TOWARD ADAPTING SPATIAL AUDIO DISPLAYS FOR USE WITH BONE CONDUCTION: THE CANCELLATION OF BONE-CONDUCTED AND AIR-CONDUCTED SOUND WAVES

A Thesis
Presented To
The Academic Faculty

By

Raymond M. Stanley

In Partial Fulfillment
Of the Requirements for the Degree
Master of Science in Psychology

Georgia Institute of Technology

December, 2006

TOWARD ADAPTING SPATIAL AUDIO DISPLAYS FOR USE WITH BONE CONDUCTION: THE CANCELLATION OF BONE-CONDUCTED AND AIR-CONDUCTED SOUND WAVES

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. Elizabeth T. Davis
School of Psychology
Georgia Institute of Technology

Date Approved: August 18, 2006

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

π   The value pi, approximately 3.1416

cc   Cubic centimeters

SUMMARY

Virtual three-dimensional (3D) auditory displays utilize signal-processing techniques to alter sounds presented through headphones so that they seem to originate from specific spatial locations around the listener. In some circumstances bone-conduction headsets (bonephones) can provide an alternative sound presentation mechanism. However, existing 3D audio rendering algorithms need to be adjusted to use bonephones rather than headphones. This study provided anchor points for a function of shift values that could be used to adapt virtual 3D auditory displays for use with bonephones. The shift values were established by having participants adjust phase and amplitude of two waves in order to cancel out the signal and thus produce silence. These adjustments occurred in a listening environment consisting of air-conducted and bone-conducted tones, as well as air-conducted masking. Performance in the calibration condition suggested that participants understood the task, and could do this task with reasonable accuracy. In the bone-to-air listening conditions, the data produced a clear set of anchor points for an amplitude shift function. The data did not reveal, however, anchor points for a phase shift function – the data for phase were highly variable and inconsistent. Application of shifts, as well as future research to establish full functions and better understand phase are discussed, in addition to validation and follow-up studies.

CHAPTER 1

INTRODUCTION

There are a variety of reasons for using sound to convey information to a listener. These include conveying speech signals, as well as conveying information to a person whose eyes are busy or to a person who is visually impaired. Regardless of the application, auditory stimuli (sounds) are typically presented to a listener through air, using loudspeakers or headphones. Headphones allow private presentation of high-fidelity dichotic (stereo) sounds to a listener, without the perception changing as a person moves and turns, all in a portable package. On the other hand, there are problems with headphones. Covering the ears with headphones deteriorates detection and localization of ambient sounds in the environment. Those external sounds may be of particular interest in augmented reality and tactical situations, as well as for visually impaired users who rely on environmental audio cues as their primary sense of orientation. Furthermore, headphones do not allow auditory display to occur simultaneously with most types of hearing protection. 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; Tonndorff, 1972), bone conduction may lead to an acceptable solution. Although bone conduction of sounds occurs naturally in listening to one's own voice and to loud external sounds, it can also be directly transmitted through the skull via mechanical transducers. Presenting auditory information to listeners through bone conduction by placing vibrators on the skull can afford the same privacy and perceptual constancy that standard headphones offer, yet leave the ear canal and pinna

uncovered. This may facilitate improvement in the detection and localization of environmental sounds, and allows the display of auditory information even when hearing protection is inserted into the ear canal. Bone-conduction devices also cater to the preferences of users who would rather not have their ears occluded (Walker, Stanley, & Lindsay, 2005b).

The use of bone conduction transducers to deliver sound is not new. Because bone-conducted sound bypasses the middle ear and directly stimulates the cochlea, bone conduction is typically used in clinical audiology settings to assess the locus of hearing damage in patients. Developed for such clinical purposes, most bone-conduction transducers in production are not suitable for use in an auditory display: They typically consist of a single transducer, which is bulky and requires special equipment to drive it. Recently, compact binaural bone-conduction headsets have become available. Due to their potential for stereo presentation of sounds, their small size, comfort, and standardized input jack, these "bonephones" are much more suitable for implementation in auditory displays. The transducers of the very latest bonephones rest on the mastoid, which is the raised portion of the temporal bone located directly behind the pinna. The mastoid is a preferable transducer location relative to the forehead or temple (used in previous bone-conduction devices) because it contains the inner ear, is relatively immune to the interference associated with muscle tissue operating the jaw, and allows dichotic presentation of sounds.

Bone Conduction and Spatial Audio

Most of the psychoacoustics research and virtually all of the human factors research on auditory displays has assumed the conduction of sound through air (i.e., from

speakers or headphones), and thus has overlooked the alternative acoustic pathway of bone conduction. Because sound design guidelines established for air conduction do not necessarily apply to bone conduction, auditory display design needs to be re-evaluated for bone conduction. One type of auditory display that requires extensive research to implement with bone-conducted audio is a virtual three-dimensional (3D) auditory display. In this type of display, sounds are typically presented through headphones, after being processed to make them sound like they are originating from specific spatial locations outside the head (i.e., they are "spatialized"). Virtual 3D audio displays have gained recent popularity, due to their ability to increase detectability of signals amidst distracters and noise (e.g., Brungart & Simpson, 2002), as well as provide orientation cues in cases of vision loss (e.g., Walker & Lindsay, 2006). Spatializing audio signals for virtual 3D auditory displays is a complex process, based on considerable psychophysical research investigating how to manipulate acoustic cues to produce a reliable percept of sounds originating from different locations (see Blauert, 1983). Because spatialized audio is typically delivered through air-conduction via traditional headphones (which cover the ears), the perception of environmental sounds deteriorates. As a result, a tradeoff must occur between hearing spatialized audio and hearing external sounds when using regular headphones. This tradeoff is 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). The SWAN is just one example of a system that could benefit from presentation of 3D audio via bone conduction. However, there is little research on whether bonephones can

effectively replace headphones for the display of spatial audio, and how the audio would need to be processed to produce virtual sound source locations.

One approach to evaluating the potential effectiveness of bonephones for spatial audio is to just replace headphones with a pair of bonephones, use standard spatial audio filters developed for headphones, and just see how well people can perform the spatial audio task. Although this approach has shown that bonephones can produce some degree of spatial audio (Walker & Lindsay, 2005), higher performance and greater perceptual fidelity may be achieved if the processing applied to sounds for spatialization is customized for the bonephones. A substantial difference in optimal acoustic parameters between air and bone conduction is likely, given the very different mechanisms through which those pathways transmit sound to the cochlea. Air-conducted signals are filtered by the pinna and ear canal, as well as by the workings of the tympanic membrane (eardrum), and the ossicles in the middle ear. The ossicles connect to the cochlea at the oval window; this results in standing waves on the basilar membrane, which are converted into neural impulses by the hair cells (Sekuler, 2002). For the bone-conducted signal, however, the majority of the perception results from the signal traveling directly through the skull and shaking the cochlea to set up standing waves on the basilar membrane (Békésy, 1960; Yost, 1994). Further, the bone-conduction pathway does not need to accomplish the impedance-matching that air-conduction does, since bone is a denser material than air (Tonndorff, 1972).

The goal of the research presented in this document is to take an initial step towards identifying techniques for processing sounds presented through headphones so

that they can be customized for bonephones. With this information, spatial audio

displays can be tuned to be more effective with bonephones.

Relevant Research: Threshold Measurement and Related Discussion

A formal evaluation of using bonephones to present spatial audio requires

understanding the basic properties of the bone-conduction hearing mechanisms, spatial

audio cues, and how the virtual 3D audio is created. As with air-conduction hearing,

some basic information about thresholds of audibility has been gathered for bone

conduction. In particular, most research on bone conduction has focused on establishing

threshold norms for clinical testing of middle ear disorders. This clinical research has

yielded threshold curves such as those shown in Figure 1. The methods and implications

of this research can inform the design of research aimed at using bonephones in spatial

audio displays.

Figure 1. Bone-conduction thresholds from several researchers. The y-axis is in units that are the bone-conducted equivalent to the decibel measurement used for air-conduction (see text). These thresholds are used to define "normal" hearing in order to screen people for whether they have a middle ear disorder.

The y-axis units (dB) in Figure 1 are not exactly the same as the units used to describe hearing thresholds (i.e., dB SPL), because bone-conduction intensity levels cannot be measured by simply placing a sound level meter with a microphone up to the transducer. Rather, a standardized mechanical coupler that simulates the impedance of a human mastoid, an "artificial mastoid," picks up the vibration from the bone-conduction transducer. Within the artificial mastoid, the vibration is picked up by piezoelectric discs, which convert the vibration into a voltage that can be measured by the electronics in the sound level meter. Decibels are a ratio between two intensities, with the measured intensity in the numerator and the reference intensity in the denominator. The ratio of voltages from the artificial mastoid creates a decibel metric for bone conduction, just as a ratio of voltages sent from the microphone creates a decibel metric for air conduction. This makes it possible to directly compare decibel values between bone conduction and

air conduction. How those values match up depends greatly on the reference intensity chosen.

Long before standardized bone-conduction thresholds for clinical purposes were developed, Georg von Békésy (1960) completed some of the initial investigations into hearing through bone conduction. In addition to thresholds, Békésy's investigations included a wide variety of related topics: the specification of the nodes formed in the skull when vibrated, measurement of the linearity of sound transmission through skin, the resonant frequency of the ossicles, and the speed of sound through the skull. For threshold measurement, Békésy had listeners alter the phase and amplitude of waveforms until they cancelled each other out and produced silence. Specifically, listeners adjusted the phase and amplitude of a pure-tone wave presented through air-conduction until it cancelled out a static bone-conducted signal from a vibrator on the forehead. The change in amplitude needed to cancel out the wave in air was then taken as the threshold value for bone conduction. The phase adjustments were not reported in his publications (at least not the ones written in English). The focus of Békésy's interpretation of his findings was that cancellation could be done between air and bone, which suggested that air and bone conduction shared similar mechanisms, at least at some level.

Some modern threshold research has been more focused on applications to auditory displays. Specifically, the threshold curve for the bonephones has been plotted under a variety of listening conditions: Walker and Stanley (2005) conducted an applied assessment of how much relative power needs to be driven into bonephones for a listener to hear a sound at a variety of frequencies in various practically relevant listening conditions. Figure 2 shows the relative intensities of sounds sent to the bonephones for

the listener to detect, for each frequency and listening condition. Note that the top of the y-axis is 0 dB attenuation, which represents the maximum intensity sound. As the position on the y-axis descends from this maximum, the magnitude of the attenuation increases. Thus, lower points on the y-axis indicate that a quieter sound could be detected. Also note that because sound intensity was specified at the level of the input into the bonephones, this threshold curve represents the combined frequency response for both the bone-conduction hearing mechanisms and the bonephones device.

Figure 2. Threshold of Audibility: Masked, Open, and Plugged. Shown are threshold curves measured by Walker & Stanley (2005) with bonephones when ears were open, plugged, or masked with 45 dB pink noise. On the y-axis, lower position represents better sensitivity (i.e., detection of softer sounds). The attenuation specifies the input into the bonephones, where more attenuation means a less intense sound sent to the bonephones could be detected. The error bars represent one standard error above and below the mean.

These sensitivity specifications are useful because they can be used to optimize audio for the bonephones under the various listening conditions. These curves suggest that for equal detection performance, low and high frequencies need to be more intense than middle frequencies (i.e., 570 – 1850 Hz). Indeed, subjective listening experience with unaltered sound played through the bonephones suggests that low-frequency sounds are typically not loud enough and the midrange frequencies sounds are too loud. Essentially, a different equalization setting is needed, due to the differences between air and bone conduction. These differences include disparities in the auditory mechanism through

9

which sound travels as well as the physical properties of the device used to deliver the audio.

A description of the relative intensities sent to the bonephones in order to detect a signal (Walker & Stanley, 2005) is helpful in understanding purposeful spectral changes that can be made to sounds as part of processing them for spatialization. That research was the first in a potential series of investigations that could lead to a detailed description of the signal processing that needs to be applied for the spatialization of sounds played through the bonephones.

