitted from specific locations in the free field. Then the Fourier transform of this response is divided by the Fourier transform of the signal, which creates the HRTF. The HRTF characterizes all the primary cues for localization - the spectral filtering achieved by the pinna, as well as interaural differences in time and intensity.

The critical issue remaining for V3DADs on BCTs is how the sound could be

adjusted to optimize HRTFs for BCTs. This issue is the focus of the studies presented here. There is no way to measure HRTFs for bone conduction in an analogous way to how they are measured for air conduction, because spatial audio does not occur naturally through bone conduction. It is possible, however, to understand how to alter a given bone conduction signal so that it sounds like an air-conduction signal. One result of understanding this relationship would be a filter specifying amplitude shift values across frequencies that equate the two signals. Such a function could then be used to adapt the spectral features of the HRTFs so that they are more suitable for use with BCTs. I will call this function a "bone-adjustment function," or BAF. It is hypothesized that a sound could be passed through an HRTF and BAF, then presented via BCTs, causing the sound source to be localized at a particular point in space outside the listener's head. The current studies present a methodology for measuring, implementing, and assessing the efficacy of BAFs.

1.2. Overview of Studies

The first study described in this document (Study 1) measures BAFs through a loudness-matching paradigm between bone-conduction and air-conduction sounds. This method of defining BAFs includes the frequency response of the transducer, which can be subtracted from the BAF after the measurement. In Study 2, the efficacy of the BAFs are tested by measuring localization ability with V3DADs through air (the "air" condition), V3DADs through bone with only a transducer correction applied (the "unadjusted bone" condition), and V3DADs through bone with a BAF applied (the "adjusted bone" condition).

1.3. Theoretical Implications of Proposed Studies

In a preliminary exam paper by the author, a "Limited-Fidelity Model of Bone-Conducted Spatial Audio" was described, based on the aforementioned and other literature. That model predicts results of the studies described in this document, and these studies further inform the model. In the context of this research, fidelity is defined as the extent to which simulated spatial audio implemented on BCTs can match the perceptual saliency and localization performance shown with simulated spatial audio delivered through headphones. This model asserts the feasibility of using BCTs to deliver spatial audio while also highlighting factors limiting the ultimate fidelity. The Limited-Fidelity Model of Bone-Conducted Spatial Audio consists of four components:

A) Theory of leaky dichotic bone-conduction hearing: there is a greater degree of binaural separation possible through bone conduction than is typically thought (i.e., none), but there is a maximum limit that is less than the separation possible through air conduction.

B) Air-conduction filters are ill-suited for bone conduction pathway: Air-conduction filters do not allow the best possible performance to be gained from BCTs.

C) Measurable pathway: the bone-conduction pathway is measurable using a variety of methods, so that adjustments can be made to the air-conduction filters.

D) Imperfect filter effectiveness: Despite an ability to measure adjustment filters, the maximum effectiveness of the filters will be limited by several factors inherent to the bone-conduction pathway. These include inconsistency within a person because of sensitivity to movements, transducer placement, cardiac system changes, and respiratory system changes, variability in the resulting cochlear

stimulus across people, co-occurring independent bone-conduction hearing mechanisms, antiresonances, and the "leaky" aspect of dichotic stimulation.

1.4. Major Methodological Decisions

1.4.1. Adjustment Function Methodology

A variety of methods can be used to understand the relationship between bone-conducted and air-conducted sounds. One method is to measure the relationship behaviorally through a bone-to-air matching paradigm. Previous work aimed to test the feasibility of using the bone-to-air cancellation paradigm to measure BAFs (Stanley, 2006). In this paradigm pioneered by Békésy (Békésy, 1932 as cited in Békésy, 1960), a participant adjusts the phase and amplitude of a signal (e.g., air-conducted) in one ear so that it cancels out the other signal (e.g., bone-conducted) in the same ear, thus producing silence (or at least a significant reduction in volume). In the Stanley (2006) work, the amplitude and phase shifts required at a particular frequency to reach this cancellation were used to define a "bone-to-air shift". The eventual goal was to establish a set of bone-to-air shifts for a large number of frequencies that could then be combined to build a BAF. Stanley (2006) measured bone-to-air amplitude and phase shifts at three frequencies. The data provided a preliminary (sparse) set of bone-to-air amplitude shifts that could be used to build BAFs.

There were difficulties associated with the cancellation paradigm in the Stanley (2006) work. These included sensitivity of cancellation to head movement, variability between people in the degree of cancellation achieved, and the length of time required to complete the task. Each session consisted of 80 trials and took between 45 minutes and 2 hours to complete. In addition to the difficulties associated with the method, the phase-

shift data differed greatly between participants. The amplitude data, however, did produce stable results. Consideration of binaural synthesis literature suggests that listeners are not sensitive to aspects of the phase spectrum within a transfer function applied to a single ear (Kistler & Wightman, 1992; Kulkarni, Isabelle, & Colburn, 2001; Wightman & Kistler, 1997). Phase variability between ears, of course, affects the ITD cue, which would have an effect on localization.

Together, these findings suggested that a loudness-matching paradigm devoid of any phase adjustments may still be useful for building BAFs. The studies described in this document tested that hypothesis. The present studies measured BAFs using a method-of-adjustment equal-loudness paradigm, which provided an efficient way of obtaining judgments for 33 frequencies at each ear, with a wide range of loudness levels available to the participant.

One study has investigated the point of equal loudness for bone- and air-conducted stimuli (Stenfelt & Håkansson, 2002). Stenfelt and Håkansson had listeners adjust the loudness of a bone-conducted tone until it matched the loudness of an air-conducted tone. This was done at six frequencies ranging from 250 Hz to 4 kHz, and at six levels ranging from 30 dB to 80 dB HL. These data are not sufficient for developing a BAF, however, because measuring equal loudness for a BAF requires testing more frequencies, including those above 4 kHz.

1.4.2. Selection of HRTFs

As discussed above, HRTFs can be measured for an individual in order to characterize the spectral filtering and interaural differences that occurs for a free-field sound source. Although individualized measurement gives the highest fidelity V3DAD, it

requires considerable technical expertise and equipment that were not available in the Sonification Lab. For an initial study considering the effectiveness of bone-conducted spatial audio and adjustment filters, a more efficient method was better suited.

Several approaches have been taken to avoid having to measure HRTFs for each individual participant but retain localization performance. These approaches include finding a "generalized" HRTF (e.g., Møller, Sørensen, Jensen, & Hammershøi, 1996), and selecting a "best-fit" HRTF from a public database of HRTFs, based on objective (Ahmad, Stanney, & Fouad, 2008) and subjective (Seeber & Fastl, 2003) selection routines. Recent work has identified specific acoustic characteristics of HRTFs that govern how well an HRTF works for an individual. Middlebrooks (Middlebrooks, 1999a; Middlebrooks, Makous, & Green, 1989) observed that the position of spectral features along the frequency axis varies between listeners for directional components of HRTFs, called "directional transfer functions" (DTFs). DTFs are a variant of HRTFs; more details on DTFs can be found in Section 3.1.3. Middlebrooks (Middlebrooks, 1999a; Middlebrooks, et al., 1989) further showed that the position of spectral features can be aligned between participants by applying frequency scaling to the DTFs, and that the scaling factor was related to size of the participants' head and external ears. In addition to showing differences in these acoustic features between individuals, Middlebrooks (1999b; 2000) showed that a scaling factor can be applied to non-individualized DTFs to improve virtual localization for a particular listener. Furthermore, Middlebrooks and colleagues (2000) showed that applying a scaling factor determined by perceptual judgments or by anthropometry could typically reduce quadrant errors (errors greater than 90 degrees in the vertical or front/back dimension) by half, relative to non-

individualized DTFs. In some cases, the number of errors even approached that of performance measured with individualized DTFs. Middlebrooks' customization routine (Middlebrooks, 1999a, 1999b; Middlebrooks, et al., 2000) was used in the present studies.

1.4.3. Measurement of Localization Judgments

Investigators measuring localization abilities with V3DADs have used many different methods for their listeners to indicate where in space they hear a sound in both the horizontal and median planes. These include: pointing with a wand that has an electromagnetic sensor on its tip (Brungart, Rabinowitz, & Durlach, 2000); pointing with their hand and taking a picture of its location in two planes (Oldfield & Parker, 1984a); pointing with their nose and monitoring the head's location with a head tracker (Makous & Middlebrooks, 1990); verbal report of spherical coordinates (Wightman & Kistler, 1989b); placing a position sensor at a location on a manikin head that corresponds to the perceived location relative to the listener's head (Gilkey, Good, Ericson, Brinkman, & Stewart, 1995); and indicating location on a computer screen displaying a two dimensional model of a head (MacDonald, et al., 2006; Møller, et al., 1996). Each of these methods has its own unique advantages and disadvantages, and all have some amount of error.

The present studies had participants indicate the position of a sound source with a computer mouse by clicking on a set of two-dimensional drawings of a head on a computer screen. Further details about this response method are available in Section 3.1.2. This method avoids the slowness of verbal response and the technical and financial resources required for response systems involving sensors. This method borrows

techniques similar to Møller (1996) and MacDonald (2006), and uses head models adapted from Middlebrooks (2000). Pilot testing ensured that participants could understand and use this interface after minimal practice.

CHAPTER 2

STUDY 1 - THE BONE-CONDUCTION ADJUSTMENT FILTER

2.1. Method

2.1.1. Participants

The participants were ten paid participants from the graduate and undergraduate community of the Georgia Institute of Technology (six male, four female; mean age = 26.2; age range = 22 to 30). They were required to have normal hearing (sensitivity to 20 dB HL pure tones), as tested by an audiometer.

2.1.2. Apparatus

The air-conducted stimuli were presented through Etymotic ER-1 insert headphones (see Figure 1), which have a diffuse-field frequency response (see Appendix D). These insert headphones avoid the variability in stimulation that occurs with earbuds, avoid the interference with the mastoid area that occurs with supraaural and circumaural headphones, and have been used in other studies for delivering V3DADs (e.g., Carlile & Best, 2002). Bone-conducted stimuli were delivered through a Teac HP-F100 Filltune bone conduction transducer (see Figure 2). This is a newly released transducer that is built for everyday use by the general public for applications such as listening to music. It was used rather than the more common RadioEar B-71 built for audiometry because of its relatively flat frequency response that extends up to 18 kHz (see Figure 3) compared to the B-71's steep drop-off at 4 kHz and maximum frequency of 10 kHz. The Teac transducer was applied to the mastoid location of the head directly behind the external ear.

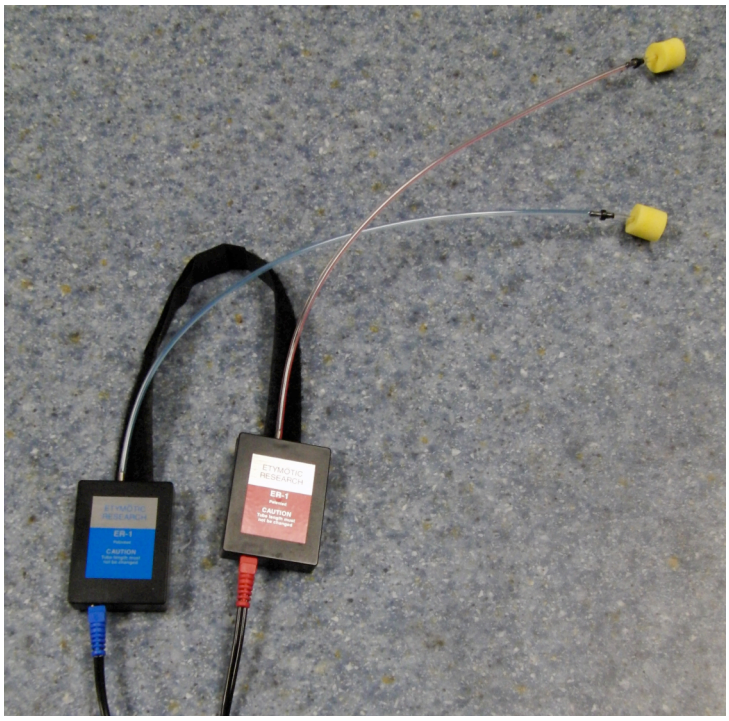

Figure 1. Photograph of Etymotic ER-1 insert headphones used in the present studies.

Figure 2. Photograph of Teac HP-F100 "Filltune" bone-conduction transducers used in the present studies.

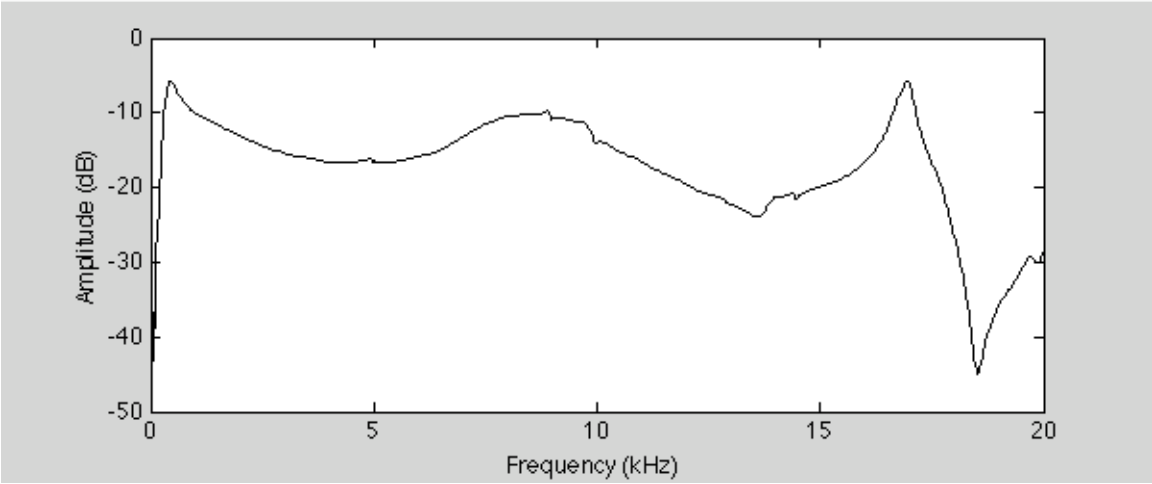

Figure 3. Frequency response of Teac Filltune bone conduction transducer, as measured in Sonification Lab by Bruel & Kjaer PULSE system analyzer and Bruel & Kjaer Type 4930 Artificial Mastoid. The input sent through the bone conduction transducer was a swept sine wave. This frequency response was measured for each transducer (left and right), and averaged across them.

Participants used a computer keyboard and mouse to indicate their responses and receive supporting experiment information through the visual display. Stimuli delivery, response collection, stimuli creation, and data analysis were all done via scripts written in Matlab.

Auditory stimuli were sent out digitally through an Apple iMac computer (2 GHz Intel Core 2 Duo) to an M-Audio Firewire 410 external sound card. The air conduction channels were sent through a Furman HA-6AB headphone amplifier before being sent to the headphones. The bone-conducted stimuli were sent from the line outputs of the sound card directly into the battery-powered amplifier that comes with the Teac HP-F100 Filltune bone conduction transducers, which passed the signal on to the transducers. During measurements of the distortion point of the bone-conduction transducers (which will be described below), a dependence of levels (especially second harmonics) on battery power was observed. Specifically, fundamental levels began to decrease and the

relative level of the second harmonic began to increase after the battery level dropped below a specific level on a voltmeter. Therefore, all batteries were tested before each session began and not used if they were below this level.

Before each session began and the participant arrived, the experimenter measured the output from the line level of the sound card (for the bone conditions) and the output from the headphone amplifier (for air conditions) with a voltmeter. This value was monitored for consistency to ensure that no changes had been made in the levels. After the participant arrived, the experimenter mounted the bone-conduction transducer onto the participant, and the participant inserted the headphones. Before the main task of the experiment began, a test sound was played through the left and right transducer for both air and bone. This ensured that the devices were mounted properly (i.e., the channels were not reversed) and were functioning correctly. After the sensitivity to headphone insertion depth was realized, pictures of each ear and its surrounding area were taken with the headphones inserted and BCTs mounted, before each session began.

Loudness was matched between bone and air using the interface shown in Figure 4. The participant commenced each trial by pressing the "Play Sound" button. This initiated a loop consisting of a monaural tone being played through the air-conduction headphones, 500 milliseconds of silence, and then an ipsilateral monaural tone being played through the bone-conduction transducer. This loop continued while the participant adjusted the slider, until the participant pressed the "Submit Value" button (see Figure 4). The bone-conducted (test) tone started at a soft level, and the participant adjusted its volume by moving the slider with the mouse. Pressing the "Submit Value" button recorded the amplitude value of the test tone and advanced parameters to the next pure

tone to match, which was again be initiated by the “Play Sound” button.

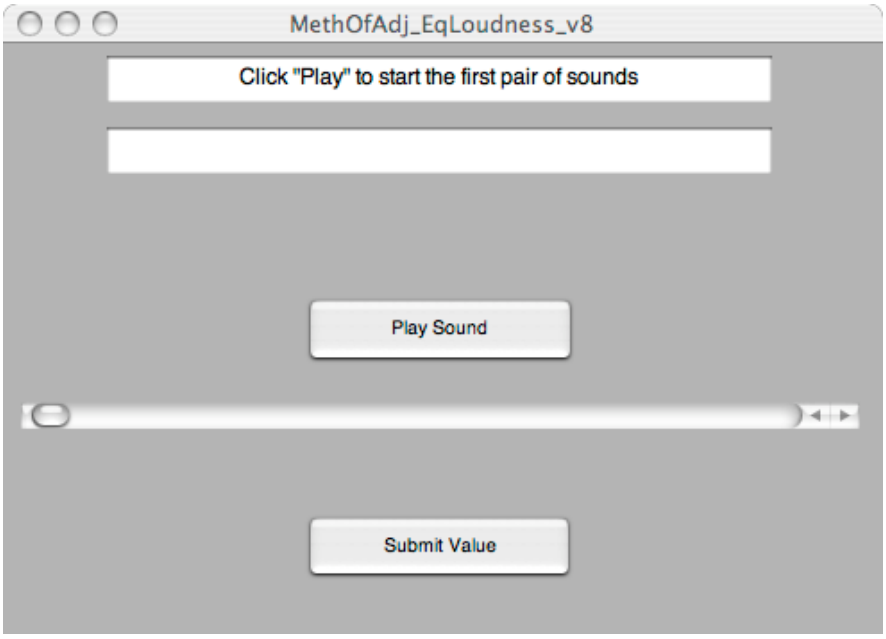

Figure 4. Interface used for making equal-loudness matches, using the method of adjustment.

2.1.3. Stimuli

Auditory stimuli were delivered through a single bone-conduction or air-conduction transducer on the same side of the head during any given trial. The stimuli were pure tones at the 33 critical band centers lying between 250 and 16500 Hz (the bandpass limits of my final virtual 3D audio targets). The critical bandwidths were calculated using a combination of equations. The critical bandwidths were calculated using the equation provided in Glasberg and Moore (1990),

$$ERB_N = 24.7(4.37F+1),$$

where ERB_N = the normal equivalent rectangular bandwidth, and F = the center frequency. These bandwidths were used to determine the upper and lower bound of each

critical band, and then the center for each critical band was computed using the geometric mean,

$$M_{geometric} = \sqrt{U * L},$$

where $M_{geometric}$ = the geometric mean, U = the upper bound of the critical band and L = the lower bound of the critical band. The upper bound, lower bound, and centers of resulting from these computations can be seen in Table 1.

Table 1

Lower bound, center, and upper bound of critical bands calculated based on Glasberg and Moore (1990). Pure tones for the equal loudness judgments were at each of the center frequency values.

Lower (Hz)	Center (Hz)	Upper (Hz)
14789	15621	16500
13253	14000	14789
11874	12545	13253
10637	11238	11874
9525	10066	10637
8528	9013	9525
7632	8067	8528
6828	7219	7632
6106	6457	6828
5458	5773	6106
4877	5159	5458
4354	4608	4877
3886	4113	4354
3465	3669	3886
3087	3271	3465
2748	2913	3087
2444	2592	2748
2171	2303	2444
1925	2044	2171
1705	1812	1925
1508	1603	1705
1330	1416	1508
1171	1248	1330
1028	1097	1171
900	962	1028
784	840	900
681	731	784
588	633	681
505	545	588
430	466	505
362	395	430
302	331	362
248	273	302

The stimuli had the same duration and amplitude envelope as in the subsequent virtual audio localization test (Study 2): 250 milliseconds long, with 10-millisecond raised-cosine onset and offset ramps.

The air-conducted stimuli were delivered at a maximum narrowband level of 75 dB SPL. The standardized acoustic couplers for measuring levels coming out of the ER-1 insert headphones were not available. Homemade versions of such couplers were built and attempted for use, but resulted in measurements that were grossly inaccurate (as indicated by the author's perceptual listening tests and output calculations based on voltage sensitivity). Thus, sensitivity specifications measured at the factory on the exact pair of headphones used in the study were utilized to compute the sound pressure level output. The sensitivity values give the output sound pressure level for a given input voltage into the system, at a specific frequency. The transducer's calibration charts showed that given an input of 320 millivolts, the output at 2500 Hz was 107 dB SPL (2500 Hz was the peak of the diffuse-field frequency response). Because of the proportionality property of a linear system, a logarithmic reduction in voltage will result in an equal logarithmic reduction in output. A 32 decibel reduction in transducer output (down from 107 dB) was achieved by a 32 decibel reduction in input voltage, which equaled 8 millivolts ($20*\log_{10}(320/8) = 32$ dB), producing a peak output of 75 dB SPL at 2.5 kHz. The SPL measured across frequency (a broadband measurement) would be lower, given the diffuse-field frequency response of the headphones used in this study. The value of 75 dB was chosen so that the weaker frequencies (as defined by the diffuse-field response) were still audible, yet the stronger frequencies were still in safe ranges,

while leaving space for the bone-conducted tone to be adjusted to be louder than the air-conduction tone.

There was a minimum (lower bound) and maximum (upper bound) level to which the bone-conducted stimuli could be adjusted. The lower-bound bone-conduction levels were determined by pilot testing, set to a level that was audible but much softer than the test stimulus. The upper-bound bone-conduction levels were initially determined by piloting, set to a level that was noticeably louder than the reference stimulus while still at a comfortable level. These upper-bound values had to be adjusted because of distortion that occurred in the bone-conduction transducer – harmonics above the pure tones could be heard, which would effect the loudness adjustments. The point of unacceptable distortion was operationally defined as the highest adjustment point at which the highest amplitude harmonic was greater than or equal to 30 dB down from the fundamental. These values were determined using a Bruel & Kjaer Type 4930 Artificial Mastoid coupled to a Bruel & Kjaer PULSE system analyzer, using one-third octave band analysis. The maximum allowable adjustment level was taken as the minimum adjustment level causing the point of unacceptable distortion, taken across five measurements for each transducer. The upper-bound bone-conduction adjustment level was set to the piloted value or the distortion point, whichever was the lowest level.

2.1.4. Procedure

Participants ran themselves through trials to match the loudness of the bone-conducted tone to the air-conducted tone using the aforementioned software apparatus. The ear that the stimulus was delivered to was randomly selected, and the order of frequencies administered for this ear was randomized. After all frequencies in one ear

were complete, the order of frequencies was randomized again. The total study time (including procedures described below) was about 75 minutes.

At the higher frequencies, sometimes the participants could not hear the reference tone. In this case, the participants were instructed to nudge the slider just from the minimum, and submit this value. The nudge was required because the slider had to be moved for the value to be submitted, to prevent accidental submission. For two participants the 14 kHz tone was inaudible in one or both ears, and for three participants the 15.6 kHz tone was inaudible in one or both ears.

For the first few participants, some of the bone-conducted upper-bound adjustment values were not high enough to achieve equal loudness, and participants had to come back to re-do their adjustments, with a higher upper bound. The participants' equal-loudness adjustments had to be greater than 3 dB attenuation from the top for the adjustment to be deemed acceptable. Acceptability of adjustments was determined after the session terminated.

After the equal-loudness points used to form the BAFs were measured, participants made equal-loudness judgments that would later be used to determine the volume of the unadjusted bone condition. The air-conducted stimulus was a 250-millisecond broadband noise bandpassed between 250 and 16500 Hz (the same frequency cutoffs as the virtual audio localization study); the bone-conducted stimulus was identical but was also filtered by the inverse transducer frequency response (as described in Section 3.1.3). The same interface shown in Figure 4 was used and again the air-conducted noise was always the reference stimulus.

2.1.5. Customization of DTFs

After the loudness matches were completed, measurements were taken to determine the participant's appropriate scale factor, using the anthropometric procedure described by Middlebrooks (1999a). Middlebrooks (2000) found that a person's head width and pinna height were good predictors of a person's appropriate scale factor (r = .82). The head width was measured with a set of calipers, just in front of the tragus (see Figure 5). The pinna height was measured with a set of calipers, extending from the inter-tragal notch to the rim of the helix (see Figure 5). These values were then entered into the equation provided by Middlebrooks (1999a) to determine the appropriate scale factor:

$$\log_2(\text{scale factor}) = 0.340 * \log_2(\text{pinna}_{meas} / \text{pinna}_{ref}) + 0.527 * \log_2(\text{head}_{meas} / \text{head}_{ref})$$

where pinna_{meas} and head_{meas} are the measured pinna height and head width of the participant, and pinna_{ref} and head_{ref} are the pinna height and the head width of the person whose DTFs were used in this study ("medium" (S44) DTFs from Middlebrooks (2000)). The anthropometric data and calculated scale factors can be seen in Table 2.

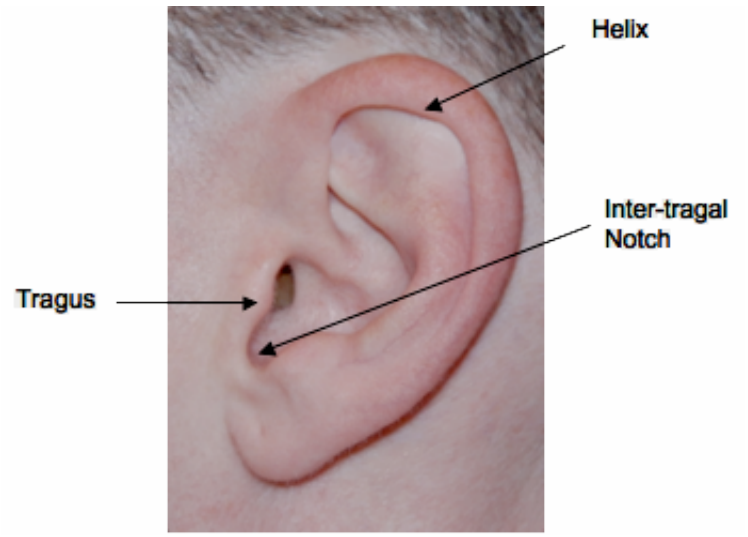

Figure 5. Ear anatomy, showing landmark features used for anthropometric measurements: tragus, helix, and inter-tragal notch. Landmarks determined based on Middlebrooks (1999a).

Table 2

Anthropometric data for the ten participants in these studies and for the DTF "base" participant.

Subject	Pinna Height (mm)	Head Width (mm)	Scale Factor
1	44	126	100.13
2	45	125	100.47
3	37	125	94.03
4	45	127	101.31
5	46	135	105.41
6	45	135	104.63
7	46	143	108.65
8	50	134	108.00
9	50	146	112.99
10	56	139	114.41
Base	40.7	132	100.00

2.1.6. Building Bone-Conduction Adjustment Filters

After Study 1 was completed, the data were used to build the BAF used in Study 2. Using the equal-loudness points measured at each frequency, 1025 values (corresponding to a 2048 tap filter) were interpolated to form a high-resolution BAF extending from 0 Hz to 22.05 kHz, as a frequency-domain filter (see Figure 6). The amplitude of the BAF was ramped up from zero at frequencies between 0 Hz and 223 Hz and down to zero at frequencies between 17959 and 22,050 Hz. The 223 Hz and 17959 Hz values were chosen because they were one sixth octave away from the band limits of the final stimuli (250 and 16500 Hz). Filters were built for each ear, and applied separately to each ear. The filters were also built and applied separately for each individual, thus creating individualized BAFs. Further details on implementation of the BAF can be found in Section 3.1.3.

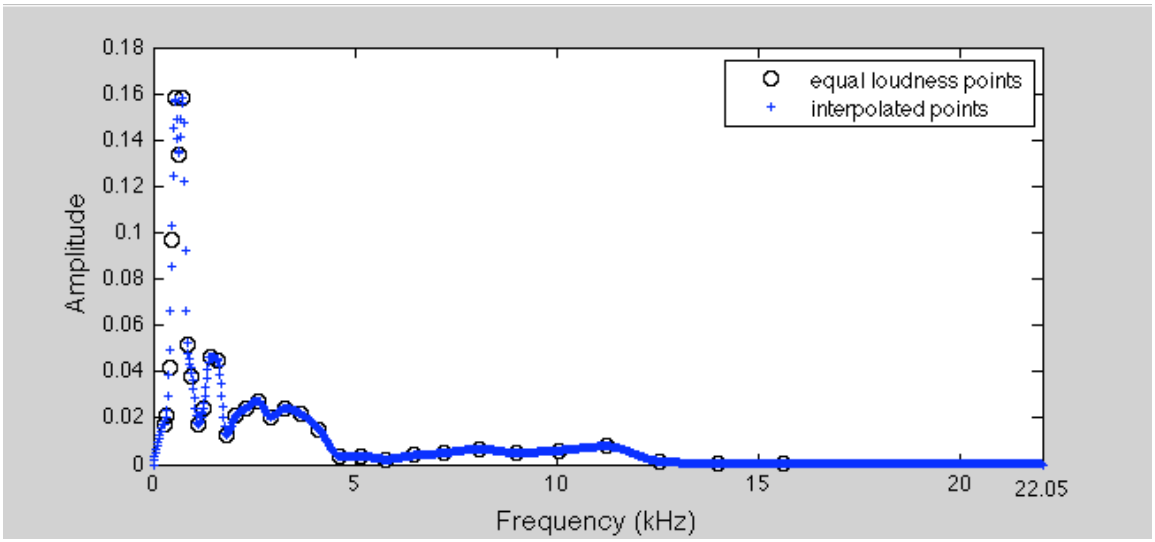

Figure 6. Sample equal-loudness points (measured in Study 1), with 1025 interpolated points to form frequency-domain BAF. This filter was applied to stimuli in Study 2.

2.2. Results

Appendix A shows the BAF and equal loudness points for each participant (and each ear) in separate plots. Figure 7 shows the BAFs of all participants, averaged across ears. Figure 7 also shows the average BAF (the BAF averaged across participants, and shifted upwards by 15 dB for graphical clarity) and the transducer frequency response (shifted by 5 dB for graphical clarity). Looking at the raw BAFs alone, they appear to be quite variable, but there are a few consistent trends. First, there is a global decrease after about 5 kHz. There are also some grouping of peaks around 750 Hz and 3.5 kHz, and some grouping of dips around 7 kHz. This is consistent with the shape of the average BAF.

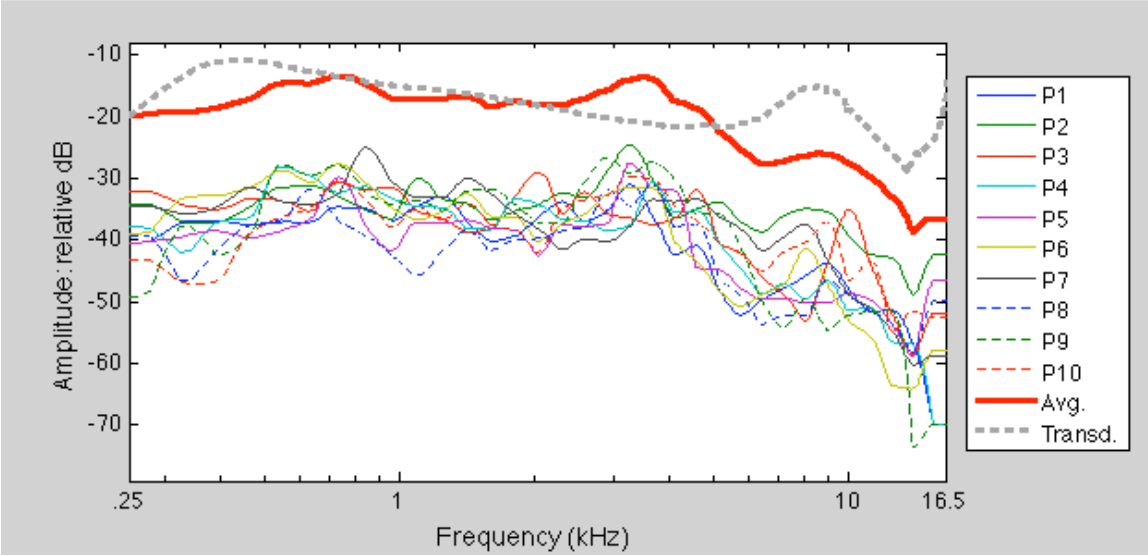

Figure 7. Bone-conduction adjustment functions (BAFs) averaged across ears, for each participant. Also shown is the BAF averaged across all participants (indicated by the "Avg." series), and the transducer frequency response ("Transd." series). For graphical clarity, the average BAF was shifted upwards by 15 dB and the transducer frequency response was shifted upwards by 5 dB.

The equal-loudness adjustments were done with an un-equalized bone-conduction transducer, and thus include the frequency response of the transducer. Figure 8 shows the individual BAFs and average BAF (again shifted upwards by 15 dB for graphical clarity) with the BCT frequency response removed from the BAF. The transducer frequency response was removed by subtracting the BCT frequency response averaged across transducers from the BAF averaged across ears. Figure 8 shows that with the transducer response removed, there are global peaks at about 750 Hz and 3 kHz as well as many local peaks and dips between 500 Hz and 4kHz, and then some again after 10 kHz. There is little variation in the average BAF after about 6.5 kHz, and the individual BAFs always have the strongest relative shape changes below 7 kHz. It should be kept in mind that at the highest two frequencies, the low amplitude values in some BAFs are caused by the participant not hearing the initial tone and therefore making the lowest amplitude value judgment (as described in Section 2.1.4).

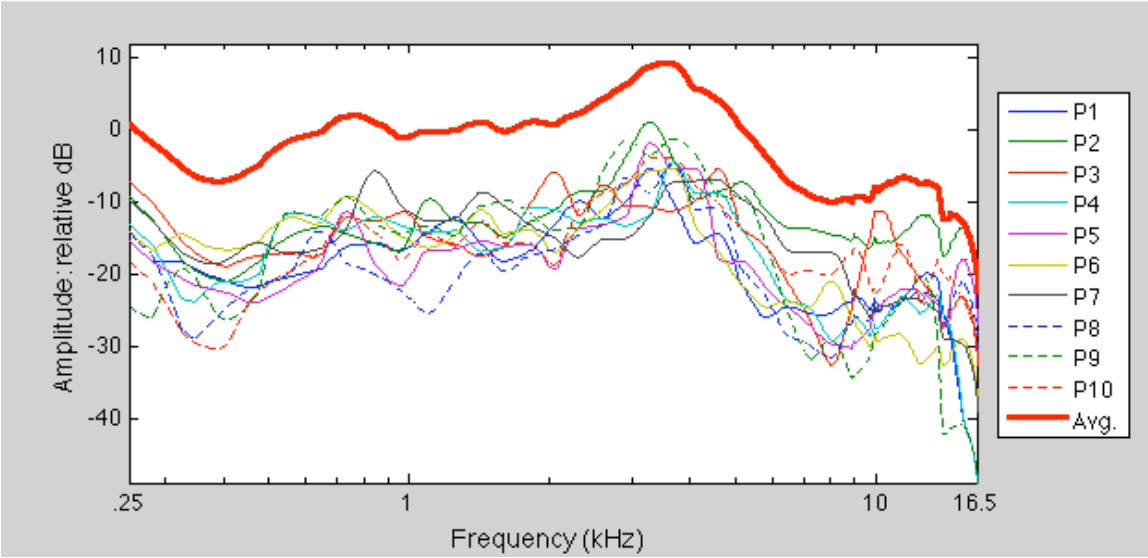

Figure 8. Bone-conduction adjustment functions (BAFs) averaged across ears for all participants, with BCT frequency response (averaged across transducers) removed. The "Avg." series depicts the BAF averaged across all participants, and was linearly shifted by 15 dB to be above individual BAFs for graphical clarity.

During the equal-loudness procedure that was used to form the BAF, participants indicated that they would sometimes hear one of the tones in the middle or opposite side of the head as the other tone. This was true despite verifications showing that the transducer on the same side of the head were active, and originated from the bone-conduction tone.

CHAPTER 3

STUDY 2 - LOCALIZATION EXPERIMENT

3.1. Method

3.1.1. Participants

The same participants used in Study 1 were used in Study 2.

3.1.2. Apparatus

The same audio delivery equipment (computer, transducers, sound cards, etc) and software platform (MATLAB) used in Study 1 were used in Study 2. In addition, the voltage check and left/right sound check conducted at the beginning of Study 1 was also conducted at the beginning of every session for Study 2. For the last four participants that participated in these studies, pictures of the ear were also taken before each session began in Study 2, to monitor headphone insertion depth.

Responses in the virtual audio localization procedure were made by indicating the position of a sound source with a mouse, by clicking on a location relative to a set of two-dimensional models of a head on a computer screen. To allow a measurement of localization in both planes, there were models on the screen for each plane of localization – a "top view" and "side view" of a model head. The response for each plane was split up into two discrete judgments. The first judgment was the azimuth judgment (labeled "Top View" for the participant), as seen in Figure 9. The azimuth judgment screen displayed for 500 milliseconds before the initial stimulus presentation occurred. The participants made the azimuth judgment by clicking on the response screen, and pressed the space bar to submit the response. After the first judgment was submitted, the elevation judgment screen (see Figure 10) displayed for 500 milliseconds, and the same stimulus was played

again. The elevation response was again submitted with the press of the space bar, which marked the end of the trial. For the elevation judgment, the participants were instructed to make the judgment as if they had turned their head to face the sound – this was required for the representation to make sense for sounds not delivered in the median plane. The azimuth task was done first so that the participant could focus solely on elevation when facing the target for the elevation judgment, and not get confused about the front/back location. Participants did occasionally report that there was some cognitive interference between the angle of the vector in the first and second judgment.

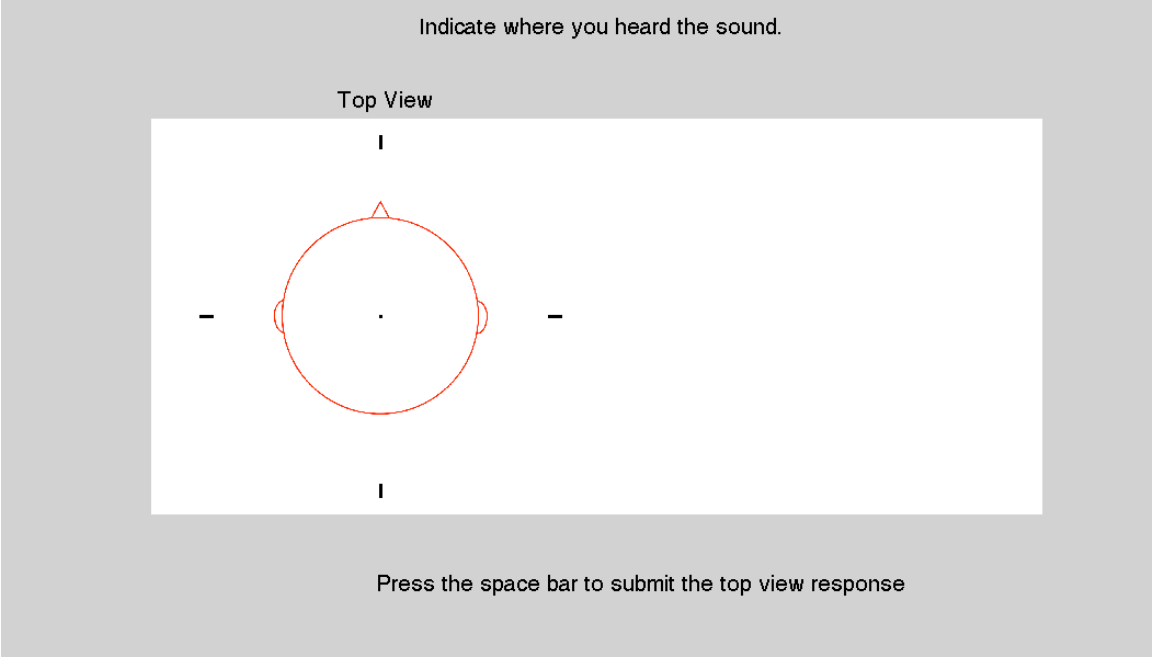

Figure 9. Audio localization response screen. First response - azimuth judgment.

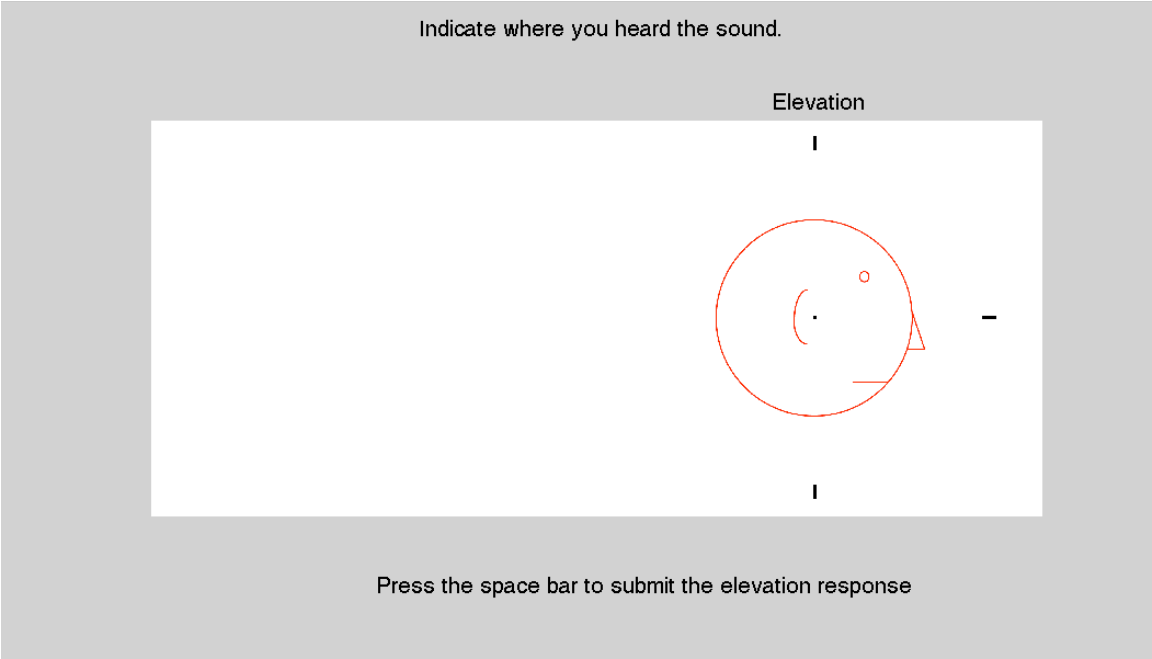

Figure 10. Audio localization response screen. Second response - elevation judgment. For this judgment, participants were instructed to make the judgment as if they had turned their head to face the sound.

At the end of each run, a screen appeared in order to assess the participant's overall impression of the sounds, the goal being to measure general aspects not captured in the response screen that was administered on every trial. This screen asked for ratings regarding two subjective sound quality aspects of the sounds just heard: externalization and diffuseness (see Figure 11**)**.

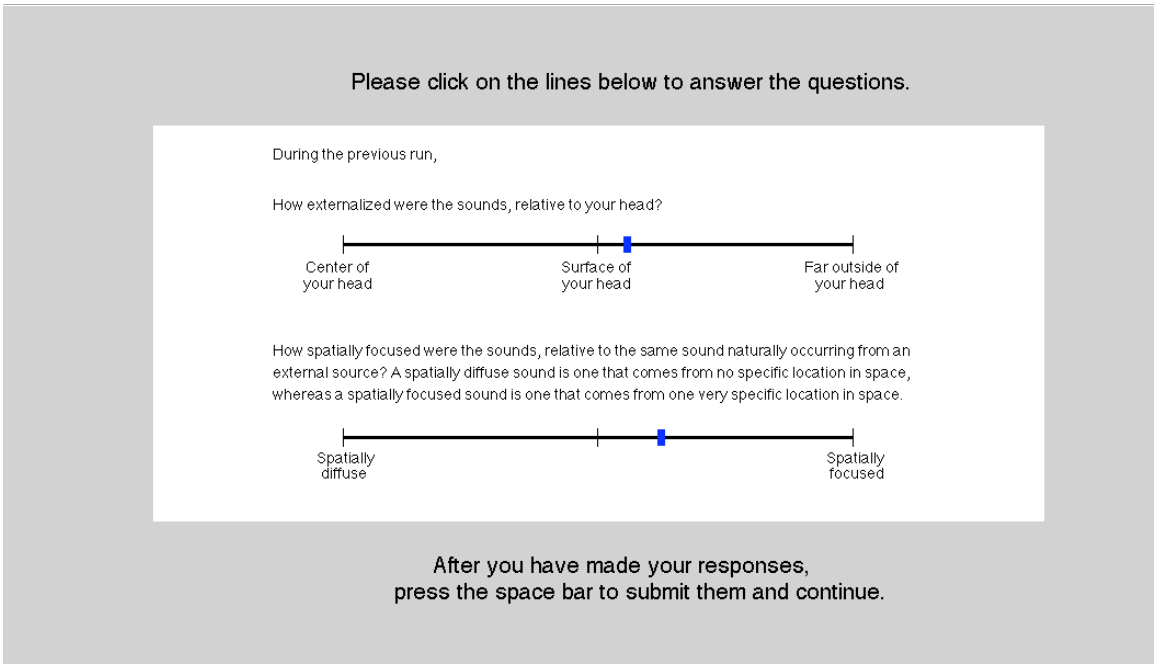

Figure 11. Subjective response screen. Requests rating of externalization and diffuseness at end of every run. Blue markers are sample judgments, and did not appear on the screen until the participant clicked on the scale.

3.1.3. Stimuli

Virtual Audio Localization Input Stimuli

The stimuli consisted of three 250-millisecond Gaussian noise bursts with 10-millisecond raised-cosine onset and offset ramps, separated by 300 milliseconds of silence. Pilot testing determined that having more than one pulse assisted in the localization judgments. Although all stimuli were broadband noise bursts, the specific magnitude spectrum of the noise varied randomly from trial to trial, to prevent listeners from becoming familiar with spectral characteristics of any given stimulus (as is traditionally done in virtual localization research - see Langendijk & Bronkhorst, 2002; Middlebrooks, 1999b; Wightman & Kistler, 1989b). Air-conducted stimuli were amplified at the same level as the air-conducted stimuli in Study 1. Based on the shape of

the diffuse field response and a peak of 75 dB at 2500 Hz, the broadband level is estimated to be 65 dB SPL. The bone-conducted stimuli were matched to this level via the bone-conduction adjustment function.

Filtering and Equalization

Table 3 shows the filters applied to test different conditions tested in Study 2. Table 3 shows that the spatial audio filters (the DTFs) were applied in every condition. The headphones used for the air condition had a built-in diffuse-field response, the adjusted bone condition had no transducer correction, and the unadjusted bone condition had a filter representing the inverse of the BCT's frequency response applied. The BAF was only applied in the adjusted bone condition. Table 3 also shows that all conditions included a 250 Hz to 16.5 kHz band-pass filter. This avoided having the very low and high frequencies in the stimuli, where filtering artifacts can occur (Wightman & Kistler, 1989b). It also ensured that frequencies outside the range of the BCT would not be used (see Figure 2). Removing spectral content above 16.5 kHz should not effect localization because there are not any important spectral cues for localization above 16 kHz. Specifically, up-down spectral cues are primarily located in the 6-12 kHz band, and front/back cues in the 8-16 kHz band (Langendijk & Bronkhorst, 2002).

Table 3

Filtering applied for each condition in this study.

Condition	Directional Transfer Function (DTF)	Transducer Correction	Bone-conduction Adjustment Filter	Bandpass Filter
Air (baseline)	√	Built-in Diffuse	--	√
Adjusted Bone	√	None	√	√
Unadjusted Bone	√	BCT Inverse Frequency Response	--	√

The spatial audio filters applied for this experiment followed the previously described methods of Middlebrooks (Middlebrooks, 1999a, 1999b; Middlebrooks, et al., 2000), using only the components specific to sound-source location (DTFs), without the spectral content common to all locations. The component common to all locations is calculated by averaging the sound pressure across all locations (via root-mean-square), for each frequency in the digital filter (Middlebrooks, 1999a). The DTF is then computed by dividing the HRTF for each location by the common component. The common component removed by this process is added back in to the signal by use of the diffuse-field response of the headphones. Thus, DTFs provide a solution that avoids having to measure individualized headphone-to-eardrum transfer functions (Middlebrooks, 1999b). The "medium" (S44) base DTFs used in Middlebrooks (2000) were obtained from the author and were used in this study. These DTFs were frequency-scaled for an individual

participant via the anthropometric customization procedure described in Section 2.1.5.

Similar to Wightman (1989b), 36 virtual audio sound source locations were chosen from the full set of DTF locations. The 36 locations were chosen so that the full range of azimuths and elevations would be represented. These locations were drawn from a set that was equally distributed (in spherical space) across 360° of azimuth and -70 to +90° in elevation. The locations were chosen in a proportional manner, such that the proportion of locations in each location category in the sample drawn equaled the proportion in the population of locations available in the full set of DTF locations. The location categories were the conjunction of low, middle, and high elevations with front, back, and side azimuths, as used in Wightman (1989b). The side locations were evenly split between left and right. The locations were randomly drawn once, and all participants experienced the same locations in all sessions. The locations used in this study can be seen in Table 4; a visualization of the locations can be seen in Figure 12.

Table 4

Locations used for virtual audio localization study, and their spatial classifications.

Azimuth (°)	Elevation (°)	Classification
-21	-20	front-mid
-30	-50	front-low
-40	-60	front-low
-40	10	front-mid
42	-20	front-mid
-34	30	front-mid
-22	30	front-mid
0	0	front-mid
13	40	front-high
-20	60	front-high
-160	-41	back-low
160	-41	back-low
-135	-11	back-mid
-140	9	back-mid
-180	-11	back-mid
-150	19	back-mid
170	19	back-mid
-138	39	back-high
165	69	back-high
-135	49	back-high
-90	-40	side-low (L)
-103	-21	side-mid (L)
-74	20	side-mid (L)
-110	9	side-mid (L)
-70	10	side-mid (L)
-53	-20	side-mid (L)
-90	69	side-high (L)
-124	49	side-high (L)
120	-50	side-low (R)
64	-20	side-mid (R)
116	-1	side-mid (R)
80	10	side-mid (R)
50	10	side-mid (R)
74	-20	side-mid (R)
120	69	side-high (R)
101	49	side-high (R)

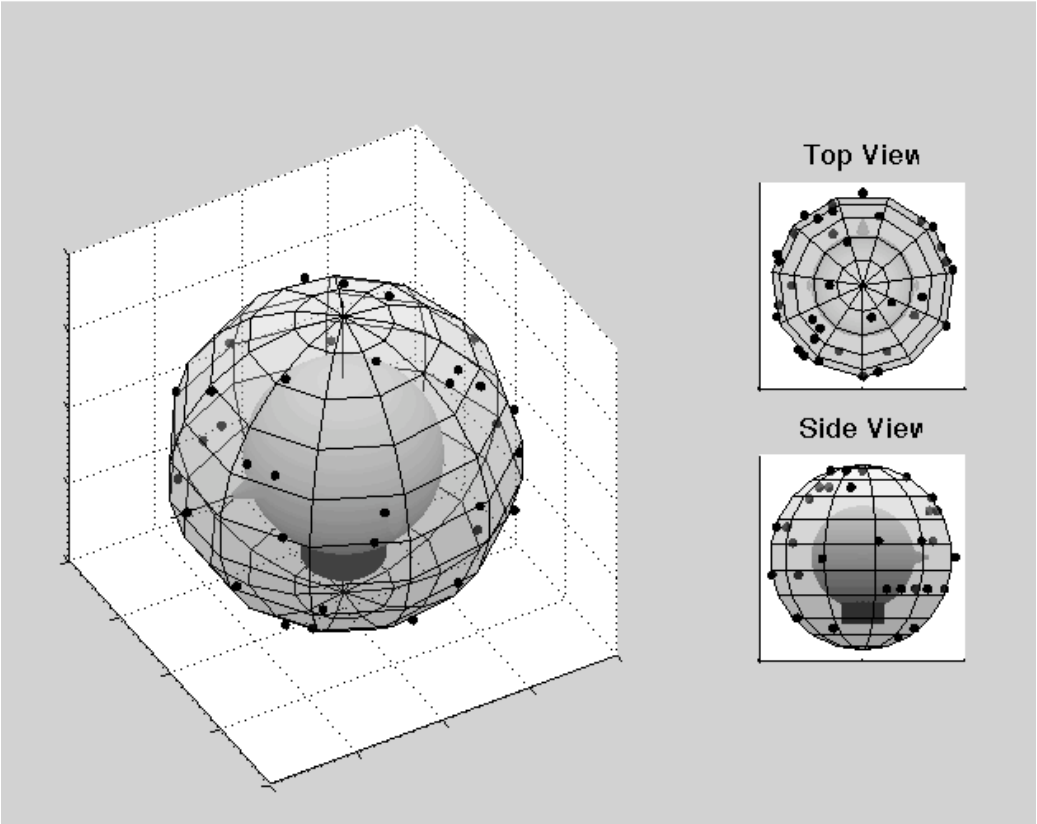

Figure 12. Visualization of locations used for virtual audio localization study.

The filters shown in Table 3 were applied using a system of cascading filters applied via frequency-domain multiplication. A simplified schematic can be seen in Figure 13, and details of the signal processing can be found in Appendix B. First, 500 milliseconds of Gaussian noise was generated. This noise was then filtered by the frequency-scaled DTFs, using frequency-domain multiplication. This output was then normalized so that it had a maximum value of one. What happened to the signal next depended on the condition of the stimuli being generated.

For the adjusted bone condition, each channel of the DTF filtering output was filtered by a linear-phase BAF. The BAF was also windowed in the time domain (with a Hamming window). Filtering was again achieved by multiplication in the frequency

domain. The ends of the resulting time-domain signal were then trimmed to remove the remnants of the filter.

For the unadjusted bone condition, the DTF filtering output was filtered by the inverse of the BCT frequency response measured in the Sonification Lab (see Figure 3). The frequency response was measured separately for each transducer (left and right), and applied to the appropriate channel in the filtering process. The frequency response was converted to a minimum phase version and windowed in the time domain with the last half of a Hanning window before the inverse was computed. Filtering was again conducted by multiplication in the frequency domain. The filtered signal was truncated to the length of the original input signal. After this was done, the amplitude of the time-domain filtered signal was multiplied by the adjustment value determined in the unadjusted bone equal-loudness trials.

The output from the DTF filtering (for the air condition), BAF filtering (for the adjusted bone condition), or inverse transducer response (for the unadjusted bone condition) was then bandpass filtered with a 4-pole butterworth filter. After this filtering, the gain that was caused by the bandpass filtering was removed.

After the bandpass filtering was complete, the final time-domain preparations were done. First, the signal was truncated to 250 milliseconds. Then a raised-cosine window was applied to the first and last 10 milliseconds of the signal. Once the stimulus burst was in its final form, it was appended with 300 milliseconds of silence three times to produce the pulse train, which was written to a wave file (Fs = 44100, N = 16) that was used for stimuli presentation.

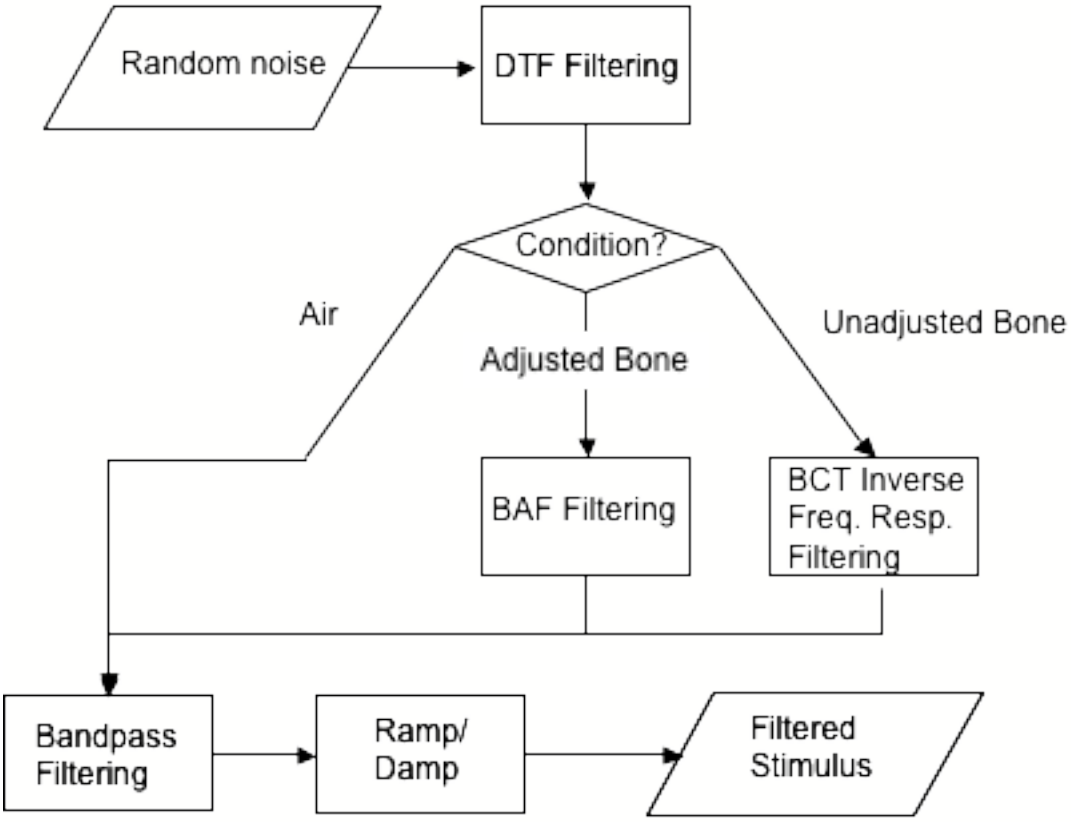

Figure 13. Simplified schematic of digital signal processing to create stimuli for virtual audio localization.

3.1.4. Procedure

Each run in this study involved judgments on the same 36 audio target locations, each presented once, in random order. There was first one practice session consisting of nine runs, three per condition. All of the practice trials were excluded from further analyses. Practice trials are important because previous research has shown learning effects for localization (e.g., Møller, Hammershøi, Jensen, & Sørensen, 1999). At the end of each run, the sound quality response screen appeared.

The experiment consisted of ten runs for each condition, for a total of 30 runs (1080 trials). These runs were equally distributed across three days of testing, each session separated by no less than one night and no more than three nights. Each session

included ten runs (three runs for two conditions, and four runs for the remaining condition) in randomized order. Each session lasted about 75 minutes long. Changing of conditions (and selection of the transducer that the signal was coming through) was automated through the computer program. Participants wore both the bone-conduction transducers and the headphones throughout the experiment.

3.1.5. Study 2b: Replication With Individualized HRTFs

Individualized HRTFs were measured for participant 6 at the Auditory Localization Facility in the Air Force Research Laboratory at Wright Patterson Air Force Base. Stimuli were delivered inside of a geodesic sphere in which the listener stood during testing. The geodesic sphere had 277 vertices equally spaced in the spherical space, where Bose speakers were mounted. The sphere was located inside of an anechoic chamber. During HRTF measurement, the listener wore an Intersense IS-900 ultrasonic headtracker, which was used to correct for any head movement during the measurement procedure. Further details on the facilities can be found in Simpson, Brungart, Gilkey, Iyer, and Hamil (2007). The blocked-ear recording technique was administered by embedding Knowles FG 3329 microphones into Westone Oto-dam ear dams. The listener stood in the sphere as a periodic chirp train played across arrays of speakers.

The impulse responses were created by dividing (in the frequency domain) the recorded signal by the presented signal. The 2048-point impulse responses were then treated to a 401-point Hanning window and padded with 400 zeros to a final length of 2448 points. After these processed HRTFs were obtained, some further processing occurred to obtain the DTFs, as described in Middlebrooks (1999a). Before the DTFs were computed, a brickwall filter had to be applied to the HRTF to remove frequencies

below 100 Hz and above 15 kHz. Without this filter, there was a very strong boost above 15 kHz in the DTFs, presumably because there was no content above 15 kHz in the common component. This boost caused the DTF impulse responses to have ringing. The amplitude component common to all locations was found by computing the root-mean-squared magnitude spectrum across all locations. The phase component common to all locations was found by computing the Hilbert transform of the logarithm of the amplitude component. These were then combined to form the complex common component in the frequency domain. Each HRTF was then divided by the common component to form the frequency-domain DTF, and then transformed into the time domain to the DTF impulse response.

The locations measured at Wright Patterson Air Force Base did not exactly match the 400 original locations measured for the DTFs measured by Middlebrooks (2000) that were used in the main set of studies. To get a set of locations from the Wright Patterson HRTFs that were similar to the subset of locations used in the main set of studies using DTFs from Middlebrooks (2000), the arc length (on the great circle) between each of the locations used in the main study and all the locations measured at Wright Patterson Air Force Base was computed. Then the locations that were the closest to each location used in the main study were used for the individualized study. A list of locations can be found in Table 5.

Table 5

Locations used for main study (with generalized HRTFs), individualized HRTF replication, and their location classifications.

Generalized		Individualized		Classification
Azimuth (°)	Elevation (°)	Azimuth (°)	Elevation (°)	
-21	-20	-24.4	-17.8	front-mid
-30	-50	-24.7	-50.0	front-low
-40	-60	-45.1	-54.3	front-low
-40	10	-42.5	14.8	front-mid
42	-20	46.1	-15.5	front-mid
-34	30	-32.5	25.2	front-mid
-22	30	-20.8	32.1	front-mid
0	0	3.2	-1.0	front-mid
13	40	19.1	36.2	front-high
-20	60	-25.4	57.2	front-high
-160	-41	-159.6	-34.2	back-low
160	-41	167.6	-36.7	back-low
-135	-11	-136.1	-16.6	back-mid
-140	9	-136.6	8.5	back-mid
-180	-11	-179.7	-16.7	back-mid
-150	19	-154.4	15.9	back-mid
170	19	170.6	21.1	back-mid
-138	39	-134.1	33.4	back-high
165	69	144.3	71.9	back-high
-135	49	-140.1	53.4	back-high
-90	-40	-91.6	-36.6	side-low (L)
-103	-21	-101.7	-27.0	side-mid (L)
-74	20	-73.6	26.6	side-mid (L)
-110	9	-111.0	8.9	side-mid (L)
-70	10	-69.4	7.4	side-mid (L)
-53	-20	-52.1	-19.0	side-mid (L)
-90	69	-98.5	67.4	side-high (L)
-124	49	-128.5	44.5	side-high (L)
120	-50	110.0	-50.1	side-low (R)
64	-20	59.5	-15.0	side-mid (R)
116	-1	119.0	-0.7	side-mid (R)
80	10	74.9	8.1	side-mid (R)
50	10	48.7	9.2	side-mid (R)
74	-20	72.4	-13.5	side-mid (R)
120	69	105.3	71.5	side-high (R)
101	49	102.9	47.0	side-high (R)

The subsequent filtering algorithms were identical to those with the frequency-scaled DTFs, except the individualized DTFs were used rather than the non-individualized frequency scaled DTFs. Study 1 and study 2 were then replicated to generate the data.

3.2. Results

3.2.1. Overview

The primary analysis of data for the present studies will be the review of patterns in each participant or group of participants. Conventional inferential statistics (i.e., ANOVA) will also be performed on aggregate data. It is important to note that these statistics must be interpreted with caution, because of the small number of participants and high number of trials in this study (as is typical in psychophysics). The pooling of data across trials and small number of participants results in a test with low statistical power that can only detect quite large effects. Considered within the signal-detection response framework, inferential statistics still provide important information about effects present in our sample that occur in the population ("hits"), but may have a high "miss" rate of detecting condition effects.

The repeated-measures analysis of variance (ANOVA) has three key assumptions: independence of scores, sphericity, and normally distributed populations (Keppel, 1991). The randomization of trials within a condition and randomization of condition order in the present studies ensured that the scores were independent. The sphericity assumption includes the homogeneity of variance assumption that is shared with between-subjects designs, but also includes an assumption about the correlations between pairs of

conditions (Howell, 2002). To address this assumption, an inferential evaluation of sphericity will be conducted in all the ANOVAs presented here.

The normal population distribution assumption is a more complicated issue in the present studies. If the population is not normal, but is symmetrical and the sample size is greater than 12, then there are no meaningful consequences of violating this assumption (Clinch & Keselman, 1982; Tan, 1982 as cited in Keppel, 1991). In the present studies, however, the sample size is only seven, which means it is not protected from normality violations. Furthermore, assessing whether a population is normal based on a histogram of seven points is a futile exercise. A consideration of the central limit theorem (CLT) is relevant in a discussion of the shape of an underlying population distribution.

The CLT states that the distribution of a population of means approaches normal as *n*, the sample size, increases (Howell, 2002). In the most direct interpretation, the population of means is the population underlying each summary statistic (e.g., Pearson's *r*, average azimuth error), and *n* is the number of empirical data points attained for each condition (one for each person) in this study. This would suggest that with an *n* of only seven, the central limit theorem does not help the situation at hand. However, the CLT can also be conceptualized in terms of the raw data used to compute the summary statistics and the physiological system underlying the localization judgments. Specifically, the population underlying the summary statistics can be conceptualized as the azimuth and elevation data (which are the basis for the summary statistics). Furthermore, the sample underlying the azimuth and elevation judgments can be thought of as the neurons and neural systems that determine the behavioral response. In this case, the size of the sample is quite large and the underlying population certainly approaches a

normal distribution; therefore the ANOVA assumption of normality is certainly met. Nevertheless, the shape of the summary statistic population is not known, and only seven aggregate data points makes assessing this assumption difficult. Thus, the unknown nature of the population distribution's shape will remain as a caveat for the ANOVAs conducted for the present studies. As the statistics are reviewed, those that are unlikely to have a normal underlying population will be pointed out.

The sequence of data presented below will be as follows: First the raw data plots of stimulus-response data will be reviewed, for the azimuth and elevation planes, for each condition. A regression line will be fit to the elevation raw data sets, and characteristics of the regression results will be described for each participant. Then the front/back reversals, up/down reversals, and coordinate system artifacts will be resolved, and the raw data points plus regression line will again be presented. Next, the rate of front/back and up/down reversals will be considered across conditions and participants. Then trends in summary localization performance statistics, with the reversals resolved, will be reviewed. The summary localization statistics will include measures of error, bias, and variability. Participants' subjective ratings given at the end of each block will be subsequently described. Then the data collected with the individualized HRTFs on participant 6 will be reviewed, in comparison with the data collected with generalized HRTFs on participant 6.

3.2.2. Raw Data and Stimulus-Response Trends

Raw Scatter Plots and Line Fitting

All responses were computed in terms of the conventionally used single vertical pole coordinate system (e.g., Wightman & Kistler, 1989b). In this coordinate system (see Figure 14), azimuth gives the front-to-back and left-to-right position of the sound source. Negative numbers indicate a stimulus on the left and positive numbers indicate a stimulus on the right, with 0 degrees being direction in front of the listener, -90 degrees being directly to the left of the listener, +90 degrees being directly to the right of the listener, and +/-180 degrees being directly behind the listener. For elevation, 0 degrees is straight ahead of the listener, -90 degrees is directly below the listener , and +90 degrees is directly above the listener.

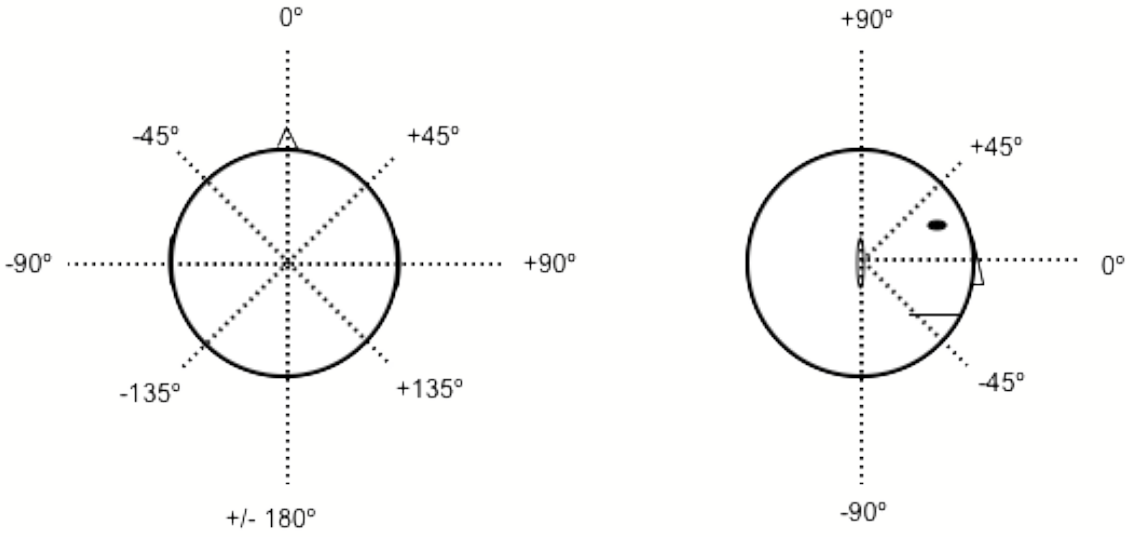

Figure 14. Single vertical pole system used in this set of studies, with azimuth and elevation.

Scatter plots of the audio localization data for each participant and condition, for azimuth and elevation, can be seen in Figure 17 through Figure 26. An orientation to the

scatter plot and important data patterns can be found in Figure 15 and Figure 16. For both azimuth and elevation plots, the target position is on the x-axis and the judgment position on the y-axis. Thus, perfect performance would result in data points falling on the diagonal. For azimuth, there are three additional patterns of data on the scatter plot: front/back reversals, two-cluster response pattern, and wrap-around artifacts (see Figure 15 and Figure 16). Figure 15 shows the two areas where front/back reversals occur, which are perpendicular to the diagonal. Front/back reversals occur when a sound is presented in the front hemi-field and perceived in the back, or vice versa. The two-cluster response pattern can also be seen – this typically occurs when listeners respond with many different azimuths at varied azimuths around their head, but are only correctly judging whether the stimulus is on the left or right, and never hear the sound in front. A variant of the two-cluster response pattern is the binary response pattern, in which there is no judgment variability, and listeners simply respond that the sound is on the left or right. Figure 15 also shows the two regions for "wrap-around" artifacts, which occur when a stimulus would be closer to the diagonal if it was expressed in terms of the other (left/right) hemisphere's coordinate system. These will be subsequently discussed in more detail. For elevation, there is one important pattern on the scatter plot besides perfect performance, and that is the region perpendicular to the diagonal where up/down reversals occur (see Figure 16).

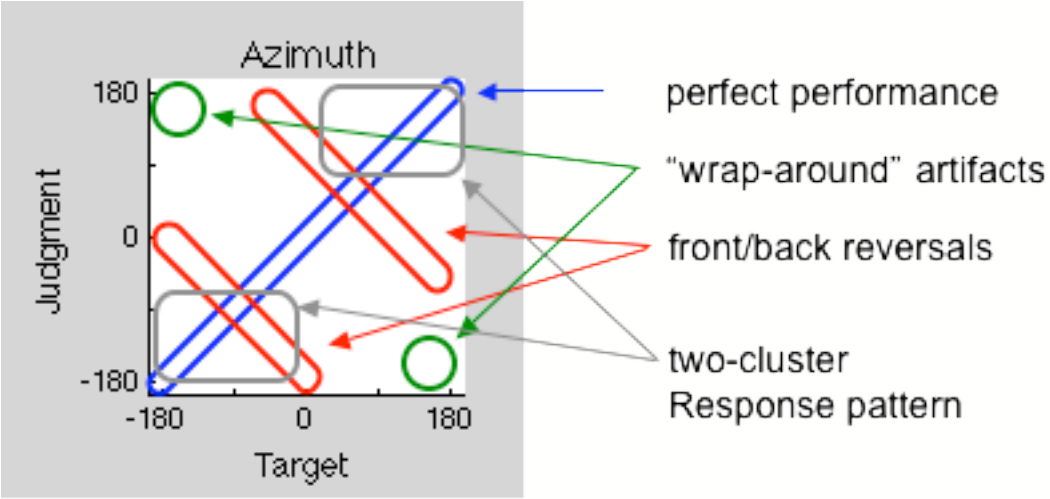

Figure 15. Azimuth scatter plot with orientation to important data patterns: perfect performance, "wrap-around" artifacts, two-cluster response pattern, and front/back reversals.

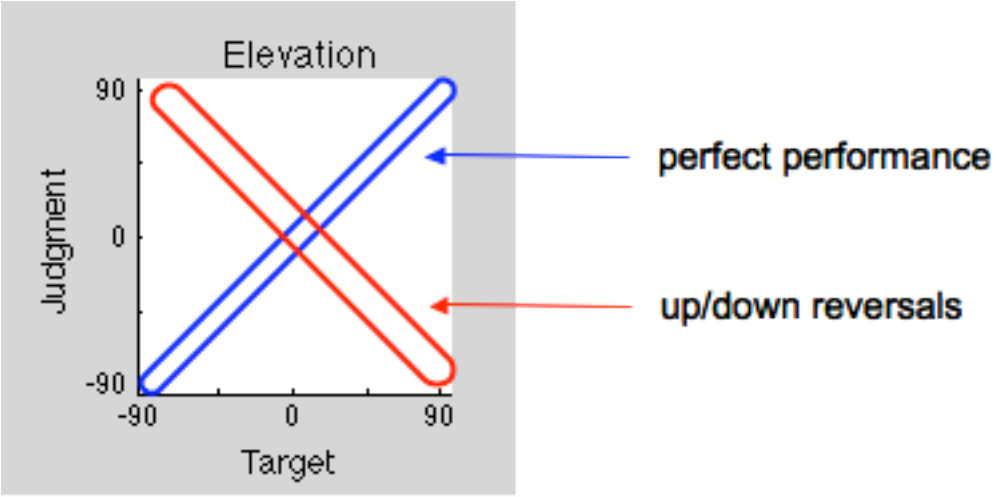

Figure 16. Elevation scatter plot with orientation to important data patterns: perfect performance and up/down reversals.

Lines of best-fit were computed for each elevation scatter plot. Simple linear regression was performed on the raw data for elevation to determine the line of best fit. Simple linear regression takes a set of two variables (in the present case stimuli and

response), and produces an equation that describes the best-fitting line, by minimizing the sum of squared errors between the actual and predicted y values (Howell, 2002).

The line can be defined by the simple following equation:

$Y = bX + a$

Where

Y = the predicted value of Y

b = the slope of the regression line

a = the y-intercept of the line

X = the value of the predictor variable

(Howell, 2002).

The slope of a regression line describes "the rate of change in Y units per X unit" (Cohen & Cohen, 1983, pp. 42-43). With stimulus position as X and stimulus response as Y, a slope of one is expected for perfect localization performance. A reduction in slope from a value of one indicates a decrease in accuracy, and possibly a compression in the response space.

The R^2 statistic indicates "the proportion of the variance of Y linearly associated with x" (Cohen & Cohen, 1983, p. 47). Thus, it serves as a measure of the degree to which the regression line fits the data. In the case of the present data, a lower R^2 value indicates more responses distant from the regression line, and thus a less accurate response.

The raw azimuth data were not suitable for regression analyses because of the unsuitability of the circular azimuth coordinate system to linear (or curvilinear) analysis.

In a subsequent section, an analysis with corrections for these issues will be presented. One problem with the azimuth data is its "wrap-around" artifact (see Figure 15 above). For example, if the stimulus is presented at +178 degrees azimuth, and the response is made at -178 degrees azimuth, this will influence the regression slope and R^2 by a disproportionate amount. Specifically, it will be treated in the regression analysis as an error of 356 degrees, when it is really best described as an error of only 4 degrees. Furthermore, this data point presents itself as an outlier, which heavily influence correlations (Bordens & Abbott, 2005). Another complication with performing a regression analysis on the azimuth data are front-back confusions. They present themselves as outliers, when they are perfectly expected to occur (even in free-field listening), and thus will influence the regression slope and R^2 a disproportionate amount.

The scatter plot of data for azimuth and elevation across conditions will now be described (see Figure 17 through Figure 26). The scatter plots are laid out in a matrix with three rows (one for each condition) and two columns (one for each plane: azimuth and elevation). The black rings indicate the participants' raw responses, the blue diagonal line (darker line when printed in grayscale) indicates perfect performance, and the red line (lighter line when printed in grayscale) indicates the line of best fit by simple regression. The regression equation and R^2 value for the elevation data can be found to the left of each elevation graph. As the set of scatter plots is described for each participant, the fit of the data to the diagonal and for the elevation the slope of the line that the data fall on will be described. Terms such as low, moderate, and high will be used for analysis of the elevation regression data. For these analyses, a low slope is less than .3, a moderate slope is between .3 and .6, and a high or steep slope is greater than .7.

For the elevation fit (as indexed by R^2), a low fit is less than .2, a moderate fit is between .2 and .3, and a high fit is greater than .3 (r >.55). These values were determined based on the range of data present in the current sample.

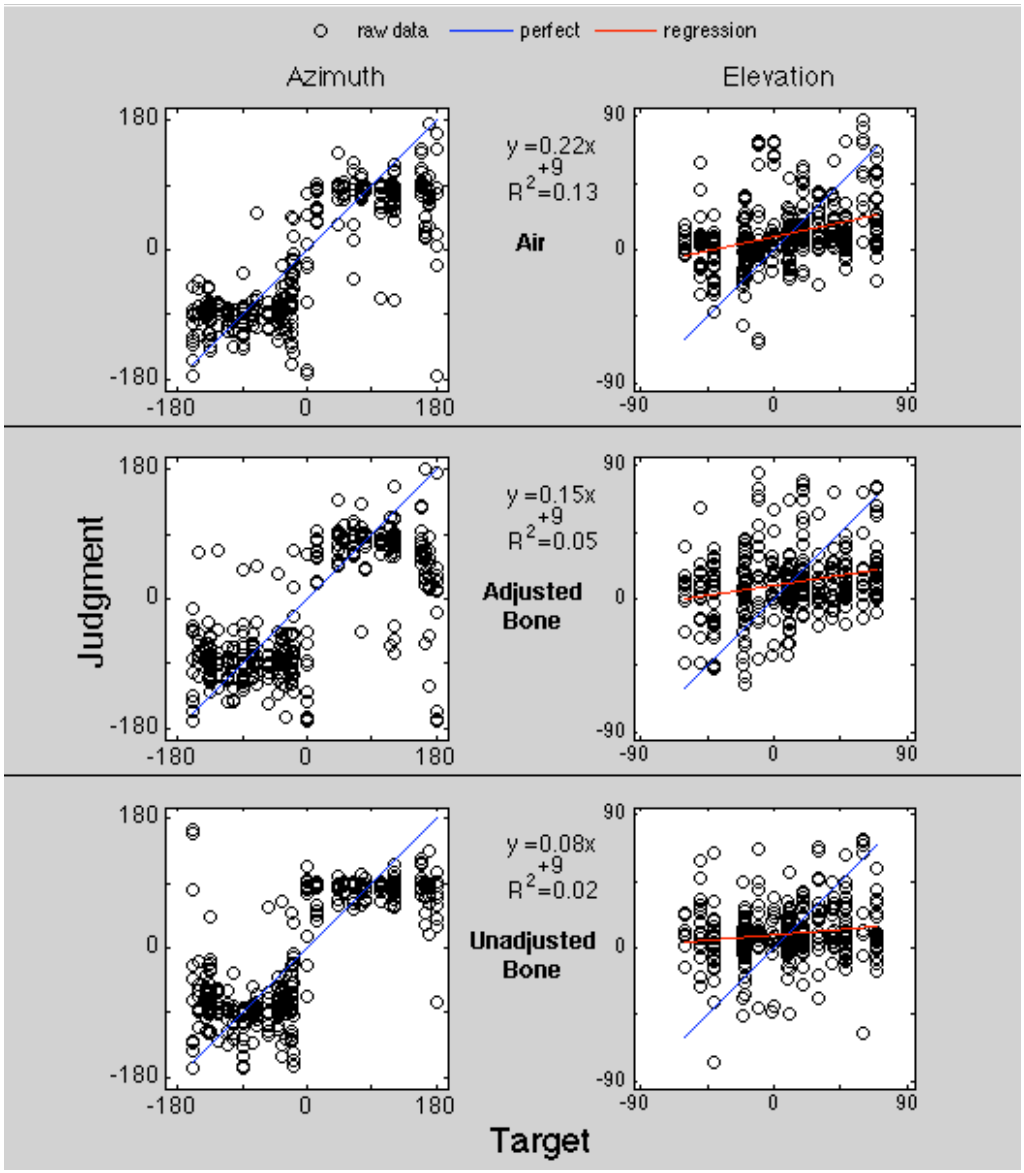

Figure 17. Scatter plots for participant 1. For azimuth, the air condition baseline shows many points that fall close to or on the diagonal, with a large amount of front/back reversals on the cross-diagonal. For the adjusted bone azimuth, data points begin to fall off the diagonal more, with less stimuli perceived in front (the center of the plot), beginning to take on a "two-cluster" response pattern. For the unadjusted bone azimuth, the responses begin to form into two clusters, but still have some stimuli perceived in front. For elevation, the air condition baseline data fall on a moderately sloped diagonal, with a moderate amount of spread from that diagonal. This is echoed in the shallow slope of the regression line (.22) and low fit of the data to that line (R^2 = .13). When going from the air condition to the adjusted bone then unadjusted bone condition, the data points fall on a progressively shallower sloped line and become more spread out around that line, as indicated by the monotonically decreasing regression slope and R^2 value.