Acoustic Cues For Spatial Separation

The two acoustic cues producing the perceptual experience of lateralization[1] are interaural level differences (ILDs) and interaural time differences (ITDs) (Yost & Hafter, 1987). In order to implement spatial audio with the bonephones, sensitivity to these basic spatial audio cues must exist. Until recently, many researchers have assumed that spatial audio with bone conduction is not possible, because the interaural attenuation, and thus the maximum ILD, was not considered sufficient (Blauert, 1983; Goldstein & Newman, 1994).

On the other hand, Audiology handbooks indicate that bone-conducted interaural attenuation (BC IA) may be greater than zero, and as much as 20 dB, though audiologists often assume its lower bound estimate of zero dB (e.g., Katz, 2002). There have been few investigations into BC IA, and these have focused on the possibility of detecting a tone in

---

[1] Lateralization involves space in only one dimension (spatial separation within the head), whereas spatialization involves space in three dimensions (invoking the percept of a sound outside the head and in the vertical dimension). Lateralization is a logical first step toward 3D audio via bonephones, because if lateralization is not possible, then spatialization is not possible.

the ear that is contralateral to the ear being tested (e.g., Hood, 1960; Liden, Nilsson, &

Anderson, 1959). The language of these resources suggests that in audiometry, for BC

IA, the "worst-case scenario" is more important than what the empirical evidence alone

reveals. This conservatively-biased estimate of BC IA is appropriate for clinical purposes

where erring on the side of caution is preferred. For the purposes of adapting spatial

audio filters for air-conduction so that they are suitable for bonephones, however, a

neutral approach is more suitable.

New information about sensitivity to interaural differences delivered through

bone conduction gives a different perspective than typical audiology guidelines on the

level of BC IA. In a direct assessment of sensitivity to interaural differences, Kaga,

Setou, and Nakamura (2001) found that the subjective report of image lateralization

systematically depended on interaural differences delivered through binaural application

of clinical bone-conduction vibrators. The researchers showed sensitivity to ILDs and

ITDs in children with normal hearing, as well as in children with abnormalities of the

middle and outer ears. Furthermore, in participants with normal hearing, these

sensitivities were not significantly different from ITDs and ILDs assessed through air-

conduction (See Table 1 for threshold values).

Table 1

ILD and ITD Discrimination Thresholds for Otologically Normal Children, from Kaga et al. (2001).

Type of Receiver	ILD threshold (dB)	ITD threshold (μs)
Bone Conduction	4.9 ± 2.3	171.7 ± 72.3
Air Conduction	*5.5 ± 2.1*	*186.7 ± 50.1*

The air-conduction interaural thresholds found by Kaga et al. (2001) are higher than many other estimates by psychoacoustics researchers. The reason for the discrepancy between Audiology and Psychoacoustics research may be due to differences in methods. Kaga and colleagues' methods used a Bekesy-tracking procedure for the detection of lateralization from center. In this procedure, the ILD or ITD was constantly increasing until the participant pressed a button. The average point at which this button is pressed over repeated observations is taken as the threshold of lateralization. Those doing psychoacoustics research, on the other hand, often ask listeners to make a discrimination between a test and reference stimulus, at varying ILDs or ITDs (e.g., Yost & Dye, 1988).

Nevertheless, Kaga and colleagues (2001) demonstrated that there may be more binaural separation than typically thought possible through bone conduction, even though the mechanisms underlying this binaural separation are not clear. In terms of BC IA, the bone-conducted ILD threshold of 4.9 dB suggests that BC IA is at least on the order of 5 dB. The binaural separation observed suggests that spatial (or at least lateralized) audio

with bone conduction may actually be possible.

Bonephones have also been studied with a more objective and applied task that indirectly assesses sensitivity to ILDs and ITDs. In particular, the Coordinate Response Measure (CRM) task (Bolia, Nelson, Ericson, & Simpson, 2000) was used by Walker, Stanley, Iyer, Simpson, and Brungart (2005a) to assess the efficacy of using spatial separation to enhance speech intelligibility in multi-talker communications environments. The CRM task requires listeners to correctly identify a spoken color name and number embedded in a carrier phrase, among many other similar distracter phrases. The extent to which a listener can correctly identify color-number combinations in the carrier phrase can then be interpreted as the listener's ability to selectively attend to a single channel while filtering out extraneous channels. Spatial separation of the target channels from the distracter channels improves performance on this task in a systematic manner (Brungart & Simpson, 2002). Walker and colleagues (2005a) compared performance on the CRM task across three listening conditions: headphones, bonephones with open ears, and bonephones with plugged ears. Increasing performance on the CRM task as a function of both ILDs and ITDs corresponding to lateralization of the sound suggested sensitivity to interaural cues, implying that spatial separation is possible with the bonephones (see Figure 3).

Panel a

Panel b

Figure 3. Performance on the CRM task as a Function of ILDs and ITDs from Walker et al. (2005a). Panel a shows ILDs, Panel b shows ITDs. Although performance is consistently superior with headphones, the monotonically increasing performance with bonephones as a function of both ILDs and ITDs suggests some degree of reliable segregation can be achieved. The error bars represent the 95% confidence intervals around each data point.

Figure 3 also shows that performance with headphones was consistently superior to performance with bonephones. Nonetheless, the monotonically increasing performance as a function of interaural differences in the bonephone conditions suggests that reliable segregation could be produced with bonephones. This indicates that bonephones may be suitable for displays that require spatial separation, such as multi-talker communication displays (Brungart & Simpson, 2002) and navigational aides for the blind (Walker & Lindsay, 2006). Together, these studies (e.g., Kaga et al., 2001; Walker et al., 2005a) suggest that there is sufficient interaural attenuation to facilitate interaural cues for spatial audio. This sensitivity to interaural differences suggests that there must be some interaural attenuation with bone conduction. Research on air-conducted interaural attenuation (AC IA) can provide estimates of BC IA.

Audiologists have sought out measurements of AC IA to determine when masking is needed to prevent the response of the non-test ear (NTE) being involved in the patients' response to pure tones. The NTE is the ear that should not contribute to the threshold response, and the test ear (TE) is the ear for which the threshold is being determined. In the case of a signal that exceeds AC IA, the TE could appear to have higher sensitivity than it actually does, because the NTE is contributing to the response due to cross-hearing. Audiologists estimate AC IA values to be 60 dB on average (Katz & Lezynski, 2002). Although some cross-hearing is due to air leakage and physiological crosstalk, the majority of cross-hearing with airborne signals is actually due to bone conduction (Studebaker, 1962; Wegel & Lane, 1924; Zwislocki, 1954).

Estimates of BC IA based on AC IA confirm that there may be a greater degree of interaural attenuation through bone than initially thought. Specifically, air conduction is estimated to transfer to bone conduction as the sound level exceeds 40 dB (Békésy, 1960; Palva & Palva, 1962). Given the average AC IA value of 60 dB, subtracting the amount of energy it takes for airborne sound to be conducted through bone yields a value of 20 dB. Subtracting the amount due to physiological crosstalk ("central masking") – 5 dB – yields a final estimated BC IA value of 15 dB, which is a value greater than traditionally considered for bone conduction.

Determining the Bone-to-air Shifts

The evidence discussed to this point suggests that spatial audio with bonephones may be possible. There have already been techniques established for air-conduction to produce virtual 3D audio displays. These techniques make changes in a waveform's frequency, amplitude, and phase components to produce the "spatialized" percept (this will be discussed in further detail below). A function of shift values could be used to adapt these signal-processing techniques so that they are suitable for use with bonephones. The purpose of this research was to collect an initial data set of shift values that provides anchor points for this sort of function. The purpose of collecting only anchor points was to conduct an initial investigation that established appropriate methodology for finding the shift values and determine factors that affect these shift values. The data were gathered using methods informed by the research that has been discussed up to this point. In particular, Békésy's (1960) technique for measuring thresholds was leveraged to measure bone-conducted interaural attenuation.

The method is most similar to Békésy's (1960) technique for finding thresholds and for showing that air and bone conduction share hearing mechanisms. In this technique, a participant adjusts the phase and amplitude of a signal in one ear so that it cancels out the other signal in the same ear, thus producing silence (or at least a significant reduction in volume). The multidimensional nature of this method takes into account both the phase and amplitude of waveforms interacting through the skull. This method also takes advantage of the ear's presumably higher sensitivity to concurrent changes in amplitude of a combined wave, rather than detecting the presence of a near-threshold tone.

To explain how this method yields data that can be used to optimally "spatialize" a signal through bonephones, I will first review how simulating spatial audio through air-conduction headphones is accomplished. A modern method of delivering spatial audio through air-conduction is dependent on measures of what happens to a broadband sound between its source and immediately before it arrives at the eardrum. Small microphones are placed in the ear canal of a "normalized" mannequin head and torso. The description of how the sound changes between the source and the microphones is known as the Head-Related Impulse Response (HRIR), and is measured for each ear. The Fourier transform of the HRIR specifies how to change a given sound so that it produces the same change that occurred in the impulse response. This Fourier transform is known as the Head-Related Transfer Function (HRTF). This HRTF can be convolved with any monaural sound source to yield a replica of what would happen if that sound source had come from a particular location in the space outside the head (Duda, 2000). The HRTF consists of amplitude and phase information across a continuous range of frequencies,

and thus together with another ear's HRTF, includes all the cues to localization of sound sources in a free field. These cues include ILDs, ITDs, and the spectral filtering imposed by the pinna, head, and torso. HRTFs are collected at many combinations of azimuth and elevation, all usually at about 1 meter from the head.

There is no equivalent procedure for bone conduction, because there is no real-world occurrence of bone-conducted audio that would not shake the whole head, rather than deliver bone-conducted signals from a concentrated location near the cochlea (i.e., the mastoid). Therefore, the best way to simulate a sound source's spatial location through bonephones is to find out how to alter HRTFs for air so that they provide the same perceptual experience through bonephones. This would allow externalized spatial audio to be invoked on bonephones, utilizing the numerous data sets of generalized HRTFs already available (e.g., Algazi, Duda, Thompson, & Avendano, 2001). Like all waveforms, the HRTF produces a waveform that can be defined at any given frequency component in terms of its phase and amplitude. The goal for this research is to specify how the phase and amplitude of a bone-conducted sound needs to be adjusted to match an air-conducted sound. With an understanding of how the physical parameters that define a waveform need to shift for bone conduction, a preliminary series of "bone-to-air" shifts can be defined. Once future studies establish a sufficiently large set of shifts, a full adjustment function can be mapped out. These functions can then be applied to adjust pre-existing HRTFs designed for air-conduction so that they are suitable for bonephones.

CHAPTER 2

METHOD

Explanation of Conditions

Each participant experienced two listening conditions: bone-to-air and air-to-air.

These conditions produced a set of shifts that relate bone-conducted waves and air-

conducted waves at particular frequencies. Schematics of these conditions are shown in

Figure 4.

Figure 4. The conditions administered during this experiment. In this schematic, the left ear is always the test ear (TE), and the right ear is always the non-test ear (NTE). The bone-to-air conditions were meant to yield a subset of the amplitude and phase shifts that would need to be applied to air-conducted HRTFs. The air-to-air condition served as calibration, making sure that participants understand the task. The band-stop noise was provided in the TE to mask harmonics, and the band-pass noise was provided in the NTE to efficiently and accurately remove the contribution of the NTE to responses.

The schematic of the conditions in Figure 4 shows the left ear as the test ear (TE) and the right ear as the non-test ear (NTE). Panel a shows the bone-to-air condition, in which the participant received a bone-conducted tone, an air-conducted tone, and band-stop noise in the TE. The band-stop noise was delivered in the TE to mask the harmonics that occurred outside of the pure tone frequency that was being delivered. In the NTE, the participant received band-pass noise to remove the response of the NTE from the perceptual judgment being made. For both conditions, the participant adjusted the phase and amplitude of a tone in the TE until it canceled out the other tone in the TE.

The bone-to-air condition specifies how to match a bone-conducted signal that has passed through the mastoid and arrived at the cochlea to an air-conducted signal on the same side of the head. This indicates the phase and amplitude shift that needs to occur in the HRTF at a particular frequency. For this study, only a set of three frequencies are tested. Future studies can test more frequencies, to establish full functions. These functions would relate bone conduction to air conduction in the whole range of audible frequencies, and provide a way to adjust air-conducted HRTFs so that they are more suitable for bonephones.