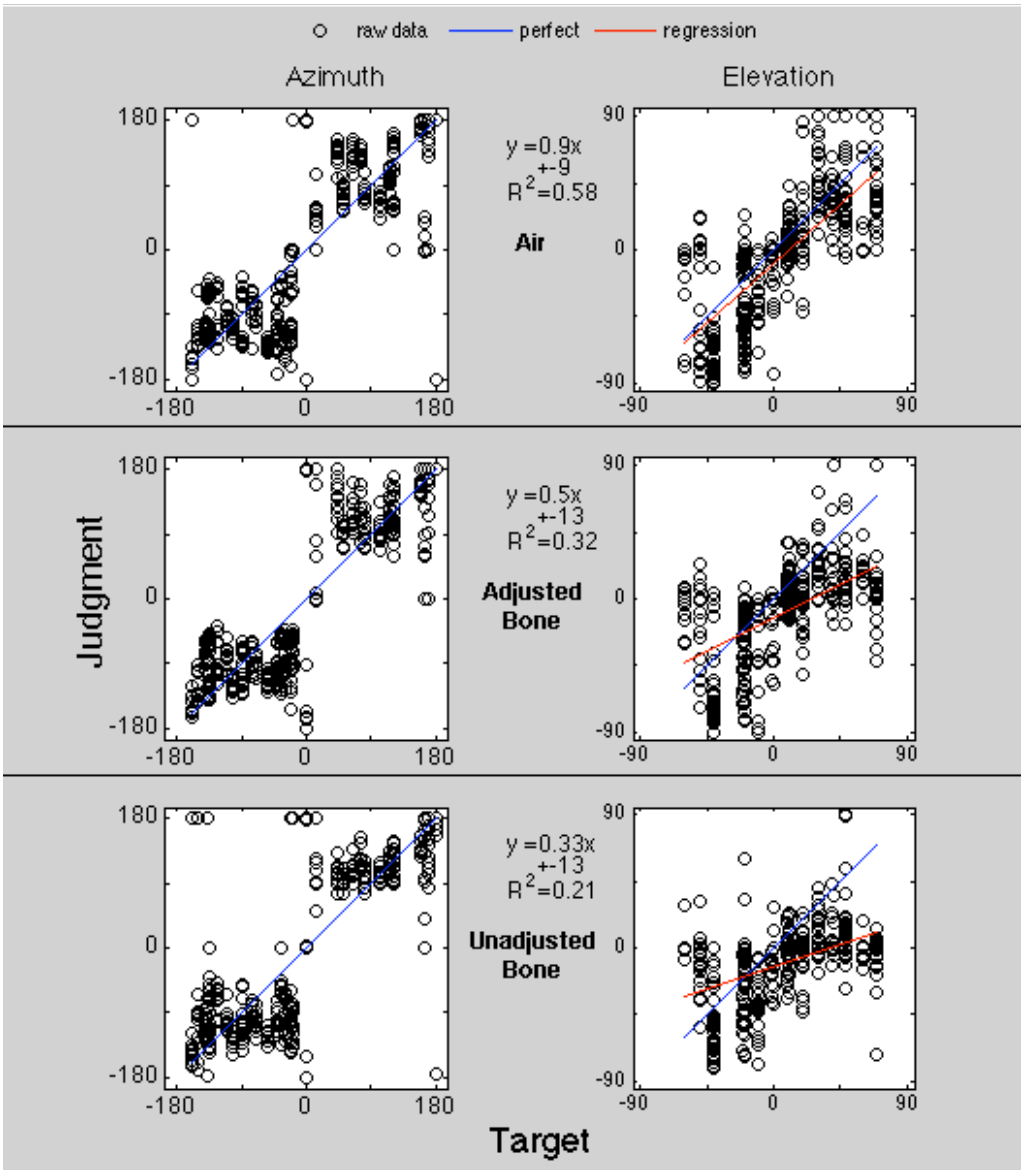

Figure 18. Scatter plots for participant 2. For azimuth, the air condition baseline shows a large number of points that fall on or close to the diagonal, with a large amount of front/back reversals on the cross-diagonal. For the adjusted bone azimuth, data points begin to fall off the diagonal more, with less stimuli perceived in front (the center of the plot), taking on a "two-cluster" response pattern. For the unadjusted bone azimuth, the pattern is relatively similar. For elevation, the air condition baseline data fall on a steeply sloped diagonal, with a low amount of spread from that diagonal. This is echoed in the near-perfect slope of the regression line (.90) and high fit of the data to that line (R^2 = .58). When going from the air condition to the adjusted bone then unadjusted bone condition, the data points fall on a progressively shallower sloped line and become more spread out around that line, as indicated by the monotonically decreasing regression slope and R^2 value.

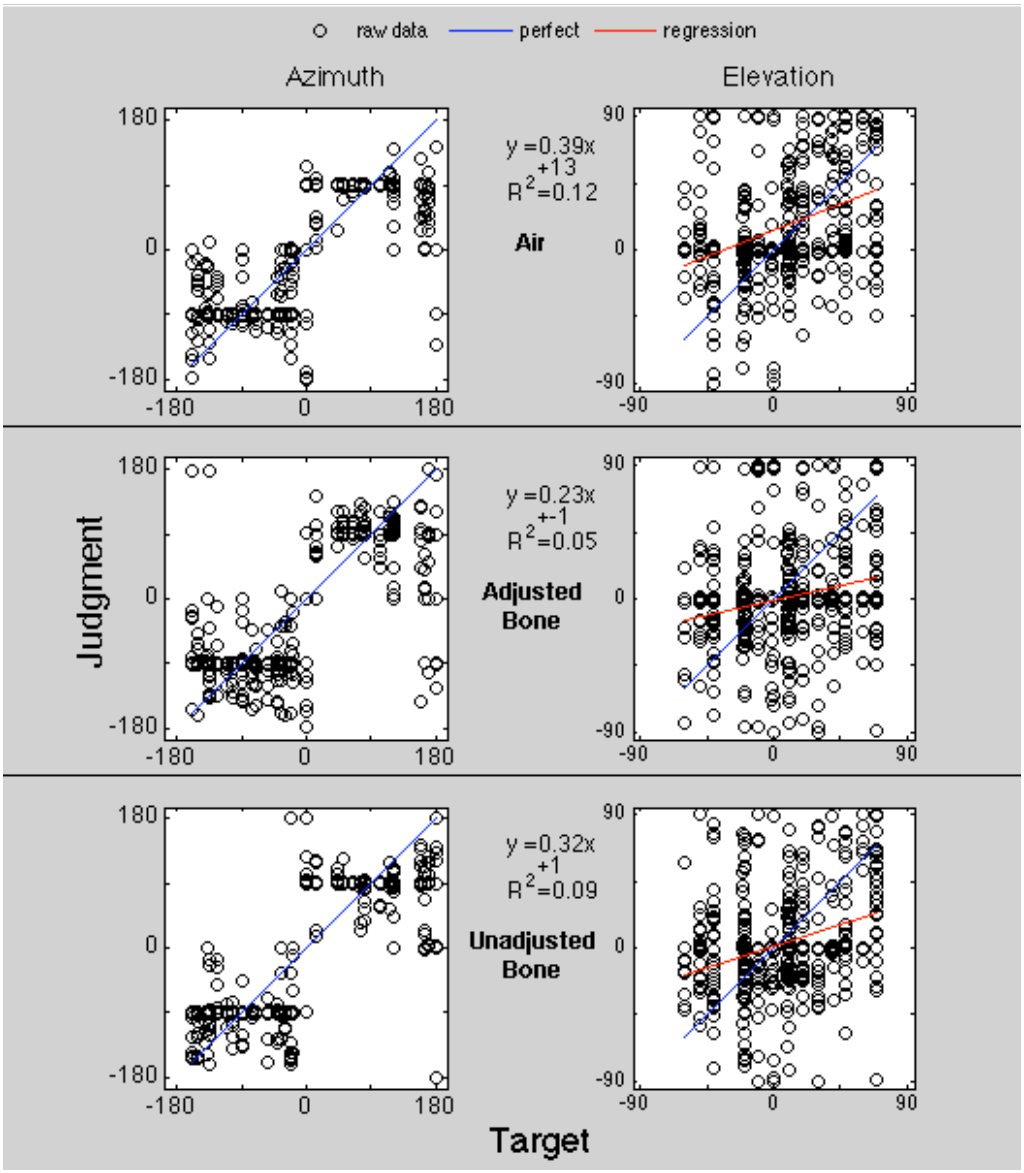

Figure 19. Scatter plots for participant 3. For azimuth, the air condition baseline shows many points that fall on or close to the diagonal, with a slight binary response pattern and some front/back reversals. For the adjusted bone azimuth, data points begin to fall off the diagonal more, with less stimuli perceived in front, and less of a binary response pattern. For the unadjusted bone azimuth, the pattern is relatively similar, but with the response pattern becoming binary again. For elevation, the air condition baseline data are quite variable, but generally fall on the diagonal. This is echoed in the moderate slope of the regression line (.39) and low fit of the data to that line (R^2 = .12). When going from the air condition to the adjusted bone condition, the data points become increasingly variable, and fall on a shallower sloped line, as indicated by the decreased regression slope and R^2 value. However, the unadjusted bone condition then gives an increased regression slope and decreased variability (increased R^2) relative to the adjusted bone condition.

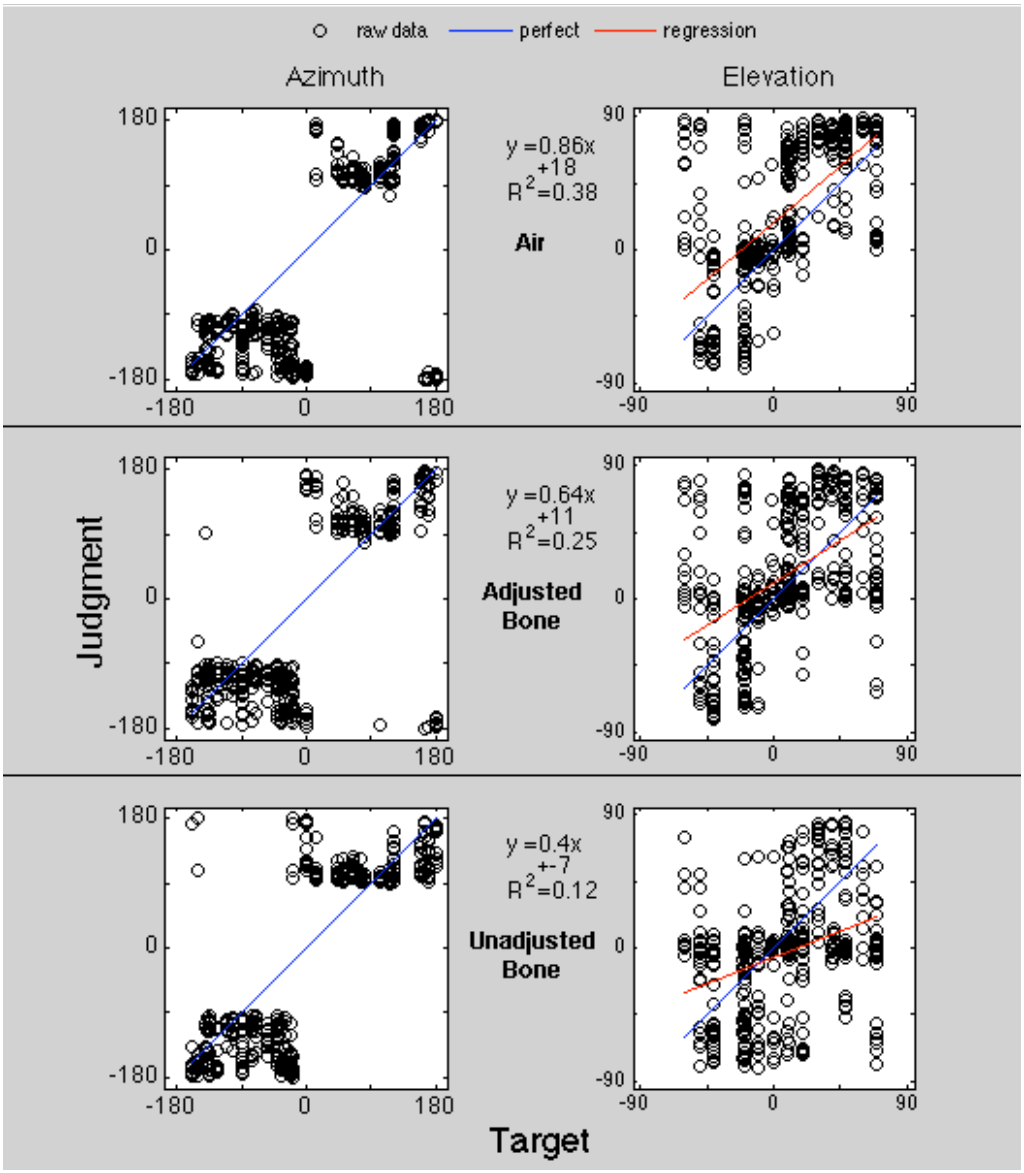

Figure 20. Scatter plots for participant 4. For azimuth, the air condition baseline shows relatively few data points falling on the diagonal, no stimuli perceived in the frontal region, and many front–to-back reversals, forming a "two-cluster" response pattern. For the adjusted bone azimuth, the data look similar. For the unadjusted bone azimuth, the pattern is still similar, but with slightly more of a binary response pattern on the right-hand side. For elevation, the air condition baseline data fall on a steeply sloped diagonal, with a relatively low amount of spread from that diagonal. This is echoed in the high slope of the regression line (.86) and good fit of the data to that line (R^2 = .38). When going from the air condition to the adjusted bone then unadjusted bone condition, the data points fall on a progressively shallower sloped line and become more spread out around that line, as indicated by the monotonically decreasing regression slope and R^2 value.

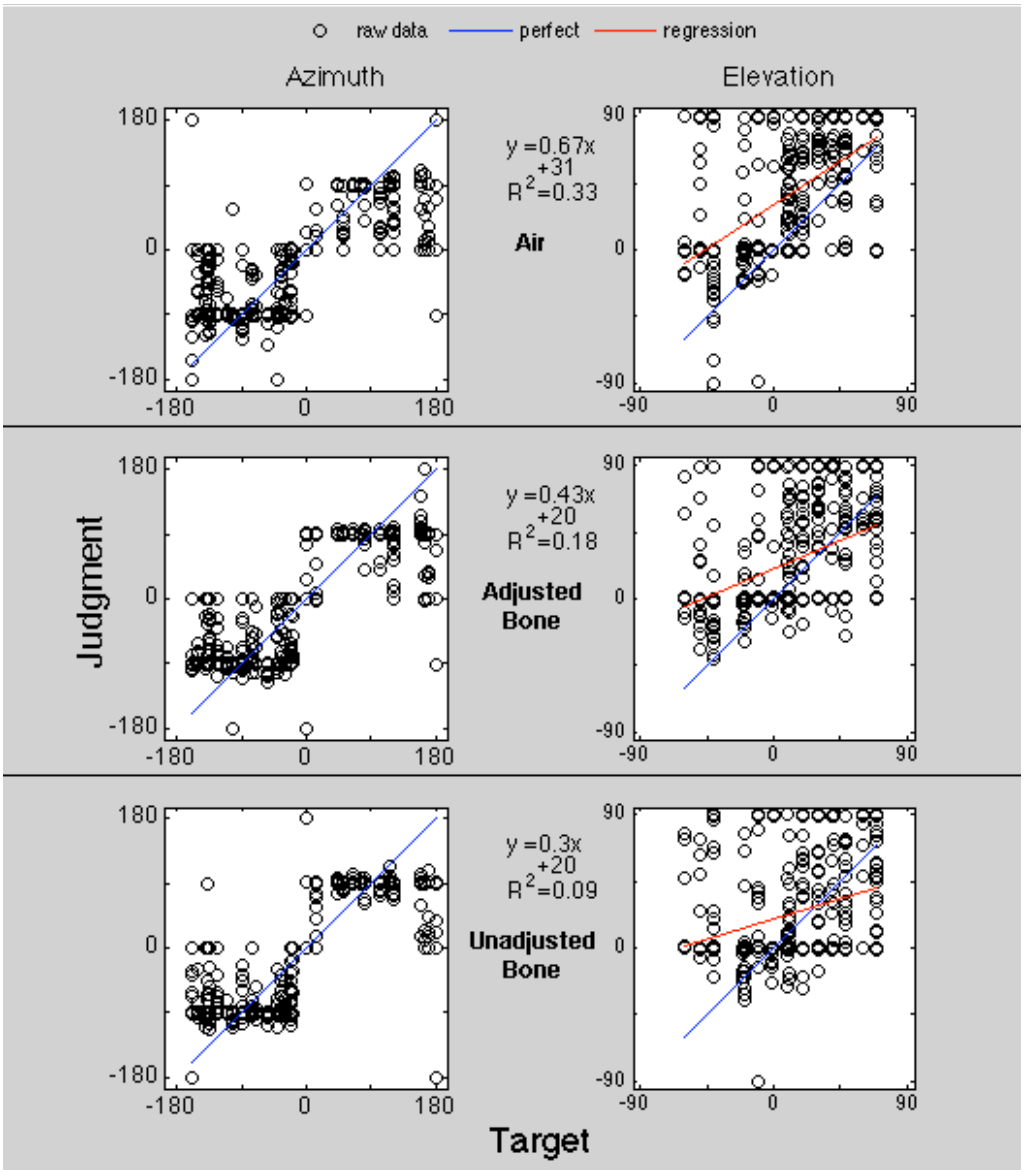

Figure 21. Scatter plots for participant 5. For azimuth, the air condition baseline shows many points that fall on the diagonal, except at the ends (few stimuli were perceived in back), and some front/back reversals. For the adjusted and unadjusted bone azimuth, the response pattern tends to become less varied and slightly binary. For elevation, the air condition baseline data fall on a moderately sloped diagonal, with a relatively low amount of spread from that diagonal. This is echoed in the slope of the regression line (.67) and fit of the data to that line (R^2 = .33). When going from the air condition to the adjusted bone then unadjusted bone condition, the data points fall on a progressively shallower sloped line and become more spread out around that line, as indicated by the monotonically decreasing regression slope and R^2 value.

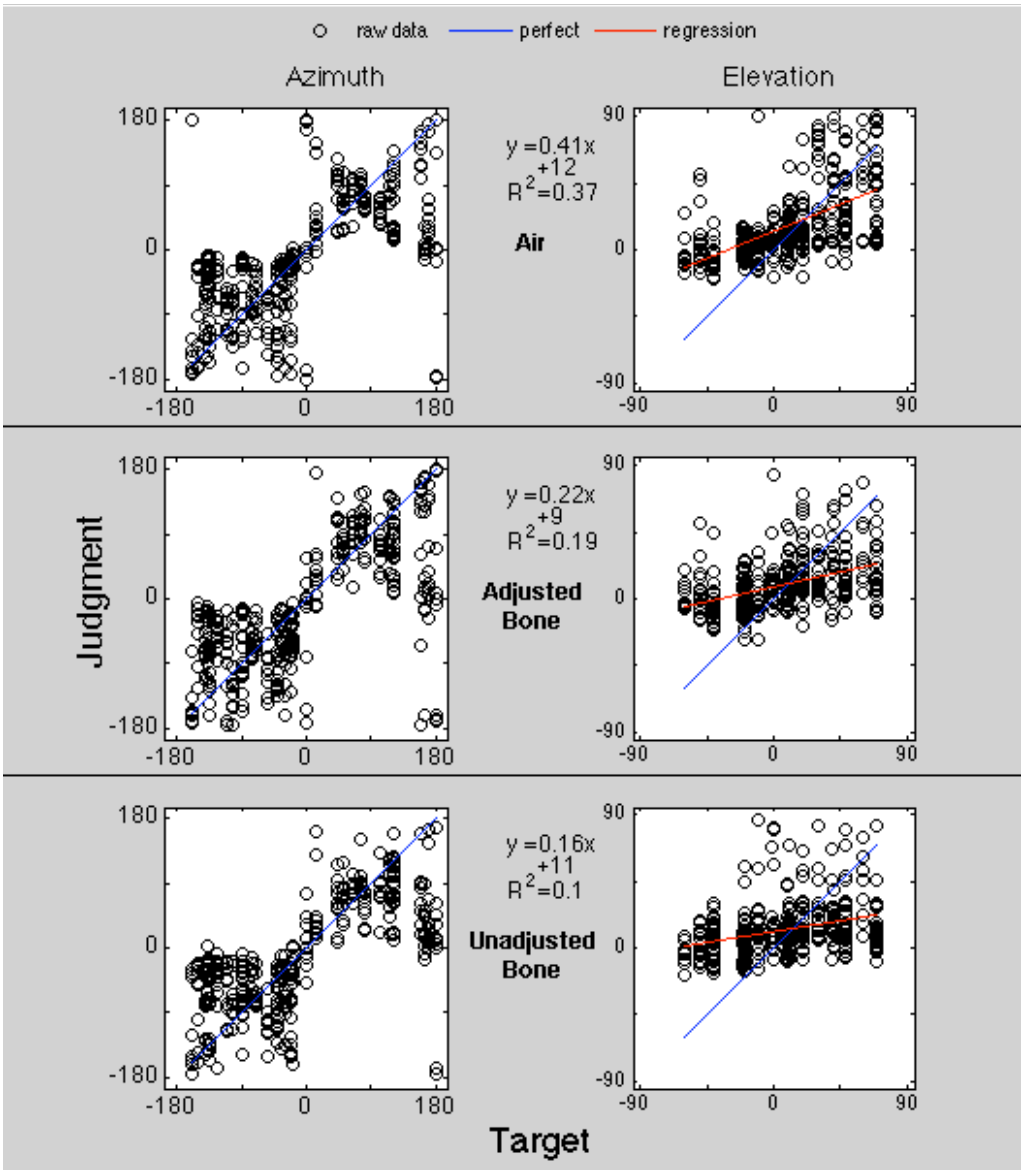

Figure 22. Scatter plots for participant 6. For azimuth, the air condition baseline shows a large number of data points that fall on the diagonal, besides the large amount of front/back reversals on the cross-diagonal. For the adjusted and unadjusted bone azimuth, the response patterns are similar but with a bit more variability. For elevation, the air condition baseline data fall on a steeply sloped diagonal in the top hemi-field, but few stimuli are perceived in the lower hemi-field. This pattern is echoed in the slope of the regression line (.41) and fit of the data to that line (R^2 = .37). When going from the air condition to the adjusted bone then unadjusted bone condition, the data points fall on a progressively shallower sloped line and become more spread out around that line, as indicated by the monotonically decreasing regression slope and R^2 value.

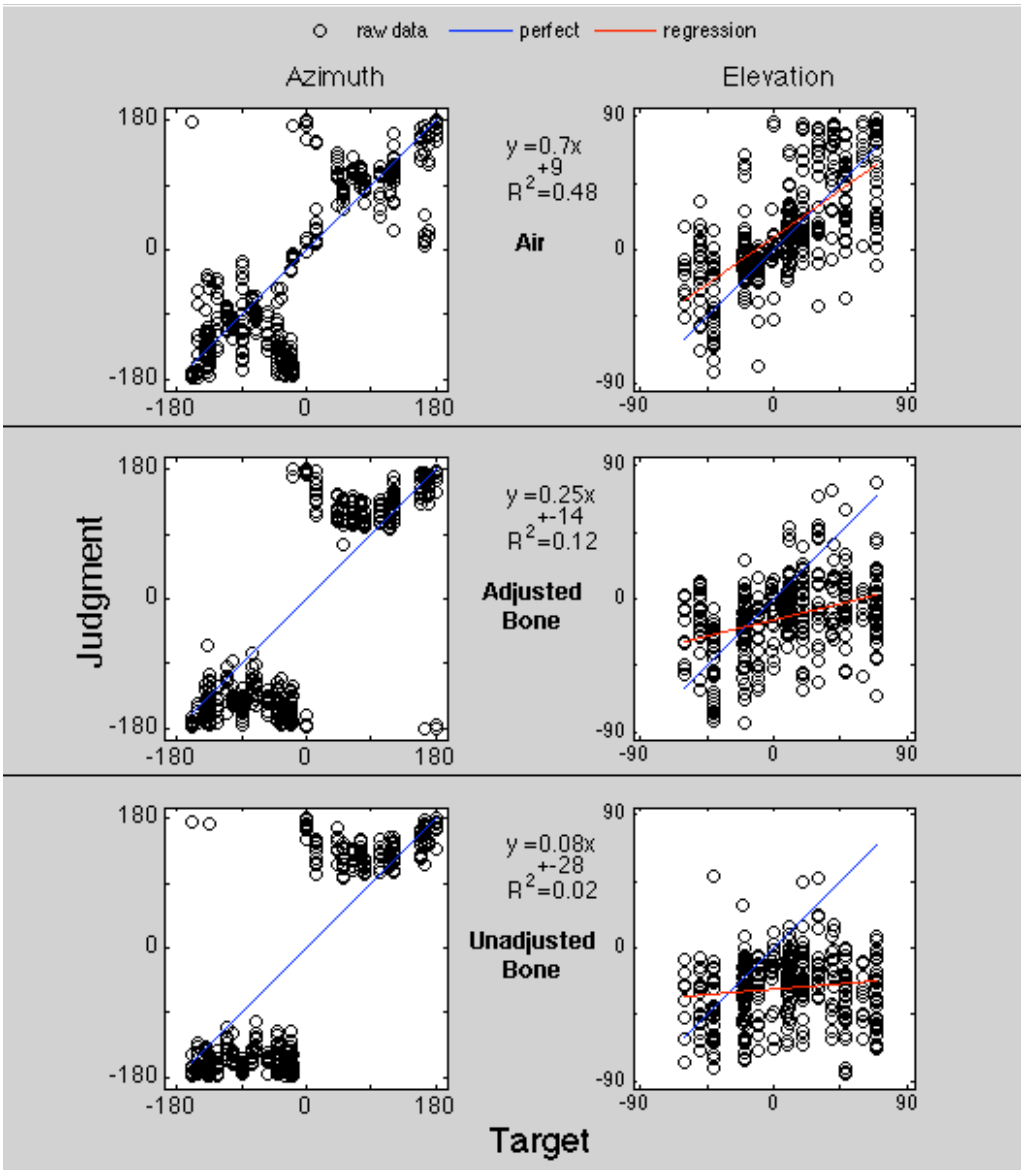

Figure 23. Scatter plots for participant 7. For azimuth, the air condition baseline shows a large number of points that fall on the diagonal, besides the large amount of front/back reversals on the cross-diagonal. For the adjusted bone azimuth, stimuli are no longer perceived on the frontal region, forming the "two-cluster" response pattern, and the front/back reversals become biased towards the back. The two-cluster response pattern becomes slightly more prominent in the unadjusted bone condition. For elevation, the air condition baseline data fall on a steeply sloped diagonal, with a low amount of spread from that diagonal. This is echoed in the steep slope of the regression line (.71) and high fit of the data to that line (R^2 = .48). When going from the air condition to the adjusted bone then unadjusted bone condition, the data points fall on a dramatically shallower sloped line and become more spread out around that line, as indicated by the monotonically decreasing regression slope and R^2 value.

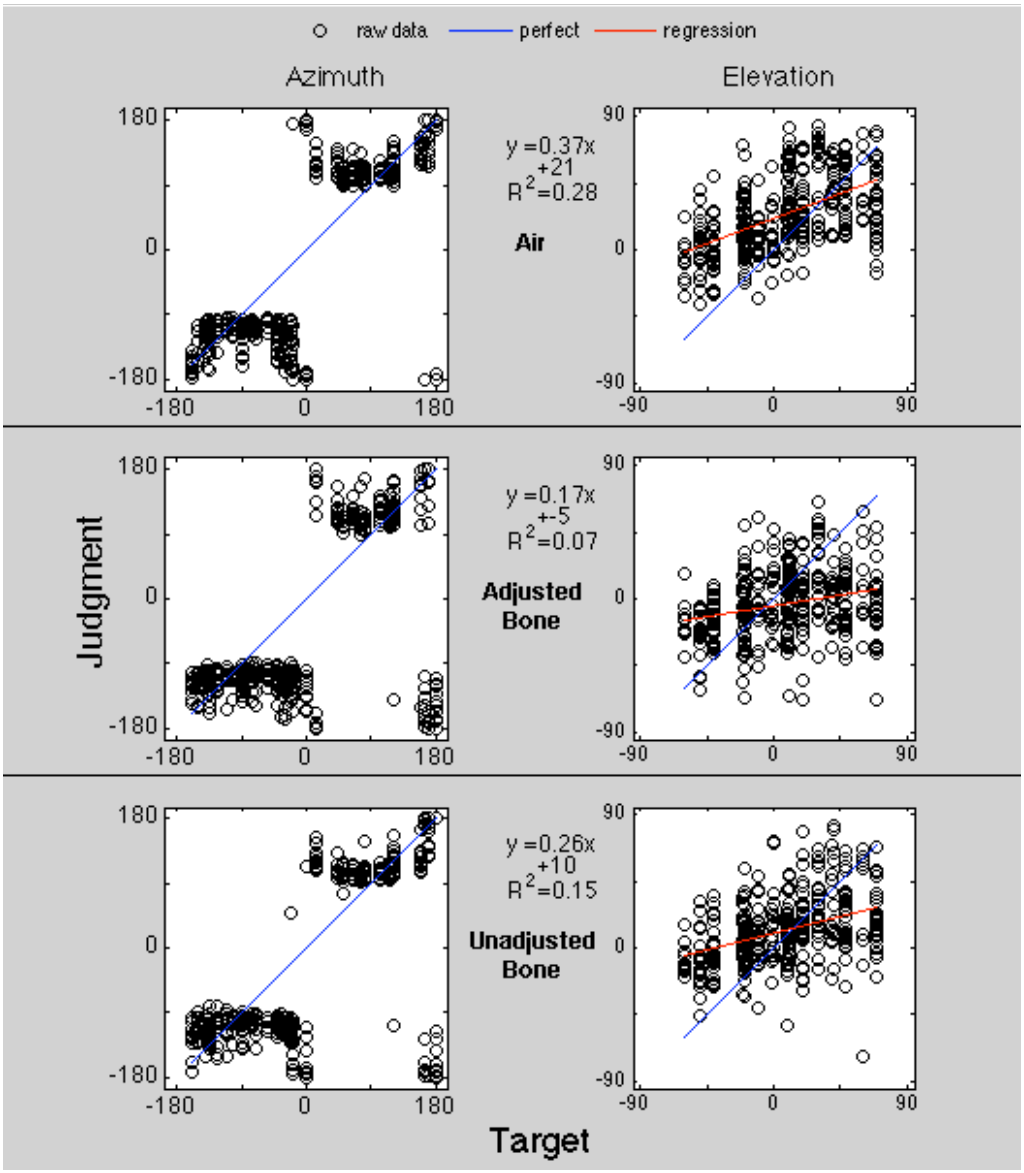

Figure 24. Scatter plots for participant 8. For azimuth, the air condition baseline shows that few points fall on to the diagonal, with no stimuli perceived in the frontal region and many front–to-back reversals, showing a "two-cluster" response pattern. For the adjusted bone and unadjusted bone condition, the data is similar but takes on a progressively more binary response pattern. For elevation, the air condition baseline data fall on a moderately sloped diagonal, with a moderate amount of spread from that diagonal, reflected in the moderate slope of the regression line (.37) and high fit of the data to that line (R^2 = .28). When going from the air condition to the adjusted bone condition, the data points become increasingly variable, and fall on a shallower sloped line, as indicated by the decreased regression slope and R^2 value. However, the unadjusted bone condition then gives an increased regression slope and decreased variability (increased R^2) relative to the adjusted bone condition.

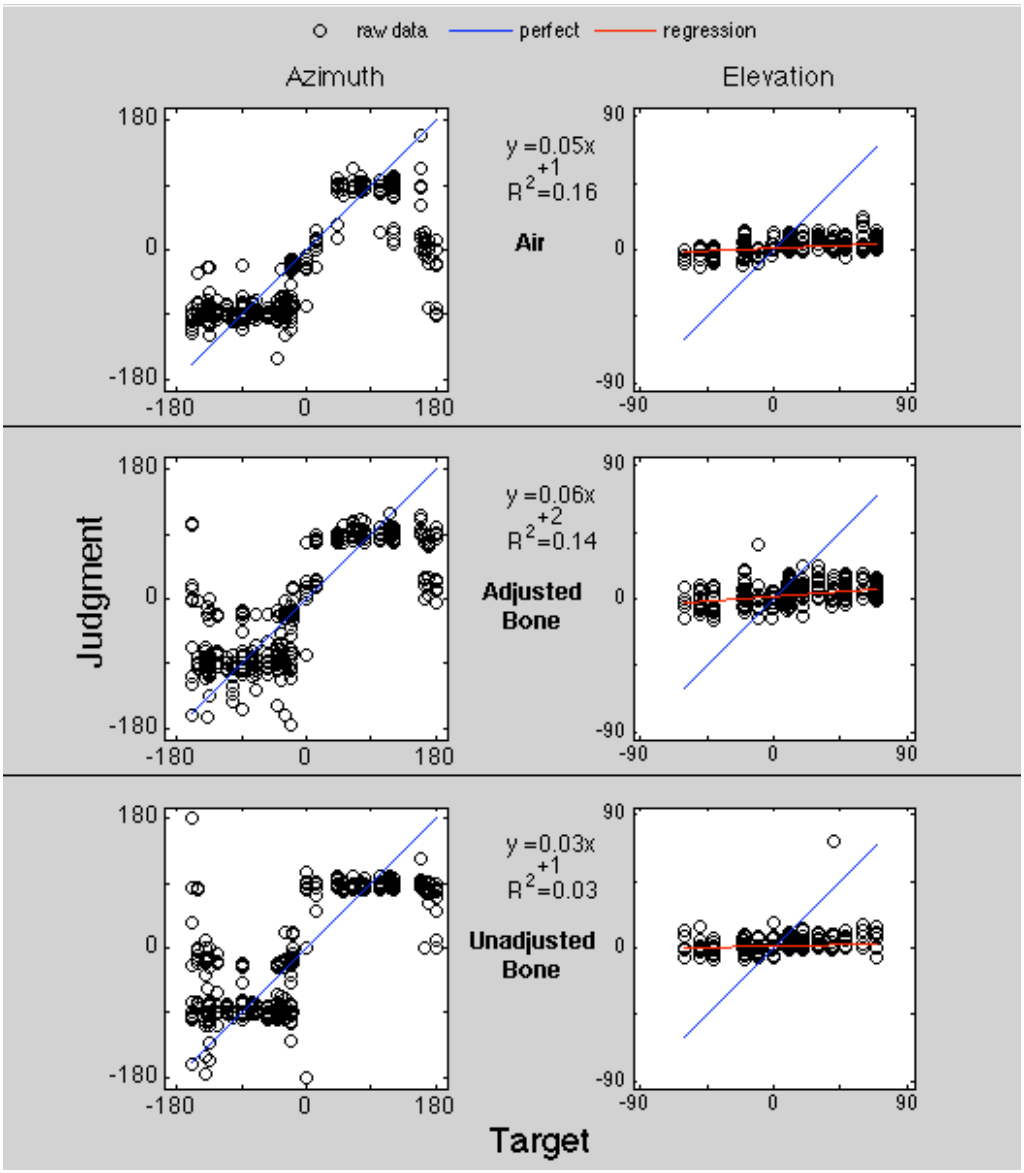

Figure 25. Scatter plots for participant 9. For azimuth, the air condition baseline shows many points that fall on the diagonal everywhere but the ends (few stimuli perceived in back), a slight binary response pattern, and some front/back reversals. For the adjusted and unadjusted bone azimuth, the response pattern remains similar, but with a bit more response variability. For elevation, the air condition baseline data fall on a very shallow-sloped diagonal, with very little range in the elevation response (nearly all responses are within +/- 20 degrees). This is echoed in the slope of the regression line (.05). When going from the air condition to the adjusted bone then unadjusted bone condition, the data points remain tightly grouped around a slightly sloping line.

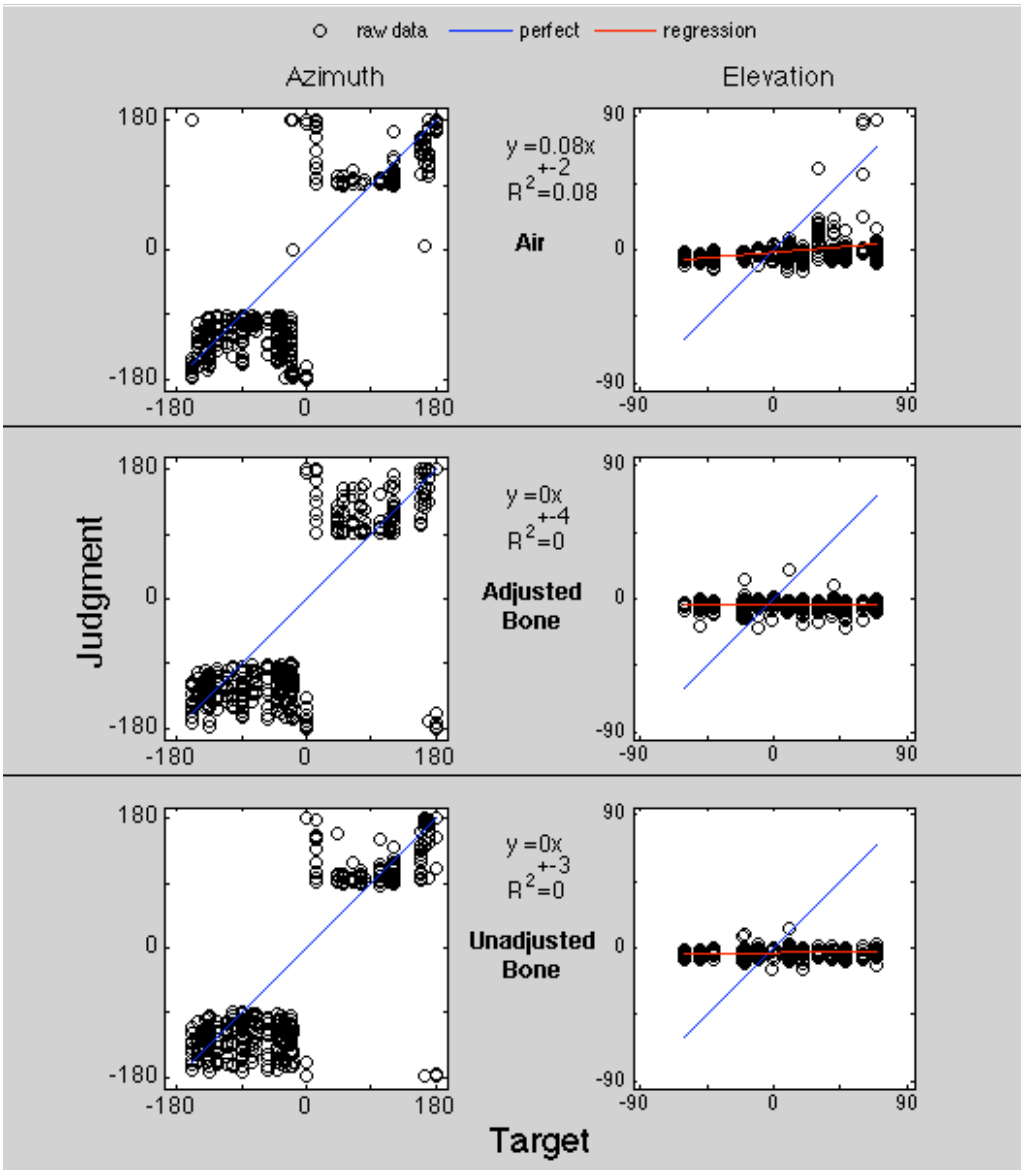

Figure 26. Scatter plots for participant 10. For azimuth, the air condition baseline shows that many points fall on the diagonal at the ends, but no stimuli are perceived in the frontal region and there are many front–to-back reversals, with a mix of a "two-cluster" and binary response pattern. For the adjusted bone and unadjusted bone condition, the data remains similar, with a bit more response variability in the right hemi-field, taking on a "two-cluster" response pattern. For elevation, the air condition baseline data fall on a very low sloped diagonal (with exception of a few points), with very little range in the elevation response (nearly all responses are within +/- 20 degrees). This is echoed in the slope of the regression line (.08). When going from the air condition to the adjusted bone then unadjusted bone condition, the data points remain tightly grouped around a slightly sloping line.

General Trends and Participants Excluded

The raw azimuth data showed quite varied patterns across participants, but there were a few consistent trends worth noting. In general, most air condition scatter plots showed a strong cross-diagonal, indicating a large amount of baseline front/back reversals. Three participants (participants 1, 2 and 7) showed a consistent trend of having many air-conduction data points fall on the diagonal, then in the bone-conduction conditions showing fewer data points that fall on the middle of the diagonal, suggesting a decrease in the number of sounds perceived to be directly in front of the listener, as well as a two-cluster response pattern. Several other participants started with very few stimuli perceived in front in the air condition (participants 4, 5, 8 and 10).

With the exception of participants 3, 8 and 10, all participants showed a trend of the raw elevation regression line slope decreasing when going from air conduction to adjusted bone conduction to unadjusted bone conduction. This same trend occurred for the R^2 values, indicating a spread in the pattern of responses from that regression line, and thus decreased accuracy.

For four participants, the difference in slope between the air and adjusted bone condition was larger than the difference in slope between adjusted bone and unadjusted bone conditions. For two participants, the differences between the two sets of conditions were similar. For one participant, the difference in slope between the bone conditions was greater than the difference between the air and adjusted bone condition. A similar pattern occurred for the R^2 statistic.

Table 6

Regression Slope (b) and R^2 values for raw azimuth and raw elevation data.

	Slope (*b*)			R^2		
Participant	Air	Adjusted Bone	Unadjusted Bone	Air	Adjusted Bone	Unadjusted Bone
1	0.22	0.15	0.08	0.13	0.05	0.02
2	0.90	0.50	0.33	0.58	0.32	0.21
3	0.39	0.23	0.32	0.12	0.05	0.09
4	0.86	0.64	0.40	0.38	0.25	0.12
5	0.67	0.43	0.30	0.33	0.18	0.09
6	0.41	0.22	0.16	0.37	0.19	0.10
7	0.70	0.25	0.08	0.48	0.12	0.02
8	0.37	0.17	0.26	0.28	0.07	0.15
9	0.05	0.06	0.03	0.16	0.14	0.03
10	0.08	0.00	0.00	0.08	0.00	0.00
Average	0.47	0.27	0.20	0.29	0.14	0.08
Standard Deviation	0.31	0.20	0.14	0.17	0.10	0.07

Participants 9 and 10 had a restricted range of elevation judgments, such that nearly all responses in the air conduction condition were less than +/-20 degrees. Thus, they showed virtually no sensitivity to elevation in the condition that serves as the baseline reference (air conduction). Thus, it is nearly impossible to evaluate the effect of the pathway and filter on their elevation performance. Because elevation is such an important part of the evaluation of the pathway and filter effects, and this indicates a general difficulty in localization with these HRTFs, participants 9 and 10 were excluded from all future analyses.

Participant 8 presented the opposite trend from most other participants in elevation slope and R^2 values within the bone conduction conditions, with the unadjusted bone conduction values being higher than the adjusted bone conduction values. Participant 8 had particular difficulty getting the ER-1 headphones inserted to an appropriate depth in the early sessions (more than any other participant), which included the equal loudness adjustments for building the BAF. Participant 8's results in light of this issue are not surprising – the user guide that accompanies the ER-1 headphones indicate that the frequency response resulting at the eardrum is highly dependent on insertion depth. The insertion depth was not corrected upon initial observation for two reasons: 1) it seemed that any further depth would cause discomfort for the participant, and 2) the participant was run early on in data collection, before the degree of sensitivity to insertion depth was realized. To verify that participant 8's BAF was ill-formed, the BAF was measured again after completion of Study 2, this time with a slightly deeper insertion that came from experience in the experiment sessions. A plot of the initial BAF and post-experiment BAF can be found in Appendix A, Figure A8. The post-experiment BAF shows peaks and notches not present in the original BAF, giving further confirmation that the BAF was not appropriately measured for this participant. Given all this evidence that the atypical results were caused by an anomaly associated with the acoustic apparatus, participant 8 was excluded from all future analyses.

Participant 3 showed the same "opposite" trend that deviated from most other participants. Participant 3 also had difficulty getting the headphones inserted to an appropriate depth, but no more definitively than the other few participants who had difficulty but did not show the same trend. Participant 3 even had one session where the

elevation slope and R^2 indicated better performance for the unadjusted bone condition than for the air condition. This would not be surprising given improper insertion, which would render inappropriate air-conduction spectral cues. In the session that participant 3 had the best air-conduction elevation performance (indicated by slope and R^2), their elevation performance followed the trend of the other participants (air was better than adjusted bone, which was better than unadjusted bone). This could have been one session when the insertion was appropriate, rendering the spectral cues appropriately in the air- and adjusted bone condition. Because participant 3's difficulty with insertion could not be definitively shown to be worse than others whose data are not excluded, participant 3 was included in subsequent analyses.

Inferential Treatment of Slopes and Fit

A one-way repeated measures ANOVA was conducted on the raw elevation slope values to see if the pattern of regression slopes found in this sample can be generalized to the population. There was an omnibus effect of condition on regression slope ($F(2,12) = 18.64$, $p < .001$). Tukey's Honestly Significant Difference ("HSD") posthoc analyses were used here and in all other follow-up tests to find which pairwise comparisons were statistically significant. The difference in elevation slope between air conduction and adjusted bone conduction, as well as between air conduction and unadjusted bone conduction, was statistically significant, but the difference between the adjusted bone conduction condition and unadjusted bone conduction condition did not quite reach statistical significance (see Table 7).

Table 7

Post-hoc tests for effect of condition on elevation regression slope. Critical q(3,12) = 3.77 at .05 significance level.

Comparison	q[1]
Air – Adjusted Bone	7.01*
Air – Unadjusted Bone	6.25*
Adjusted Bone – Unadjusted Bone	3.61

[1]these and al subsequent q values were computed based on t values from a paired-samples t-test conducted in SPSS. q values were computed using the following formula: $q = t^* \sqrt{2}$ (Howell, 2002). This q value was then compared to the critical q value from a lookup table (hence exact p values were not available).
*$p < .05$

A similar analysis was done on the pattern of fit to the regression line (indexed by Pearson's r). r, rather than R^2 is used because there is a transformation specifically designed for r that makes it suitable for ANOVAs. Because r is directly related to R^2, ($r^*r = R^2$) it still serves as an index of how well the regression line fits the data. In addition, r alone can serve to describe the degree of linear relationship (Cohen & Cohen, 1983).

It is a somewhat unusual situation to have a correlation coefficient as a score for each participant, but this certainly does occur in research, when using Q-sorts in personality and clinical psychology areas (Cohen & Cohen, 1983; e.g., Teti & McGourty, 1996). To do an inferential test on the correlation coefficient, a transformation is needed to make r "more likely to satisfy the normality and equal variance assumptions when used as a dependent variable" (Cohen & Cohen, 1983, p. 271).

Fisher's *r* to *z'* transformation can be defined by:

$z' = \frac{1}{2} [\ln(1+r) - \ln(1-r)]$

where ln is the natural logarithm (Cohen & Cohen, 1983, p. 53).

A one-way repeated measures ANOVA was conducted on the transformed elevation *r* (i.e., *z'*) values to see if the pattern of correlation values found in this sample can be generalized to the population. Mauchley's test of sphericity showed that equal variances could not be assumed ($W(2) = 0.249$, $p = .031$). Geisser-Greenhouse corrections were made to the degrees of freedom to correct for the positive alpha bias that occurs with unequal variances. With this correction the ANOVA showed an omnibus effect of condition on *r,* $F(1.14,6.85) = 20.34$, $p = .002$. Tukey's HSD posthoc analyses showed statistically significant differences between all comparisons: air conduction - adjusted bone conduction, air conduction - unadjusted bone conduction, and adjusted bone conduction - unadjusted bone conduction (see Table 8).

Table 8

Post-hoc tests for effect of condition on fit to regression line (indexed by Pearson's r, transformed to Fisher's z'). Critical q(3,12) = 3.77 at .05 significance level. Degrees of freedom adjusted by Greenhouse-Geisser correction.

Comparison	*q*
Air – Adjusted Bone	7.13*
Air – Unadjusted Bone	6.49*
Adjusted Bone – Unadjusted Bone	4.29*

*$p < .05$

Resolved and Corrected Scatter Plots

As previously mentioned, the circular geometry of the azimuth coordinates prevents regression analyses. To provide azimuth data that were suitable for regression analyses, a "resolved azimuth" data set was formed. The first operation done to the azimuth data was the resolution of the front/back reversals (where the stimulus is presented on the front hemifield, and is perceived in the rear hemifield, or vice versa). Gross errors in hemisphere (reversals) are a separate class of errors than smaller localization "blur" (Blauert, 1997). To deal with this separate class of errors, reversals are often counted and then resolved (recoded into the correct quadrant) for subsequent summary statistics, which avoids overstating the typical localization error (Wightman & Kistler, 1989b)).

Front/Back reversals were determined and resolved in a manner similar to Wightman & Kistler (1989b). For each stimulus/response pair, the response was reflected across the coronal plane (vertical plane passing through ears). This is done by putting the response in the opposite front/back hemisphere and in the same left/right hemisphere, at the same distance from the sagittal plane (vertical plane passing from nose to center of back of head; see Figure 27). If the error between stimulus and response was smaller with this reflection, then a front/back confusion was counted (the count data were used for subsequent computations of reversal rates), and the response was "resolved" by using its reflected value for all analyses.

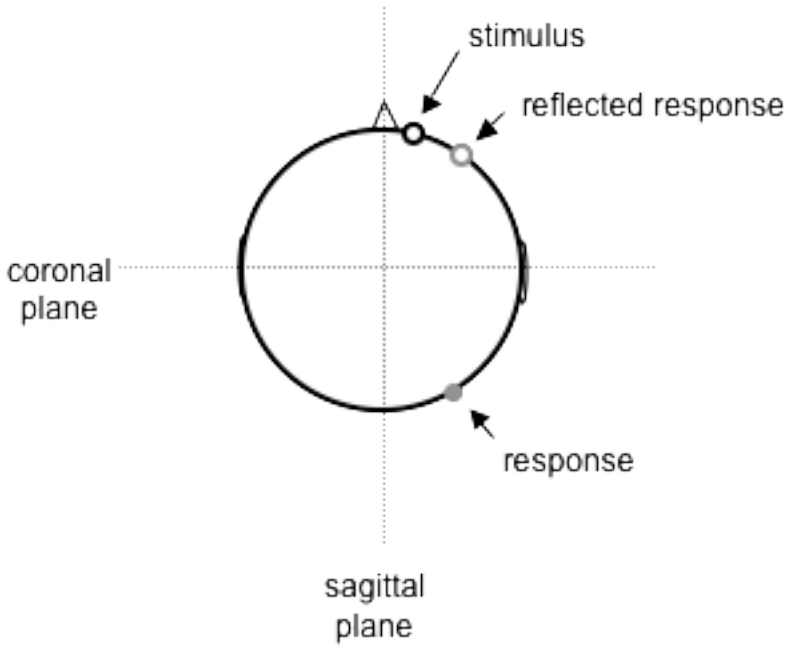

Figure 27. Example of a prototypical front/back reversal. Note how the reflected response is closer to the stimulus than the original response. This response would be scored as a front/back reversal and the reflected value would be used in all subsequent analyses.

The convention established by Wightman & Kistler (1989b) was modified by adding an exclusion range for the reversal computations. If the stimulus or response was within 15 degrees of the coronal plane, it was not eligible for a reversal. The value of 15 degrees was chosen based on the assertion that anything within a range of 30 degrees around the left or right ear should be categorized as local error, rather than be eligible to be counted as a reversal. This more conservative definition of a reversal was utilized to restrict the resolution of reversals to more prototypical reversals, and avoid resolutions that result in only small reductions in error (see Figure 28). With a 15 degree front/back exclusion range, a trial was not eligible for a reversal if the stimulus or the response was between +/- 75 and 105.

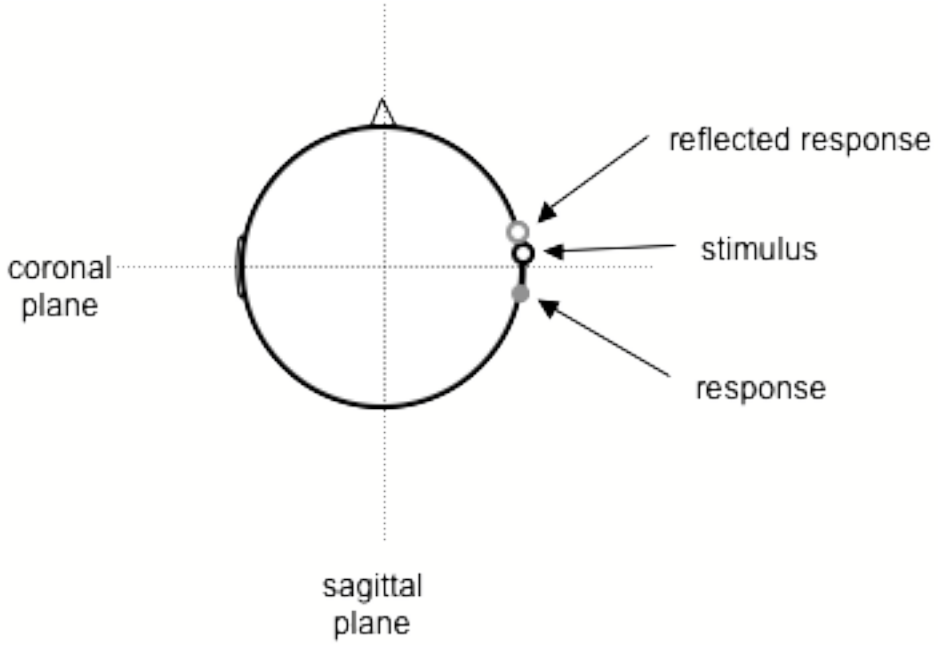

Figure 28. Example of what could be classified as a front/back reversal without an exclusion range. Note that this response is better described as localization blur than a gross error in hemisphere. An exclusion range of 15 degrees was used to avoid stimulus-response patterns such as this one being classified as a reversal, when they are more appropriately grouped with small errors in localization.

A procedure analogous to the one just discussed was also done for confusions in the up/down dimension. In this case, the response was reflected across the axial (horizontal) plane placed at 0 degrees elevation (see Figure 29). An exclusion range of 15 degrees was again used for this computation: if the stimuli or response was -15 to +15 elevation, the trial was ineligible for reversal.

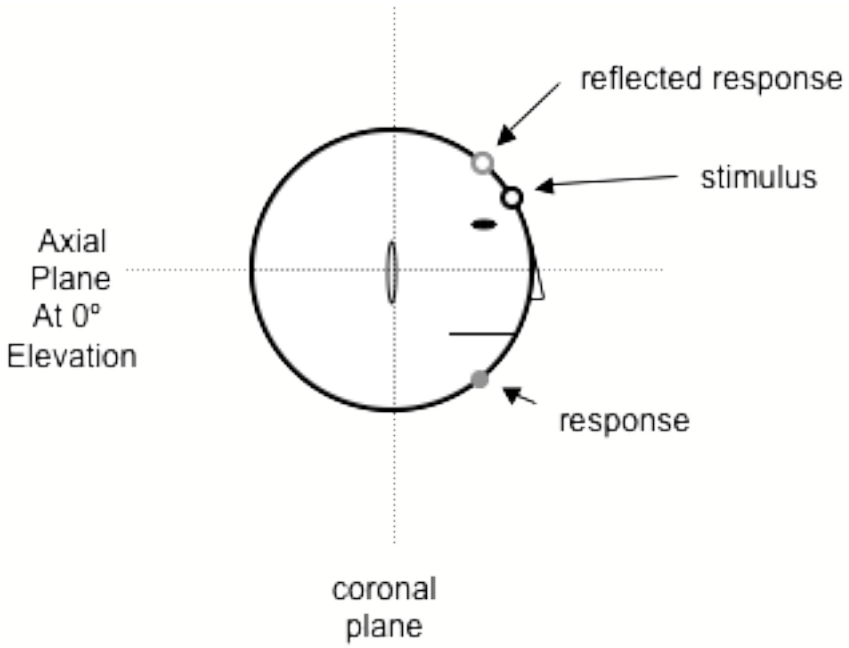

Figure 29. Example of a prototypical up/down reversal. The reflected response is closer to the stimulus than the original response, and thus would be scored as an up/down reversal. The reflected value would then be used in all subsequent analyses.

The wrap-around artifact present in the azimuth data was also addressed in this analysis. As previously shown, the wrap-around artifacts occur in the bottom-right and top-left portion of the plot, and cause the point to be treated as farther away from the stimulus than it really is. This problem can be conceived in terms of two separate azimuth coordinate systems: a left (positive) and right (negative) coordinate system. When the stimulus and response pair cross the boundary coordinate system boundary (180/-180 degrees), the linear plot of the azimuth number system makes the response look farther away (from the diagonal) than it truly is in circular space. Thus, sometimes the response would look closer to the stimulus (the diagonal) in the other hemisphere's coordinate system on a linear plot, even though it is in the same place in circular space. To correct for this artifact, every response was converted to its position in the other hemisphere's coordinate space. The distance from the stimulus was then computed for the response in

its own coordinate response system and in the other hemisphere's coordinate system. The response that led to the least distance from the stimulus was then stored as the corrected response and plotted in the graph. It is important to note that there was no change in the actual position of the response (like there was for the front/back reversal resolution), but instead the data were just put into a different coordinate system. Consider as an example a stimulus at -173 degrees and a response at 165 degrees (see Figure 30). In absolute distance on a linear plot, the response would be 338 degrees away from the stimulus (i.e., the diagonal). In circular space, it can be clearly seen that the response is really only 22 degrees away from the stimulus. The response can be figured to be 22 degrees in linear space if it is put in the stimulus's coordinate system: 165 degrees in the left hemisphere's coordinate system is - 195 degrees, which is 22 degrees away from -173 degrees in linear space. In this case, the artifact removal algorithm would cause the point would be plotted as -195 degrees, rather than the original 165 degrees. As an example of a response that would not require conversion into the other hemisphere's coordinate space, consider a stimulus at -175 degrees and a response at -163 degrees. The response in the right hemisphere's coordinates would be 197 degrees, which would be 372 degrees away from the stimulus in linear space. In this case, the artifact removal algorithm would cause the original response to be stored and plotted, with no corrections made.

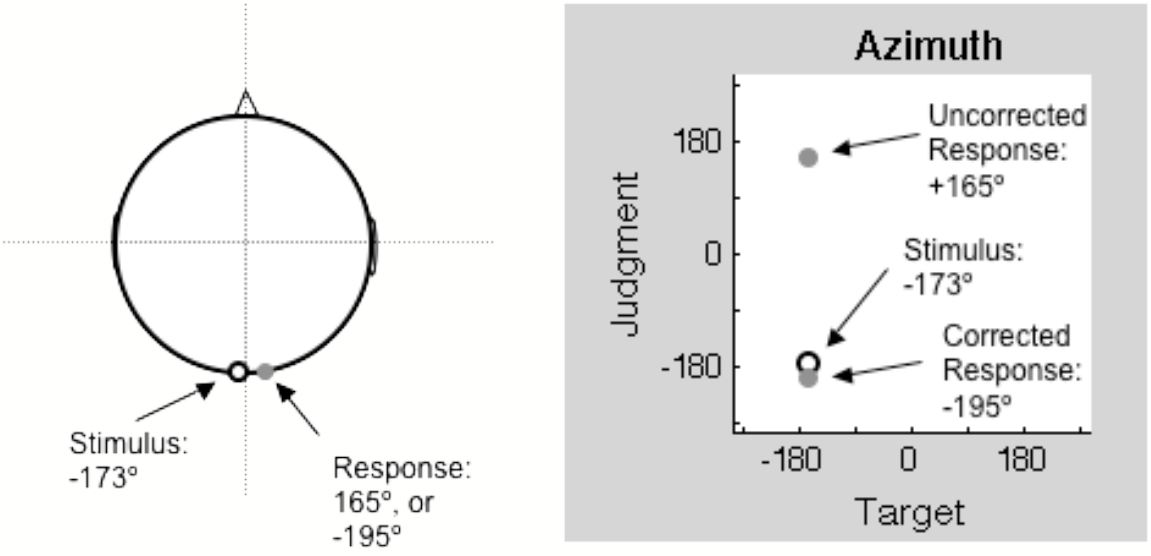

Figure 30. Wrap-around artifact that occurs for azimuth. On the left is a depiction of a response and stimulus in circular azimuth space, and on the right is a depiction of the same response and stimulus in linear space. In circular space, the stimulus and response are quite close to each other in circular space (22 degrees apart). In linear space, however, the uncorrected response is 338 degrees away from the stimulus. Correcting for the wrap-around artifact by putting the right-hemisphere response into the left-hemisphere coordinate system leads to the appropriate distance on the linear scale (22 degrees).

The data plotted with front/back reversals resolved for azimuth, as well as wrap-around artifacts removed, and up/down reversals resolved for elevation, can be found in Figure 31 through Figure 37 below. The layout is the same as Figure 17 through Figure 26, in a matrix of plots. The regression equation and R^2 value for each data subset can be found to the bottom-right of each azimuth graph, and to the top-left of each elevation graph. The scatter plots will again be described for each participant, with the regression line that the data fall on discussed for each plot. Terms such as low, moderate, and high (determined based on this sample's range of values) will again be defined for these analyses to describe regression slope and fit. For the azimuth data, there was a small range of slope and R^2 values; a high slope is defined as a value greater than or equal to .9, a moderate slope is any value less than .9, a high fit is any R^2 value greater than or equal

to .9, and a moderate fit is any value less than .9. For the elevation data, a high slope is greater than or equal to .9, a low slope is less than or equal to .5, and a moderate slope is between the two. A high fit is any R^2 value greater than or equal to .5, a low fit is less than .3, and a moderate fit is any value between the two.

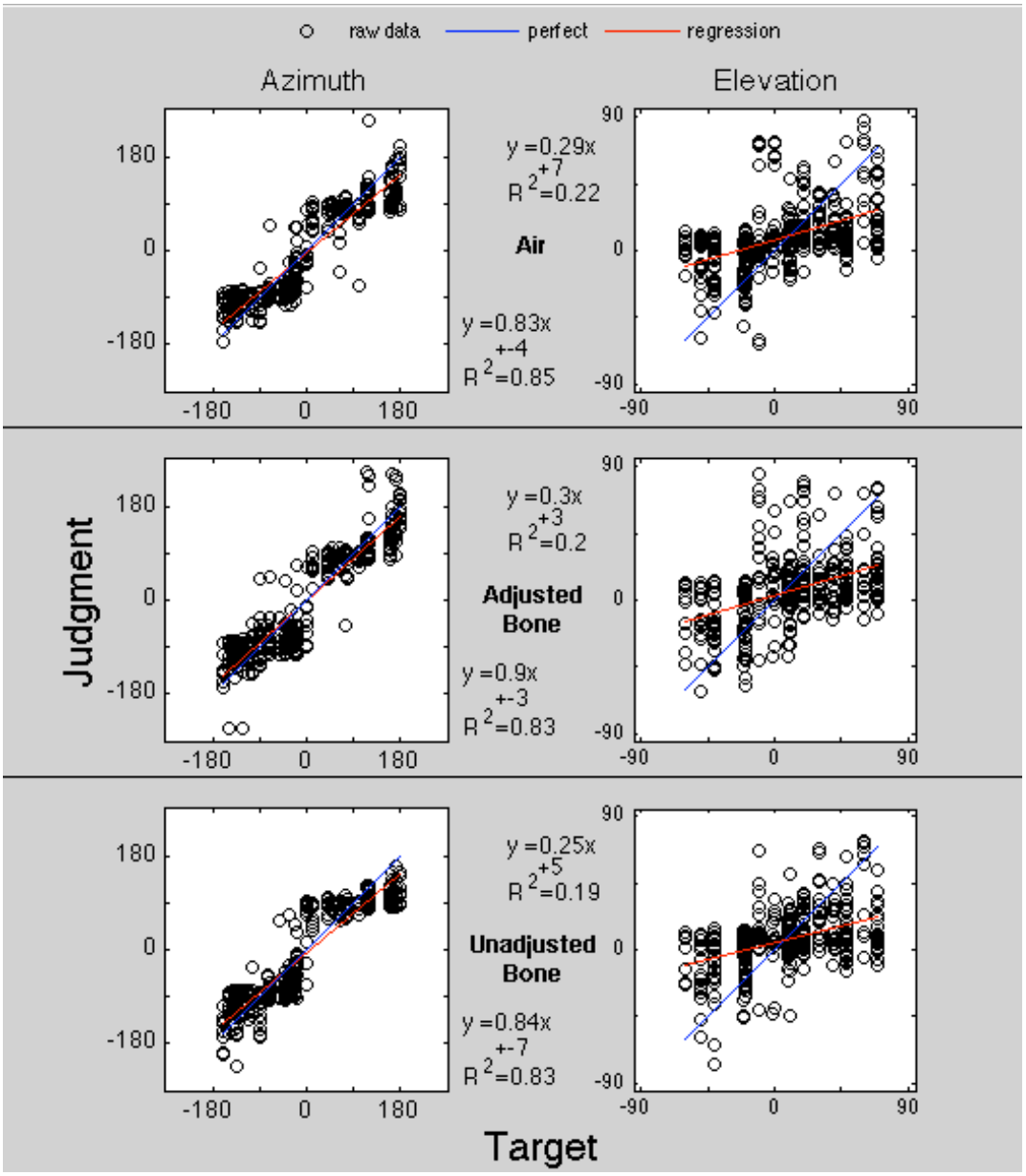

Figure 31. Scatter plots for participant 1, with front/back errors adjusted resolved and wrap-around artifact removed. For azimuth, the air condition baseline data fall on a moderately sloped regression line (*b*=.83) with a medium spread from the regression line (R^2 = .85). Relative to the air condition, the regression line for the adjusted bone condition data show an increase in slope and very slight reduction in fit. When going to the unadjusted bone condition, the slope is reduced to a level similar to the air condition and the fit remains the same. For elevation, the air condition data fall on a shallow sloped regression line (*b*=.29) with low fit to the regression line (R^2 = .22). The slope and fit stay roughly the same for the adjusted bone condition, but in the unadjusted bone condition the slope decreases slightly while the fit stays roughly the same.

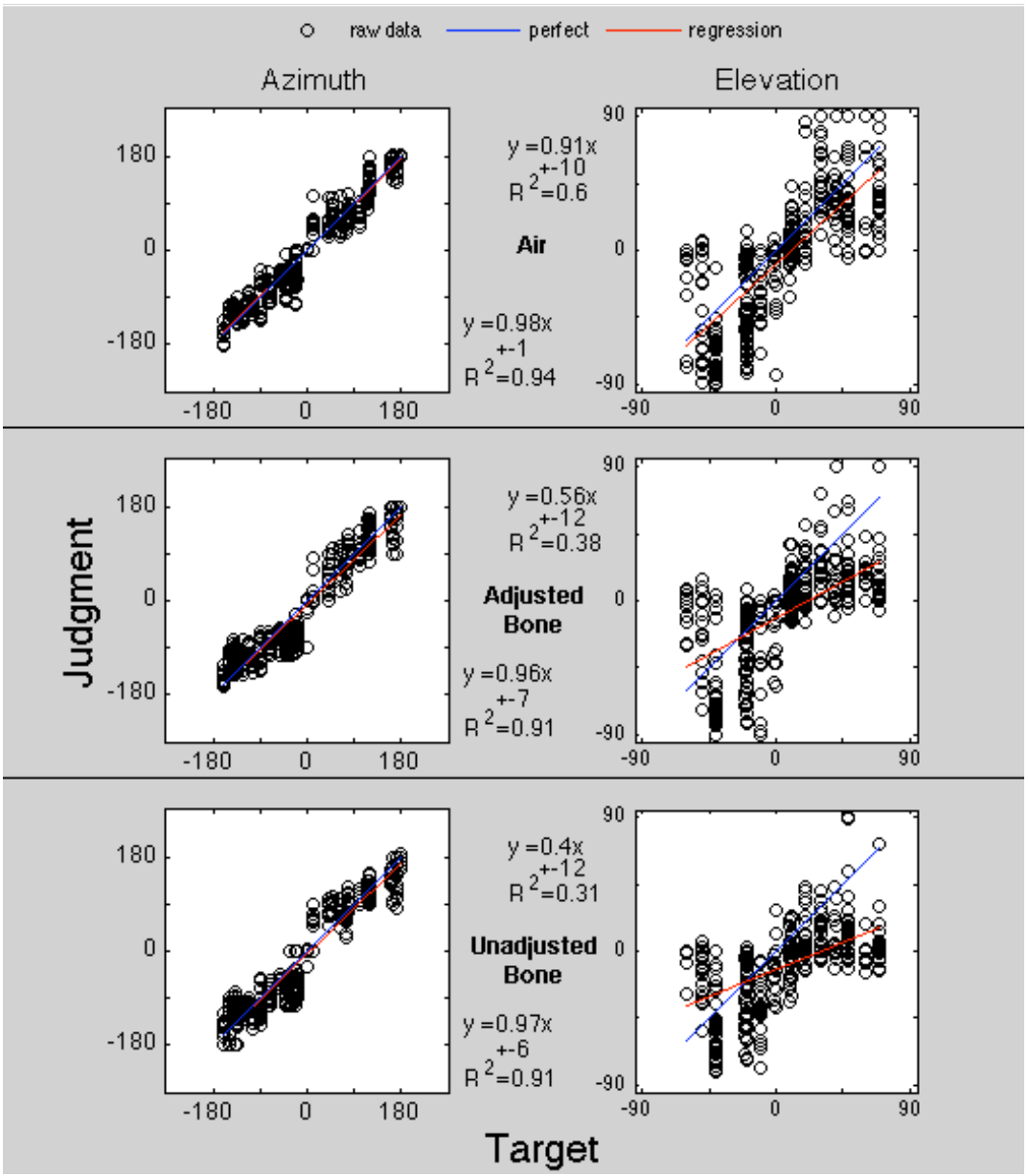

Figure 32. Scatter plots for participant 2, with front/back errors adjusted resolved and wrap-around artifact removed. For azimuth, the air condition baseline data fall on a steeply sloped regression line (*b*=.98) and high fit to that line (R^2 = .94). Relative to the air condition, the regression for the adjusted bone and unadjusted bone conditions shows similar slope and very slight reduction in fit. For elevation, the air condition data fall on a steeply sloped regression line (*b*=.91) with a high fit to that regression line (R^2 = .6). When going from the air condition to the adjusted bone then unadjusted bone condition, the data points fall on a much shallower sloped line and become more spread out around that line, as indicated by the monotonically decreasing regression slope and R^2 value.

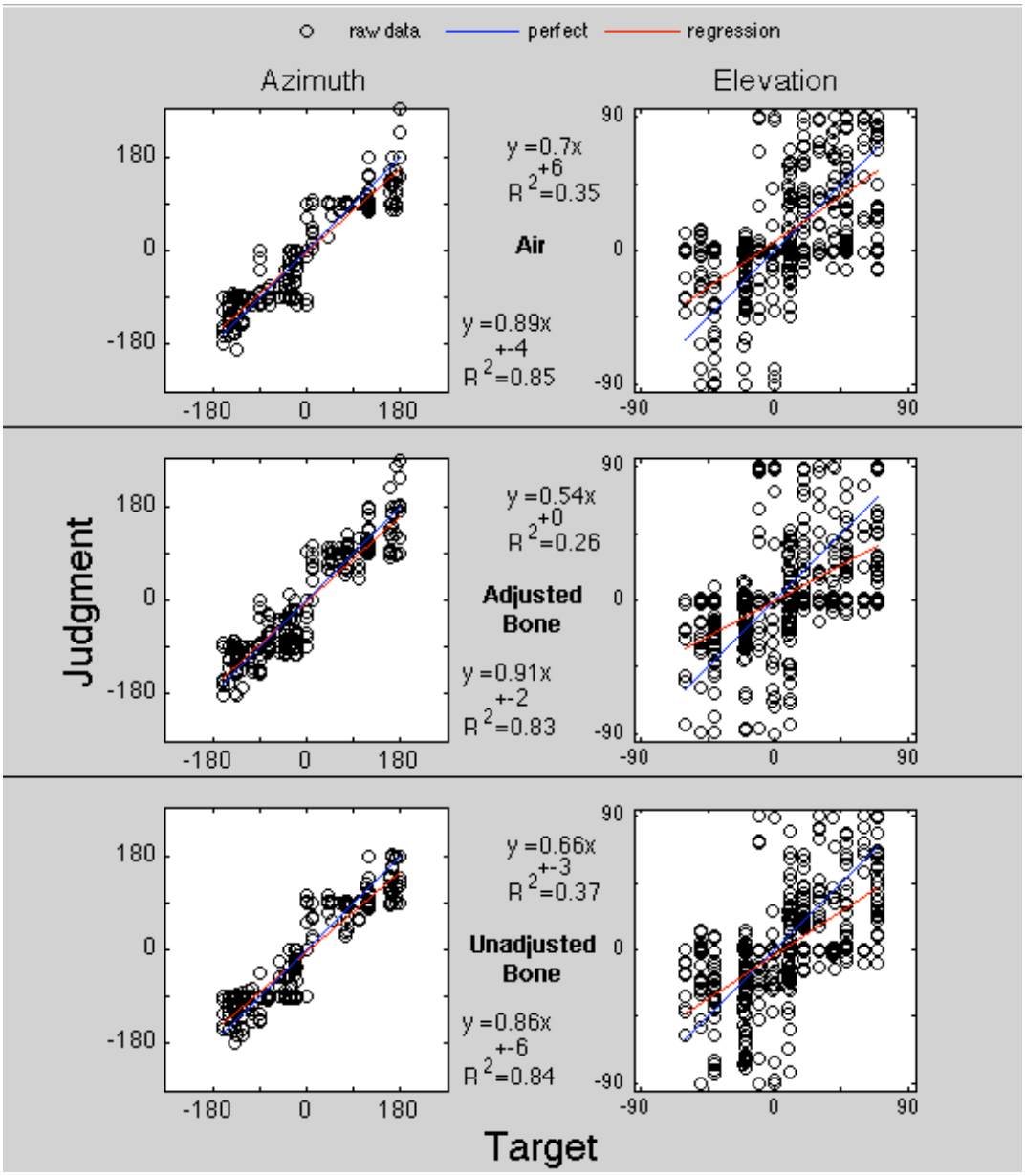

Figure 33. Scatter plots for participant 3, with front/back errors adjusted resolved and wrap-around artifact removed. For azimuth, the air condition baseline data fall on a steeply sloped regression line (*b*=.89) with a moderate fit to that line (R^2 = .85). Relative to the air condition, the regression for the adjusted bone shows similar slope and fit. The unadjusted bone condition then shows a slight reduction in slope and similar fit. For elevation, the air condition data fall on a regression line with moderate slope (*b*=.7) and have a moderate fit to that line (R^2 = .35). Relative to the air condition, the adjusted bone condition data fall on a shallower sloping regression line with a lower fit to that line. Then the unadjusted bone condition data fall on a steeper sloping regression line with a higher fit to that line.

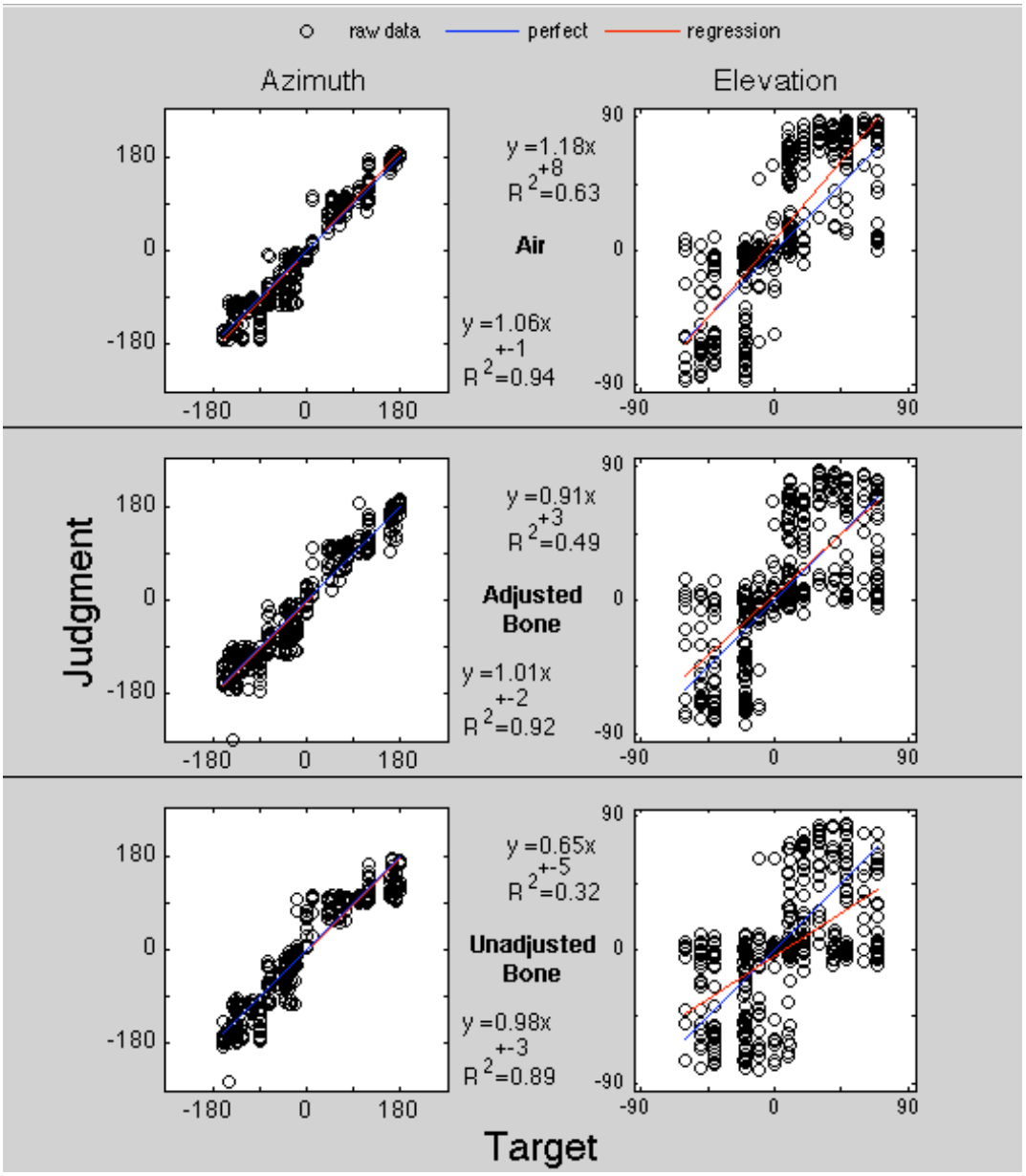

Figure 34. Scatter plots for participant 4, with front/back errors adjusted resolved and wrap-around artifact removed. For azimuth, the air condition baseline data fall on a steeply sloped regression line (b=1.06) with a high fit to that regression line (R^2 = .94). Relative to the air condition, the regression line for the adjusted bone condition shows a small reduction in slope and similar fit to that line, and the unadjusted bone condition data shows a further reduction in slope and fit to the regression line. For elevation, the air condition data fall on a regression line with steep slope (b=1.18) and have a high fit to that line (R^2 = 63). When going from the air condition to the adjusted bone then unadjusted bone condition, the data points fall on a shallower sloped line and become more spread out around that line, as indicated by the monotonically decreasing regression slope and R^2 value.

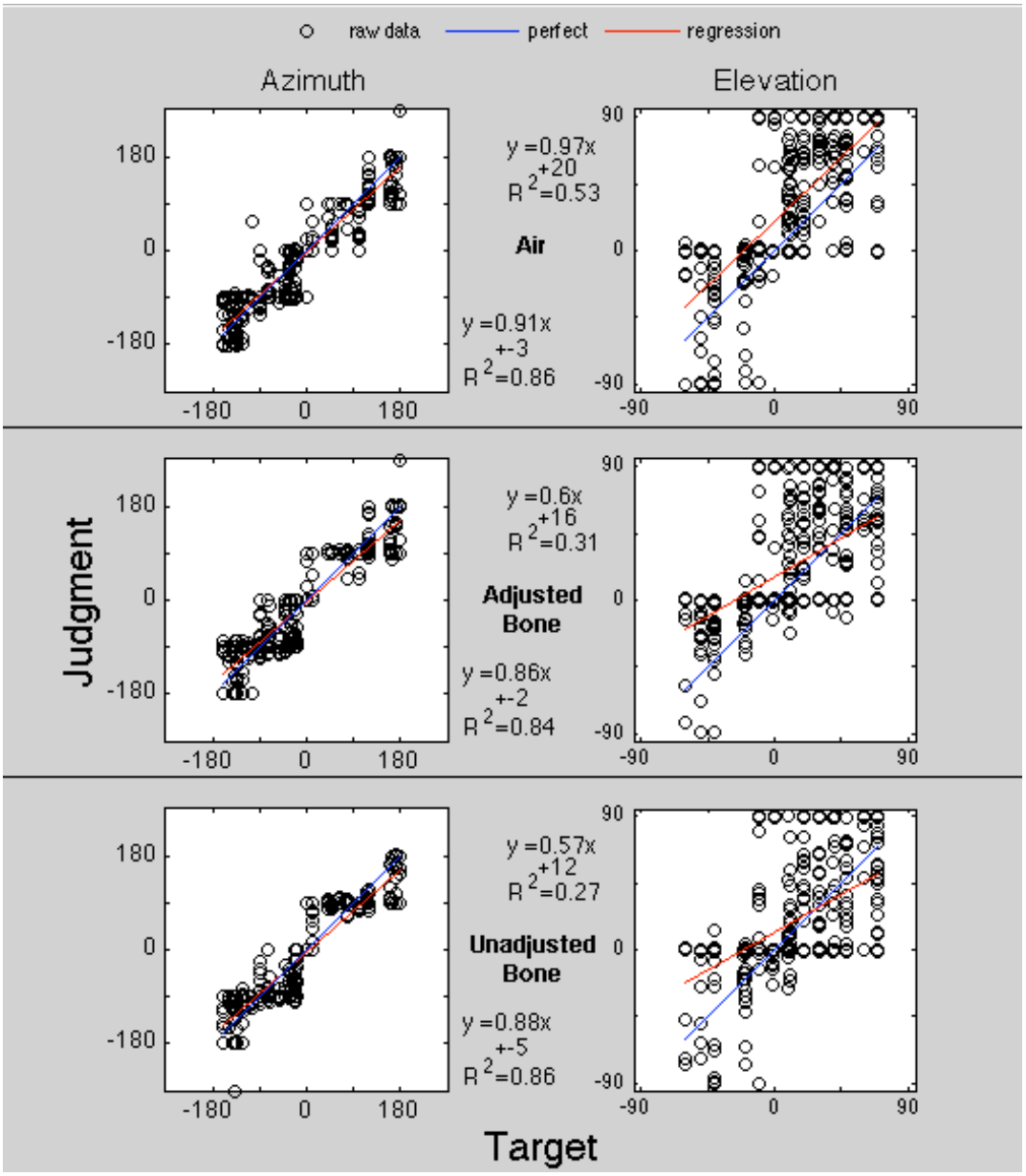

Figure 35. Scatter plots for participant 5, with front/back errors adjusted resolved and wrap-around artifact removed. For azimuth, the air condition baseline data fall on a steeply sloped regression line (*b*=91) with a moderate fit to that regression line (R^2 = .86). Relative to the air condition, the regression for the adjusted bone shows slightly reduced slope and similar fit. The unadjusted bone condition then shows a similar slope and fit. For elevation, the air condition data fall on a regression line with steep slope (*b*=.97) and have a high fit to that line (R^2 = 53). When going from the air condition to the adjusted bone then unadjusted bone condition, the data points fall on a shallower sloped line and become more spread out around that line, as indicated by the monotonically decreasing regression slope and R^2 value.