The air-to-air condition was administered to calibrate for participants and equipment. Two identical waves were sent to the same air-conduction earphone. Thus, they should have required equal amplitude and 180 degrees of phase shift to cancel each other out. Any deviation from these values represents error due to a participant's lack of ability to understand or complete the task, or a consistent error in the equipment. No feedback on performance was given. Band-stop noise was again delivered in the TE to

mask the harmonics. Band-pass noise was also delivered again to the NTE so that it did not contribute to the response that is assumed to be from the TE only.

Participants

There were 10 volunteer participants (six males, four females, mean age = 25.8, age range = 23-28) from the graduate student community of the Georgia Institute of Technology. They were required to have normal hearing (sensitivity to 20 dB pure tones), as tested by an audiometer.

Stimuli

Each of the two listening conditions was tested with three pure tones at the following frequencies: 500, 3150, and 8000 Hz. These frequencies were chosen because they tap several different spatial hearing mechanisms (Yost & Dye, 1988). The lowest tone, 500 Hz, is where ITDs thresholds are lowest. The 3150 Hz tone represents the frequency range where bone-conduction thresholds are low, air-conducted localization performance is weak, and in which speech sounds occur. These two frequencies (500 and 3150 Hz) also avoid the resonant frequencies of the skull, namely 400 and 2000 Hz (Zwislocki, 1954). The 8000 Hz tone represents a frequency domain where ILDs function best, the top end of a range where optimal localization performance occurs in the higher frequencies, as well as a frequency domain where spectral changes caused by particle-like reflection off the pinna and torso occur. Table 2 shows the initial sound pressure level (air) and acceleration level (bone) at which the tones were delivered. These levels were established by preliminary testing, setting the levels at the point at which the tone was first clearly audible.

Table 2

Physical Output of Tones

Center (Hz)	Air-conducted[1]	Bone-conducted[2]
500	72.9	24.8
3150	50.1	18.2
8000	48.5	19.6

[1]dB re 20µPa, also known as dB SPL   [2]dB re 3.16 cm/s^2 (acceleration)

The lower and upper bound of amplitude and phase values were established based on preliminary testing. Specifically, they were set based on putting preliminary cancellation values in the middle. The apparatus used (described under the "apparatus" heading) was such that the amplitude and phase adjustments changed in set intervals, or "steps". The interval of stimulus change was chosen based on a value that allowed relatively fast adjustment without reducing precision to a point that a cancellation point ly between two successive steps. The step size can be seen in Table 3. Physical output equivalents to the amplitude scalers can be seen in Appendix B.

Table 3

Amplitude and Phase Adjustment Step Sizes

Listening Condition	Adjusting	500 Hz	3150 Hz	8000 Hz
		Amplitude step size		
Bone-Air	Bone	0.01333	0.01000	0.03000
Bone-Air	Air	0.00300	0.00267	0.00233
Air-Air	Air	0.00147	0.00187	0.00193
		Phase step size		
Bone-Air	Bone	0.02693	0.03590	0.07181
Bone-Air	Air	0.02693	0.03590	0.07181
Air-Air	Air	0.07181	0.07181	0.07181

The pure tones were played in a cyclical on-off pattern: on for one second and off for one second. A visual indicator in the software interface (see "apparatus" section) showed when the tones were playing and when they were not. A non-continuous duration was chosen to ensure that the tones were arriving at the TE simultaneously, before the phase adjustment was made to cancel out the waves. This periodic pattern also made it easier for participants to detect whether or not the tone was audible. The on-off pattern played until the participant had finished adjusting their phase and amplitude for cancellation.

The maskers had the ANSI-defined 1/3 octave stop- or pass-band centered on the pure tone frequency being tested (ANSI, 2004). Table 4 shows the upper and lower bound of the maskers' frequency bands. The bandwidth of the maskers is larger than the critical band to avoid confusing signal and masker, but narrow enough to avoid the loudness associated with broadband noise (Katz & Lezynski, 2002). White noise was filtered through 4-pole Butterworth filters built in Matlab, producing a new wave file with the desired spectral components. The sound pressure level output of the maskers can also be seen in Table 4. These levels were established by setting a general range based on previous literature to avoid cross-masking and threshold shifts, and then using preliminary testing to fine-tune to a level that was comfortable yet provided enough pressure to mask an audible tone.

Table 4

1/3 Octave Bands For Narrow-band Masking Noise, as Specified in ANSI S3.6-2004

Center (Hz)	Lower Limit (Hz)	Upper Limit (Hz)	Stop dB_A	Pass dB_A
500	445	561	66.1	58.0
3000	2670	3370	59.5	57.9
8000	7130	8980	60.2	58.5

Note: See Appendix A for more details on measurement of stimuli at the physical level.

Measuring in decibels for bone conduction required the artificial mastoid apparatus described earlier. The output of both the headphones and bonephones were measured at a variety of levels, so that a function could be plotted between input values (specified by the software) and the device output values. The software-specified input values that the participant submitted were recorded in a file. These input data were processed by computing the resultant output value, based on the functions obtained during measurement. See Appendix A for detailed description of the measurement procedure completed, and the resulting measurements of physical output coming from the bonephones and air-conduction headphones.

Apparatus

The bone-conducted tones were delivered through a pair of HG-28 stereo bone-conduction headsets, or "bonephones" produced by Temco, which place the vibrators on the mastoid. The air-conducted tones were delivered through Sennheisser MX400 earbud-style headphones inserted into the ear.

Participants adjusted the amplitude and phase of tones by way of a Griffin Powermate rotary knob input device. The rotary knob altered amplitude or phase parameters that were passed to the online generation of sine waves within SuperCollider™, a sound programming language for real-time audio synthesis running on the Macintosh OS X operating system. SuperCollider™ was used to create a graphical user interface (see Figure 5) that allowed participants to submit their final phase and amplitude adjustments that led to cancellation, as well as provide controls for the experimenter to manipulate which condition was being tested. The graphical interface also provided sliders indicating the adjusted phase and amplitude values relative to the

maximum and minimum values, but did not have an indication of the absolute phase or

amplitude. Initial testing done without visual indicators of phase and amplitude values

suggested that the task was much too difficult to complete without the familiarity of a

slider that people are used to having when interacting with modern computers.

Figure 5. SuperCollider™ interface provided to participant. The slider for amplitude
adjustment can be seen on the left in a vertical orientation. The slider for phase can be
seen at the bottom left in a horizontal orientation. At the top of the interface is a start/stop
button; in the middle is a visual indicator of the tone playing; below that is a toggle for
controlling phase or amplitude with the Powermate rotary knob. All buttons for the user
were controlled by labeled keys on the numeric pad of a keyboard. Experimenter controls
are the gray buttons and drop-down menus at the right of the interface.

The masker in the TE was also delivered through the computer, with a wave file

being played by Quicktime. The sound delivery apparatus, set up for delivering the bone-

to-air condition, can be seen in Figure 6. The two channels from an Apple G5 PowerMac

26

computer were sent out through the sound card optically to an M-Audio SuperDAC 2496 digital-to-analog converter, and then to a Behringer Powerplay PRO-8 HA-8000 headphone amplifier. From there, the two channels were directed to their appropriate channel on the appropriate device for the condition being measured (refer to Figure 4). The masker in the NTE (a third channel) was generated by a separate Sony DVP-NS575P compact disc player playing recorded band-stop noise, with a separate track for each frequency center. The output of the compact disc player was sent to the same headphone amplifier as the tones, and directed to the appropriate channel of the earphone.

Figure 6. Sound delivery apparatus used in this study to deliver the bone-to-air condition. In this schematic, the left ear is the test ear, and the right ear in the non-test ear. The tones and bandstop noise originate from the computer and are sent to the left ear (the test ear) of the appropriate transducers. The bandpass noise originates from a CD player and is sent to the headphone on the right ear (the non-test ear).

Procedure

A schematic of a sample procedure at the top level and within a block can be seen in Figure 7, Panels a and b. The experiment consisted of two sessions, each lasting between 45 minutes to two hours. Every session included a screening for normal hearing, which was conducted with a Micro Audiometrics Corp DSP Pure Tone Audiometer.

28

Hearing was monaurally tested with a 500 Hz, 3000 Hz, and 8000 Hz tone at each ear. Testing at each ear and frequency began by playing the tone through the audiometer into its TDH headphones at a setting indicating an output level of 20 dB HL. The participant raised their hand in the ear that they heard the sound. The intensity of the tone was then decreased in 5 dB increments until they did not raise the appropriate hand again. The last tone at which the participant raised their hand in the correct ear was taken to be the participant's audiometric threshold. Every participant tested had a threshold of 20 dB HL or less at every frequency and ear.

The hearing test was followed by one block of calibration via the air-to-air condition (see Figure 7, Panel b), and then a block of the bone-to-air condition (see Figure 7, Panel a). Each block began with 10 practice trials at a constant but randomly chosen frequency (from the set of those being tested) and constant but randomly chosen test ear (left or right). For practice and experiment trials, the participant was instructed to adjust amplitude until a slight increase in loudness had occurred, adjust phase until the combination tone was softest, and then go back and tweak amplitude and/or phase until the sound was as quiet as possible. The phase and amplitude each began at zero. The values began at zero because preliminary testing revealed that starting participants anywhere else, or randomizing the starting point, confused participants about the amplitude and phase manipulation they were controlling. In addition, starting participants at a phase of zero encourages finding the phase shift for cancellation that lies closest to zero, in case there are multiple cancellation points.

Figure 7. The procedure for a sample participant is shown in Panel a; Panel b shows a sample block. Panel a shows that the experiment consists of two sessions, each consisting of a calibration block and the bone-to-air condition. In one session, the bone-to-air condition involved the participant adjusting the air-conducted tone while the bone-conducted tone remained constant. In the other session, the bone tone was adjusted. Panel b shows that each block began with ten practice trials and then proceeded for six runs, each consisting of five adjustments for one test ear at a given frequency. The first test ear in each pair of runs was randomly selected and was followed by a run with the same frequency but with the other ear as the test ear. After a run pair was completed, a new frequency was randomly selected along with the first test ear for the next run pair.

Once the amplitude and phase had been adjusted so that the resultant tone was as soft as possible, the participant submitted the values by pressing a button on the keyboard. This marked the end of a trial, terminating the sound. The data trials consisted of five phase/amplitude adjustments at each ear, for three frequencies, yielding a total of 30 phase/amplitude adjustments per block. The TE was blocked and counterbalanced, and both ears were tested before moving to the next randomly-selected frequency. In one session, the bone-to-air condition involved the participant adjusting the air-conducted tone, while the bone-conducted tone remained at 0° phase and a constant amplitude (the physical output at this constant amplitude can be seen in Table 2). In the other session, the bone-to-air condition involved the participant adjusting the bone-conducted tone, while the air-conducted tone remained at 0° phase and a constant amplitude (again, the physical output at this constant amplitude can be seen in Table 2). This adjustment manipulation (air or bone) was done to make sure that there were not significant changes in phase and amplitude adjustments as a result of which tone the participants were adjusting. The order of sessions was counterbalanced between participants.

CHAPTER 3

RESULTS

Air-to-Air: Overview

The air-to-air condition was administered to assure that participants could do the task, and to assess the degree of error associated with their judgments. This condition had the unique quality of having a physically correct answer that corresponded to a subjective experience, unlike the remainder of this study, which relied on reports of subjective experience. Error was defined in terms of the deviation from the physically correct values for cancellation of two waves passing through the same medium: equal amplitude and 180 degrees phase. For both amplitude and phase, the error was standardized by the interval that the slider moved on. Thus, the error was defined in terms of the number of steps away from the correct answer. This standardization equated errors across frequencies and accounted for the different points between steps that participants could navigate to with the mouse[2].