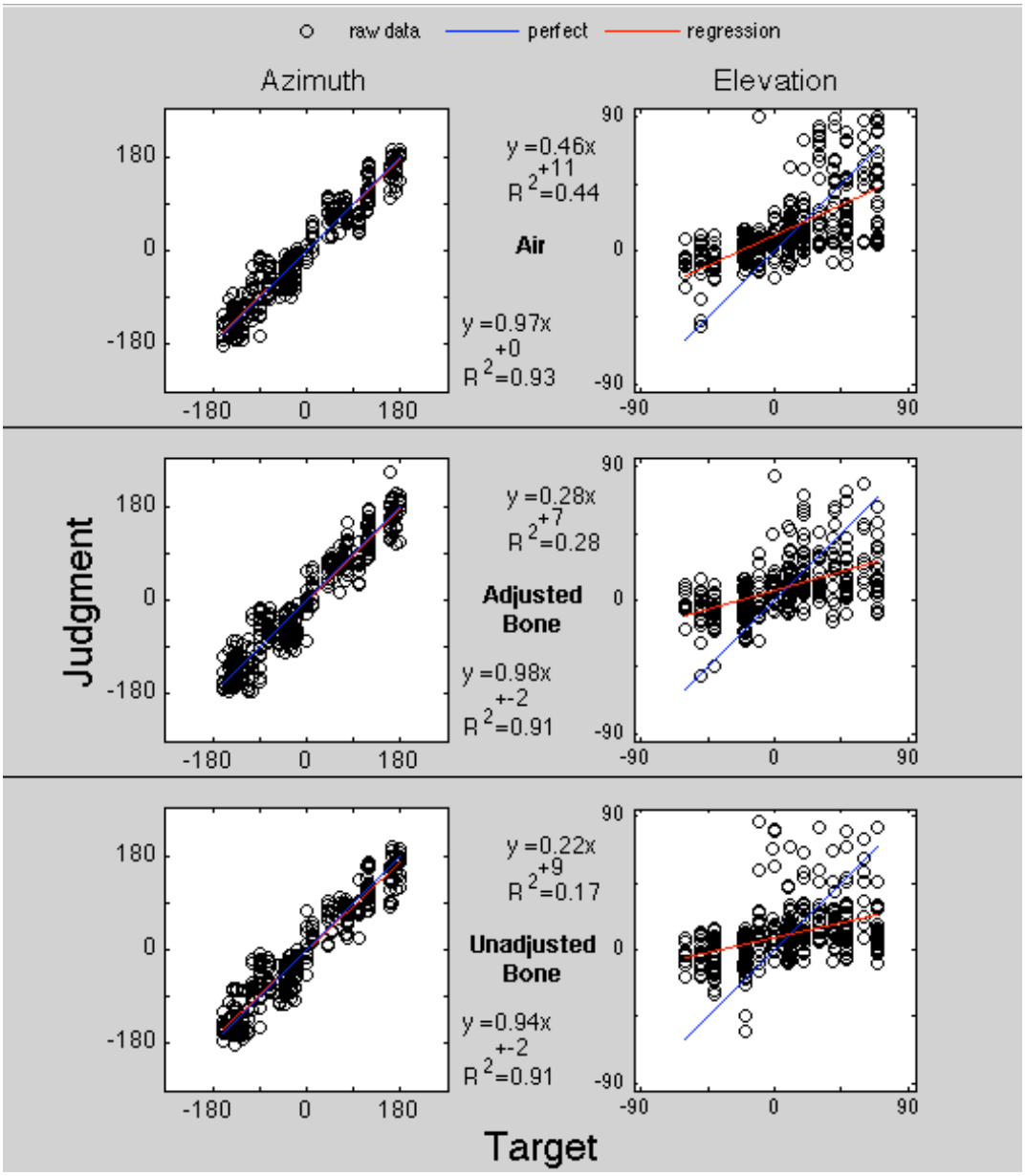

Figure 36. Scatter plots for participant 6, with front/back errors adjusted resolved and wrap-around artifact removed. For azimuth, the air condition baseline data fall on a steeply sloped regression line (*b*=.97) with a high fit to that regression line (R^2 = .93). Relative to the air condition, the regression for the adjusted bone shows similar slope and fit. The unadjusted bone condition then shows a reduced slope and similar fit. For elevation, the air condition data fall on a regression line with low slope (*b*=.46) and have a moderate fit to that line (R^2 = .44). When going from the air condition to the adjusted bone then unadjusted bone condition, the data points fall on a shallower sloped line and become more spread out around that line, as indicated by the monotonically decreasing regression slope and R^2 value.

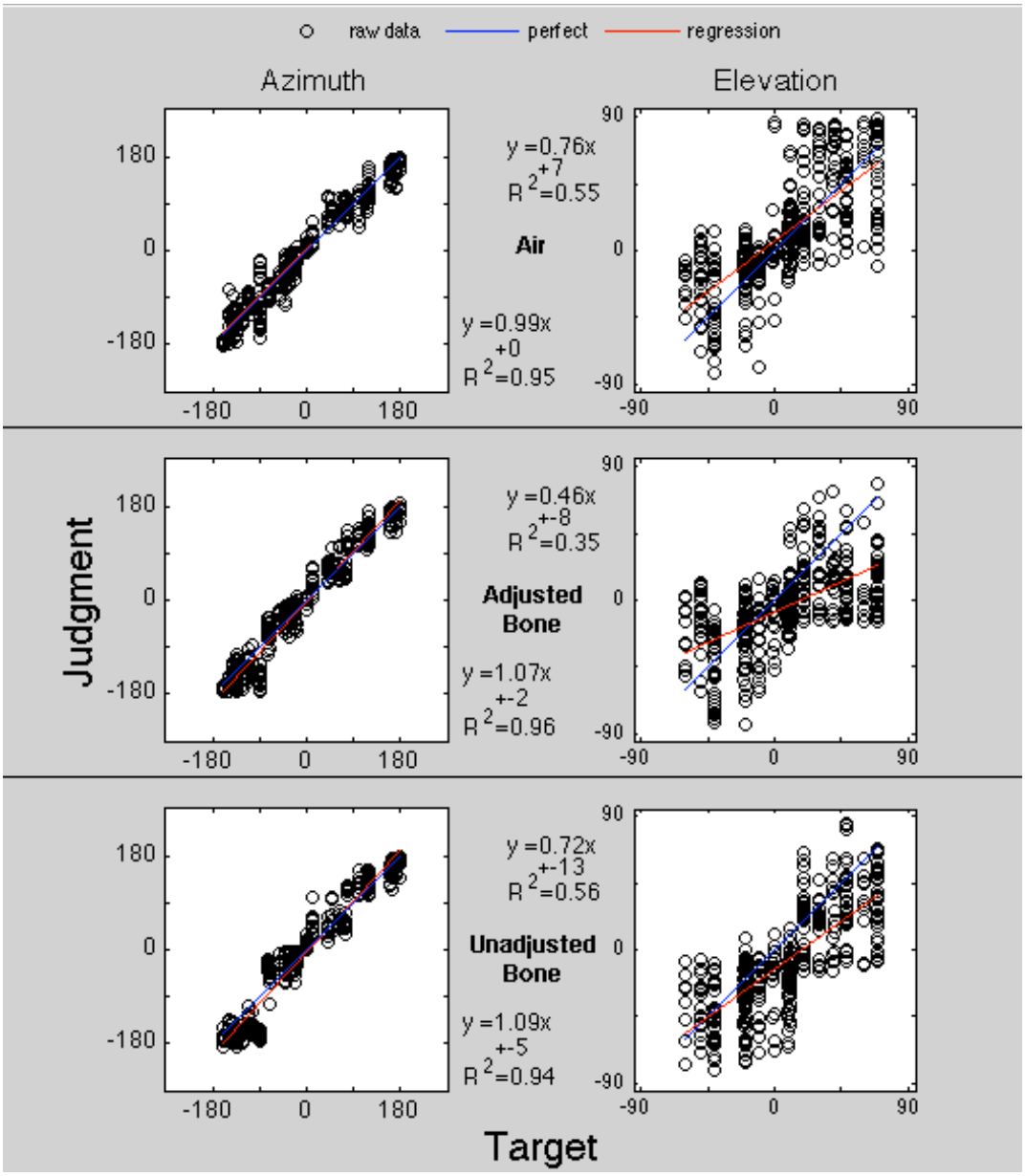

Figure 37. Scatter plots for participant 7, with front/back errors adjusted resolved and wrap-around artifact removed. For azimuth, the air condition baseline data fall on a steeply sloped regression line (*b*=.99) with a high fit to that regression line (R^2 = .95). Relative to the air condition, the regression for the adjusted bone shows a steeper slope similar slope and fit. The unadjusted bone condition then shows a similar slope and fit. For elevation, the air condition data fall on a regression line with moderate slope (*b*=.76) and have a high fit to that line (R^2 = .55). Relative to the air condition, the adjusted bone condition data fall on a shallower sloping regression line with a lower fit to that line. Then the unadjusted bone condition data fall on a steeper sloping regression line with a higher fit to that line.

General Trends and Inferential Treatment

Table 9 replicates the slope and R^2 values for both resolved azimuth and resolved elevation for every participant. The azimuth data and regression lines generally showed a steep slope and high fit relative to perfect performance, with little variation between participants. Within this small variation, the patterns of slope across conditions did differ between participants, but there were no consistent trends across participants. The fit to the regression line (R^2), however, did have one consistent trend: the fit was always best for the air-conduction condition (except for participant 7), and the two bone conditions had similar fit.

The elevation data and corresponding regression lines showed the same trend as the raw data for four participants (participants 2, 4, 5, and 6), with a reduction in regression line slope and fit to that regression line when going from air conduction to adjusted bone conduction to unadjusted bone conduction. For two other participants (participants 3 and 7), the air condition data fell on the steepest slope, followed by the unadjusted bone condition and then adjusted bone condition. For these participants, the fit to the regression line was similar for air and unadjusted bone conditions, but decreased for the adjusted bone condition. For one participant the air and adjusted bone condition fell on a similarly sloped line, and the unadjusted bone condition data fell on a shallower sloped line, with similar fit to those lines in all conditions.

Table 9

Regression Slope (b) and R^2 values for resolved azimuth and resolved elevation data.

Azimuth

	Slope (*b*)			R^2		
Participant	Air	Adjusted Bone	Unadjusted Bone	Air	Adjusted Bone	Unadjusted Bone
1	0.83	0.9	0.84	0.85	0.83	0.83
2	0.98	0.96	0.97	0.94	0.91	0.91
3	0.89	0.91	0.86	0.85	0.83	0.84
4	1.06	1.01	0.98	0.94	0.92	0.89
5	0.91	0.86	0.88	0.86	0.84	0.86
6	0.97	0.98	0.94	0.93	0.91	0.91
7	0.99	1.07	1.09	0.95	0.96	0.94
Average	0.95	0.96	0.94	0.90	0.89	0.88
Standard Deviation	0.08	0.07	0.09	0.05	0.05	0.04

Elevation

	Slope (*b*)			R^2		
Participant	Air	Adjusted Bone	Unadjusted Bone	Air	Adjusted Bone	Unadjusted Bone
1	0.29	0.3	0.25	0.22	0.2	0.19
2	0.91	0.56	0.4	0.6	0.38	0.31
3	0.7	0.54	0.66	0.35	0.26	0.37
4	1.18	0.91	0.65	0.63	0.49	0.32
5	0.97	0.6	0.57	0.53	0.31	0.27
6	0.46	0.28	0.22	0.44	0.28	0.17
7	0.76	0.46	0.72	0.55	0.35	0.56
Average	0.75	0.52	0.50	0.47	0.32	0.31
Standard Deviation	0.30	0.21	0.21	0.15	0.09	0.13

ANOVA tests were conducted on the resolved azimuth and elevation data to see if any patterns present in this sample could be generalized to the population. A one-way repeated measures ANOVA conducted on the azimuth slope values revealed that there

was no effect of condition on resolved azimuth regression slope ($F(2,12) = .592$, $p = .568$). A similar analysis was done on the pattern of fit for azimuth data to the regression line, indexed by Pearson's *r*, after undergoing the Fisher's *r* to *z'* transformation. A one-way repeated measures showed that there was an omnibus effect of condition on azimuth *z',* $F(2,12) = 6.56$, $p = .012$. Tukey's HSD posthoc analyses showed that the difference in azimuth *z'* between air conduction and adjusted bone conduction, as well as between air conduction and unadjusted bone conduction, was statistically significant. The difference between the adjusted bone conduction condition and unadjusted bone condition, however, was not statistically significant (see Table 10).

Table 10

Post-hoc tests for effect of condition on fit of azimuth data to regression line (indexed by Pearson's r, transformed to Fisher's z'). Critical q(3,12) = 3.77 at .05 significance level.

Comparison	*q*
Air – Adjusted Bone	4.07*
Air – Unadjusted Bone	4.62*
Adjusted Bone – Unadjusted Bone	1.46

*$p < .05$

A one-way repeated measures ANOVA conducted on the elevation slope values revealed that there was a significant omnibus effect of condition on resolved elevation regression slope ($F(2,12) = 8.83$, $p = .004$). Tukey's HSD posthoc analyses showed that the difference in elevation slope between air conduction and adjusted bone conduction, as

well as between air conduction and unadjusted bone conduction, was statistically significant. The difference between the adjusted bone conduction condition and unadjusted bone condition, however, was not statistically significant (see Table 11).

Table 11

Post-hoc tests for effect of condition on elevation slope. Critical q(3,12) = 3.77 at .05 significance level.

Comparison	*q*
Air – Adjusted Bone	6.46*
Air – Unadjusted Bone	4.36*
Adjusted Bone – Unadjusted Bone	0.58

*$p < .05$

A one-way repeated measures ANOVA conducted on the elevation *z'* fit values (transformed *r* values) revealed that there was a significant omnibus effect of condition on resolved elevation *z'*, $F(2,12) = 7.64$, $p = .007$. Tukey's HSD posthoc analyses showed that the difference in elevation slope between air conduction and adjusted bone conduction, as well as between air conduction and unadjusted bone conduction, was statistically significant. The difference between the adjusted bone conduction condition and unadjusted bone condition, however, was not statistically significant (see Table 12).

Table 12

Post-hoc tests for effect of condition on fit of elevation data to regression line (indexed by Pearson's r, transformed to Fisher's z'). Critical q(3,12) = 3.77 at .05 significance level.

Comparison	*q*
Air – Adjusted Bone	7.58*
Air – Unadjusted Bone	3.96*
Adjusted Bone – Unadjusted Bone	0.27

*p < .05

3.2.3. Summary Localization Performance Statistics

Reversals

In the determination of front/back reversals and up/down reversals, the number of reversals was counted to compute the percentage of reversals. Percentages were computed for each condition using the following equation:

$$100 * \frac{n_{reversals}}{n_{eligible}}$$

where $n_{reversals}$ is the number of reversals counted, and $n_{eligible}$ is the number of stimuli eligible for reversals (i.e., not including those within the exclusion range).

Conducting ANOVAs directly on proportion data presents problems because "the scale is bunched up at its extremes relative to its center", among other issues (Cohen & Cohen, 1983, p. 266). An arcsine transformation (2*arcsin($\sqrt{p}$)) makes the data suitable for ANOVA treatment by stretching out the tails of the distribution. All reversal proportions were given an arcsine transformation before ANOVAs were conducted.

Figure 38 shows the rate of front/back reversals for each condition and every participant, and Table 13 shows the numerators and denominators used to compute the percentages. There were not any apparent effects of condition on front/back reversal rate that were consistent across participants, and all values were close to chance performance. A one-way repeated measures ANOVA conducted on arcsine-transformed front/back reversals showed no significant main effect of condition on arcsine transformed front/back reversal rates ($F(2,12) = .660$, $p = .535$).

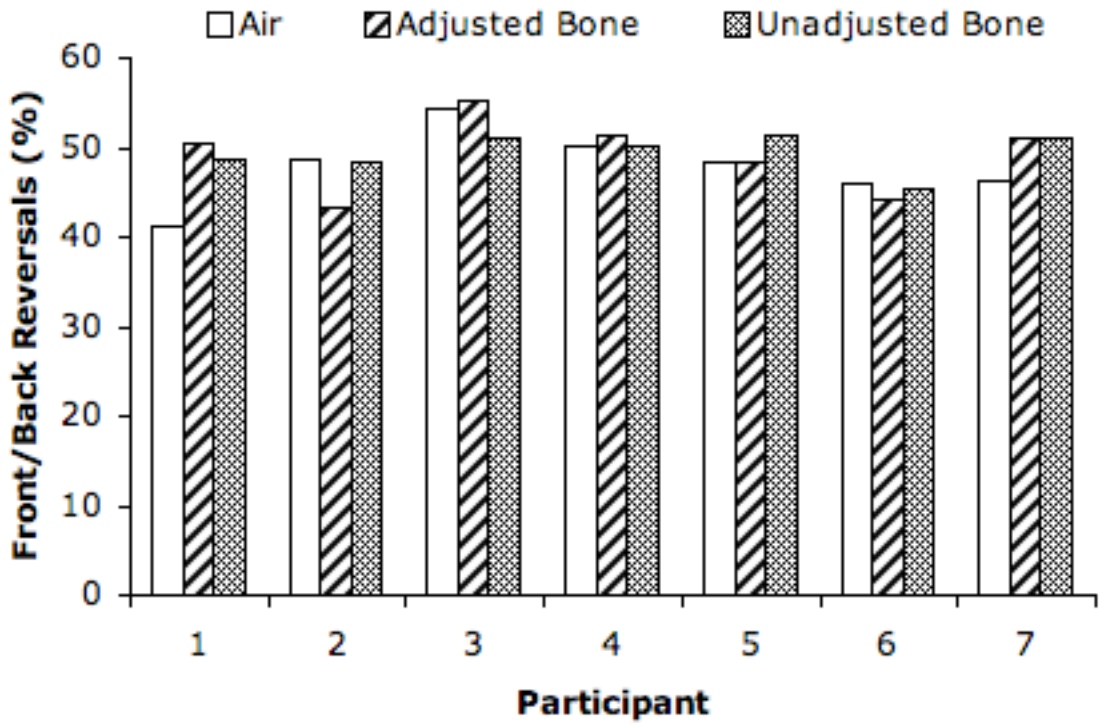

Figure 38. Front/back reversals for each participant and condition.

Table 13

Numerators and denominators used to compute front/back reversals. 360 trials were available in each condition to be eligible for reversals.

Participant	Air	Adjusted Bone	Unadjusted Bone
1	62/151	83/165	69/142
2	127/261	95/220	107/221
3	57/105	65/118	47/92
4	117/233	118/230	110/219
5	78/161	54/112	52/101
6	119/259	113/256	112/247
7	118/255	153/299	156/305

Figure 39 shows the rate of up/down reversals for each condition and every participant, and Table 14 shows the numerators and denominators used to compute the rates. For all but participant 5, up/down reversal rates were the lowest for air conduction, higher for adjusted bone conduction, and highest for unadjusted bone conduction. Participant 5 had similar up/down reversals for air conduction and adjusted bone conduction. For some participants (1,2,3), the reduction in up/down reversals in the adjusted bone condition relative to the unadjusted bone condition was very slight, and there was a large increase in errors in the bone conduction conditions relative to the air conduction condition. For other participants (4 and 5), the reduction in reversals in the adjusted bone condition relative to the unadjusted bone condition was quite large, and there was only a small increase in reversals for the adjusted bone condition relative to the air condition.

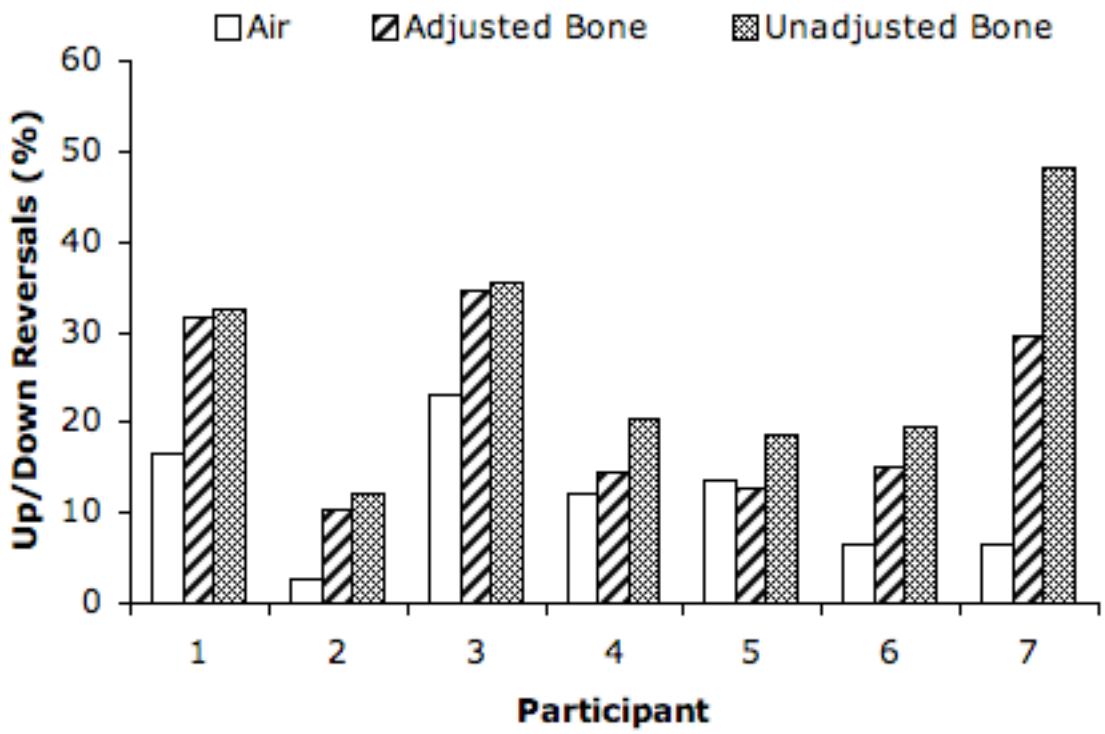

Figure 39. Up/down reversals for each participant and condition.

Table 14

Numerators and denominators used to compute up/down reversals. 360 trials were available in each condition to be eligible for reversals.

Participant	Air	Adjusted Bone	Unadjusted Bone
1	17/102	37/117	32/98
2	5/189	14/137	13/107
3	35/152	50/144	59/167
4	24/199	26/178	29/143
5	23/168	16/125	22/118
6	7/106	13/87	15/77
7	11/171	41/139	94/195

A one-way repeated measures ANOVA conducted on arcsine-transformed up/down reversal rates showed a significant main effect of condition on arcsine

transformed up/down reversal rates ($F(2,12) = 12.11$, $p = .001$). Tukey's HSD posthoc analyses showed statistically significant differences between all comparisons: air conduction - adjusted bone conduction, air conduction - unadjusted bone conduction, and adjusted bone conduction - unadjusted bone conduction (see Table 15).

Table 15

Post-hoc tests for effect of condition on arcsine transformed up/down reversals. Critical q(3,12) = 3.77 at .05 significance level.

Comparison	*q*
Air – Adjusted Bone	-4.78*
Air – Unadjusted Bone	-5.18*
Adjusted Bone – Unadjusted Bone	-3.83*

*p < .05

Error

An array of absolute azimuth error values for each condition was created by finding the absolute difference between each stimulus azimuth and response azimuth. The mean and standard error of each array was then calculated. The elevation error was computed in an analogous manner.

Figure 40 shows that for all participants the mean azimuth error was more for the bone conditions than for the air conditions. The pattern of results between the two bone-conduction conditions was highly variable across participants: Some had similar error for both (1,2,3,6), others had more for the unadjusted bone condition (4,7), and one participant had more for the unadjusted bone condition (5).

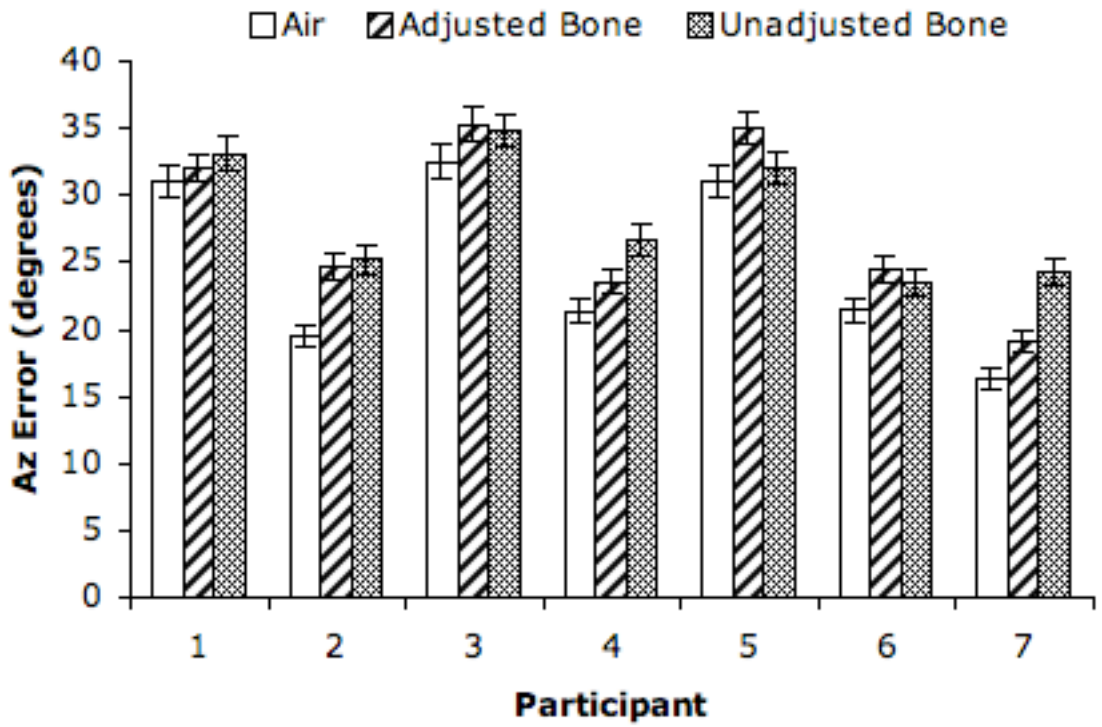

Figure 40. Azimuth error (degrees) for each participant and condition. Error bars reflect one standard error above and below the mean.

A one-way repeated measures ANOVA conducted on azimuth error showed a significant main effect of condition ($F(2,12) = 10.54$, $p = .002$). Tukey's HSD posthoc analyses showed statistically significant differences for the air conduction - adjusted bone and air - unadjusted bone comparisons, but not the adjusted bone - unadjusted bone comparison (see Table 16).

Table 16

Post-hoc tests for effect of condition on azimuth error. Critical q(3,12) = 3.77 at .05 significance level.

Comparison	*q*
Air – Adjusted Bone	-8.52*
Air – Unadjusted Bone	-5.53*
Adjusted Bone – Unadjusted Bone	-1.07

*$p < .05$

Figure 41 shows the average elevation error for each condition and every participant. There was not any apparent effect of condition that is consistent across participants. A one-way repeated measures ANOVA conducted on arcsine-transformed front/back reversals showed no significant main effect of condition on arcsine transformed front/back reversal rates ($F(2,12) = 2.68$, $p = .109$).

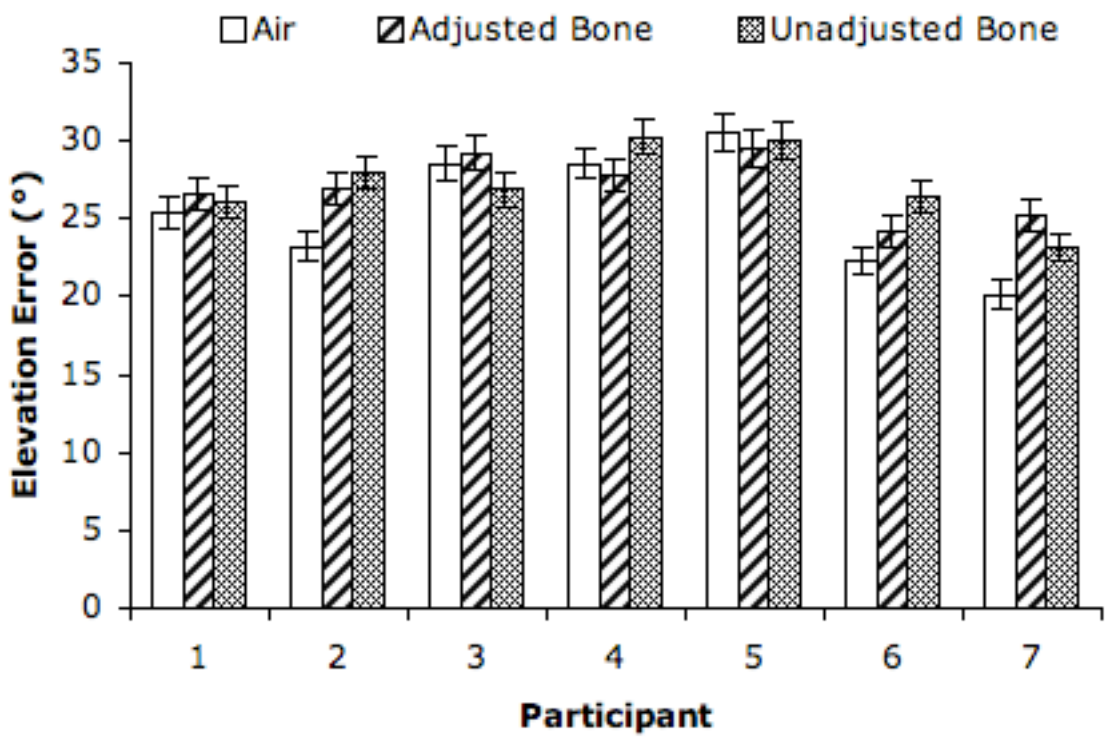

Figure 41. Elevation error (degrees) for each participant and condition. Error bars reflect one standard error above and below the mean.

Bias

A metric for bias was computed for each plane of judgment (azimuth and elevation) to understand any tendencies to respond more or less lateral than the stimulus, and more or less elevated than the stimulus. To compute the signed lateralization error, first the angular distance between the stimulus and the nearest sagittal plane azimuth (0 degrees for front, 180 degrees for back) was determined. Then this was repeated for the response. The signed lateralization error value was computed by subtracting each response distance from each stimulus distance, giving a difference in distance from the closest sagittal plane. This direction of subtraction creates a negative value for responses that are less lateralized than the stimulus, and a positive value for responses that are more lateralized than the stimulus, regardless of whether the stimulus and response are in the

same hemisphere (recall that trials with stimuli or responses within 15 degrees were not eligible for reversal resolution). The signed lateralization error was computed for every stimulus-response pair, to create an array of values for each condition. The mean and standard error of each array was then computed.

Figure 42 shows that except for participant 7, the participants had signed lateralization error in the unadjusted bone condition that was similar to or higher than (more lateralized) the adjusted bone condition, and the air condition signed lateralization error was typically lower than in the bone conditions. Irrespective of the trend, nearly all signed lateralization error values were positive except for participant 7, and thus there was a general bias to make responses more lateral than the stimulus. Participant 7 showed an opposite trend of the others' general trend: the air condition had the lowest absolute signed lateralization error, the error increased in the negative direction for the adjusted bone condition, and increased further in the negative direction for the unadjusted bone condition. Thus, participant 7's responses were less lateralized than the stimulus for the bone-conduction conditions, and the bias was strongest for the unadjusted bone condition.

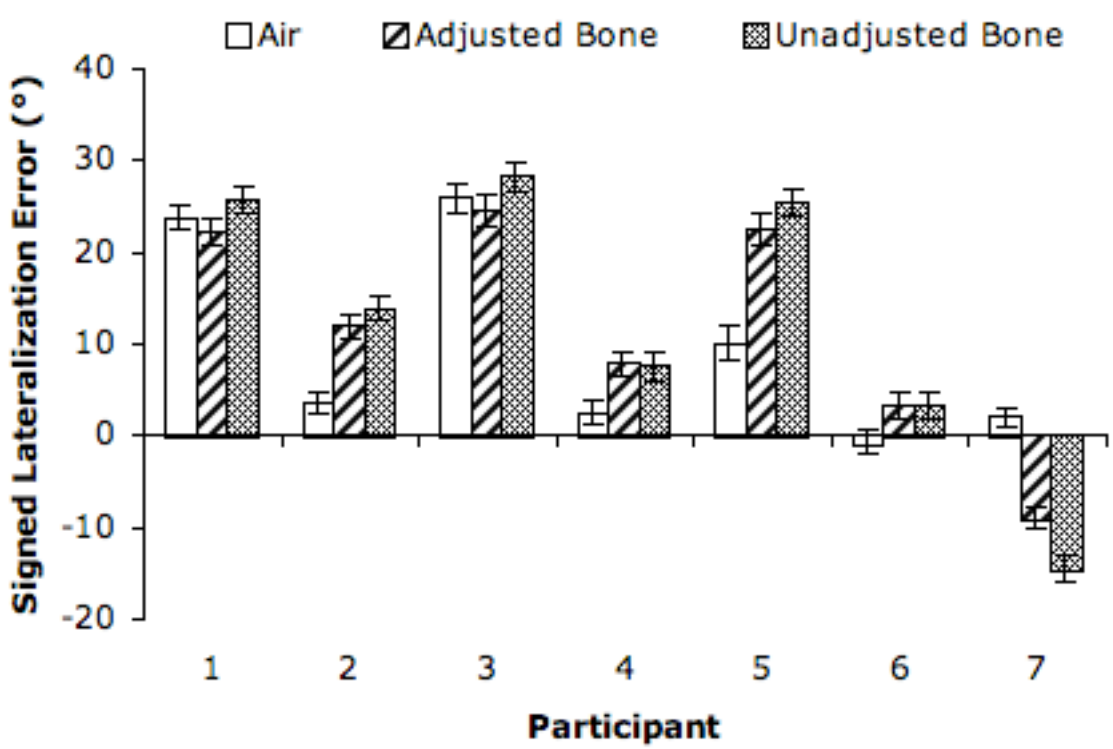

Figure 42. Signed lateralization error (degrees) for each participant and condition. Error bars reflect one standard error above and below the mean.

A one-way repeated measures ANOVA conducted on signed lateralization error showed no effect of condition ($F(2,12) = .71$, $p = .444$). Because of the categorically different trend for participant 7, the ANOVA was run again with this participant excluded. This analysis showed a statistically significant main effect of condition ($F(2,10) = 6.71$, $p = .014$) on signed lateralization error. Tukey's HSD posthoc analyses showed statistically significant differences for the air conduction - unadjusted bone and adjusted bone - unadjusted bone comparisons, but not the air - adjusted bone comparison (see Table 17).

Table 17

Post-hoc tests for effect of condition on signed lateralization error, with participant 7 excluded. Critical q(3,10) = 3.88 at .05 significance level.

Comparison	*q*
Air – Adjusted Bone	-2.90
Air – Unadjusted Bone	-4.32*
Adjusted Bone – Unadjusted Bone	-3.89*

* $p < .05$

The signed elevation error was computed by subtracting the stimulus elevation from response elevation. This creates an error signed positive if the response is biased upward from the stimulus, and an error signed negative of the response is biased downward from the stimulus, regardless of the hemisphere. The signed elevation error calculation was done for every stimulus-response pair, to create an array of values for each condition. Then the mean and standard deviation of this array was computed.

Figure 43 shows that the air conduction condition had higher signed elevation error than the bone conditions for all participants. For most participants (2,3,4,5,7), the signed elevation error was the highest for air conduction, lower for adjusted bone conduction, and lowest for unadjusted bone conduction. This shows a relative downward bias, where the response is consistently lower than the stimulus. For participants 1 and 6, the unadjusted bone signed elevation error was less than the adjusted bone elevation error.

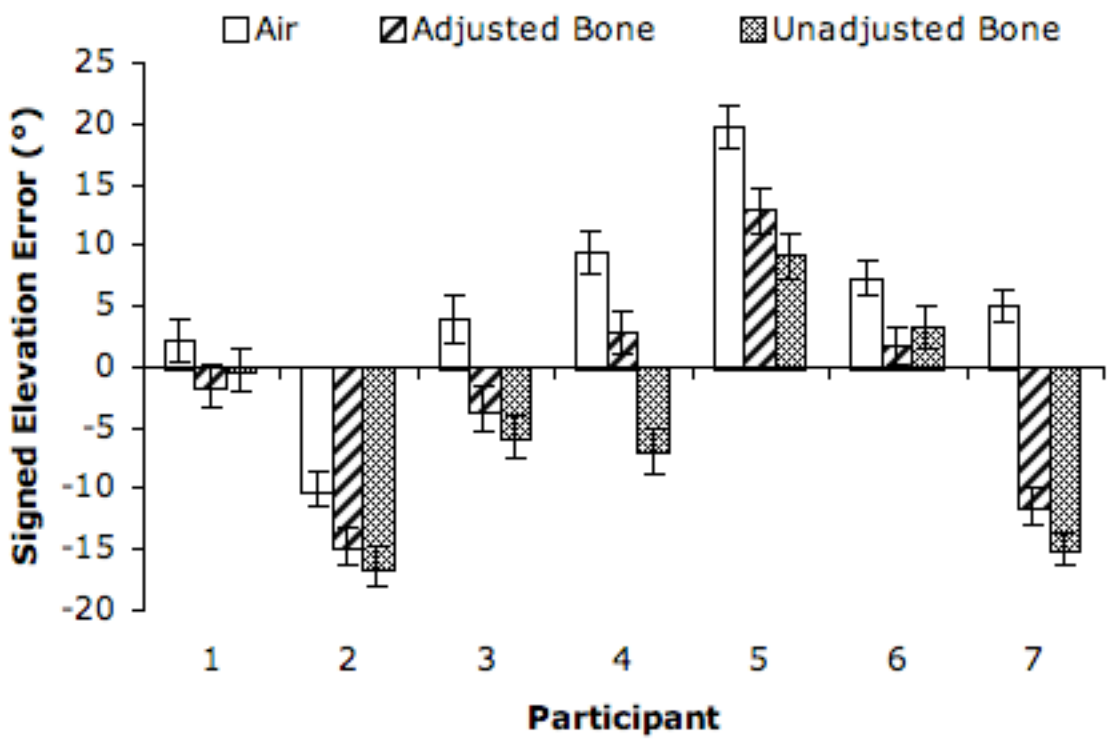

Figure 43. Signed elevation error (degrees) for each participant and condition. Error bars reflect one standard error above and below the mean.

A one-way repeated measures ANOVA conducted on signed elevation error showed a statistically significant main effect of condition ($F(2,12) = 15.01$, $p = .001$). Tukey's HSD posthoc analyses showed statistically significant differences for the air conduction - adjusted bone and air - unadjusted bone comparisons, but not the adjusted bone - unadjusted bone comparison (see Table 18).

Table 18

Post-hoc tests for effect of condition on signed elevation error. Critical q(3,12) = 3.77 at .05 significance level.

Comparison	*q*
Air – Adjusted Bone	6.48*
Air – Unadjusted Bone	5.77*
Adjusted Bone – Unadjusted Bone	2.50

* $p < .05$

Variability

The azimuth standard deviation was computed by taking the standard deviation of azimuth responses for a given stimulus location, across trials at that location, within a condition. The azimuth standard deviation was computed for every stimulus location, to create an array of values for each condition. Then the mean and standard error of each array was computed. The elevation standard deviation was computed with elevation responses in an analogous manner.

Figure 44 shows the average azimuth standard deviation for each condition and every participant. Given the large error bars, there is no apparent effect of condition that is consistent across participants. A one-way repeated measures ANOVA conducted on average azimuth standard deviation showed no significant main effect of condition ($F(2,12) = 2.71$, $p = .107$).

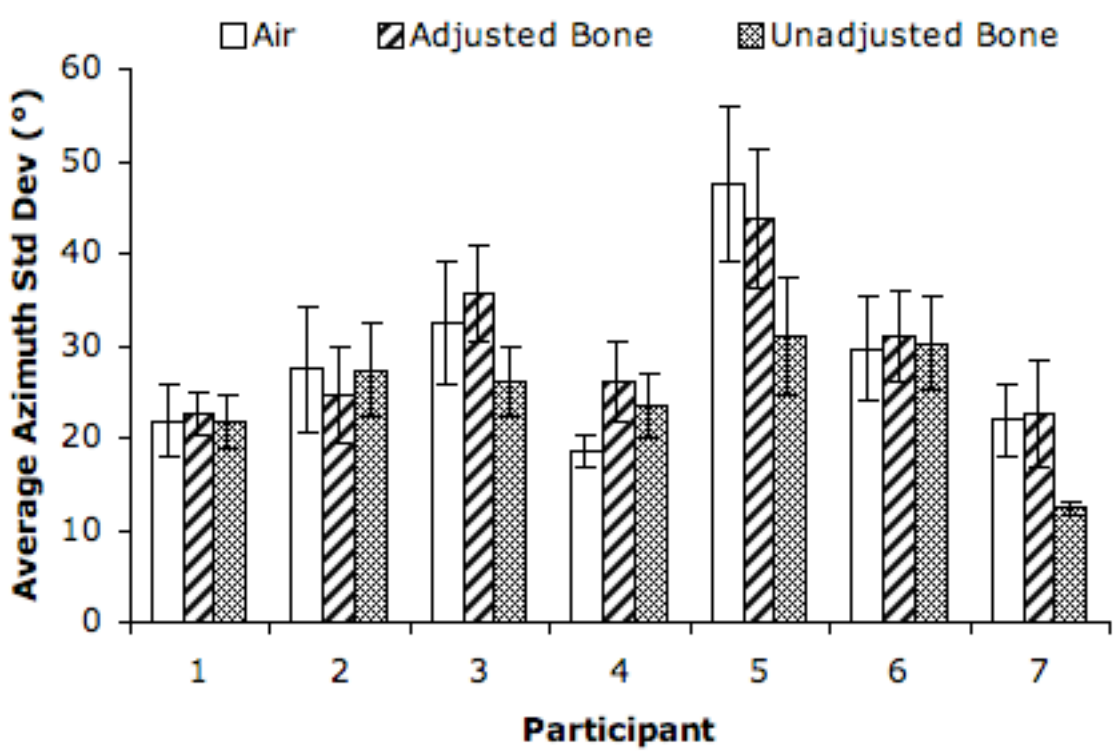

Figure 44. Average azimuth standard deviation for each participant and condition. Error bars reflect one standard error above and below the mean.

Figure 45 shows that there is no apparent effect of condition on elevation standard deviation that is consistent across participants. A one-way repeated measures ANOVA conducted on average elevation standard deviation showed no significant main effect of condition ($F(2,12) = 0.62$, $p = .554$).

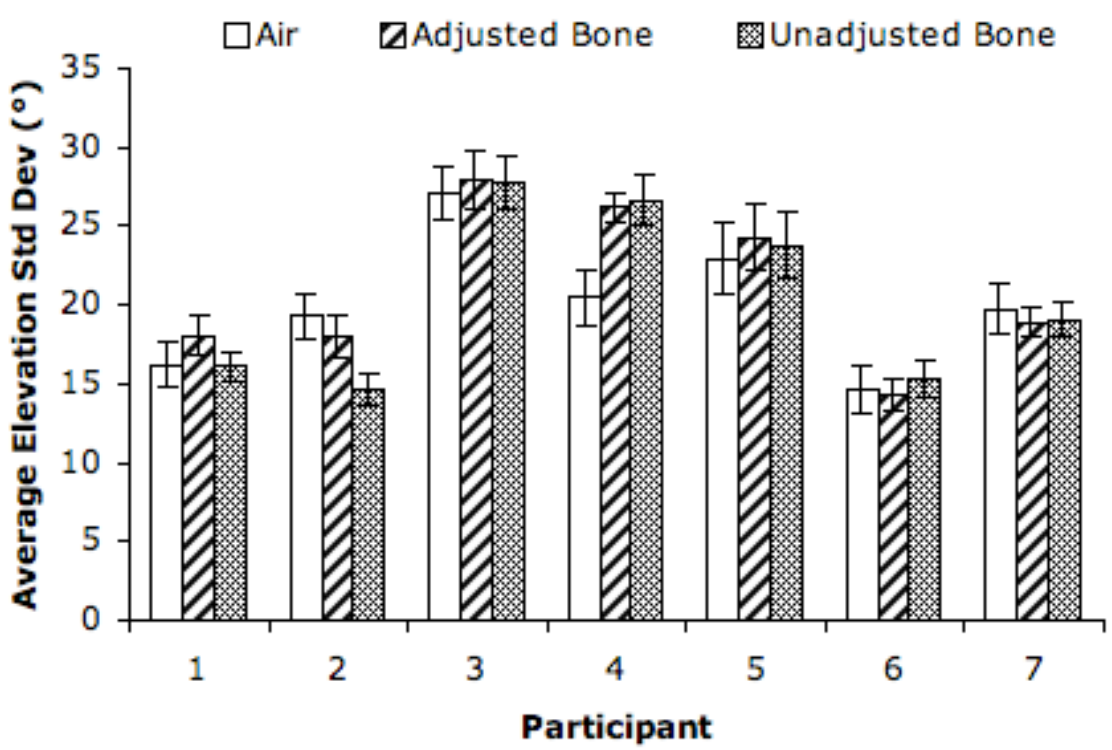

Figure 45. Average elevation standard deviation for each participant and condition. Error bars reflect one standard error above and below the mean.

Maximum Judgments

The maximum lateralization statistic was the absolute value of the most lateral azimuth judgment given across all trials within a condition. Figure 46 shows that there was virtually no variation in the maximum azimuth judgment between conditions. The maximum lateralization values for all participants were exactly or close to 90 degrees, the maximum lateralized stimulus. Note that the possible maximum lateralization scale ranges from zero to 180 degrees, because of the azimuth coordinate system (left versus right hemisphere was ignored). Thus, any value below or above 90 degrees is less than maximally lateralized, and values greater than 90 degrees decrease in lateralization as their distance from 90 degrees increases.

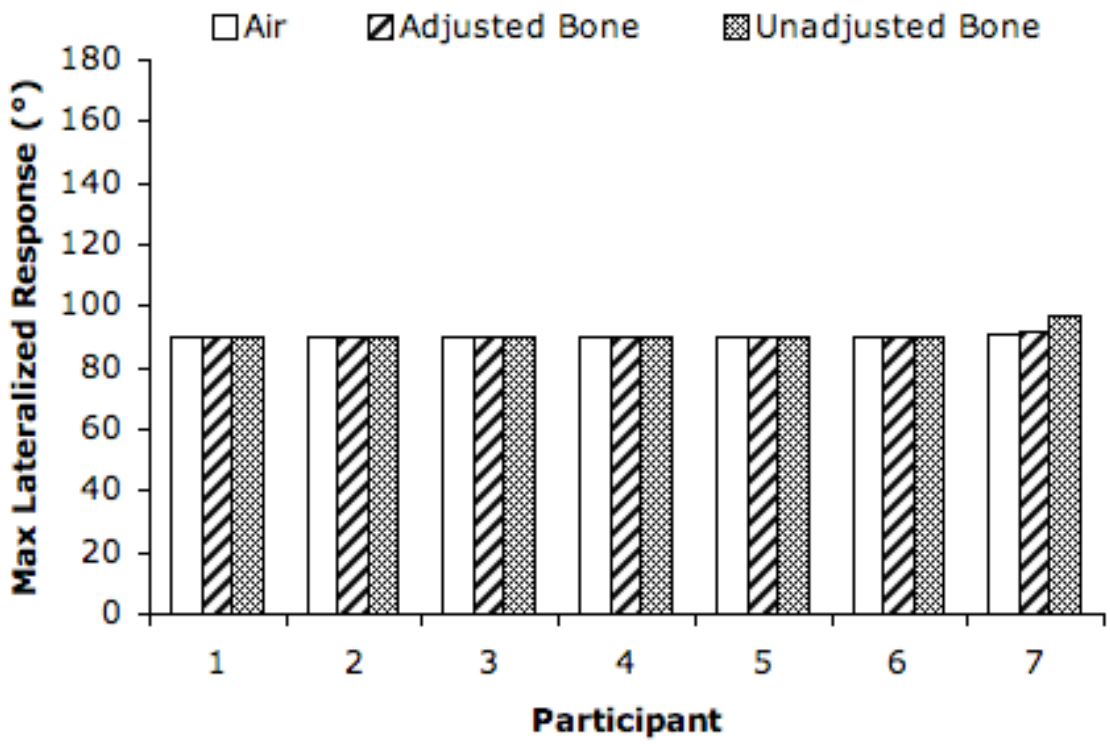

Figure 46. Maximum lateralized azimuth response for each participant and condition.

The maximum elevation was the highest elevation response across all trials within a condition. Figure 47 shows the maximum elevation for each condition and every participant. The maximum elevation is similar across conditions for three participants (2,3,5), and shows a decreasing trend going from air to adjusted bone to unadjusted bone for three participants (1,4,7). Participant 6 shows a similar decrease in maximum elevation for the bone conditions, but with similar maximum elevation for the two bone conditions. There seems to be the most variability in maximum elevation across participants for the unadjusted bone condition.

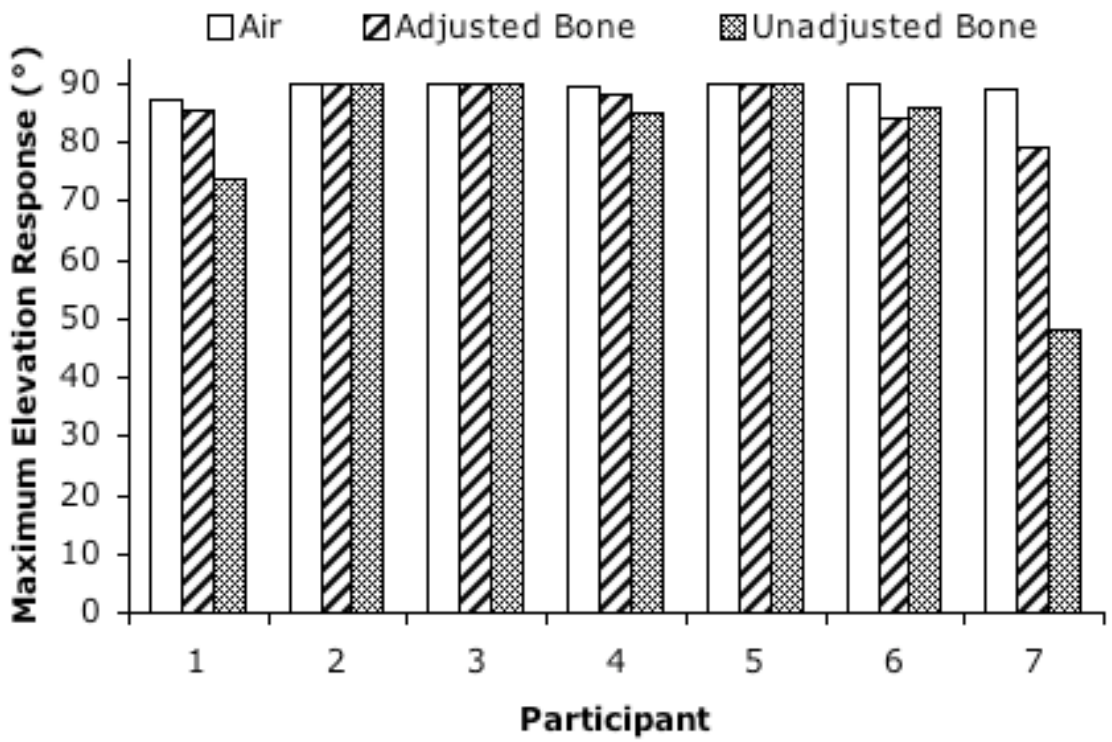

Figure 47. Maximum elevation response for each participant and condition.

A one-way repeated measures ANOVA was conducted on maximum elevation, and Mauchley's test of sphericity showed that equal variances could not be assumed ($W(2) = .071$, $p = .001$). Geisser-Greenhouse corrections were made to the degrees of freedom to correct for the positive alpha bias that occurs with unequal variances. With this correction the ANOVA showed that there was not a statistically significant effect of an omnibus effect of condition ($F(1.04,6.22) = 2.41$, $p = .17$). Although this is a null effect, this is one statistic where the population is likely to be saturated around 90 degrees, and thus not be normally distributed.

The minimum elevation was the lowest elevation response across all trials within a condition. Figure 48 shows that there were not any apparent effects of condition on minimum elevation that was consistent across participants. A one-way repeated measures

ANOVA was conducted on minimum elevation, and Mauchley's test of sphericity showed that equal variances could not be assumed ($W(2) = .293$, $p = .047$). Geisser-Greenhouse corrections were made to the degrees of freedom to correct for the positive alpha bias that occurs with unequal variances. With this correction the ANOVA showed that there was not a statistically significant effect of an omnibus effect of condition $F(1.17,7.03) = 0.88$, $p = .398$. This is again a situation where the population's statistic is likely to be saturated around 90 degrees, and thus not be normally distributed – although again this is a null effect.

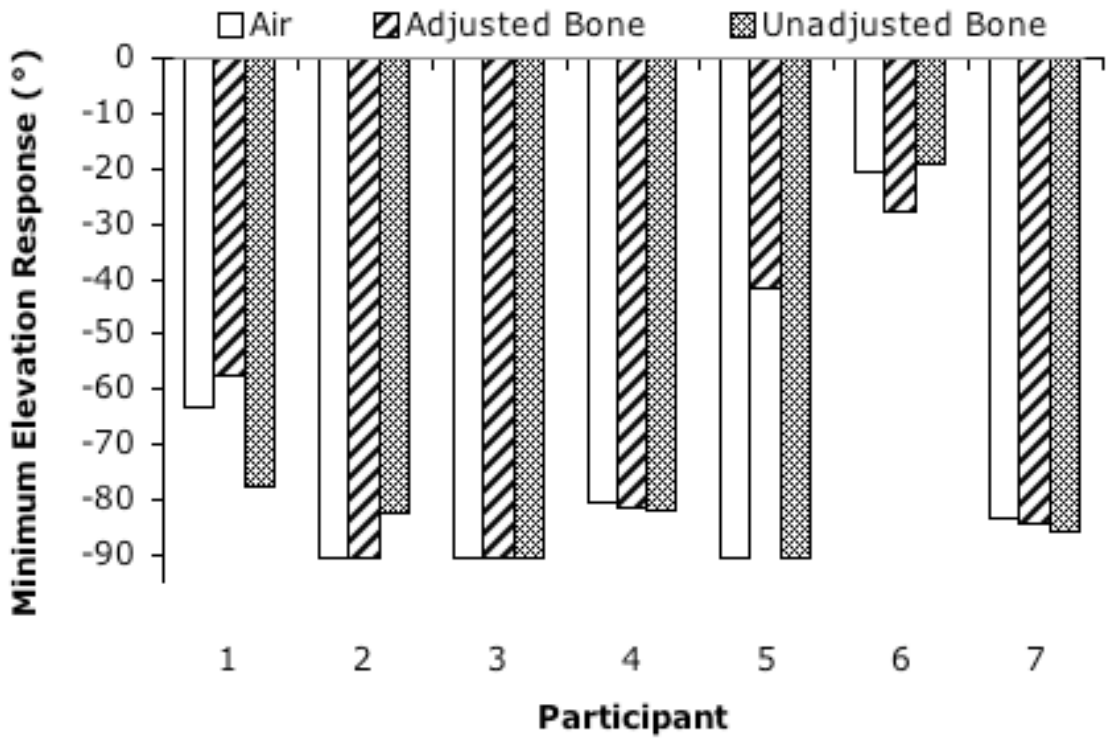

Figure 48. Minimum elevation response for each participant and condition.

3.2.4. Subjective Ratings

The subjective ratings were computed by recording the horizontal pixel position of the cursor selected on the continuous rating scale, and subtracting from it the horizontal pixel position at the minimum judgment. This created a scale that ranges between 0 (the left-most position on the line) and 464 (the right-most position on the line).

Figure 49 shows that there were not any apparent effects of condition on externalization rating that was consistent across participants. A one-way repeated measures ANOVA was conducted on externalization rating, and Mauchley's test of sphericity showed that equal variances could not be assumed ($W(2) = .148$, $p = .008$). Geisser-Greenhouse corrections were made to the degrees of freedom to correct for the positive alpha bias that occurs with unequal variances. With this correction the ANOVA showed that there was not a statistically significant effect of condition $F(1.08,6.48) = 1.09$, $p = .340$.

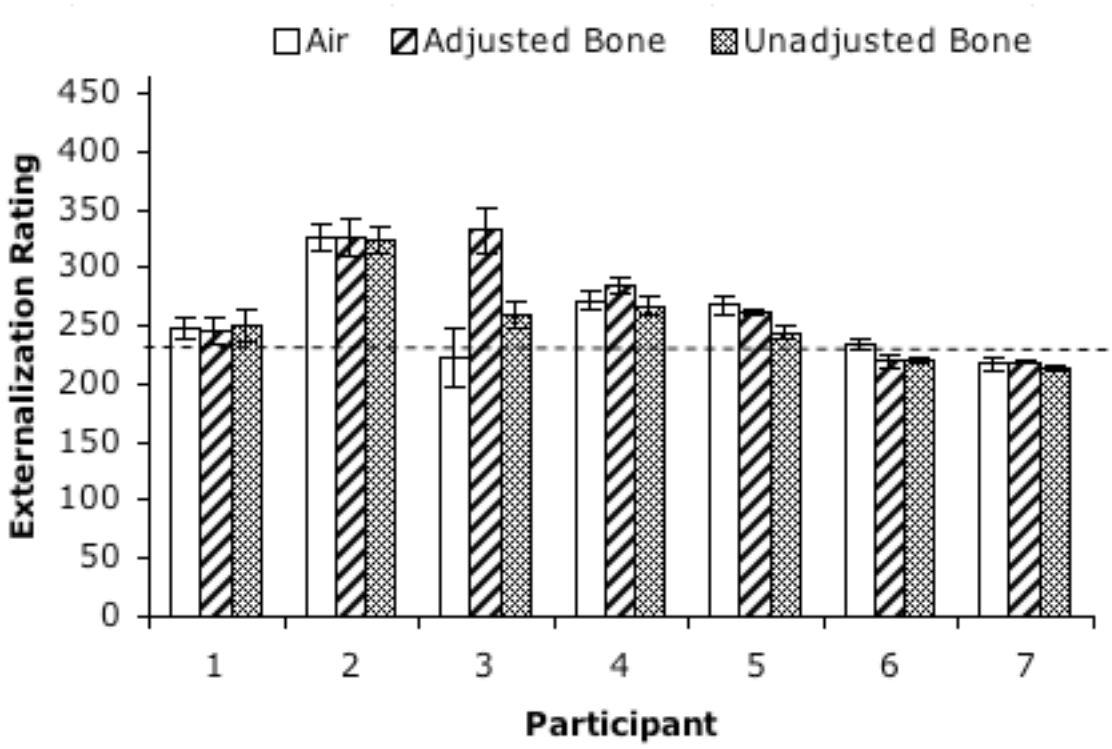

Figure 49. Subjective externalization rating for each participant and condition. Dotted line shows middle of response scale (labeled "surface of your head"). Error bars reflect one standard error above and below the mean.

Figure 50 shows that there were not any apparent effects of condition on externalization rating that was consistent across participants. A one-way repeated measures ANOVA conducted on diffuse rating showed that there was not a statistically significant effect of condition $F(2,12) = 1.08$, $p = .371$.

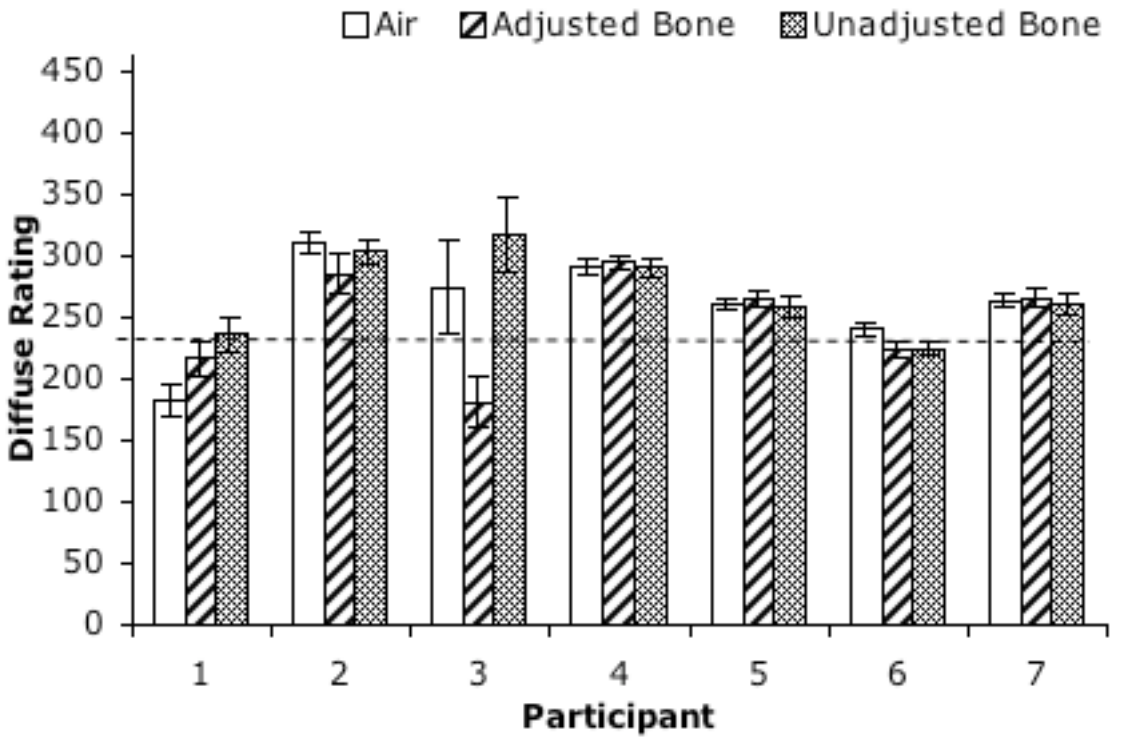

Figure 50. Subjective diffuse rating for each participant and condition. Dotted line shows (unlabeled) middle of response scale. Error bars reflect one standard error above and below the mean.

3.2.5. Individualized HRTFs

Scatter Plot

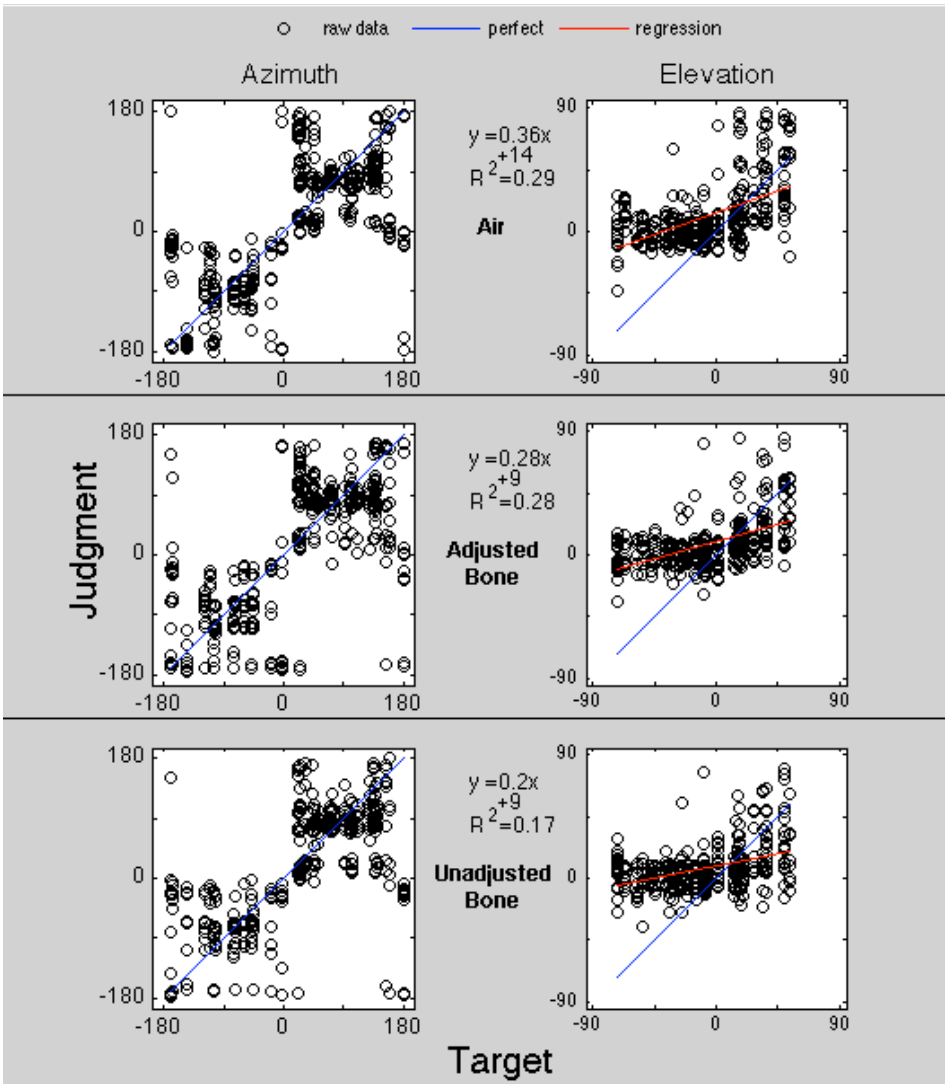

Figure 51. Scatter plot for P6, with individualized HRTFs. For azimuth, the air condition baseline shows a strong fit to the diagonal, besides the large amount of front/back reversals on the cross-diagonal. This pattern seemed to have more points fall closer to the diagonal than the pattern seen with the generalized HRTFs (see Figure 22). For the adjusted and unadjusted bone azimuth, the response patterns are similar but get progressively more variable, with a binary response pattern starting to appear in the unadjusted bone condition. This progressive decline in performance is more severe than the pattern seen with generalized HRTFs. For elevation, the air condition baseline data fall on a steeply sloped diagonal in the top hemi-field, but few stimuli are perceived in the lower hemi-field (just as with the generalized HRTFs). This pattern is echoed in the slope of the regression line (.36) and fit of the data to that line (R^2 = .28) – both of these values are reduced from the slope and R^2 values with generalized HRTFs (.41 and .37, respectively). When going from the air condition to the adjusted bone then unadjusted bone condition, the data points fall on a slightly shallower sloped line and become more spread out around that line, as indicated by the monotonically decreasing regression slope and R^2 value – following the trend of the generalized HRTFs (although the individualized trend is not as strong).

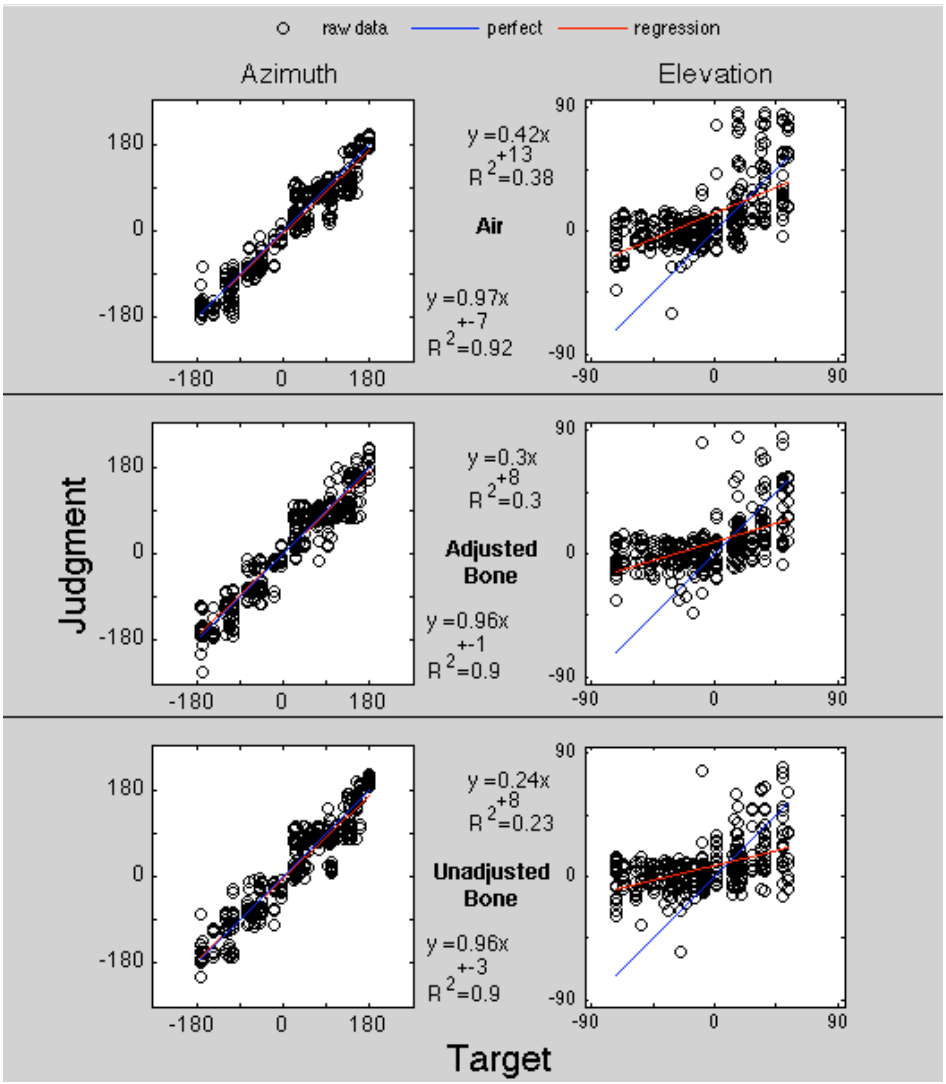

Figure 52. Scatter plot for P6, using individualized HRTFs, with front/back errors adjusted resolved and wrap-around artifact removed. For azimuth, the air condition baseline data fall on a steeply sloped regression line (*b*=.97) with a high fit to that regression line (R^2 = .92), similar to the results seen with the generalized HRTFs. Relative to the air condition, the regression for the adjusted bone and unadjusted bone shows similar slope and fit. This departs from the generalized results, where the unadjusted bone condition showed a slight decline in slope. For elevation, the air condition data fall on a regression line with low slope (*b*=.42) and have a moderate fit to that line (R^2 = .38). These values are slightly lower than those seen with the generalized HRTFs. When going from the air condition to the adjusted bone then unadjusted bone condition, the data points fall on a shallower sloped line and become more spread out around that line, as indicated by the monotonically decreasing regression slope and R^2 value. This is the same trend seen with the generalized HRTFs.

Summary Localization Performance Statistics

Descriptive statistics that showed trends with the generalized data, show a trend for the individualized, or show a difference between individualized and generalized HRTFs, will be plotted and discussed here. Other descriptive statistics will be provided in Table 19 at the end of this section. The individualized results were generally similar to the generalized results, but the subtle differences will be discussed here.

Figure 53 shows the front/back reversals for participant 6, when using individualized and generalized HRTFs. With the individualized HRTFs, the adjusted bone reversal rate was higher than the air and unadjusted bone reversal rates, which were similar. With the generalized HRTFs, however, the rates were quite similar across conditions. In addition, the air condition (baseline) reversal rate was slightly elevated for with the generalized HRTFs, compared to the air condition with the individualized HRTFs.

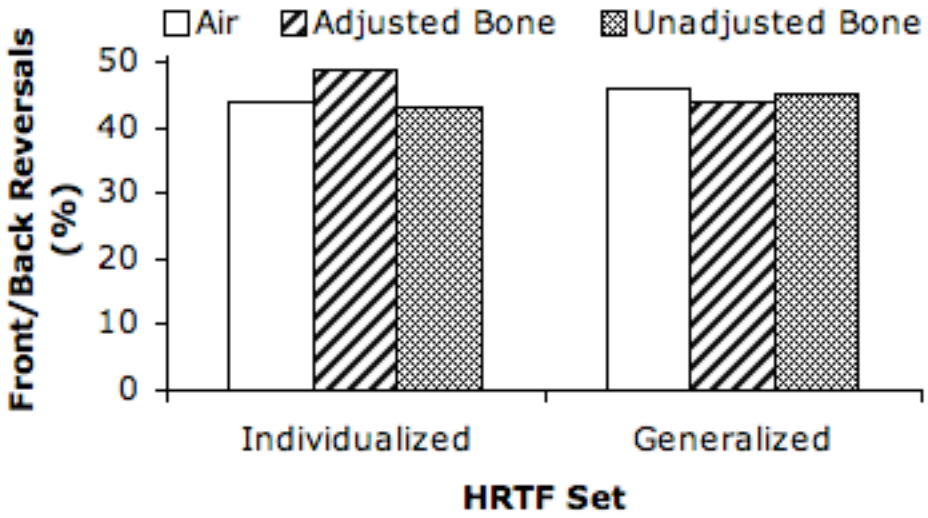

Figure 53. Front/back reversals for P6, with individualized and generalized HRTFs.

Figure 54 shows the up/down reversals for participant 6, when using individualized and generalized HRTFs. Although the generalized results show an increasing up/down reversal trend going from the air to adjusted bone to unadjusted bone condition, the individualized results show similar up/down reversal values for air and unadjusted bone condition, and less for the adjusted bone condition. The air condition (baseline) reversal percentage was higher with the individualized HRTFs than with the generalized HRTFs.

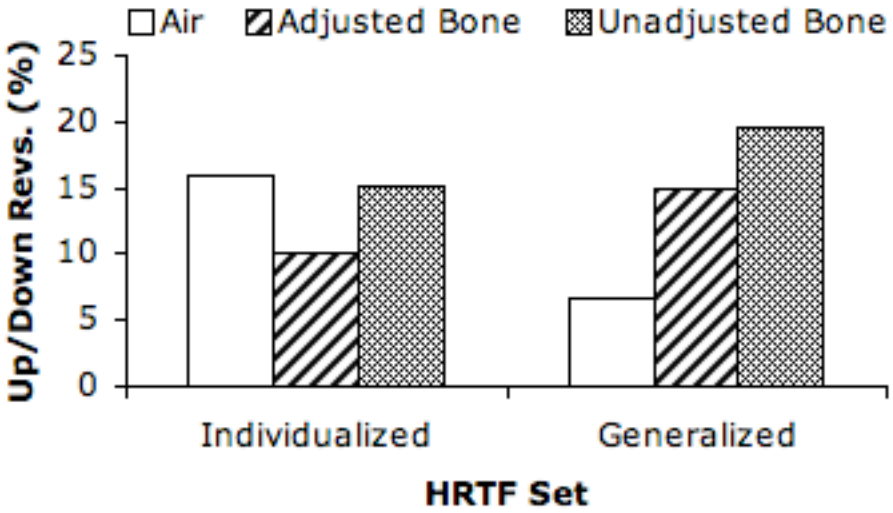

Figure 54. Up/down reversals for P6, with individualized and generalized HRTFs.

Figure 55 shows that with both the individualized and generalized HRTFs, the air condition had a reduced azimuth error relative to the two bone conditions, which had similar azimuth error. With individualized HRTFs, there was a slightly increased azimuth error in the air condition relative to the air condition azimuth error with the generalized HRTFs.

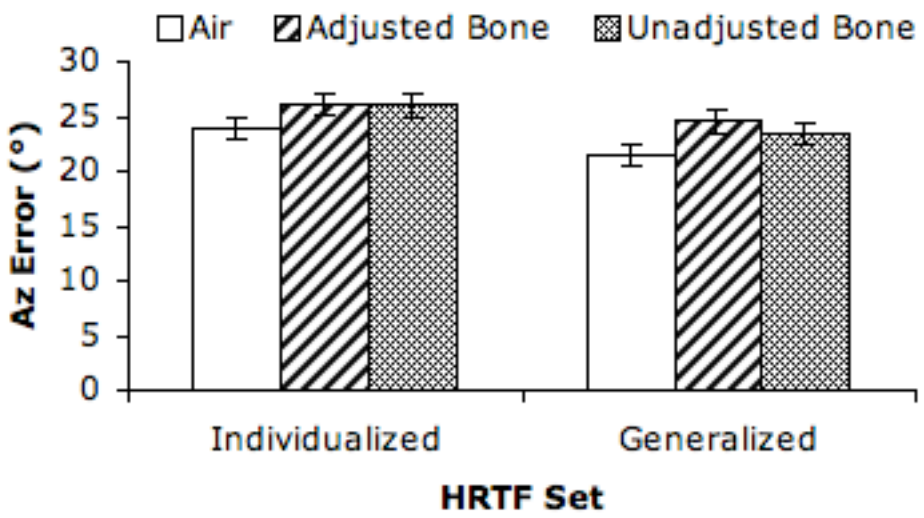

Figure 55. Azimuth error for P6, with individualized and generalized HRTFs. Error bars reflect one standard error above and below the mean.

Figure 56 shows that with the generalized HRTFs, the elevation error showed an increasing trend going from the air to adjusted bone to unadjusted bone condition, but the individualized results show similar elevation error across conditions. The elevation error values were in similar ranges for individualized and generalized HRTFs.

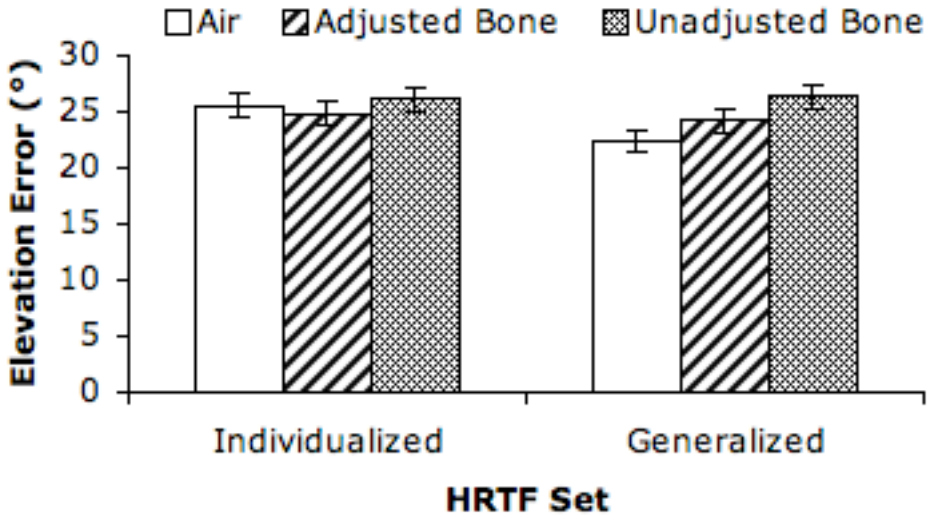

Figure 56. Elevation error for P6, with individualized and generalized HRTFs. Error bars reflect one standard error above and below the mean.

Figure 57 shows that for both HRTF sets, the adjusted bone and unadjusted bone condition had similar signed elevation errors, and the air condition had increased signed elevation error. Regardless of condition, the signed elevation error was always higher for the individualized HRTFs than for the generalized HRTFs.

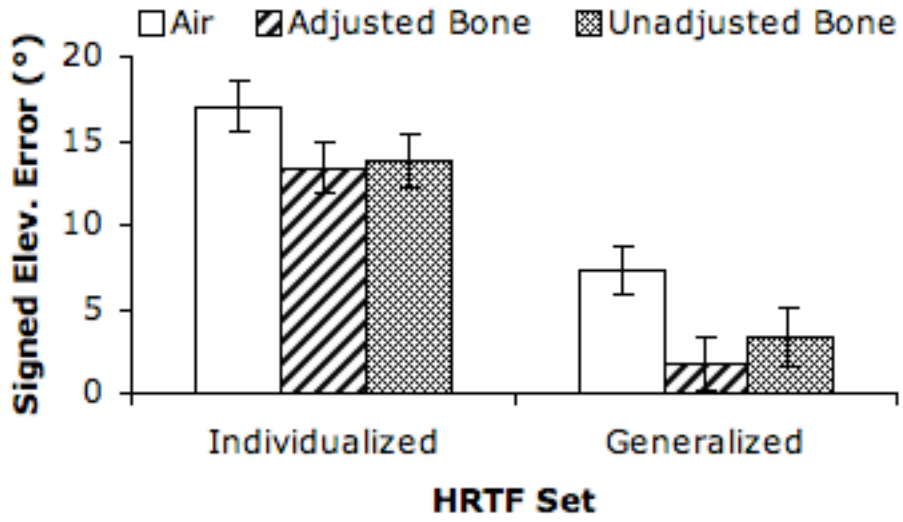

Figure 57. Signed elevation error for P6, with individualized and generalized HRTFs. Error bars reflect one standard error above and below the mean.

Figure 58 shows that signed lateralization error was relatively similar across conditions for the individualized HRTF, but that the signed lateralization error was considerably lower in the air condition, compared to the bone conditions for the generalized HRTF.

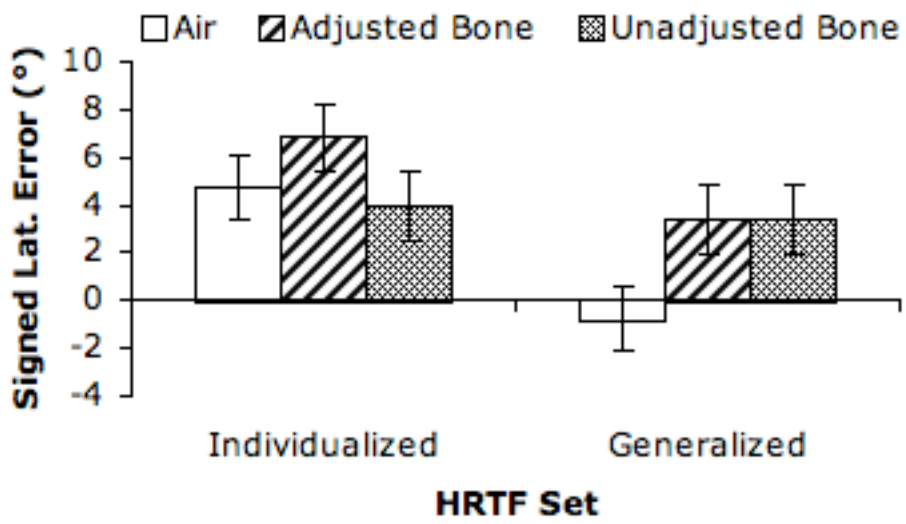

Figure 58. Signed lateralization error for P6, with individualized and generalized HRTFs. Error bars reflect one standard error above and below the mean.