The error was reported at the level of each participant, because of the large amount of variability between participants in the level of error, both for amplitude and

---

[2] The steps away from the correct value were not always whole numbers. This is due to the ability of participants to use the mouse to move large distances. When the slider was moved with the mouse, it was not restrained to the intervals that the rotary knob input was. However, the participants were instructed to begin with gross unrestrained adjustment with mouse, and then finish by making set-interval adjustments via the rotary knob. Because of this combination of variable-interval and set-interval adjustment, any value less than half a step away from the correct point was the closest to correct that a participant could get. If the difference between the adjusted value and the correct value was greater than .5, the participant could have moved the slider to a consecutive or previous step with the knob and been closer to the correct value. Thus, there was some error induced by the method of adjustment itself.

phase. Practice trials were not considered in the data, and data were collapsed across sessions and all trials (resulting in 60 trials total).

Air-to-Air: Amplitude

The amplitude adjustment (in scaler units) was converted to a standardized "step" value by multiplying the amplitude adjustment by the step size (see Table 3 for step size values). The step size was computed by dividing the total scaler range by the total number of steps (30). The correct standardized step value was 15, which was halfway between the top and bottom of the slider, and the point at which the scaler for each tone was equal. The step error was computed by subtracting the adjusted value from 15. Then to compute the aggregate root-mean-square (RMS) step error metric, the step error values were squared, averaged across trials, and then the square root of the mean for each participant was computed. The RMS standard deviation (RMS SD) error metric was computed by taking the standard deviation of the squared step errors across trials, and then taking the square root of that value.

The RMS step error for amplitude, for each participant, can be seen in Table 5. Eight out of the ten participants had RMS step errors less than one, and the remaining two had RMS step errors less than three. The standard deviation values indicated that participants' error was generally consistent: With exception to participants 7 and 8, people were on average less than one step away from their RMS step error.

Table 5

Air-to-Air Amplitude Error (RMS Step Error)

Participant	RMS Step Error	RMS SD Step Error
1	0.1	0.4
2	0.0	0.0
3	0.4	0.6
4	0.6	1.0
5	0.2	0.4
6	0.1	0.2
7	1.3	2.0
8	1.6	2.7
9	0.2	0.3
10	0.0	0.0

Air-to-Air: Phase

Calculations similar to the RMS step error for amplitude were completed to create an RMS step error metric for phase. First, the error was computed for each data point by subtracting the adjusted phase value from the correct value (positive or negative π radians). Because cancellation could occur at positive π radians or negative π radians, the absolute values of the phase adjustments were used. This error was then divided by the step size to produce the step error metric. The step size was measured by computing the

difference between consecutive phase values recorded by the software when adjusting the rotary knob (see Table 3 for step size values). The step error values were then all squared and subsequently averaged across trials within each participant. The square root of this mean was then computed to yield the final RMS step error metric. The RMS standard deviation (RMS SD) error metric was again computed by taking the standard deviation of the squared step errors across trials, and then taking the square root of that value.

There was a range of 88 steps on each side (positive and negative) of the phase slider. The RMS step error for phase for each participant can be seen in Table 6. The RMS step error for half the participants was less than one, and for the other half less than five. In terms of the variability of participants' error, all but two were on average nine or fewer steps away from their RMS step error.

Table 6

Air-to-Air Phase Error (RMS Step Error)

Subject	RMS Step Error	RMS SD Step Error
1	0.4	0.5
2	2.9	8.0
3	5.0	12.0
4	4.9	11.5
5	2.0	5.5
6	0.4	0.5
7	0.8	1.4
8	3.9	9.0
9	0.3	0.4
10	0.4	0.3

Bone-to-Air

Subjective Report of Cancellation

The subjective report of cancellation was solicited from participants during the experiment. A detailed description of perceptual experience is difficult given the qualitative nature of these data. In general, however, they reported cancellation comparable to that of the air-to-air condition. Participants reported (without prompt) that cancellation in the bone-to-air condition was very sensitive to head movements. There was some variability in the degree to which people said they achieved cancellation. In

36

particular, there would be an occasional run where the participant had difficulty canceling the waves. In these cases, the experimenter confirmed that the participant was at least reaching a combination of phase and amplitude where if either was adjusted to a different point, the tone got louder (i.e., a local minimum in the loudness space). One participant (beyond the 10 participants whose data was analyzed) failed to find this "trough" of loudness at one point in the experiment. The experimenter discontinued that participant's experiment session, and that participant's data were excluded from analysis.

Amplitude Adjustments

Measurement

The amplitude value at which cancellation occurred was recorded whenever the participant pressed the "submit" button. This amplitude value was specified in terms of scaler input into the bonephones at the level of the software. The function relating the scaler input to the physical output of the bonephones was measured for each frequency and transducer. The unit of physical measurement was acceleration. The units of acceleration were dB re 3.16 cm/s^2. This reference value (3.16 cm/s^2) is the British standard for the human threshold of bone-conduction hearing (Bruel & Kjaer, 1974). This reference value was chosen in an attempt to establish a scale that was comparable to the standard scale used for air-conduction (where the reference value is also thought to be the threshold for hearing). The function relating the scaler input and the physical output of the earphones was also established, in the standard units of pressure used to measure air-conducted sounds: dB re 20 microPascals (Rossing, Moore, & Wheeler, 2002). Details of the measurements can be found in Appendix A. The phase and amplitude cancellation values were converted to a physical measurement of the output using the scaler-output

functions established in measurement. The final amplitude shift value was computed by subtracting the Bone dB from the Air dB. Subtraction was done in this direction to yield positive shift values, because the values of Bone dB at cancellation were always less than the values of Air dB. Reasons for this difference will be discussed later in this paper.

Description of Results

A 2x2x3 within-subjects ANOVA was conducted on amplitude shift, with pathway adjusted (bone or air), ear (left or right), and frequency (500 Hz, 3150 Hz, or 8000 Hz) as within-subjects independent variables. A standard alpha level of 0.05 was used throughout the analyses for this study. The results of the ANOVA can be seen in Table 7. This analysis revealed a statistically significant main effect of pathway and frequency, and a statistically significant two-way interaction between frequency and ear. The three-way interaction (pathway x ear x frequency) and the two-way interactions (pathway x ear, pathway x frequency) did not reach statistical significance. There was no main effect for ear.

Table 7

Analysis of Variance for Amplitude Shift

Source	df_{Effect}	df_{Error}	MSE_{Effect}	MSE_{Error}	F	p
Pathway (P)	1	9	84.50	10.77	7.85*	.02
Frequency (F)	2	18	5001.72	56.09	89.17*	< .01
Ear (E)	1	9	10.62	26.35	0.40	.54
P X F	2	18	18.53	8.53	2.17	.14
P X E	1	9	0.92	6.73	0.14	.72
F X E	2	18	192.79	15.96	12.08*	< .01
P X F X E	2	18	2.19	13.00	0.17	.85

*$p < 0.05$

The main effect of pathway on amplitude shift was such that the amplitude shift was consistently higher when adjusting air ($M = 32.1$, $SE = 0.86$) than when adjusting bone ($M = 30.4$, $SE = 0.92$). The interaction between frequency and ear can be seen in Figure 8. This figure shows that the relationship between the left and right ear amplitude shift depended on frequency.

Figure 8. Interaction between ear and frequency on amplitude shift. Amplitude shift values shown here are averaged across pathway adjusted. The error bars represent one standard error above and below the mean.

The main effect of frequency can be seen in Figure 9, which suggests that the amplitude shift was greater at 500 Hz than the other two frequencies, but that 3150 Hz and 8000 Hz had similar amplitude shifts. Post-hoc comparisons confirmed this trend. Specifically, paired comparison *t*-tests were run and then assessed using Tukey's *q* statistic. There was a statistically significant difference between 500 Hz and 3150 Hz, and between 500 Hz and 8000 Hz, but no statistically significant difference between 3150 Hz and 8000 Hz (see Table 8).

Figure 9. Main effect of frequency on amplitude shift. The error bars represent one standard error above and below the mean.

Table 8

Post-hoc Tests for the Main Effect of Frequency on Amplitude Shift.

Comparison	df	t	p[1]
500 - 3150	9	12.24*	< .05
500 – 8000	9	9.26*	< .05
3150 - 8000	9	-0.04	> .05

[1] exact *p* values are not available because a critical *t*-value from a look-up table was used, rather than an exact calculation by a computer program. This look-up method was necessary in order to do a Tukey adjustment for paired comparisons. This adjustment involved looking up the critical *q* value for 3 means and 9 degrees of freedom at the 0.05 alpha level ($q = 3.95$), and then converting it to an equivalent t-value ($t = q/\sqrt{2}$): 2.792.
*$p < 0.05$

Phase Adjustments

Data Processing

      The phase value at which cancellation occurred was specified in terms of radians at the level of the software. The median phase value across trials for each combination of pathway, ear, and frequency was computed. The median was used because it avoided several pitfalls of using the mean: it avoided the heavy influence of outliers, averaging across positive and negative values that would create an unrepresentative mean value, and misrepresenting distributions of responses that were not unimodal. Since the median value was often negative, the values were shifted by a constant number to avoid averaging negative numbers. The constant number's value was eight radians. This value was the smallest whole number that could assure that only positive values greater than one would occur. After inferential statistics were run on phase in terms of radians, the phase was converted to degrees for descriptive statistics.

Description of Results: Degrees

      Figure 10 shows the average median phase shift across ears, separated by pathway, for each frequency. These figures show little difference in phase shift between adjusting air and adjusting bone ("pathway") at 500 Hz, and complementary values for 3150 Hz and 8000 Hz. Most importantly, they show that there is a large difference in the amount of variability between frequencies.

Figure 10. Average median phase shift across ears, separated by pathway. The error bars represent one standard error above and below the mean.

Indeed, when a 2x2x3 (pathway x ear x frequency) within-subjects ANOVA was conducted on phase in terms of degrees, Mauchly's test of sphericity showed that sphericity could not be assumed for the frequency main effect, $W(2) = 0.11$, $p < .01$, the frequency by ear interaction, $W(2) = 0.13$, $p < .01$, or the pathway by frequency by ear interaction, $W(2) = 0.13$, $p < .01$. Mauchly's test was not statistically significant for the Pathway x Frequency effect, $p > .05$, showing that equal variances can be assumed for that effect. Mauchley's W cannot be computed for independent variables with 2 levels (df = 0). Thus, the main effects of pathway and ear, as well as the interaction between pathway and ear, could not be tested for sphericity.

Unequal variances in repeated measures designs increases the probability of a type I error (rejecting the null hypothesis when it is true). It creates a positive bias,

resulting in a significance value that is greater than the alpha value specified (Keppel,

1991). Geisser-Greenhouse adjustments were made to correct for this positive bias. This

adjustment corrects for variance heterogeneity by adjusting the degrees of freedom for F-

table lookup. This test is a very aggressive correction, assuming maximal heterogeneity.

Greenhouse-Geisser corrections were made for the factors that Mauchley's test

determined sphericity could not be assumed. Despite apparent differences in Figure 10,

the corrected analysis showed no interactions and no main effects (see Table 9).

Table 9

Analysis of Variance for Phase in Terms of Degrees

Source	df_{Effect}	df_{Error}	MSE_{Effect}	MSE_{Error}	F	p
Pathway (P)	1	9	0.09	3.95	0.02	.88
Frequency (F)[1]	1.30	9.54	1.22	8.74	0.14	.73
Ear (E)	1	9	1.31	1.57	0.84	.39
P X F	2	18	2.60	3.43	0.76	.48
P X E	1	9	1.45	3.71	0.392	.55
F X E[1]	1.07	9.64	4.59	4.20	1.09	.33
P X F X E[1]	1.07	9.62	8.36	7.18	1.16	.31

[1]Mauchly's test of sphericity showed that sphericity could not be assumed, so Geisser-Greenhouse adjustments were made to correct for the positive bias that results from unequal variances.