Figure 59 shows that with the individualized HRTFs, the minimum elevation judgment was similar for the two bone conditions and lower for air. With the generalized HRTFs, however, the minimum elevation judgment was similar for air and unadjusted bone, and lower for adjusted bone. The minimum elevation values were generally lower for the individualized than for the generalized HRTFs.

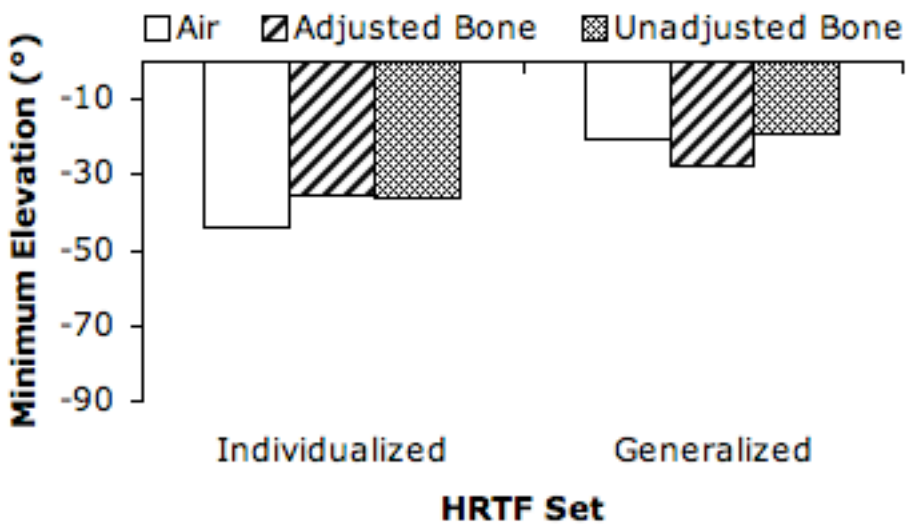

Figure 59. Minimum elevation for P6, with individualized and generalized HRTFs.

Figure 60 shows that the patterns of maximum elevation judgments across conditions were slightly different with individualized and generalized HRTFs. With individualized HRTFs, the maximum elevation judgment was similar for the air and adjusted bone condition, and slightly lower for unadjusted bone. With the generalized HRTFs, however, the maximum elevation judgment was similar for the two bone conditions, and slightly higher for the air condition. The maximum elevation values were slightly higher with the generalized HRTFs than with the individualized HRTFs in the air and unadjusted bone conditions, and slightly lower for the adjusted bone condition.

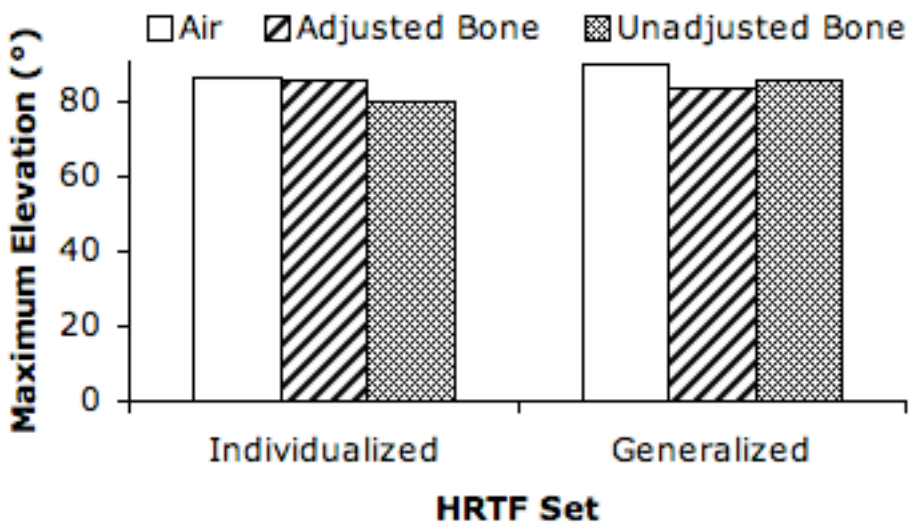

Figure 60. Maximum elevation for P6, with individualized and generalized HRTFs.

Figure 61 shows that there were only very subtle differences between the externalization rating for individualized and generalized HRTFs, but with very little variability in judgments. The ratings were similar across conditions with the individualized HRTFs, but for the generalized HRTFs the air condition rating was slightly higher than the bone conditions (which were rated similarly).

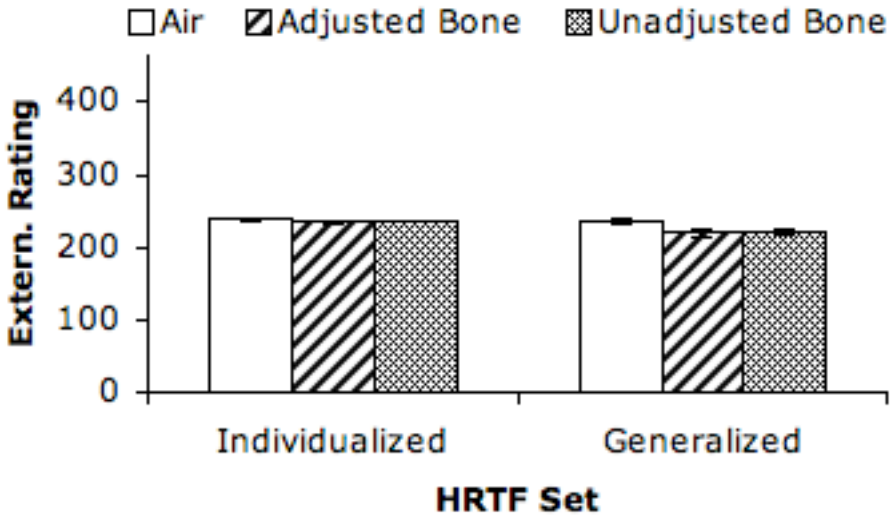

Figure 61. Externalization rating for P6, with individualized and generalized HRTFs. Error bars reflect one standard error above and below the mean.

Figure 62 shows that the results for the diffuse rating were similar to the results for the externalization rating: the diffuse ratings were similar across conditions with the individualized HRTFs, but the air condition diffuse rating was slightly higher than the similarly rated bone conditions.

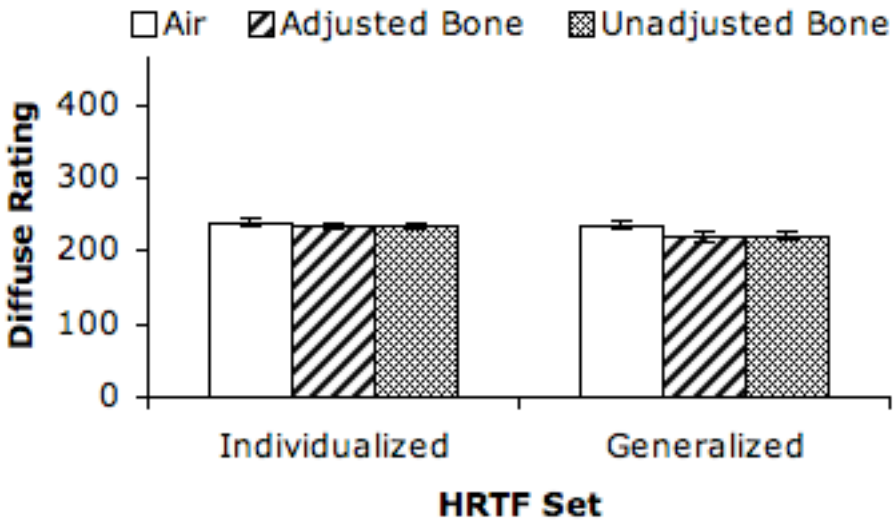

Figure 62. Diffuse rating for P6, with individualized and generalized HRTFs. Error bars reflect one standard error above and below the mean.

Table 19

Additional descriptive statistics for individualized HRTFs

Condition	Azimuth Standard Deviation			
	Mean*		Standard Error†	
	Individualized	Generalized	Individualized	Generalized
Air	28.8138	29.6315	4.7981	5.6249
Adjusted Bone	35.7809	31.0134	5.0103	5.0374
Unadjusted Bone	33.2750	30.1946	4.5457	5.0633

	Elevation Standard Deviation			
	Mean*		Standard Error†	
	Individualized	Generalized	Individualized	Generalized
Air	13.3070	14.7136	1.3065	1.5209
Adjusted Bone	12.6403	14.3173	1.1407	0.9776
Unadjusted Bone	12.4854	15.2812	1.1207	1.1301

	Maximum Lateralization Judgment	
	Individualized	Generalized
Air	90	-90
Adjusted Bone	90	90
Unadjusted Bone	90	90

*This is the mean of the azimuth standard deviation, across locations; see Section 3.2.3 for a detailed explanation.
†This is the standard error of the azimuth standard deviation, across locations; see Section 3.2.3 for a detailed explanation.

CHAPTER 4

DISCUSSION

4.1. Predictions and Theoretical Basis

Predictions about the effect of the conditions can be made based on the Limited-Fidelity Model of Spatial Audio Through Bone Conduction proposed in the author's preliminary exam paper. This model states the feasibility of using BCTs to deliver spatial audio, with some limitations, through four components:

A) The bone-conduction pathway is a leaky dichotic system

B) Air-conduction filters are ill suited for bone conduction pathway

C) The bone-conduction pathway is measurable

D) Any bone-conduction adjustment function will be imperfect in its effectiveness, caused by properties of the bone-conduction pathway (e.g., sensitivity to head and jaw movements).

Component B, that air-conduction filters are ill-suited for the bone-conduction pathway, predicts that in study 2 the unadjusted bone condition would have the worst performance for aspects of localization in the up/down and front/back dimension. Specifically, the spectral cues designed to be delivered through the air-conduction pathway would be altered by the bone-conduction pathways, and thus produce filters that were not effective. This would produce an increase in localization error for aspects of localization (higher rate of up/down and front/back reversals, and increased elevation and azimuth error).

Component C, that the bone-conduction pathway is measurable, predicts that study 1 would give a measurement of the bone-conduction pathway for each participant

that could be used to make adjustments for the pathway, thus restoring the inappropriate spectral cues invoked by the bone-conduction pathway. This leads to the prediction that the adjusted bone condition would have improved performance over the unadjusted bone condition for the error aspects of localization. However, component D, that any BAF will have imperfect effectiveness, predicts that the adjusted bone condition would still not have as good of performance as the air condition on all aspects of localization. This includes all the error metrics of localization, as well as greater variability in the bone-conduction conditions than in the air-conduction condition. Component A, that the bone-conduction pathway is a leaky dichotic system, contributes to the prediction by component D, and also predicts that the maximum lateralization will be lower for both bone conditions. Together, these predictions result in the general prediction that performance will be best in the air condition, decreased somewhat in the adjusted bone condition, and worst in the unadjusted bone condition.

4.2. Effects Caused by Spectral Cues

The effect of delivering spatial audio through the bone-conduction pathway and the effect of the BAF was most clearly seen in the elevation responses, and the elevation data will be discussed before any azimuth data are considered. The pattern of the air condition having the best elevation performance, followed by adjusted bone condition, and then finally unadjusted bone condition, manifested itself in several ways. First, the regression line fit to the raw elevation responses showed a decrease in slope and R^2 values when going from air to adjusted bone to unadjusted bone for all but one participant. The shared decrease of slope and R^2 values showed that, on average, the responses became less dependent on the stimulus when going from air to adjusted bone to unadjusted bone.

A decrease in slope with constant R^2 values would have indicated a compression of elevation space. However, with both values decreasing, the slope effect is likely driven at least partially by the same cause of the reduction in R^2 : a decrease in accuracy.

For most participants, the difference in slope between the air and adjusted bone condition was larger than the difference between adjusted bone and unadjusted bone conditions. The same was true for the R^2 statistic. For other participants, the difference in slope and R^2 between the two bone conditions was greater, or the differences were similar.

When the up/down confusions were resolved and then the regression lines were fit to the elevation data, the air > adjusted bone > unadjusted bone condition pattern of slope and R^2 values held for four participants. This showed that for these participants, local error in elevation judgments (separate from the up/down reversals) was at least a partial cause of the slope and fit regression trends. For the other participants, the effect no longer occurred, suggesting it was primarily driven by up/down reversals.

The decrease in localization performance demonstrated by the raw data and regression analyses were mostly supported in the elevation aggregate statistics. For all but one participant, the up/down reversal rate was highest for the unadjusted bone condition, lower for the adjusted bone condition, and then lowest for the air condition. For most participants, the difference in reversal rate between the air and adjusted bone conditions was greater than the difference in reversal rate between adjusted bone and unadjusted bone conditions. The elevation error showed no consistent effects of condition consistent across participants, suggesting that the majority of elevation effects were caused by reversals rather than local error. The general finding of the worst elevation

performance in the unadjusted bone condition is consistent with Langedijk and Bronkhorst's (2002) findings that removing spectral cues increases elevation error (up/down reversals were not analyzed separately by Langedijk and Bronkhorst). Presumably administering the unadjusted bone condition is acting on the same mechanisms investigated by Langendijk and Bronkhorst, but is altering the shape rather than removing the cues. The elevation findings are also consistent with preliminary tests of localization with BCTs done by Schonstein and colleagues (2008). Together, these findings show that the BAF was effective in restoring the spectral cues degraded by delivering spatial audio through the bone-conduction pathway.

It was noted that in most cases the effect of the bone conduction pathway was greater than the effect of the BAF (as evidenced by the difference between air and adjusted bone performance being greater than the difference between adjusted and unadjusted bone performance). In other (less common) cases, the effect of the BAF was greater than or similar to the effect of the bone-conduction pathway (as evidenced by the difference between the two bone conditions being equal to or greater than the difference between the air condition and adjusted bone condition). By itself, the finding of greater performance cost associated with the bone-conduction pathway could suggest that aspects of the bone-conduction pathway that cannot be addressed with spectral filtering are the cause for this effect. The lack of effects for variability and maximum lateralization (which will be discussed in more detail later), however, prevent the conclusion that the performance cost is caused by aspects of the pathway that cannot be addressed with spectral filtering. Thus, this finding suggests that the biggest performance cost is caused by spectral cue adjustments not captured by the BAF, and that a more accurate BAF

could produce an even greater improvement in localization performance in the elevation plane. Relatedly, perhaps the participants that had a greater effect caused by the BAF happened to produce more accurate BAFs. On the other hand, some aspects of the BAF shape suggest that a perfectly accurate BAF may still produce similar results. Specifically, the biggest changes in spectral shape within participants and on average were not in the frequency range most important for spatial hearing, between 8 and 16 kHz (Langendijk & Bronkhorst, 2002).

In addition to error, the difference in spectral cues implemented by the conditions also led to an effect on judgment bias in the elevation plane. Specifically, most participants showed a downward bias when going from the air to adjusted bone to unadjusted bone condition, presumably caused by the degradation in spectral cues. Relatedly, three participants showed a trend of a reduction in the maximum elevation perceived when going from air to adjusted bone to unadjusted bone condition, again presumably caused by degradation in spectral cues. How spectral cues could cause a monotonic trend in downward bias and limits in maximum elevation is not clear. In his seminal work on the "boosted band" theory of median localization, Blauert (1969/1970) showed that the frequency of narrowband noises (independent of location) can drive localization in the median plane. Perhaps a distortion of the spectral cues that was consistent across participants tapped into frequency bands corresponding to stimuli lower in the median plane. Despite the relative downward trend found in the present study, all participants showed a positive elevation bias in the air-conduction condition, confirming the "elevation illusion" with generalized HRTFs shown by Folds (2006).

Despite all of these effects of condition on elevation that are presumably driven

by spectral cues, few effects were found in the azimuth plane of judgment. The raw azimuth plot and front/back reversal rate showed no effect of condition, but this was likely caused by the near-chance baseline levels of front/back reversals. The raw azimuth plot did show a trend for some participants of less stimuli being perceived in front on the bone conditions, relative to the air condition. Although this perceptual effect probably had something to do with a difference in spectral features (caused by some BAF measurement error), it is not clear why this specific effect occurred. Once front/back reversals were resolved, there was also less error in the air condition relative to the two bone conditions. This effect could be caused by either spectral cues or differences in lateralization driven by the pathway.

It may be noted that there were effects in the regression data not present in the aggregate data. For example, there were still some effects of condition on elevation after up/down reversals were resolved in the regression analysis, but not in the aggregate analysis. In another case, there was an effect of condition on azimuth after front/back reversals were resolved in the regression analysis, but not in the aggregate analysis. This can be explained by the stimuli locations, which were distributed equally in spherical space. Plotted in linear space, such a distribution of locations shows unequal spacing along the diagonal (see Figure 63). When regression analyses were used, a linear scale was assumed and the slope and fit were more sensitive to the endpoints of the diagonal. The summary statistics, on the other hand, were averaged across locations without regard to their position on the diagonal.

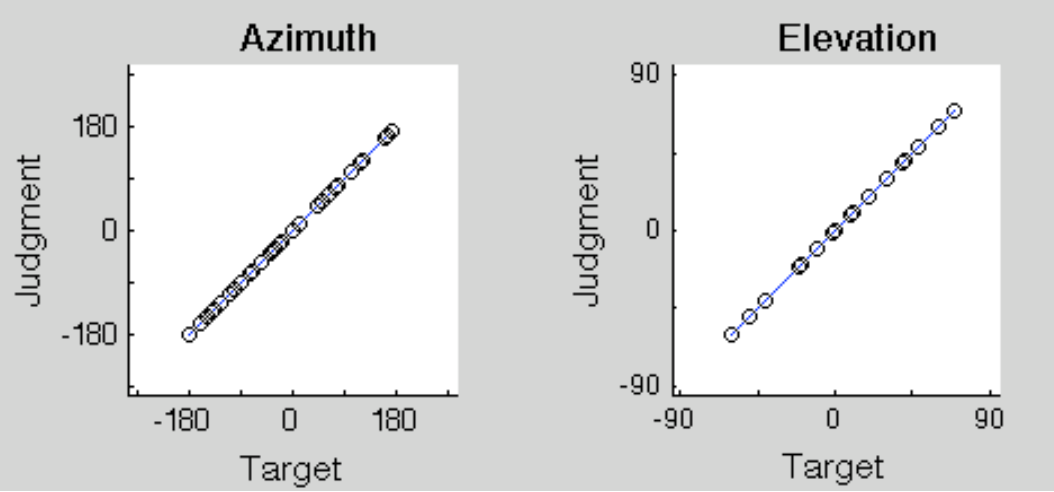

Figure 63. Plot of stimuli locations. Note the unequal spacing along the diagonal, which could explain the slight differences found between regression analyses and aggregate statistics.

The change in spectral cues invoked by the conditions should have produced changes in subjective ratings of externalization and diffuseness, since these aspects of the stimulus are likely to be affected by the spectral shape of the stimulus. A complication with these ratings is that a single judgment was made at the end of a block that required a lot of concentration to complete. There could have often been differences between trials in these aspects of the stimulus, which would be difficult for the participant to "average" across. Furthermore, the cognitive resources required to complete the localization task from trial to trial could have prevented participants from accessing any aggregate information about the externalization or diffuseness of the stimuli.

4.3. Effects Caused by the Inherent Properties of Bone Conduction

In addition to error and bias, an effect of condition on variability and maximum judgments was predicted, caused by properties of the bone conduction pathway. There was no evidence in the aggregate variability data for instability in the pathway causing changes in the physical cues at the cochlea. There was also no evidence in these data for a

limit in maximum lateralization caused by reduced interaural attenuation in the bone-conduction pathway. Moreover, all but one participant had a bias towards perceiving the response to be more lateralized than the stimulus in the bone-conduction conditions, relative to the air condition. Despite the ears being plugged, there is a possibility that this was partially caused by air-conducted leakage. However, the attenuation provided by the earplugs was probably greater than the air-conducted leakage, which makes this unlikely. Another possible cause of the lateralization bias was the air-conduction component of the bone-conduction pathway. It is important to note that this conceptualization of lateralization assumes the localization space has some sort of spherical shape. If there were a one-dimensional non-spherical localization space (i.e., a line), a correct vector judgment of stimuli with different interaural differences would always be +/- 90 degrees.

4.4. Individualized HRTF Replication

Studies 1 and 2 were replicated with individualized HRTFs on participant 6, which was designed to test the effect of the conditions in the context of the highest-fidelity V3DAD display possible. The individualized HRTF test did not provide a higher-fidelity baseline to test with these conditions, and the effects were largely the same as with the generalized HRTFs. There are several possible reasons for not attaining the performance improvement found in the literature with individualized HRTFs. First, the directional-transfer function method of creating V3DADs relies on the diffuse-field response of the headphones to replicate the common HRTF component. By relying on what has been standardized to be a "diffuse field response," the DTF system is effectively relying on a generalized common HRTF component function. This is in contrast to the more conventional HRTF method, which includes an individualized measurement of pathways

captured by the common component (e.g., headphone-to-eardrum response). Plotting of the diffuse-field frequency response and the common HRTF component in Appendix D shows that there were considerable differences between the common component and the diffuse-field frequency response at frequencies above 5 kHz. This disparity was likely to be a large contributor to the low performance with the individualized HRTFs. Related to the common component correction is the fact that the speaker and microphone responses were thought to be approximately flat and thus were not corrected for. The degree to which these responses deviated from a perfectly flat response could have contributed to the low individualized HRTF performance.

Second, participant six had never been previously tested with individualized HRTFs, so the listener's best possible performance is not known. Third, although the response method seemed to work well, this specific implementation has not been used in previous research, so it could have influenced the results. Overall, there are numerous other unknown effects and interactions that could have been caused by using a combination of measurements, processing, response, and listener that have not been previously verified. The specific HRTF measurement system used, for example, was not previously used with the DTF methodology, the same listener, or the computer-model response method.

4.5. Baseline Performance Relative to Previous Literature

It may be noted that the air condition baseline results were not as good as the Middlebrooks studies would have predicted. However, the analysis shown in those papers departed greatly from the analysis shown here. First, the horizontal polar system was used by Middlebrooks and colleagues (1999b; 2000). Relatedly, data analysis was

confined to 30 degrees from the vertical midline, to avoid the compression of polar error at far laterals (Middlebrooks, 1999b). This resulted in a lower count of front/back confusions than other conventional definitions (Middlebrooks, 1999b). Furthermore, reversals were not counted unless there was a difference between polar (front/back, up/down) stimulus and response that was greater than 90 degrees ("quadrant errors") (Middlebrooks, 1999b, 2000). These differences in analyses prevent direct comparison of baseline values. In addition, there were differences in the methodology of the studies: Middlebrooks' studies had a different response method (the listener pointed their nose to the stimulus), and did not have a bandpass filter for the virtual stimuli.

These baseline reversal rates can still be compared to investigations that used similar coordinate systems and response metrics that investigated generalized HRTFs. Many of the studies investigating generalized HRTFs used ordinal responses that prevent comparison with the present data, or did not explicitly state front/back reversal rates. A study by Wightman and colleagues (1993), however, provides a suitable comparison. Using HRTFs from another individual Wightman and colleagues found that inexperienced listeners (like the ones in the present study) had on average 31 percent front/back reversals, with values ranging from 10 to 50 percent. This is generally a lower rate than what we found in our study. One obvious reason for this difference was that they did not have an exclusion range like we did. Analysis on the data from the present study without an exclusion range, however, still an average front/back reversal rate of 48 percent. Other causes for the differences include different locations used in Wightman et al. (1993), the bandpass cutoff frequencies (200 Hz and 14 kHz), the response method (verbal response), and the specific HRTF used. Another difference was that they used the

original HRTF subjects' headphone-to-eardrum transfer function, rather than the diffuse field response used in the present study. Wenzel and colleagues' (1993) study showed up/down reversal rate was on average 18 percent, ranging from 7 to 32 percent. This is generally similar than the findings in our study, which used an elevation exclusion range of 15 degrees. Without that exclusion range, though, the average up/down reversal rate was 23 percent, ranging from 14 to 34 percent.

4.6. Contributions, Conclusions, and Implications

The primary contribution of this research was the first known test showing that a bone-conduction adjustment function can be used to improve spatial audio performance. More generally, this research provided a unique comprehensive test of auditory localization through bone conduction. This study showed that a BAF based on equal-loudness adjustments can restore spectral cues and allow effective localization in the elevation plane through the reduction of up/down reversals. These findings can now be added to the body of evidence against the popular belief that spatial audio cannot be effectively delivered through bone-conduction hearing. This study also showed, however, that there is a performance cost associated with using bone conduction that may not be able to be overcome through use of a BAF. In addition, this study also provided a test of a unique response method that allows continuous judgment of both azimuth and elevation using only a personal computer and a widely-used software program (MATLAB). Although a full evaluation of this method would require side-by-side comparisons with other more established response methods, these data show that the method is likely to be suitable for a variety of applications. This response method can provide an easy way for other researchers to rapidly collect localization responses.

The findings from the present study extend the findings of previous research investigating related aspects of bone-conducted spatial audio. The present findings replicate and extend the findings showing lateralization of sounds delivered through bone conduction (e.g., Jahn & Tonndorf, 1982; Kaga, et al., 2001; Stanley & Walker, 2006; Walker & Lindsay, 2005; Walker, Stanley, Iyer, et al., 2005). The present studies show in addition to lateralization, changes in elevation are also possible. The present studies also extend the findings of Macdonald and colleagues (2006), showing that once frequencies where spectral cues lie are included, and noise is not delivered in the environment, there indeed is a difference between air-conducted and bone-conducted spatial audio (as shown by Schonstein, et al. (2008)).

This has implications for the design of BCTs that are intended to deliver spatial audio, showing that a built-in correction filter could improve their use. The present studies also extend Stanley's (2007) initial investigation into building a BAF, by testing an alternate method and verifying the efficacy of the BAFs. The present findings also show that the difficulties associated with the phase component of the BAFs can be avoided while still obtaining an effective BAF, allowing a more efficient measurement of a BAF. It is of course conceivable that the addition of phase manipulation to the BAF could produce an added benefit – but it is not necessary for a measurable benefit in performance.

The results of this study provided theoretical contributions by testing the Limited-Fidelity Model of Bone-Conducted Spatial Audio proposed by the author. The strong reduction in elevation performance when going from air to adjusted bone conduction to unadjusted bone conduction gave support for three components of the Limited-Fidelity

Model of Bone-Conducted Spatial Audio. The worst performance occurring in the unadjusted bone condition showed that air-conduction filters are ill-suited for the bone-conduction pathway (component B). The relative improvement in performance seen in the adjusted bone condition showed that the bone-conduction pathway is measurable (component C) – specifically, through equal-loudness adjustments of pure tones at frequencies centered on critical bands. The fact that performance in the adjusted bone condition was still not as good as performance in the air condition did show some evidence that there is imperfect filter effectiveness, but not by the predicted factors. The lack of differences in variability across conditions in fact suggested that dynamic properties of the pathway were not an issue. One possible cause for the apparent limitation of the bone-conduction pathway is that perceptual judgments are required to form the BAF, rather than the perhaps more accurate acoustic measurements that occur with air-conducted HRTFs. The lack of differences in lateralization between conditions and the bias towards perceiving stimuli to be more lateralized in the bone condition provided strong evidence against a leaky dichotic theory of bone conduction (component A), suggesting that there may be no limitations in the maximum lateralized stimulus delivered through bone conduction.

Finally, these findings provide insights into applied use of bone-conduction transducers. These findings show that BCTs could be used with V3DAD technology to provide a way to deliver controlled spatial audio while leaving the ears unoccluded for optimal access to ambient sounds, or allowing the ears to be plugged to block out dangerously loud ambient sounds. The present study showed this with plugged ears, but the findings should generalize to open ears as well (although the specific BAF would

certainly be different). This study also provided an estimate of the performance cost associated with using BCTs, which users will have to carefully weigh against the advantages of having the ears unoccluded or plugged. In addition, these studies also provided an average ("generalized") BAF that can be used by the public to improve V3DADs on BCTs. However, the variability of the BAFs for each individual shown in this study suggested that BAFs may have to be measured for each person (that is, be "individualized"). If individualized BAFs are justified, the present studies demonstrated a fast and effective way of measuring a BAF.

It was clear from these data that air-conduction performance was predictive of bone-conduction performance: those who were better at air were better at bone, and those were worse at air were worse at bone. Given the additional effort associated with measuring the BAF, air-conduction localization performance could be used as a screening criterion for bone-conduction performance. Pearson's *r* correlations between the conditions in performance metrics were calculated to quantify the predictive power of air-conduction performance for estimating bone-conduction performance (see Appendix C).

The correlations between air and adjusted bone performance were greater than .8 for many metrics: azimuth error, elevation error, raw elevation slope, and raw elevation R^2. The correlations between air and unadjusted bone were generally less strong, but all greater than .3. Thus, air-conduction performance is highly predictive of bone-conduction performance with a BAF, but less predictive of bone conduction performance without a BAF. The strong correlations for the metrics that showed considerable differences between conditions are particularly valuable (e.g., raw elevation slope and R^2). They indicate that despite a difference in performance between conditions, the air condition

performance is highly predictive of bone-conduction performance with a BAF. Overall, these correlation data suggest that air-conduction localization performance could be a useful screening criterion for bone-conduction performance when a BAF is used.

4.7. Future Directions

Future investigations should also seek to verify these results with a complete spatial audio system that has been used to create the high performance levels consistent with the literature using individualized HRTFs, which require data collection in a lab specializing in virtual 3D audio systems.

Future research investigating BAFs should consider the benefit of individualized BAFs relative to generalized BAFs. This will indicate the benefit gained by taking the resources to measure BAFs for individuals. If there is little or no cost associated with using generalized BAFs, then they would be preferable. Other methods of measuring BAFs should also be explored to seek a method that can create more accurate BAFs, if that is possible. For example, an alternative BAF formation method could rely on computational models of the bone-conduction pathway (e.g., Przekwas, et al., 2007; Walker, et al., 2007). Perhaps in combination with alternative BAF measurement techniques, it would also be informative to measure the added benefit of phase manipulations in the BAF.

The present studies involved measurement and test of BAFs with plugged ears. It would be interesting to understand the effect of unplugging the ears on the BAF and localization performance. In addition, bone-conduction transducers could be useful in many other contexts outside of and in conjunction with spatial audio – for example, radio communications and music listening. It would be interesting to look into the effect of

BAFs in these contexts, with the corresponding stimuli that differ from the noise stimuli used in the present studies.

APPENDIX A
BONE-CONDUCTION ADJUSTMENT FUNCTIONS

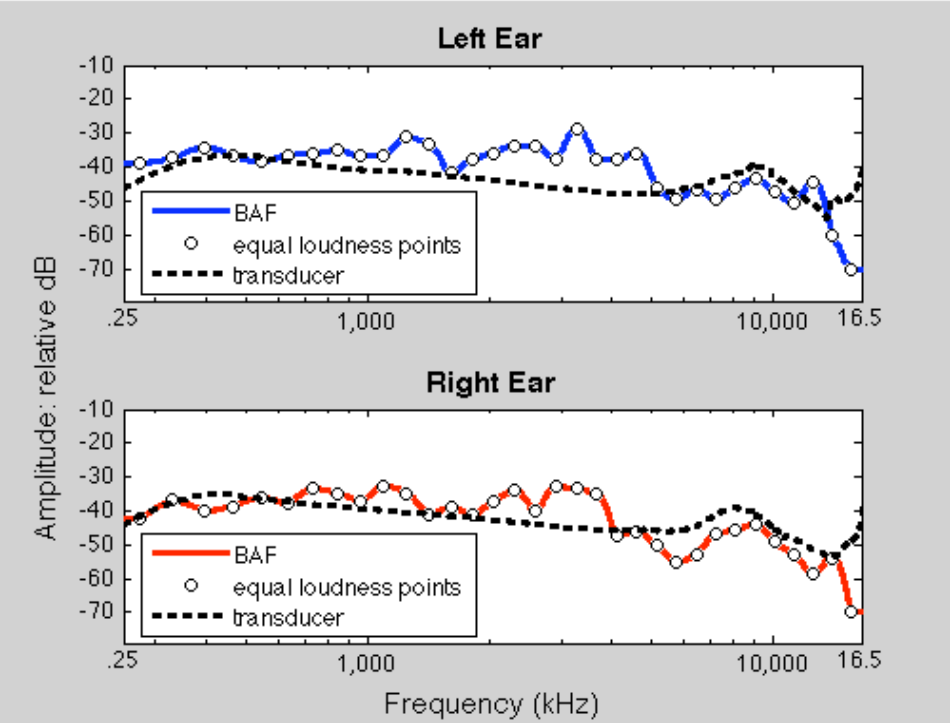

Figure 64. Bone-conduction adjustment function (BAF) for P1.

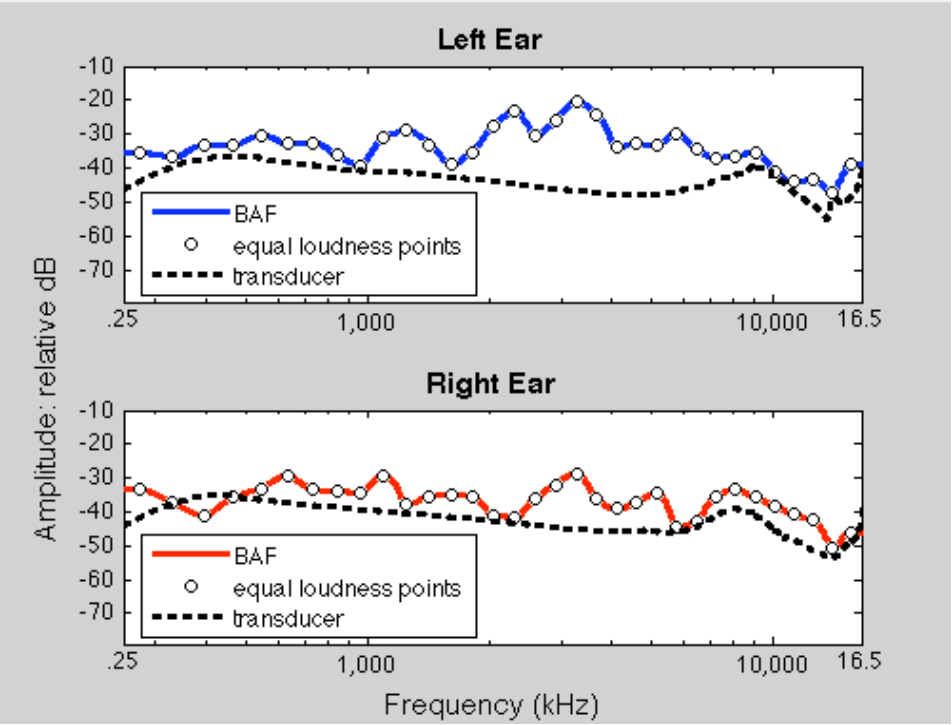

Figure 65. Bone-conduction adjustment function (BAF) for P2.

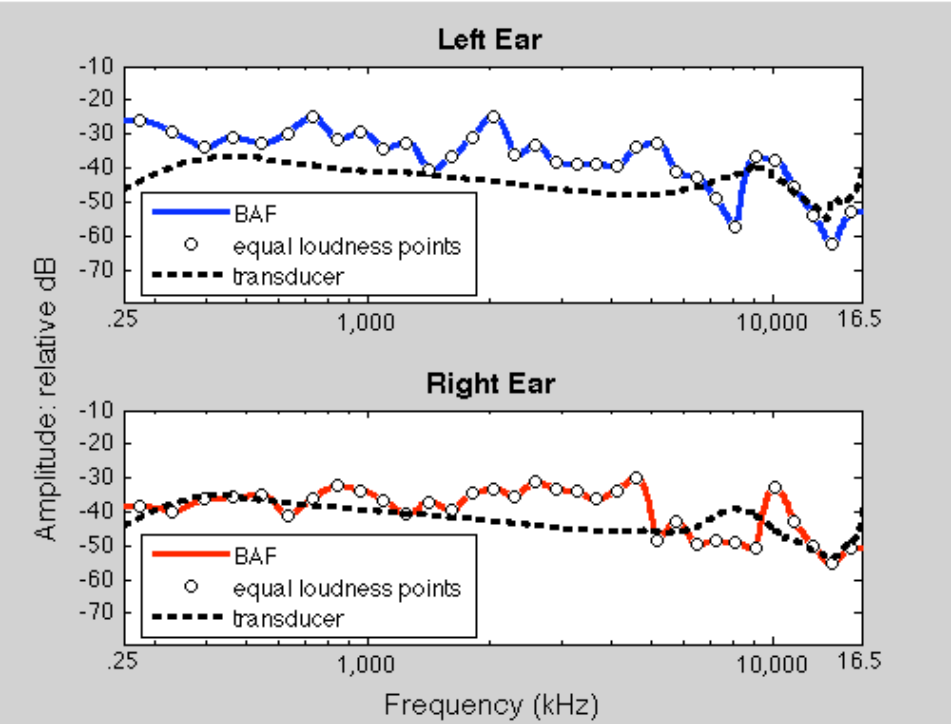

Figure 66. Bone-conduction adjustment function (BAF) for P3.

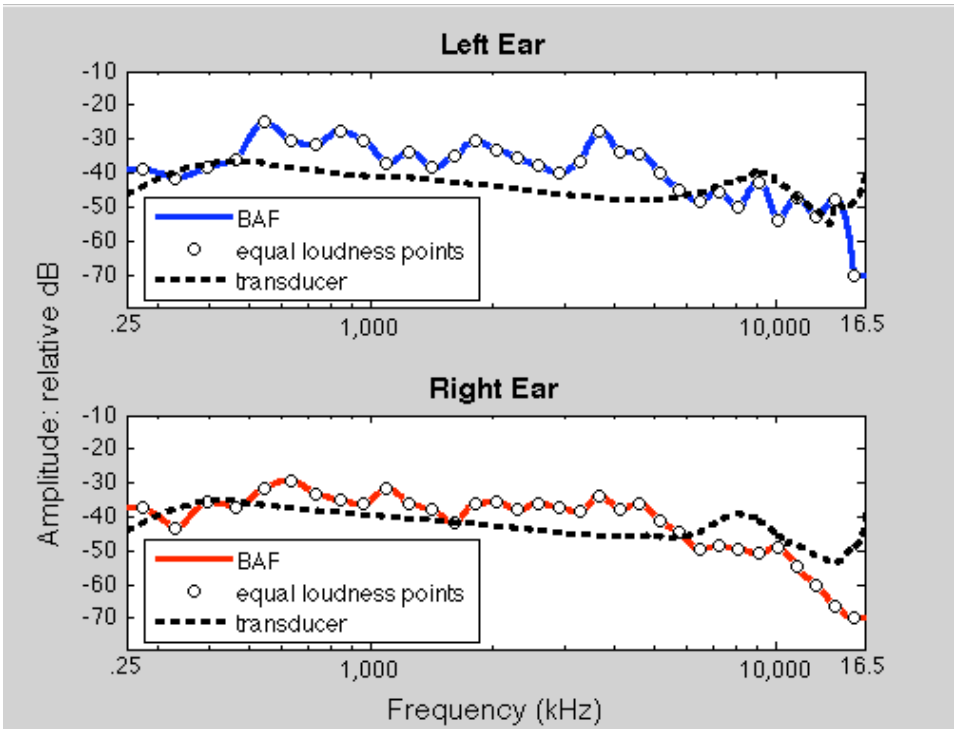

Figure 67. Bone-conduction adjustment function (BAF) for P4.

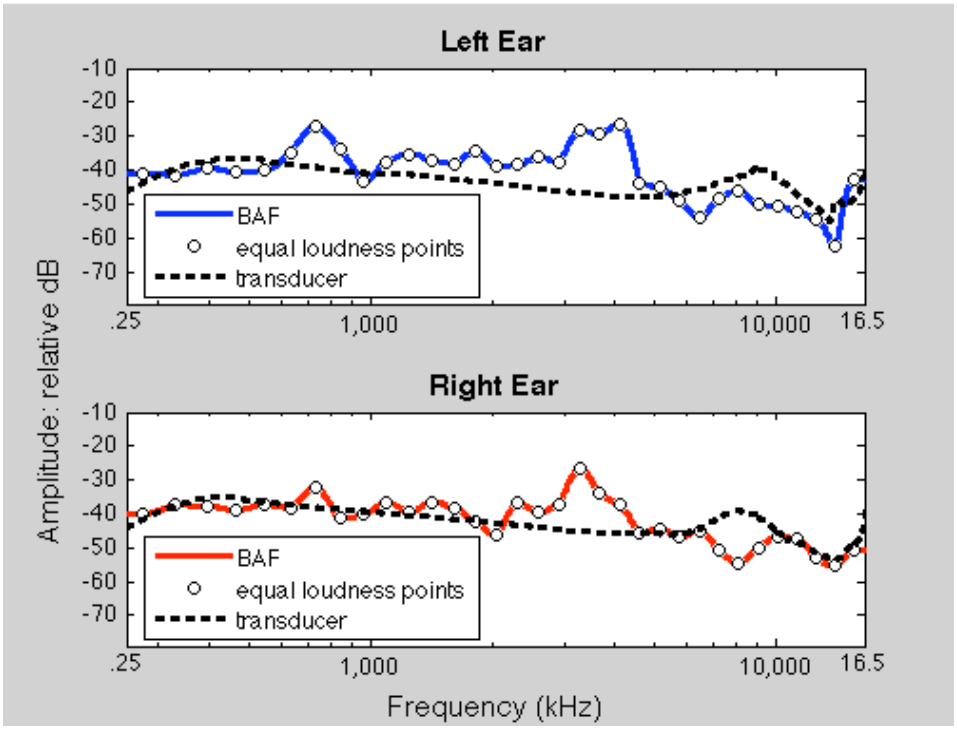

Figure 68. Bone-conduction adjustment function (BAF) for P5.

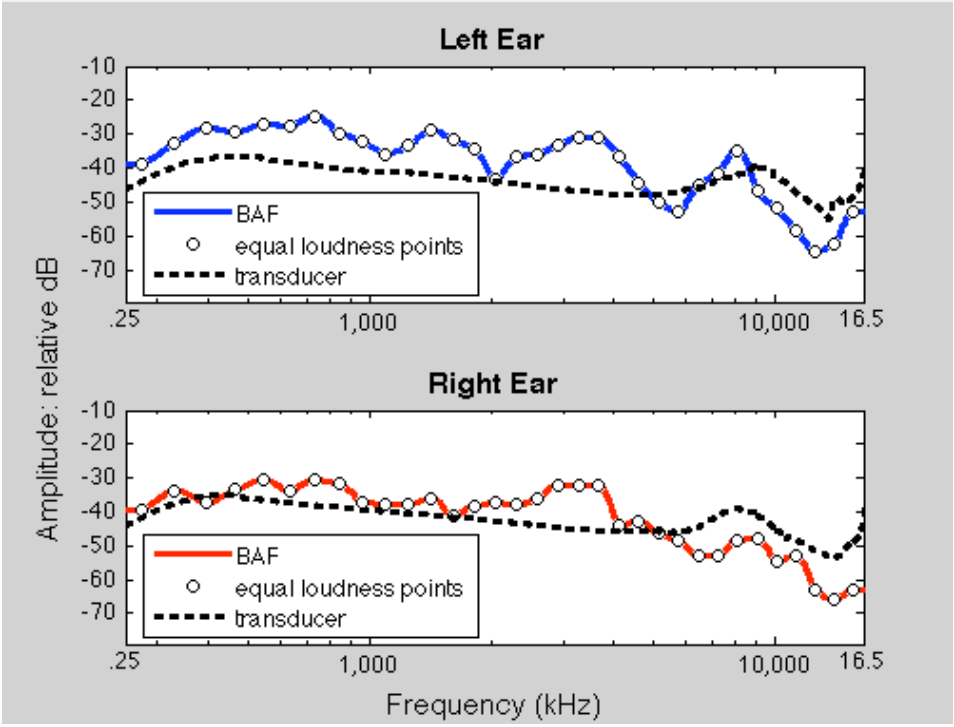

Figure 69. Bone-conduction adjustment function (BAF) for P6.

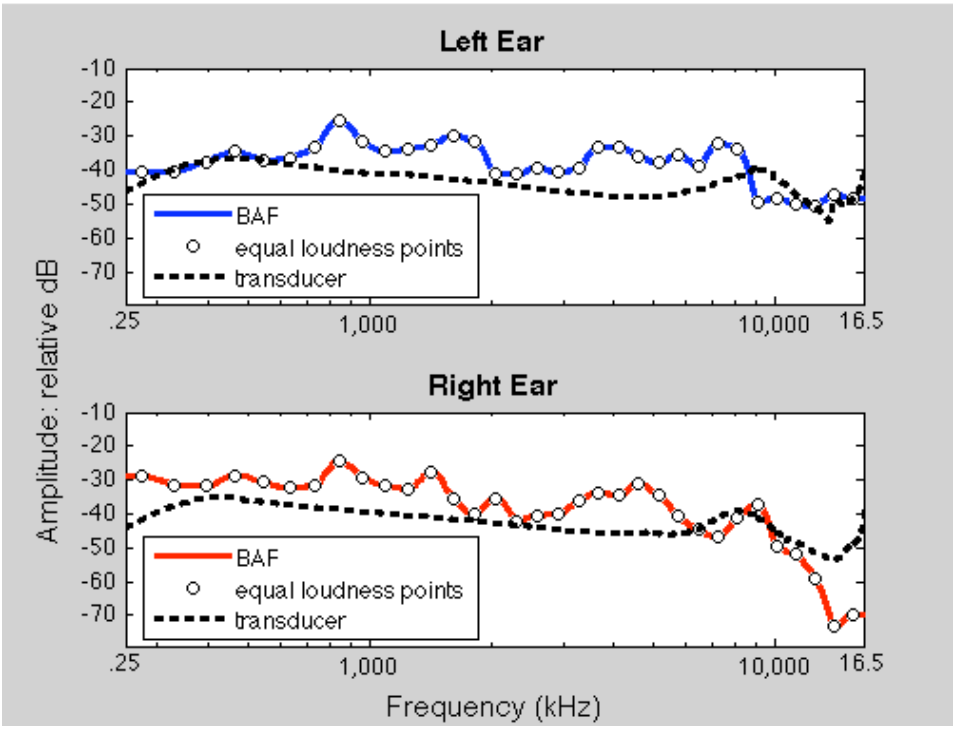

Figure 70. Bone-conduction adjustment function (BAF) for P7.

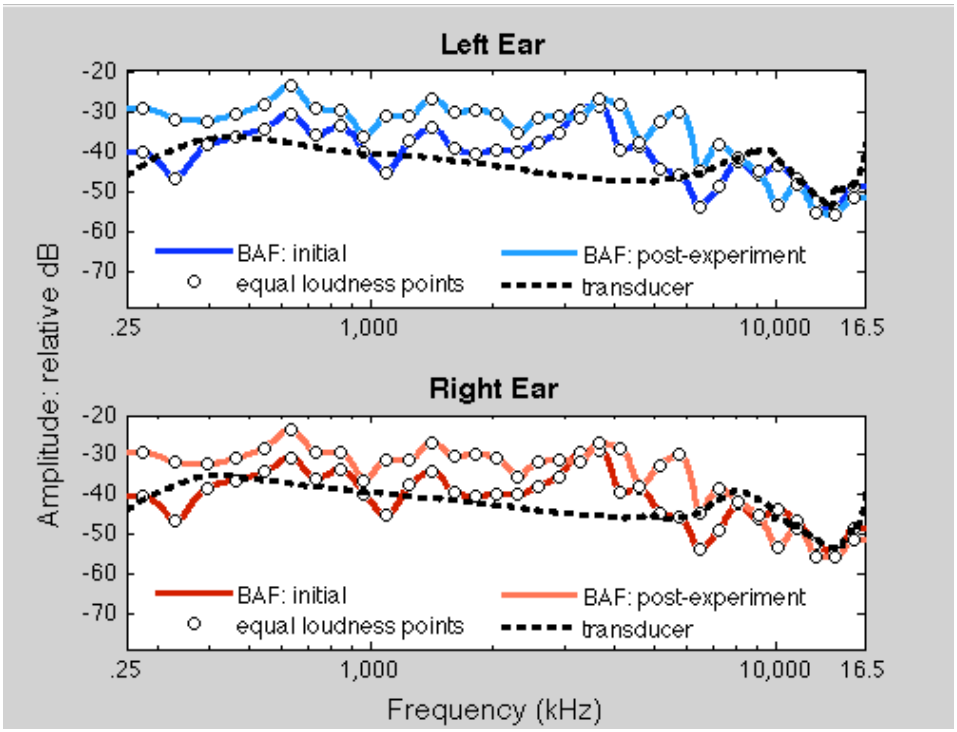

Figure 71. Initial ("pre") bone-conduction adjustment function (BAF) and re-measured ("post") BAF for P8.

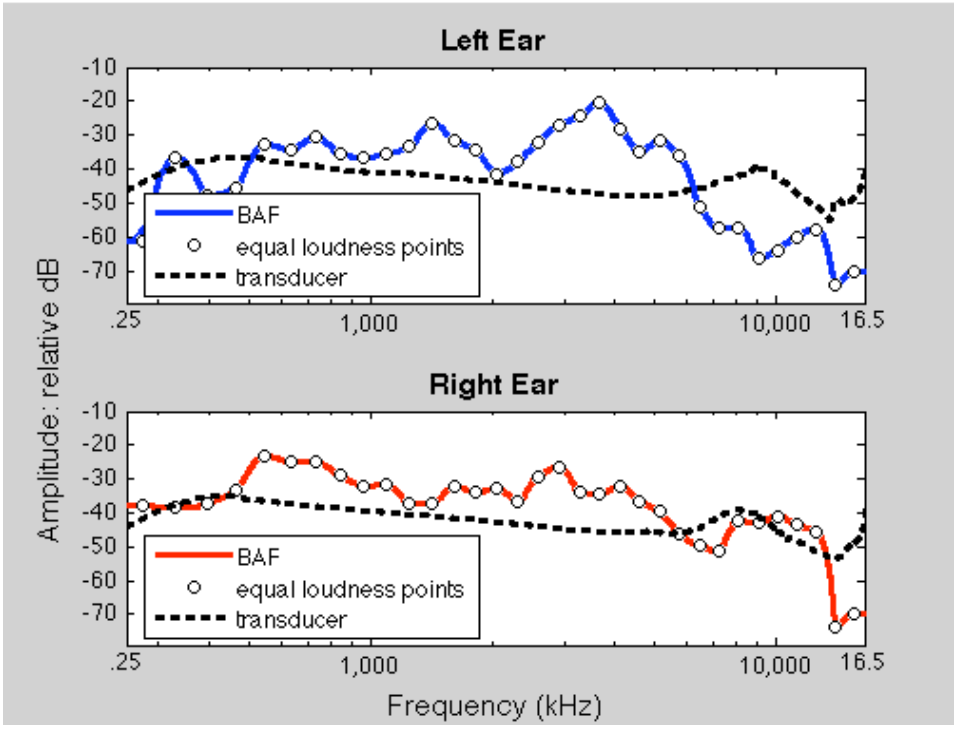

Figure 72. Bone-conduction adjustment function (BAF) for P9.

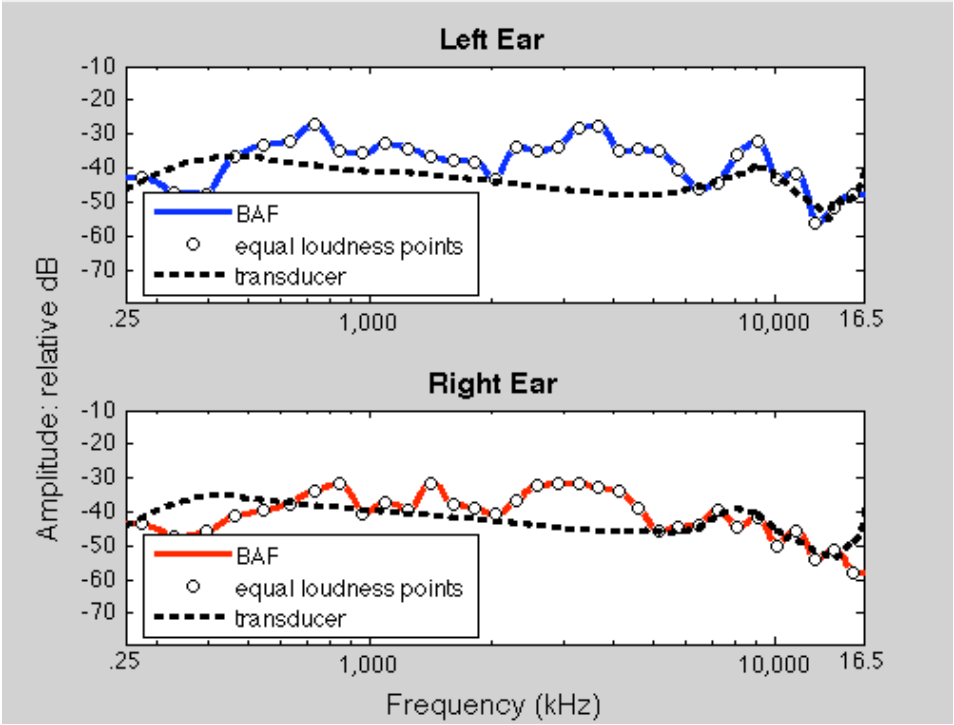

Figure 73. Bone-conduction adjustment function (BAF) for P10.

APPENDIX B

DIGITAL SIGNAL PROCESSING DETAILS

The stimuli used in Study 2 were created using a system of cascading filters applied via frequency-domain multiplication, conducted in the MATLAB environment. All transformations between the frequency domain and time domain occurred using MATLAB's "fft" and "ifft" algorithms. All values at the negative frequencies were computed by taking the complex conjugate of each value (except at the DC frequency (0 Hz) and at the Nyquist frequency (22050 Hz)), and reversing the order of values.

The initial input signal was 500 milliseconds of Gaussian-distributed noise, at a sampling rate of 44.1 kHz with 16 bits per sample. This noise was then filtered by the DTFs. Within the DTF filtering algorithm, a time-domain version of the DTF for each ear was loaded from the file provided by Middlebrooks (1999b). All subsequent steps described for the DTF filtering occurred for each ear's DTF, using the same input signal. The time-domain DTF was transformed into the frequency domain, truncated to only include the real frequencies, and padded with zeros. Then the second half of the spectrum (negative frequencies) was computed, and subsequently transferred back into the time domain. These steps accomplished time-domain interpolation. Then this interpolated time-domain signal was decimated (every n^{th} sample selected) based on the scale factor (where decimation step = scale factor/interpolation factor, and the interpolation factor equals 32). The time-domain interpolation combined with the time-domain decimation produced the frequency scaling, as described in Middlebrooks (2000). The time-domain frequency-scaled DTF was then padded with zeros and transformed back into the frequency domain (this time with the full spectrum). The input signal was then padded

with zeros and transformed into the frequency domain. The number of zeros added on to each vector was such that the length equaled the total length of the DTF plus the length of the input signal, minus one, to prevent circular convolution. The frequency-domain representation of the frequency-scaled DTF and input signal were then multiplied to accomplish the filtering. The product of this multiplication was then transformed back into the time domain. This output was then normalized so that it had a maximum value of one. What happened to the signal next depended on the condition that a stimulus was being generated for.

For the adjusted bone condition, each channel of the DTF filtering output was filtered by a BAF defined separately for each ear. The BAF was defined at 33 points by the equal-loudness study, from 273 Hz to 15621 Hz. The amplitude at 0 Hz was set to zero, the amplitude at 20 Hz was set to the amplitude defined at 273 Hz, and the amplitude at 22050 Hz was set to the same value defined at 15621 Hz. Then 2048 linearly spaced points lying between 0 and 22.05 kHz were interpolated using MATLAB's cubic interpolation algorithm (piecewise cubic Hermite interpolation). Next the frequencies between 0 Hz and 223 Hz were ramped up from zero to the amplitude value at 223 Hz. The amplitude values between 17959 and 22,050 Hz were ramped down from the amplitude value at 17959 Hz to an amplitude of zero. The frequency ramp-down points were chosen because they were one-sixth octave away from the band limits of the final stimuli (250 and 16500 Hz). The amplitude values at the negative frequencies were then computed and concatenated with the values at the real frequencies to form the full symmetrical frequency-domain representation of the BAF. The phase values were left undefined for the BAF, which left the phase of the input unaffected by the filter (thus

making the BAF a linear phase filter). The BAF was then transformed into the time domain. Next, this time-domain signal was rotated to put the peak in the middle of the record and treated to a Hamming window. Both the input signal (the output from the DTF filtering) and time-domain BAF were then padded with zeros and put back into the frequency domain. The number of zeros added on to each vector was such that the total length of each signal equaled the two times the length of the longest vector (the input signal). The frequency-domain representations of the BAF and input signal were then multiplied, and the product was put into the time domain. The time-domain filtered signal was then truncated, removing the first and last n points from the signal, where n = the half the filter length plus two. This truncated signal was then sent on to the bandpass filter.

For the unadjusted bone condition, each channel of the DTF filtering output was filtered by the inverse of the BCT frequency response. The frequency response was measured in the Sonification Lab by Bruel & Kjaer PULSE system analyzer and Bruel & Kjaer Type 4930 Artificial Mastoid, using the swept-sine wave method. The frequency response was measured separately for each transducer (left and right), and applied to the appropriate channel in the filtering process. The BCT frequency response measurement specified 800 points (in polar coordinates) between 0 and 20 kHz. To accomplish the filtering, first the file with the frequency response data was loaded, and then the values at frequencies lying between 20 and 22050 Hz (undefined by the measurement) were set to zero. The second half of the frequency domain representation of the frequency response was then computed and concatenated with the first half. Next, the frequency-domain representation of the BCT frequency response was transformed back into the time

domain, where a minimum-phase version of the impulse response was calculated using MATLAB's "rceps" algorithm. The minimum phase impulse response was then truncated to 880 points (determined by trial and error), and multiplied by the last half of a Hanning window. Next, the windowed impulse response was padded with zeros (so that the total length was twice as long as the input signal), and transformed into the frequency domain. The inverse of the minimum-phase frequency response was then computed by taking the reciprocal of each frequency-domain element. This created very high boosts at either end of the spectrum, since the parent frequency response ramped down to zero at the ends. To remove these boosts, the amplitude was ramped up from zero until 223 Hz, and ramped down to zero starting at 17959 Hz. The phase of the original inverse frequency response was preserved and added back in after the ramping. The imaginary values at the DC frequency (0 Hz) and Nyquist frequency (22050 Hz) were then removed, since all real signals should have a real value at the DC and Nyquist frequencies. Then the second half of the inverse frequency response was calculated and concatenated with the first half, and padded with zeros. Next, the input signal (the output from DTF filtering) was transformed into the frequency domain after being padded with zeros. The number of zeros added on to each vector was such that the total length of each signal equaled the two times the length of the longest vector. The product of the inverse frequency response and frequency-domain input signal was then computed and transformed back into the time domain to produce the filtered signal. Finally, the filtered signal was truncated to the length of original input signal. After this was done, the amplitude of the time-domain filtered signal was multiplied by the adjustment value determined in the unadjusted equal-loudness trials.

The output from the DTF filtering (for the air condition), BAF filtering (for the adjusted bone condition), or inverse BCT frequency response (for the unadjusted bone condition) was then bandpass filtered. Before the signal was bandpass filtered, it was padded with n zeros, where n equals two times the signal's length. Then the signal was sent through a 4-pole butterworth bandpass filter with cutoffs at 250 and 16500 Hz. The filter was built using MATLAB's "butter" command, and applied using MATLAB's "filtfilt" command, which performs zero-phase filtering by running the input through the filter in the forward direction and then again in the reverse direction. Note that this doubles the effective filter order, making for an 8-pole filtering effect. After the bandpass filtering occurred, the gain from the filtering was removed from the signal by normalizing the amplitude back to the pre-bandpass amplitude.

After the bandpass filtering was complete, the final time-domain preparations were done. First, the signal was truncated to 250 milliseconds by removing an equal number of samples off the front and back; about 250 milliseconds' worth of total samples were typically removed. All the truncation was done so that the amplitude envelope of the signal was similar to the original input signal at all stages of the cascading filters. Finally, a raised-cosine window was applied to the first and last 10 milliseconds of the signal. Now that the stimuli burst was in final form, it was concatenated with 300 milliseconds of silence three times to produce the pulse train. This pulse train was then written to a wave file, which was read during the stimuli presentation.

APPENDIX C

PERFORMANCE CORRELATIONS BETWEEN CONDITIONS

Table 20

Correlation table showing Pearson's r for azimuth error between conditions.

	Air	Adjusted Bone	Unadjusted Bone
Air	1	0.9803	0.9529
Adjusted Bone		1	0.9206
Unadjusted Bone			1

Table 21

Correlation table showing Pearson's r for front/back reversal rates between conditions.

	Air	Adjusted Bone	Unadjusted Bone
Air	1	0.352	0.4404
Adjusted Bone		1	0.6954
Unadjusted Bone			1

Table 22

Correlation table showing Pearsons' r for up/down reversal rates between conditions.

	Air	Adjusted Bone	Unadjusted Bone
Air	1	0.6348	0.3127
Adjusted Bone		1	0.8695
Unadjusted Bone			1

Table 23

Correlation table showing Pearsons' r for elevation error between conditions.

	Air	Adjusted Bone	Unadjusted Bone
Air	1	0.9008	0.8028
Adjusted Bone		1	0.6535
Unadjusted Bone			1

Table 24

Correlation table showing Pearson's r for raw elevation slope between conditions.

	Air	Adjusted Bone	Unadjusted Bone
Air	1	0.8708	0.5783
Adjusted Bone		1	0.8090
Unadjusted Bone			1

Table 25

Correlation table showing Pearson's r for raw elevation R^2 between conditions.

	Air	Adjusted Bone	Unadjusted Bone
Air	1	0.8256	0.5344
Adjusted Bone		1	0.8350
Unadjusted Bone			1

APPENDIX D

COMMON DTF COMPONENT AND ER-1 DIFFUSE FIELD RESPONSE COMPARISON

The common component removed from the HRTF to create the DTF is added back in to the signal by the diffuse-field frequency response of the headphones. Thus, the individualized common component is a replaced by a generalized estimate of the common component. The degree to which this two functions do not match can explain the lack of high performance found with individualized HRTFs with Participant 6. Figure Figure 74 shows the diffuse-field frequency response, and Figure 75 shows the common component of the individualized HRTFs measured for Participant 6. Comparing the two graphs shows that the functions have a similar shape up to the 2.5 kHz peak and the subsequent dip. The small peak at 7 kHz in the diffuse-field frequency response is much larger in the common HRTF component, and the peak is wider such that it encompasses 5 kHz. There is also an additional peak at 12,450 Hz in the common component that is not present in the diffuse-field frequency response. These considerable differences were likely to contribute to the low performance with the individualized HRTFs.

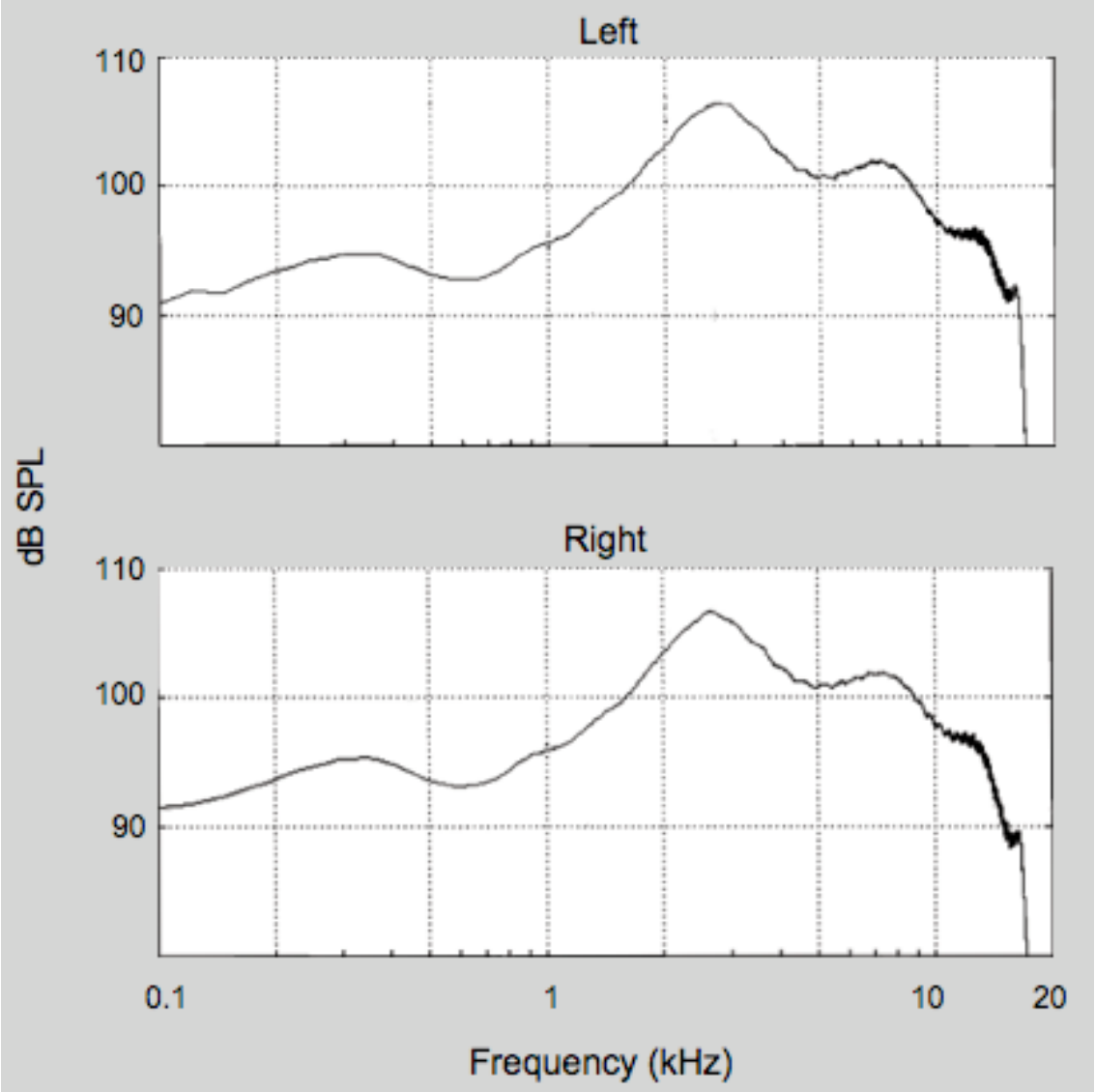

Figure 74. "Diffuse-field" frequency response of Etymotic ER-1 headphones.

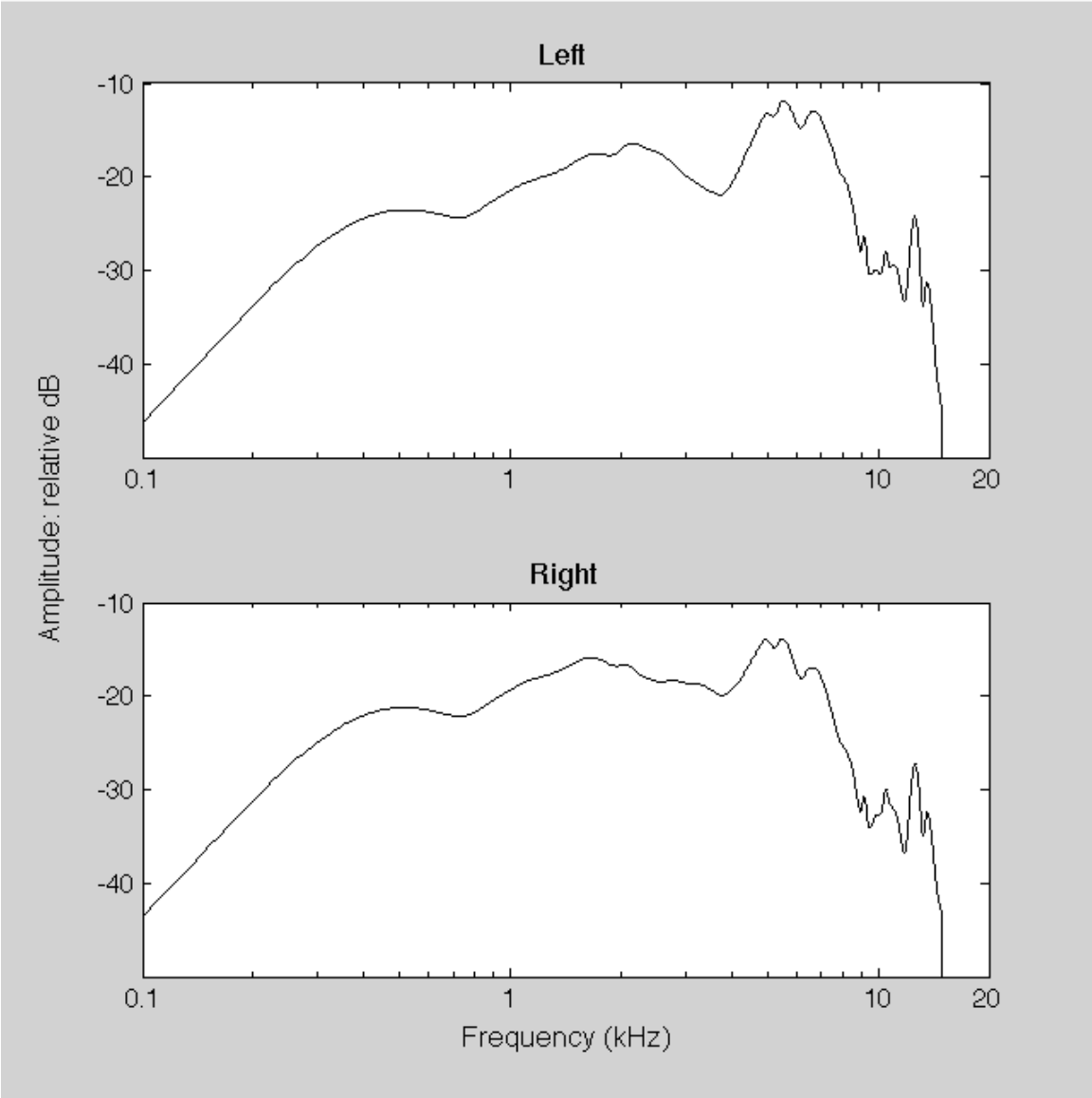

Figure 75. Common component of individualized HRTF measured on P6.

REFERENCES

Ahmad, A., Stanney, K. M., & Fouad, H. (2008). Subjective Selection of HRTFs. University of Central Florida, Office of Naval Research - Spatial Audio Optimizer Project. Unpublished Proprietary Report.

Békésy, G. v. (1960). *Experiments In Hearing* (E. G. Wever, Trans.). New York: McGraw-Hill Book Company, Inc

Blauert, J. (1969/1970). Sound localization in the median plane. *Acustica, 22*, 205-213.

Blauert, J. (1997). *Spatial hearing: The psychophysics of human sound localization* (J. S. Allen, Trans.). Cambridge, MA: The MIT Press.

Bordens, K. S., & Abbott, B. B. (2005). *Research Design and Methods* (6 ed.). New York, NY: McGraw-Hill.

Brungart, D. S., Rabinowitz, W. M., & Durlach, N. I. (2000). Evaluation of response methods for the localization of nearby objects. *Perception and Psychophysics, 62*(1), 48-65.

Brungart, D. S., & Simpson, B. D. (2002). The effects of spatial separation in distance on the informational and energetic masking of a nearby speech signal. *Journal of the Acoustical Society of America, 112*(2), 664-676.

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

Carlile, S., & Best, V. (2002). Discrimination of sound source velocity in human listeners. *The Journal of the Acoustical Society of America, 111*(2), 1026-1035.

Cohen, J., & Cohen, P. (1983). *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences* (Second ed.). Hillsdale, NJ: Lawrence Erlbaum.

Dean, C. E. (1930). Audition by bone conduction. *The Journal of the Acoustical Society of America, 2*(2), 281-296.

Doll, T. J., Gerth, J. M., Engleman, W. R., & Folds, D. J. (1986). *Development of simulated directional audio for cockpit applications*: Wright-Patterson Air Force Base, OH: Armstrong Aerospace Medical Research Laboratory.

Folds, D. J. (2006). *The Elevation Illusion in Virtual Audio.* Paper presented at the Annual Meeting of the Human Factors and Ergonomics Society, San Francisco, CA.

Freeman, S., Jean-Yves, S., & Sohmer, H. (2000). Bone conduction experiments in animals – evidence for a non-osseus mechanism. *Hearing Research, 146*, 72-80.

Gilkey, R. H., Good, M. D., Ericson, M. A., Brinkman, J., & Stewart, J. M. (1995). A pointing technique for rapidly collecting localization responses in auditory research. *Behavior Research Methods, Instruments, & Computers, 27*, 1-11.

Glasberg, B. R., & Moore, B. C. J. (1990). Derivation of auditory filter shapes from notched-noise data. *Hearing Research, 47*, 103-108.

Hallmo, P., Sundby, A., & Mair, I. W. (1992). High-frequency audiometry: Masking of air- and bone-conduction signals. *Scandinavian Audiology, 21*(2), 115-121.

Howell, D. C. (2002). *Statistical Methods for Psychology* (5 ed.). Pacific Grove, CA: Wadsworth.

Hurley, R. M., & Berger, K. W. (1970). The relationship between vibrator placement and bone conduction measurements with monaurally deaf subjects. *The Journal of Auditory Research, 10*, 147-150.

Jahn, A. F., & Tonndorf, J. (1982). Lateralization of bone-conducted sounds. *American Journal of Otolaryngology, 3*, 133-140.

Kaga, K., Setou, M., & Nakamura, M. (2001). Bone-conducted sound lateralization of interaural time difference and interaural intensity difference in children and a young adult with bilateral microtia and atresia of the ears. *Acta Oto-Laryngologica, 121*(2), 274-277.

Kelley, N. H. (1937). Historical aspects of bone conduction. *Laryngoscope, 47*, 102-107.

Keppel, G. (1991). *Design and Analysis* (Third ed.). Upper Saddle River: Prentice Hall.

Kistler, D. J., & Wightman, F. L. (1992). A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction. *The Journal of the Acoustical Society of America, 91*(3), 1637-1647.

Kulkarni, A., Isabelle, S. K., & Colburn, H. S. (2001). Sensitivity of human subjects to head-related transfer-function phase spectra. *The Journal of the Acoustical Society of America, 105*(5), 2821-2840.

Langendijk, E. H. A., & Bronkhorst, A. W. (2002). Contribution of spectral cues to human sound localization. *The Journal of the Acoustical Society of America, 112*(4), 1583-1596.

MacDonald, J. A., Henry, P. P., & Letowski, T. R. (2006). Spatial audio through a bone conduction interface. *International Journal of Audiology, 45*(10), 595-599.

Makous, J. C., & Middlebrooks, J. C. (1990). Two-dimensional sound localization by human listeners. *The Journal of the Acoustical Society of America, 87*, 2188-2200.

Middlebrooks, J. C. (1999a). Individual differences in external-ear transfer functions reduced by scaling in frequency. *The Journal of the Acoustical Society of America, 106*(3, Pt. 1), 1480-1492.

Middlebrooks, J. C. (1999b). Virtual localization improved by scaling nonindividualized external-ear transfer functions in frequency. *The Journal of the Acoustical Society of America, 106*(3), 1493-1510.

Middlebrooks, J. C. (2000). Cortical representations of auditory space. In M. S. Gazzaniga (Ed.), *The new cognitive neurosciences* (pp. 425-436). Champaign, Illinois: Publication Services.

Middlebrooks, J. C., Macpherson, E. A., & Onsan, Z. A. (2000). Psychophysical customization of directional transfer functions for virtual sound localization. *The Journal of the Acoustical Society of America, 108*(6), 3088-3091.

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

Møller, H., Hammershøi, D., Jensen, C. B., & Sørensen, M. F. (1999). Evaluation of artificial heads in listening tests. *The Journal of the Audio Engineering Society, 47*(3), 83-100.

Møller, H., Sørensen, M. F., Jensen, C. B., & Hammershøi, D. (1996). Binaural technique: Do we need individual recordings? *The Journal of the Audio Engineering Society, 44*(6), 451-469.

Nolan, M., & Lyon, D. J. (1981). Transcranial attenuation in bone conduction audiometry. *The Journal of Laryngology & Otology, 95*, 597-608.

Oldfield, S. R., & Parker, S. P. A. (1984a). Acuity of sound localisation: A topography of auditory space. I. Normal hearing conditions. *Perception, 13*(5), 581-600.

Oldfield, S. R., & Parker, S. P. A. (1984b). Acuity of sound localisation: A topography of auditory space. II. Pinna cues absent. *Perception, 13*(5), 601-617.

Przekwas, A., Tan, X., Yang, H., Harrand, V., Wilkerson, P., Walker, B. N., et al. (2007). High-fidelity modeling tools for bone conduction communication systems. *Final Report for US Army STTR Phase I Project, submitted to the Army Research Office.*

Rayleigh, L. (1907). On our perception of sound direction. *Philosophical Magazine, 13*, 214-232.

Robinson, D. W., & Shipton, M. S. (1982). A standard determination of paired air- and bone-conduction thresholds under different masking noise conditions. *Audiology, 21*(1), 61-82.

Roffler, S. K., & Butler, R. A. (1968). Factors that influence the localization of sound in the vertical plane. *The Journal of the Acoustical Society of America, 43*(6), 1255-1259.

Schonstein, D., Ferré, L., and Katz, B. F. G. (2008, July). Comparison of headphones and equalization for virtual auditory source localization. Poster session presented at the 155th meeting of the Acoustical Society of America. Paris, France.

Searle, C. L., Braida, L. D., Cuddy, D. R., & Davis, M. F. (1975). Binaural pinna disparity: Another auditory localization cue. *The Journal of the Acoustical Society of America, 57*(2), 448-455.

Seeber, B. U., & Fastl, H. (2003, July 6-9, 2003). *Subjective selection of non-individual head-related transfer functions.* Paper presented at the 2003 International Conference on Auditory Display, Boston, MA (6-9 July).

Simpson, B. D., Brungart, D. S., Gilkey, R. H., Iyer, N., & Hamil, J. T. (2007). *Localization in multiple source environments: Localizing the missing source.* Paper presented at the 13th International Conference on Auditory Display, Montréal, Canada.

Small, S. A., & Stapells, D. R. (2003). Normal brief-tone bone-conduction behavioral thresholds using the B-71 transducer: Three occlusion conditions. *Journal of the American Academy of Audiology, 14*(10), 556-562.

Snik, A. F. M., Beyon, A. J., van der Pouw, C. T. M., Mylanus, E. A. M., & Cremers, C. W. R. J. (1998). Binaural application of the bone-anchored hearing aid. *Annuals of Otology, Rhinology, and Laryngology, 107*(3), 187-193.

Snik, A. F. M., Bosman, A. J., Mylanus, E. A. M., & Cremers, C. W. R. J. (2004). Candidacy for the bone-anchored hearing aid. *Audiology & Neuro-Otology, 9*, 190-196.

Snyder, J. (1973). Interaural attenuation characteristics in audiometry. *Laryngoscope, 83*, 1847-1855.

Stanley, R. M. (2006). *Toward adapting spatial audio displays for use with bone conduction: The cancellation of bone-conducted and air-conducted sound waves.* Master's Thesis, Georgia Institute of Technology, Atlanta.

Stanley, R. M., & Walker, B. N. (2006). *Lateralization of sounds using bone-conduction headsets.* Paper presented at the Annual Meeting of the Human Factors and Ergonomics Society, San Francisco, CA.

Stenfelt, S. P. Y. (2006). Middle ear ossicles motion at hearing thresholds with air conduction and bone conduction stimulation. *Journal of the Acoustical Society of America, 119*, 2848.

Stenfelt, S. P. Y., & Håkansson, B. (2002). Air versus bone conduction: An equal loudness investigation. *Hearing Research, 167*(1), 1-12.

Stenfelt, S. P. Y., Hato, N., & Good, R. L. (2004). Fluid volume displacement at the oval and round windows with air and bone conduction stimulation. *The Journal of the Acoustical Society of America, 115*(4), 797-812.

Teti, D. M., & McGourty, S. (1996). Using mothers versus trained observers in assessing children's secure base behavior: Theoretical and methodological considerations. *Child Development, 67*, 597-605.

Tonndorf, J. (1966). Bone conduction: Studies in experimental animals. *Acta Oto-Laryngologica, Suppl. 213*, 1-132.

Tonndorf, J. (1972). Bone Conduction. In J. V. Tobias (Ed.), *Foundations of Modern Auditory Theory* (pp. 195-237). New York: Academic Press.

Walker, B. N., & Lindsay, J. (2005). *Navigation performance in a virtual environment with bonephones.* Paper presented at the International Conference on Auditory Display, Limerick, Ireland.

Walker, B. N., & Lindsay, J. (2006). Navigation performance with a virtual auditory display: Effects of beacon sound, capture radius, and practice. *Human Factors, 48*(2), 265-278.

Walker, B. N., Stanley, R. M., Iyer, N., Simpson, B. D., & Brungart, D. S. (2005). *Evaluation of bone-conduction headsets for use in multitalker communication environments.* Paper presented at the Annual Meeting of the Human Factors and Ergonomics Society.

Walker, B. N., Stanley, R. M., & Lindsay, J. (2005). *Task, user characteristics, and environment interact to affect mobile audio design.* Paper presented at the The 3rd International Conference on Pervasive Computing: Pervasive2005, Munich, Germany.

Walker, B. N., Stanley, R. M., Przekwas, A., Tan, X., Chen, Z., Yang, H., et al. (2007, September 2-7). *High fidelity modeling and experimental evaluation of binaural bone conduction communication devices.* Paper presented at the 19th International Congress on Acoustics (ICA 2007), Madrid, Spain.

Wenzel, E. M. (1992). Localization in virtual acoustic displays. *Presence: Teleoperators and Virtual Environments, 1*(1), 80-107.

Wenzel, E. M., Arruda, M., Kistler, D. J., & Wightman, F. L. (1993). Localization using nonindividualized head-related transfer functions. *The Journal of the Acoustical Society of America, 94*(1), 111-123.

Wightman, F. L., & Kistler, D. J. (1989a). Headphone simulation of free-field listening. I: Stimulus synthesis. *The Journal of the Acoustical Society of America, 85*(2), 858-867.

Wightman, F. L., & Kistler, D. J. (1989b). Headphone simulation of free-field listening. II: Psychophysical Validation. *The Journal of the Acoustical Society of America, 85*(2), 868-878.

Wightman, F. L., & Kistler, D. J. (1997). Factors affecting the relative salience of sound localization cues. In R. H. Gilkey & T. R. Anderson (Eds.), *Binaural and spatial hearing in real and virtual environments* (Vol. 795, pp. 1-23). Mahway, NJ: Lawrence Erlbaum.