Phase Variability (Microseconds)

Visual inspection of the graphs and the ANOVA sphericity violations indicate a difference in variability across conditions. To test this hypothesis more directly and thoroughly, a Brown-Forsythe procedure was carried out. This procedure involves transforming each dependent variable score by subtracting it from the median across participants within an independent variable grouping, yielding a "Z" score. Then an ANOVA can be conducted on the Z score to test for differences in variability. The values used in the ANOVA on phase were transformed into Brown-Forsythe Z scores, and then a 2x2x3 within-subjects ANOVA was conducted on these scores. Mauchly's test of sphericity again showed that equal variances could not be assumed for the pathway by frequency interaction, $W(2) = 0.15$, $p < .01$, and the three-way interaction, $W(2) = 0.13$, $p < .01$. Greenhouse-Geisser was used again to correct for lack of sphericity where Mauchly's test indicated this was necessary. Table 10 shows the results of the ANOVA analysis. The ANOVA showed a statistically significant main effect of Frequency and Ear on variability in phase adjustments. There were no other statistically significant effects (see Table 10).

Table 10

Analysis of Variance for Phase (Degrees) Variability

Source	df_{Effect}	df_{Error}	MSE_{Effect}	MSE_{Error}	F	p
Pathway (P)	1	9	0.41	0.88	0.46	.51
Frequency (F)	2	18	61.76	1.12	55.18*	< .01
Ear (E)	1	9	3.01	0.45	6.74*	.03
P X F[1]	1.08	9.75	2.59	1.48	1.75	.22
P X E	1	9	1.98	1.10	1.79	.21
F X E	2	18	1.25	0.39	3.16	.07
P X F X E[1]	1.07	9.62	2.65	1.56	1.69	.23

[1]Mauchly's test of sphericity showed that sphericity could not be assumed, so Geisser-Greenhouse adjustments were made to correct for the positive bias that results from unequal variances.
* $p < .05$

The main effect of ear was such that the average deviation from the group's median (Z) was higher for the left ear ($M = 1.3800$, $SE = 0.1236$) than it was for the right ear ($M = 1.0800$, $SE = .1272$). The main effect of frequency on the average deviation from the group's median (Z) can be seen in Figure 11. This figure shows that variability was highest for the 8000 Hz condition, followed by the 3150 Hz condition and then the 500 Hz condition. Post-hoc comparisons were conducted using the same procedure that was done for the post-hoc analysis of the frequency main effect on amplitude shift. These tests revealed that 8000 Hz was significantly higher than both 500 Hz and 3150 Hz, but

that there was not a statistically significant difference between 500 and 3150 Hz (see

Table 11).

Figure 11. Average deviation from group median (*Z*), for phase. The error bars represent one standard error above and below the mean.

Table 11

Post-hoc Follow-ups for the Main Effect of Frequency on Phase (Degrees) Variability

Comparison	df	t	p[1]
500 - 3150	9	-1.67	> .05
500 – 8000	9	-7.26*	< .05
3150 - 8000	9	-9.15*	< .05

[1] exact *p* values are not available because a critical *t*-value from a look-up table was used, rather than an exact calculation by a computer program. This look-up method was necessary in order to do a Tukey adjustment for paired comparisons. This adjustment involved looking up the critical *q* value for 3 means and 9 degrees of freedom at the 0.05 alpha level (*q* = 3.95), and then converting it to an equivalent *t*-value (*t* = q/√2): 2.792.
p < 0.05

Description of Results: Microseconds

Inferential statistics were also conducted on the shift in terms of time (which is standardized across frequencies). This was done to ensure that the differences in variability were not due to phase being directly correlated with frequency. That is, for a constant time shift, much more phase is required at higher frequencies. Thus, a constant variability in time across frequencies would be represented by markedly different variabilities in phase across frequencies. Median phase measurements were converted into microseconds and then a constant was added to prevent negative numbers being averaged in the ANOVA. A 2x2x3 (pathway x ear x frequency) within-subjects ANOVA was then conducted on these values. Once again, Mauchly's test showed that equal variances could not be assumed for some effects. Mauchly's test indicated that sphericity cannot be assumed for the pathway by frequency interaction, $W(2) = 0.14$, $p < .01$, the frequency by ear interaction $W(2) = 0.27$, $p < .01$, and the pathway by frequency by ear interaction, $W(2) = 0.29$, $p < .01$. However, sphericity could be assumed for the frequency main effect $W(2) = 0.72$, $p = 0.27$. The pathway main effect, ear main effect, and pathway by ear interaction had too few degrees of freedom to apply Mauchly's test. Greenhouse-Geisser adjustments in the degrees of freedom were used to correct for lack of sphericity where Mauchly's test indicated. With these adjustments (and without), the ANOVA revealed that none of the main effects or interactions was statistically significant (see Table 12).

Table 12

Analysis of Variance for Phase in Microseconds

Source	df_{Effect}	df_{Error}	MSE_{Effect}	MSE_{Error}	F	p
Pathway (P)	1	9	0.62	13.53	0.05	.84
Frequency (F)	2	18	1.38	1.60	0.86	.44
Ear (E)	1	9	0.62	3.56	0.17	.69
P X F[1]	1.08	9.70	1.81	27.85	0.07	.82
P X E	1	9	3.54	2.95	1.2	.30
F X E[1]	1.16	10.39	0.48	5.64	0.09	.81
P X F X E[1]	1.17	10.50	10.26	7.09	1.45	.26

[1]Mauchly's test of sphericity showed that sphericity could not be assumed, so Geisser-Greenhouse adjustments were made to correct for the positive bias that results from unequal variances.

Phase Variability (Degrees)

The tests of sphericity indicated that equal variances could not be assumed for phase in terms of microseconds. To get a more direct test of variability of phase in terms of microseconds, a Brown-Forsythe procedure was carried out again. The values used in the 2x2x3 within-subject ANOVA on phase (microseconds) were transformed into Brown-Forsythe *Z* scores, and then a 2x2x3 within-subjects ANOVA was conducted on these scores. Mauchly's test indicated that sphericity cannot be assumed for the frequency main effect, $W(2) = 0.04$, $p < .01$, the pathway by frequency interaction $W(2) = 0.28$, $p < .01$, and the pathway by frequency by ear interaction, $W(2) = 0.17$, $p < .01$.

49

However, sphericity could be assumed for the frequency by ear interaction $W(2) = 0.54$, p = 0.08. The pathway main effect, ear main effect, and pathway by ear interaction had too few degrees of freedom to apply Mauchly's test. Greenhouse-Geisser adjustments in the degrees of freedom were used to correct for lack of sphericity where Mauchly's test indicated. With these adjustments (and without), the ANOVA revealed that none of the main effects or interactions were statistically significant (see Table 13).

Table 13

Analysis of Variance for Phase (Microseconds) Variability

Source	df_{Effect}	df_{Error}	MSE_{Effect}	MSE_{Error}	F	p
Pathway (P)	1	9	1.05	1.94	0.54	.48
Frequency (F)[1]	1.02	9.19	65.33	21.37	3.06	.11
Ear (E)	1	9	0.36	0.31	1.18	.31
P X F[1]	1.17	10.49	5.03	2.28	2.21	.17
P X E	1	9	3.20	2.43	1.32	.28
F X E	2	18	0.24	0.40	0.61	.56
P X F X E[1]	1.09	9.84	2.72	3.33	0.82	.40

[1]Mauchly's test of sphericity showed that sphericity could not be assumed, so Geisser-Greenhouse adjustments were made to correct for the positive bias that results from unequal variances.

CHAPTER 4

DISCUSSION

Summary and Interpretation of Results

Air-to-Air

The air-to-air condition was administered to assure that participants could do the task, and to assess the degree of error associated with their judgments. Eight out of the ten participants had amplitude RMS step errors less than one, and the remaining two had amplitude RMS step errors less than three, out of 30 steps. The phase RMS step error for half the participants was less than one, and for the other half less than five, out of 88 steps. It is important to keep in mind the differences between amplitude and phase adjustments in the number of steps on the slider – the phase slider had many more steps than the amplitude slider. Thus, more error in phase does not necessarily mean that the phase adjustments are less accurate. From the participant's perspective, the increased number of steps on the phase slider made it much more sensitive than the amplitude slider.

There are at least two possible reasons for error within the participant: perceptual, and motor. If the errors were perceptual, then no sound was detected despite the error. That is, being five steps away from the perfect cancellation point and being zero steps away from the cancellation point may have sounded identical to the participant. On the other hand, if the error was motor, then the sound was detected and submission of the values still occurred due to accidental activation. It is important to note, however, that participants were instructed to inform the experimenter if any accidental submissions

51

were made. If the experimenter was informed of an accidental submission, the participant redid the trial.

Despite some variability between participants in the specific level of error, the air-to-air data indicate that overall participants can do the task accurately. Assuming that the degree of error is consistent across time, this also suggests that the data in the bone-to-air condition represents the cancellation point with a similar amount of small error. All further discussion of amplitude and phase will only consider the bone-to-air listening condition.

Bone-to-Air

<u>Subjective report of cancellation</u>

The participants in this study were able to achieve the perception of cancellation. Waves had to travel through quite different mechanisms before they ended up oscillating on the same basilar membrane within the same cochlea. Once the waves arrived at the cochlea, one wave's phase and amplitude had to be adjusted by the participant until the resultant wave had near zero amplitude. The fact that cancellation occurs makes an important point that often causes confusion in those who are told about bone conduction for the first time. Specifically, bone conduction is not a separate channel of hearing (or extra sense); the informational bandwidth of hearing with and without bone conduction is the same, because they share higher-level mechanisms. Air conduction and bone conduction, although they involve radically different methods to transmit the sound, become united at the cochlea.

Future studies should request that participants provide indications of the degree of cancellation they achieved, perhaps on a Likert-type scale. How this data would be used

would have to be determined. Perhaps the adjustments could be weighted by the degree to which cancellation occurred.

The high sensitivity of cancellation to head movement is probably caused by the coupling between the bonephones and the skull changing as the head moves. As this coupling changes, the time and amplitude of the bone-conducted wave arriving at the cochlea changes, which in turn alters the resultant wave on the cochlea. The resulting percept, in this case, is a loss of cancellation without changing input into the devices. The implications of this sensitivity to head movement for adapting HRTFs for bonephones are not clear. However, it seems likely that sensitivity to the presence of a tone is greater than sensitivity to deviations in the simulated location of a sound source. Thus, stable percepts of simulated spatial locations may still be possible. An empirical evaluation would be required to determine this with greater certainty.

<u>Amplitude</u>

It is curious that the values of Bone dB at cancellation were always less than the values of Air dB, given that bonephones require much more power to reach a clearly audible tone than regular headphones. A greater power required for bonephones makes sense given that the impedance-matching anatomy available in the air-conduction pathway is not available in the bone-conduction pathway. Regardless of attempts to equate the scales, there was no direct way to perfectly equate them. Attempts to equate scales were made by using standardized reference values (the denominator in the logarithmic function) for bone and air. Although these reference values were believed to estimate the threshold of hearing, the reference values for bone and air were obtained under different circumstances, procedures, and apparatus. The ANSI Reference

Equivalent Threshold SPLs (RETSPLs) provide a more accurate means of equating the two scales (ANSI, 2004), but that testing was only done for clinical apparatus such as the RadioEar B71. Regardless of the metrics used for the quantification of the bone and air-conduction physical output, the end-user of functions should be able to use the same scale to calibrate their device. The effects of variables on bone-to-air amplitude shifts will now be discussed.

There was a statistically significant interaction between frequency and ear on amplitude shift in the bone-to-air condition. This trend within a given participant could be due to differences in thresholds between ears. Across participants, however, this indicates that the right and left transducers have different responses to input, or that asymmetrical design of the bonephones differentially affected the coupling to each mastoid. Whatever the cause of the difference between transducers, the differences do not appear to be of a large magnitude. The interaction indicates that the differences between the ears depend on the frequency. Although it is interesting to note that the transducers responded differentially across frequencies, the specific pattern of response is not of interest. It is not of interest because it will most likely vary from device to device, and thus the end-user will be responsible for making adjustments for differences in transducers. If there was no specific interest in the nature of the interaction, it may seem peculiar that it was even tested. However, the interaction was tested in order to partial out the variance for this interaction in the ANOVA so that it is more sensitive to other sources that can account for the variance. Furthermore, this effect indicates that in the implementation of these filters for producing spatial audio, one source of error in the percept could be due to differences between the transducers in this study. For maximal accuracy in the

implementation of functions yielded by this line of research, differences in responses between transducers should be incorporated into the application of the filters. This would not be difficult to do since different filters are applied separately to each ear for HRTFs.

In addition to the Frequency x Ear interaction, there was also a statistically significant effect for which tone was being adjusted on the amplitude shift required for cancellation. The amplitude shift that participants adjusted was greater when adjusting bone than when adjusting air. These differences may have to do with differences in the static tone between when the bone-conducted and air-conducted tone was being adjusted. When the participant was adjusting the air-conducted tone, the bone-conducted tone started at a static amplitude, and the participant raised the amplitude of the air-conducted wave from zero to the point at which the bone-conducted tone was cancelled out. This presumably occurred when the two waves were of equal amplitude in the cochlea. In the adjusting bone condition, the opposite occurred: the participant brought in the influence of the bone-conducted tone to cancel out the static air-conducted tone. Although the static bone-conducted and air-conducted tones were perceptually estimated to have equal loudness in preliminary testing, the precise amplitude of the two static tones (bone when adjusting air, air when adjusting bone) could not be perfectly equated. Attaining this sort of precision would require measuring the stimulus at the level of the cochlea ("cochlea microphonics"), which can only be implemented in animal models.

Finally, the amplitude shift required for cancellation depended on frequency. This shows that measuring at different frequency points is important and that one amplitude shift for all frequency components of a signal cannot be applied. That is, it shows that

establishing a function, rather than a single overall shift, is important for producing the appropriate amplitude shift for adjusting HRTFs.

The amplitude shift at 500 Hz was significantly higher than 3150 and 8000, but there was no statistically significant difference in amplitude adjustment between 3150 and 8000 Hz. This indicates that the amplitude adjustment of HRTF filters should be the same at 3150 and 8000 Hz, but this adjustment should be less than the one for 500 Hz. Future work will need to consider more frequencies between these points, because it is doubtful that all frequency components between 3150 Hz and 8000 Hz travel similarly through the head.

These obtained amplitude shifts can be used to leverage the already widely-researched air-conduction spatial audio filters so that the same (matched) percept can result for bone-conduction. This is because the amplitude shift required for cancellation is the same as what would be required for matching two waves in loudness, only shifted by a certain amount of phase. So, for a bone tone at 500 Hz to match an air tone with the same frequency in loudness, for example, it should be altered so that the output of the air-conduction headphones is 45 dB (see Figure 8) more than the output of the bonephones, using the measurement standard used in this study. The empirically-based amplitude shift values specified here provide a starting point for the amplitude shifts that could be implemented in HRTFs to adjust for bonephones. Specifically, HRTF users can apply a scaler that will provide a shift in physical output that equals the values obtained in the present study.

<u>Phase: Mean Medians</u>

In addition to amplitude, phase is the other important parameter at a given frequency in HRTFs. The results of the phase adjustments did not produce as clear recommendations as the amplitude adjustments did. Specifically, inferential statistics showed no difference in phase adjustment as a function of ear, pathway, frequency, or any interaction between these factors. It is, however, interesting to note that at 3150 Hz, adjusting the bone-conducted wave resulted in a positive phase shift and adjusting the air-conducted wave resulted in a negative phase shift. If sound travels quicker through more dense materials such as bone than it does through less dense materials such as air, then when bone is being adjusted there should be negative phase shift values, and when air is being adjusted there should be positive phase shift values. The opposite trend seen in this data could be due to differences in the two apparatus. In particular, there could be inherent phase shifts in the devices, before they get to the pathway. Even though there was not a statistically significant effect of frequency on phase, one might think that there could just be a single phase shift (the grand mean phase shift) applied across frequencies to adjust HRTFs for bone.

The null effect of phase when considered in terms of microseconds shows that a single phase shift would not be a suitable solution. Specifically, null results for phase imply an equal phase shift across frequencies. An equal phase shift corresponds to unequal time shifts. This single phase shift across frequencies, if considered in terms of microseconds, would be represented by a decreasing trend as frequency increases. Inferential statistics, however, showed that there was no significant difference in microseconds shift across frequencies. This is surprising given the null results for phase

in terms of degrees. Showing a null effect when this dependent variable was considered both ways indicates a large amount of variability in time shifts – if phase in terms of degrees does not have an effect of frequency, then phase in terms of microseconds should. These data regarding adjustments in the time domain of HRTFs suggest that phase relationships between waves reaching the cochlea through the separate pathways of bone and air are not at all stable across people or between any of the independent variables tested in this study. Having large differences between people is logical, considering the pathway that waves have to travel to meet in the cochlea. Specifically, any small difference in head diameter or skull thickness could lead to differences in the amount of time it takes for a wave to travel from the transducer, through the skull, and into the cochlea.

These findings make drawing conclusions regarding phase across participants difficult. Thus, they cannot suggest HRTF phase shifts that would be effectively generalized to other listeners. This is not to say that an effective phase shift could not be found for an individual. In fact, for cancellation to occur at any given instant, there had to be a relatively stable shift in phase. Otherwise, the resultant tone of the two waves would be twice as loud as the initial wave, rather than a reduction in loudness.

This issue of generalizability for these shift values is similar to the difference between generalized and individualized HRTFs for air conduction. Generalized HRTFs are filters designed to generalize to any listener, and individualized HRTFs are filters that are custom-designed for use by a single person. Individualized HRTFs are much more effective in producing the percept of a localized sound source than are generalized

HRTFs. Similarly, it seems that an individualized shift function to adapt HRTFs for bonephones would be much more effective than a generalized shift function.

Phase: Variability

Further analyses were done to characterize the exact nature of the variability across frequencies. The Brown-Forsythe inferential tests on variability showed that the highest frequency component tested, 8000 Hz, had significantly more variability in phase adjustments than the other frequencies. One possible cause for this higher variability at 8000 Hz is that a given degrees phase shift is a much smaller time shift for 8000 Hz than for 3150 or 500 Hz. Conversely, a constant time shift would lead to a consistently increasing phase shift as frequency increases. This could show up as greater variability at higher frequencies, since a given variability in time will be represented by a greater variability in degrees at higher frequencies than at lower frequencies. A constant time error would show up on a phase graph as consistently increasing variability as frequency increases. Although the phase variability in terms of degrees was not monotonically increasing, the highest frequency had significantly more variability than the lower two frequencies.

If the differences in variability between frequencies are due to consideration of phase in terms of degrees, then these differences should disappear when phase is considered in terms of microseconds. Indeed, a Brown-Forsythe procedure showed that there was a not a statistically significant difference in phase variability in terms of microseconds, comparing across frequencies. Thus, it appears that differences in variability are due to consideration of phase in terms of degrees rather than microseconds, and not due to variability amongst people in the amount of time it takes for sound to

travel from the bonephones to the cochlea. It is important to keep in mind that the present

discussion concerns only variability, and not means. The variability amongst people in

the phase shift required for cancellation is constant across frequency when considered in

terms of microseconds. There is still not a constant time shift that can be used to adjust

HRTFs.

Implications in the Application of Adjustment Functions

In summary, this study provided initial anchor points that could be used with

other points to form a function of amplitude shifts. Future research needs to be done to

establish a full function, rather than a small set of shift values. Once a function is

established, it could then be applied to adjust HRTFs for bonephones. The phase shift

data from the present study indicates that manipulating the time-dependent aspects of the

HRTF signal processing may not be an effective way to adjust for bonephones, although

more work needs to be done to understand the nature of this problem.

It may be noted that manipulation of time parameters has produced implicit

spatial separation in other studies. Specifically, Walker and colleagues (2005a) showed

increased speech channel segregation as a function of interaural time differences

implemented on bonephones. A possible cause for the discrepancy is that the time shifts

implemented in that study were gross differences between the transducers and across the

skull. In the present study, however, participants made fine adjustments in time that

altered the resultant wave in a single cochlea. In addition, it is not clear how much of

Walker and colleagues' effect was due to air-conducted leakage and how much was due

to bone-conducted spatial audio.

I have often referred to this study's anchor points, and the function that further research could lead to. The applicability of this study's data, on its own, to existing HRTFs, is limited. Although one could apply these three shifts to three frequency ranges in the HRTFs, it is quite a crude interpolation to all the other frequencies. Nonetheless, this study's work is important for establishing methodology that can obtain the shift values, important factors that effect the shift values found, and any problem areas in the idea of finding bone-to-air shift values (i.e., phase shift variability).

Regardless of whether full functions or just a few shifts are used, the process for making the adjustments would be the same. First, end-users would select or compute the shift value, or set of shift values, for the desired frequency components. Then, the scalers that correspond to these shift values on their equipment would have to be computed, using the same standardized physical measurements used in this study. The user can then scale the HRTF filter result by these scalers. This scaling should produce an adjustment of the HRTF so that it is more suitable for bonephones. The shifts could be combined with thresholds of audibility used as equalization curves (Walker & Stanley, 2005) to optimize the signal for bonephones.

Future Research

Although modifications and additional studies have been suggested throughout the discussion, there are three major lines of research that emerge from these results. One line of research should seek to better understand what is going on with phase, and why this research did not show consistent adjustments for cancellation. Although this research suggested that phase may be too variable with bonephones, a better understanding of phase relationships between bone-conducted waves in the skull could be accomplished

with the aid of computer models. Research using simulations of the cochlea's response to sound transmission from localized transducers has been funded and is in the initial stages of planning (CFDRC & Walker, 2006). Following this modeling research, perhaps future cancellation studies can incorporate changes that allow phase to be a reliable adjustment parameter for HRTFs on bonephones. A second line of research should involve follow-up cancellation studies that present enough frequencies to approximate a complete shift function. This line of research would also benefit from more subjects to achieve greater generalization. A third line of research should verify the implementation of these HRTF adjustments. This could be done by measuring the difference between the desired and observed percept of sound source location, thus producing a measure of error. The degree of error associated with spatial audio could then be compared between adjusted HRTFs on bonephones, non-adjusted HRTFs on bonephones, and HRTFs on their intended apparatus: standard headphones. The exact order of these lines of research depends on the trade-off between developing adjustment filters and testing the results of these adjustment filters.

APPENDIX A

DOCUMENTATION OF STIMULI MEASUREMENT

Equipment Overview

Microphone

The microphone used for measurements in this study was a Bruel & Kjaer Type 4146. The microphone converts the air-conducted sound pressure waves into an electrical signal – namely, a voltage traveling through a cable that was hooked into the "preamp input" jack of the measurement amplifier (Bruel & Kjaer, BR0047-13).

2 cc coupler

A 2 cc coupler was screwed onto the end of the microphone used in this study. This coupler simulates the ear canal, providing a space between the diaphragm of the microphone and the output of the headphones. The 2 cc coupler, microphone, and connector housing can be seen coupled together in Figure 12.

Figure 12. 2 cc coupler, microphone (B&K Type 4146), and connector housing used in this study. These separate components are visually separated by different shades of metal. They are attached via matched threads.

Measurement Amplifier

The measurement amplifier used for measurements in this study was a Bruel &

Kjaer Type 2610. This measurement amplifier converts the small voltage output of the

microphone into a signal that is easier for the sound level meter to process. It also does

impedance conversion, and matches the output signal to the sound level meter's input

sensitivity (Bruel & Kjaer, BA 7675-12; BR0047-13). The AC signal from the

measurement amplifier travels out a BNC jack and through a cable to a triaxial LEMO

AC input jack on the sound level meter. A photograph of the measurement amplifier can

be seen in Figure 13.

Figure 13. Measurement Amplifier used in this study, Bruel & Kjaer Type 2610.

Sound Level Meter

The sound level meter (SLM) used for measurements in this study was a Bruel &

Kjaer Model 2260. Although many assume that an SLM always makes measurements on

its own without any other equipment, it also can be part of a larger, more sophisticated

measurement system. When the SLM is used on its own, it still has to have a

measurement amplifier and microphone (or equivalent); these parts are just all internal to

the SLM, rather than being separate hardware. This is much like a compact stereo system as compared to a component stereo system: both do the job, but the more sophisticated one does it with higher fidelity. In the setup used for this study, the SLM analyzes the output of the measurement amplifier. This analysis includes applying filtering and averaging, as well as computation of the metric on the scale chosen. The SLM then displays the results of this analysis onto a screen (Bruel & Kjaer, BR0047-13; BA 7676-12). A photograph of a 2260 sound level meter can be seen in Figure 14.

Figure 14. Sound level meter used in this study: Bruel & Kjaer Type 2260.

Artificial Mastoid

A photograph of an artificial mastoid can be seen in Figure 15. The artificial mastoid simulates many of the characteristics of the system that bone-conducted waves travel through in the skull (Bruel & Kjaer, 1974). As a system, the components of the artificial mastoid simulate the mechanical impedance of the human mastoid. It does this with standardized physical attributes to allow comparisons across instances of the

65

artificial mastoid. A diagram of the artificial mastoid's construction can be seen in Figure 16. With a standardized device that simulates the human mastoid, the output of bone-conducted headsets such as the bonephones can be measured in a manner similar to how air-conducted stimuli are measured. Specifically, decibel units of acceleration or force[1] can be measured and used as an objective quantification of the energy of the stimuli arriving at the human head. The acceleration is converted to a voltage signal via ceramic piezoelectric discs, where displacement causes a change in voltage. This voltage is sent out through a 10-32 UNF ("Microdot") connector, and through a cable to the "direct input" jack of the measurement amplifier.

Figure 15. Artificial mastoid used in this study: Bruel & Kjaer Type 4930.

---

[1] The vibrational energy delivered from a bone-conduction transducer can be measured in terms of acceleration or force. Acceleration was chosen because it avoids difficulties with the mass of other parts of the system that effect the measurement's accuracy (see Bruel & Kjaer, 1974). It also creates a measure independent of the transducer's mass, unlike force.
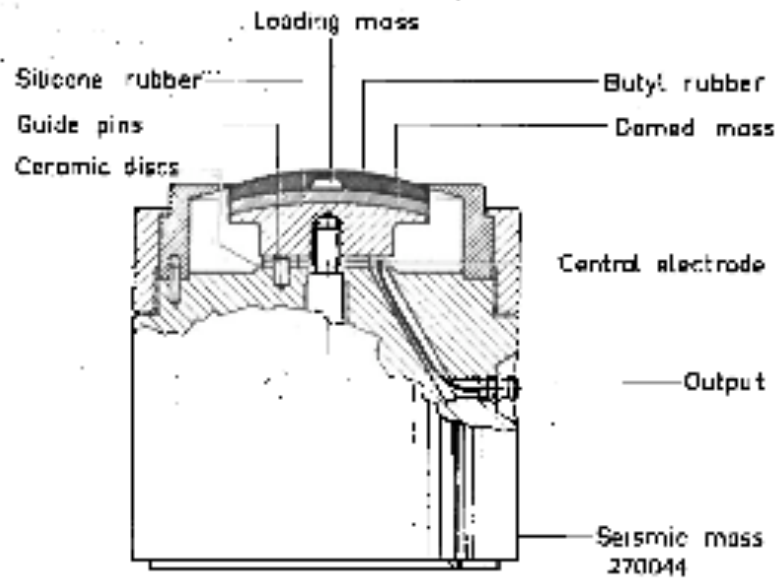
Figure 16. Diagram of artificial mastoid components. Obtained from Bruel & Kjaer (1974).

SLM Software Settings

There are many choices of measurement metrics to use on the SLM. Equivalent Continuous Sound Level with linear frequency weighting (L_{Leq}) was chosen because of its linear frequency weighting[2] and standardized time weighting algorithm (Bruel & Kjaer, BR0047-13). When tones (but not maskers) were measured, 1/3 octave band filters were used. These were used because any frequency component of the stimuli outside of the 1/3 octave band around the tone was masked by the bandpass noise. Using 1/3 octave band filters also ensured that the level measurement was referring to the frequency domain of the stimuli that was of interest, rather than a harmonic somewhere else in the spectrum of the stimuli.

---

[2] Although the metric at the level of the SLM was always linear, sometimes the A-weighting filter was applied at the level of the measurement amplifier. Further description of when this was done is included in subsequent text.

Calibration[3]

Before sound measurement can take place, the equipment must be "field calibrated." This involves minor adjustments in sensitivity of some equipment so that with a known energy source, the measurement system indicates the appropriate energy level. This process had to take place for air conduction and bone conduction separately. I will begin with a description of air-conduction calibration, which was the easiest to calibrate.

Figure 17, Panel a shows the setup for calibrating air-conduction equipment. The known power source for air-conduction is an acoustic calibrator.[4] This emits a 1000 Hz tone that has a known output of 1 pascal, which with a 20 µPa reference value, corresponds to 94 dB SPL (Bruel & Kjaer, BA 7666-11).[5] This acoustic calibrator is directly coupled to the microphone with its protective screen in place, then the signal is sent through the measurement amplifier, where it finally arrives at the SLM. The input sensitivity of the SLM must be set to the value given in the microphone calibration sheet.[6] The sensitivity adjustment screw on the measurement amplifier is then adjusted until the SLM reads 94.0 dB. This indicates that, given a sound source that should read 94.0 dB, the measurement system indeed creates a reading of 94.0 dB.

---

[3] The calibration process and its underlying physics, as well as the setup of this equipment, was tutored to this author by Dr. Adrian Houtsma at the USAARL. Mr. Marcelo Pinheiro of Bruel & Kjaer also supplemented the author's understanding of the material.
[4] Bruel & Kajer Type 4231
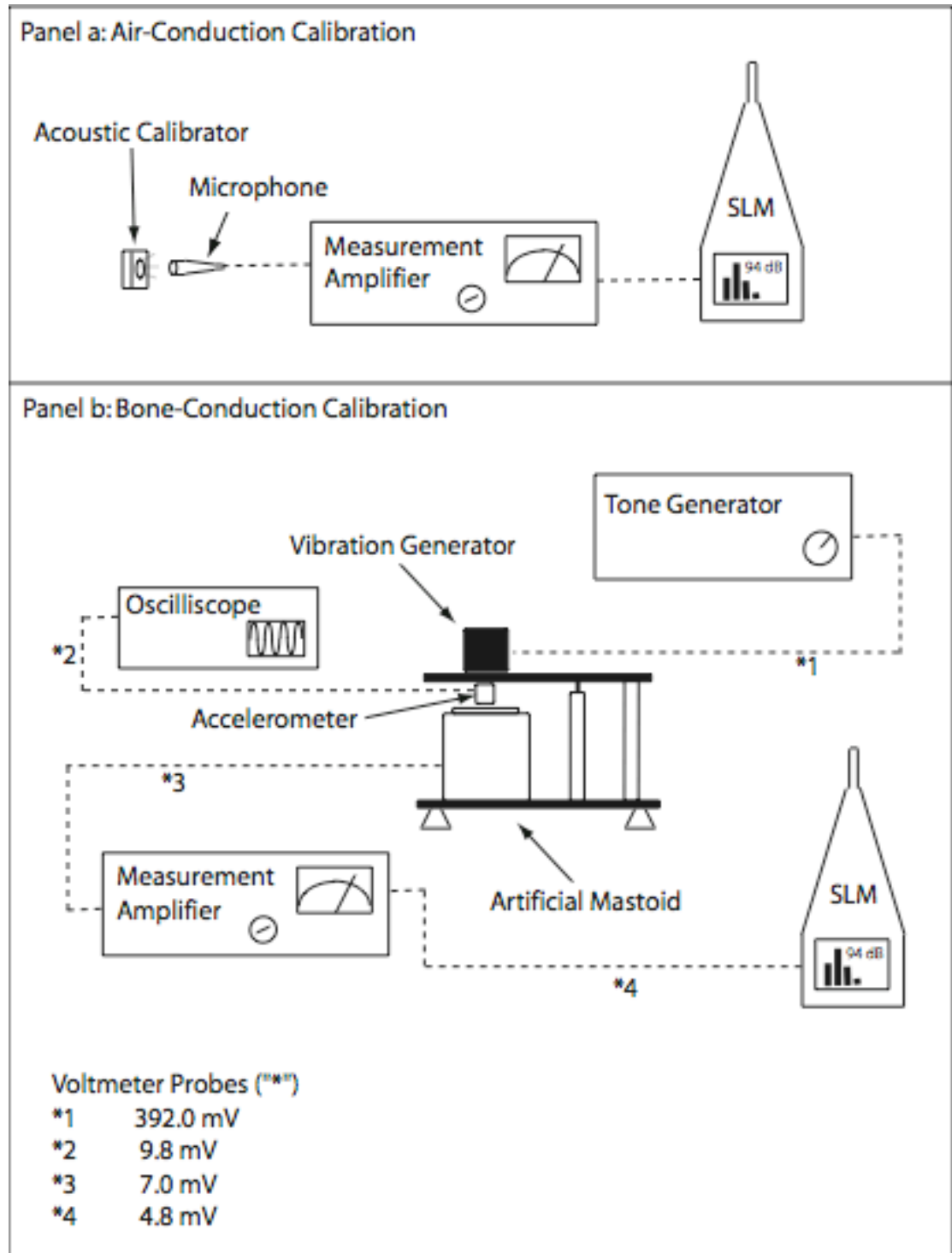[5] $20*\log(1 \text{ Pa} / 20 \text{ µPa}) = 94$ dB
[6] -26.9.0 dB re 1V/pa

Figure 17. Calibration setups for bone conduction and air conduction. "*" followed by number indicate voltage probes into system.

Figure 17, Panel b shows the setup for calibrating bone-conduction equipment. The known power source for bone conduction is more complicated than that for air conduction; it is in fact a system of devices. The first device in this system is a vibration generator known as a "mini-shaker."[7] This device takes a voltage input and turns it into a vibration. The vibration generator rests on top of a special arm on the mastoid that supports the weight of the vibration generator. The vibration generator receives input from a tone generator.[8] The vibration generator is connected to an accelerometer known as an "impedance head."[9] This connection is a physical coupling via a screw threaded at both ends. The voltage output of the accelerometer is sent to an oscilloscope. The accelerometer then rests on top of the artificial mastoid just as a bone-conduction vibrator would. The vibration generator, accelerometer, and surface of the mastoid form a physical connection together. The rest of the system is hooked up just as in the air conduction.

To achieve calibration, several adjustments need to be made. First, the output of the tone generator system that feeds into the vibration generator needs to be adjusted until a specific voltage output is displayed on the oscilliscope. The voltage output indicates what acceleration is being exerted onto the accelerometer, given a known voltage sensitivity.[10] The voltage output of 9.8 mV from the accelerometer was an easy value to create. This voltage corresponds to an acceleration of 3.20 m/s^2 exerted on the

---

[7] Bruel & Kjaer Type 4810 was used in this study.

[8] The tone generator used for this study's calibration was the system of components used to deliver sound for this experiment (see Figure 6 on page 28), playing a 1000 Hz pure tone through NCH Tone Generator software.

[9] Bruel & Kjaer Type 8000 was used in this study.

[10] The voltage sensitivity of the accelerometer used in this study was 3.16 mV/ms^{-2}. This indicates that when a one m/(s^2) force is exerted on the accelerometer, a 3.16 mV signal will be sent out.

accelerometer and physically transferred through to the mastoid. An acceleration value of

3.20 m/s^2 corresponds to 100 times the British standard for the bone-conduction

threshold.[11] In decibels, a denominator that is 100 times the reference value corresponds

to 40 dB.[12] This system of devices (the tone generator, vibration generator, and

accelerometer) acts in a manner analogous to the acoustic calibrator for air-conduction

calibration. Specifically, this system of devices provides a known power source that

should yield a specific dB output reading on the SLM: 40 dB. While the tone generator is

driving the vibration generator which in turn yields an accelerometer output of 9.8 mV, a

few adjustments need to take place: the input sensitivity of the SLM must be set to a

value that puts the output reading near 40 dB[13], then the screw on the measurement

amplifier is turned until the SLM reads out exactly 40 dB. It is important to note that any

mathematically sound combination of accelerometer voltage and its appropriate SLM

reading could be used; the values used here were for convenience. After these

adjustments, the system is now calibrated, and ready for measurement. Table 14 provides

details regarding the settings of the SLM and measurement amplifier during calibration

for both pathways.

---

[11] British standard for bone-conduction threshold = 10 dB re 1 cm/s^2; $(10^{(10/20)})/100 =$
0.0316 m/s^{-2}; 100 * 0.0316 m/s^{-2} = 3.2 m/s^{-2}.
[12] 20*log(100) = 40 dB
[13] -1.0 dB re 1V/pa

Table 14

SLM and Measurement Amplifier Settings during Calibration

Pathway	Air	Bone
Measurement Amplifier		
Input Section Gain	0 dB	-10 dB
Filters		
Ext	On	On
22.4 Hz/ L 2-200000 Hz	22.4 Hz	22.4 Hz
A[14]	On	On
Detector		
Normal/Reset	Normal	Normal
Hold/Normal	Normal	Normal
Peak/RMS	RMS	RMS
Averaging Time	Slow	Slow
Input	Preamp	Direct
Ref	Off	Off
Polarization Voltage	180 V	0 V
SLM		
Sensitivity of AC Input (dB re 1V/Pa)	-26.9	-1.0
Octave Band Analysis	Broadband[15]	Broadband[15]
Measurement Metric	L_{Leq}	L_{Leq}

[15]Broadband measurements were chosen rather than 1/3 octave band measurements because the oscilloscope voltage reading was broadband.

It may be noted in Table 14 that all broadband measurements with the air-conduction microphone had the A-weighting filter activated at the level of the

measurement amplifier. An A-weighting filter shapes the sound level measurement by weighting the signal differently across frequencies, so that the output measurement better matches with the perceived loudness of tones (Bruel & Kjaer, BR0047-13). One characteristic of an A-weighting filter is that it shifts the measurements of lower frequencies down, since the human ear is less sensitive to them. As the frequency goes lower, this shift becomes greater. This particular feature proved to be useful for the microphone used for measurement, because it was sensitive down to extremely low frequencies (less than 1 Hz!). Frequency components at sub-threshold frequencies did exist in the room, and radically affected the broadband measurements. Turning on the A-weighting filter attenuated these especially low-frequency components a great deal, giving a measurement that more accurately reflected the energy level of the audible tones. A-weighting was used for the bone-conduction calibration to be consistent.

Measurement Setup[3]

Figure 18, Panel a shows the setup for measuring air-conduction stimuli. For air conduction, measurement was identical to the calibration setup, but without the acoustic calibrator. In place of the acoustic calibrator, the earbud headphone output from the experiment apparatus (see Figure 6 in the body of this document) was coupled to the microphone. This coupling was formed by screwing a 2cc coupler directly over the microphone diaphragm. A seal between the earbud and the 2cc coupler was made via a #8 rubber O-ring.

Figure 18, Panel b shows the setup for measuring bone-conduction stimuli. For bone conduction, measurement involves less equipment than the calibration. The vibration generator, accelerometer, and oscilloscope are removed. A tensioned arm

replaces the supporting arm for the vibration generator. This tensioned arm places

downward force (5 Newtons) onto the bonephone, which sits between the arm and

artificial mastoid surface. The rest of the system is set up just as in calibration. The

settings of the SLM and measurement amplifier during measurement were similar to the

settings during calibration (see Table 14). The one exception was that for tones (both air

and bone), the 1/3 octave band filters were used and the A-weighting filter was switched

off. One-third octave band filters were used to avoid the influence of frequency

components far outside of the test tone's intended frequency components, which were

perceptually masked by bandstop noise. A-weighting was not used because it was not

needed to attenuate the low-frequency sounds, given the 1/3 octave band analysis. For

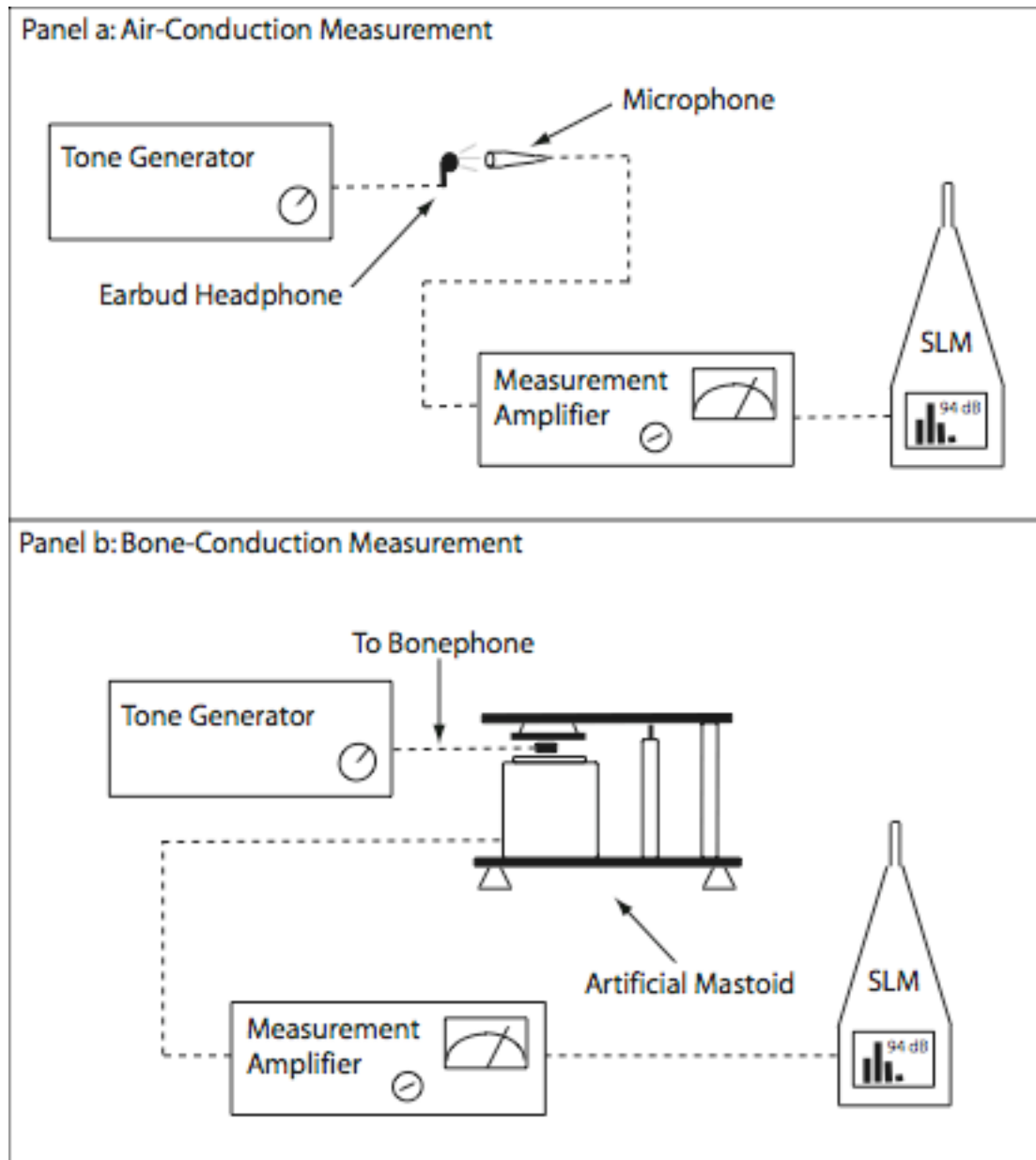maskers, all settings were identical to the calibration.

Figure 18. Measurement setups for air and bone conduction.

Measurement Data Collection

One of the main purposes of measurement was to generate a function relating the

scaler input of the tone that the participant adjusted to the physical output of the

headphones or bonephones. The procedures for establishing these functions in air-

conduction and bone-conduction measurement were identical. Each transducer in each pathway at each frequency was measured in three repeated sessions. Software identical to that used in the experiment was used, but the code was modified so that the stimuli did not pulse. For each measurement, the sound was started on the apparatus, the "reset" button on the SLM was pushed, and then a reading was taken from the SLM. Measurements began at the first nonzero scaler value. Each subsequent adjusting tone measurement was made after adjusting the slider up one interval via the arrow keys on the keyboard. After all measurements were collected, they were averaged across measurement sessions. A scatterplot relating the scaler input to the physical output was generated. Then a logarithmic function (with a base of e) was fit to this data in Excel. This function was then used to compute the physical output that corresponded to the scaler value that was chosen and submitted upon cancellation of the sound waves. A sample measurement function can be seen in Figure 19. Full results of the adjustment tone measurements can be seen in Appendix B.
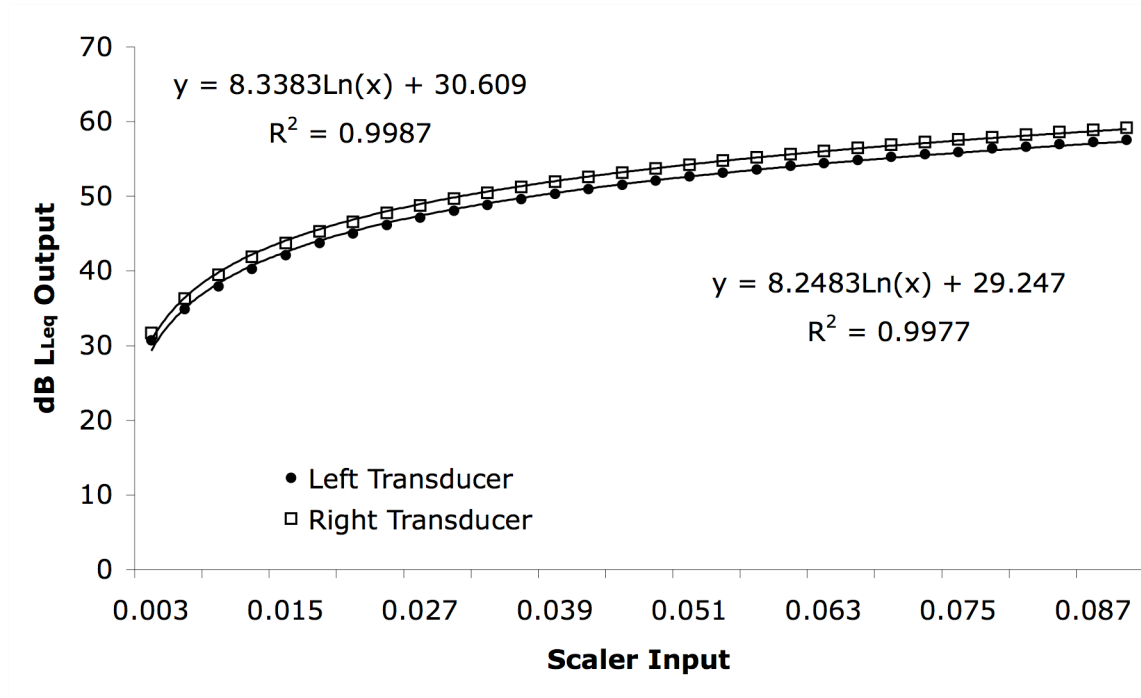
Figure 19. Sample measurement function for bonephones at 8000 Hz. The equations indicate the logarithmic function that was used to calculate output. The equation matches with the function to which it is closest. The R^2 values indicate the fit between the equation and the measurement data collected.

The static tones also had to be measured to calculate the air-bone metric. These were also measured in three sessions, using the same non-pulsing experimental apparatus. Before each measurement, the "reset" button on the SLM was pushed. The output of the pathway not being adjusted was then measured on the appropriate device. Measurements of the air-conducted maskers were also measured across three sessions, and each measurement was again separated by a press of the "reset" button. The results of the static tone and masker measurements can be seen in the body of this document. Specifically, Table 2 and Table 4 in the body of this document show the measurement results.

Scaler-dB for Adjustment Tones

Figures 20 through 25 depict the functions obtained through measurement detailed in Appendix A. These functions relate the scaler input to the dB output of the devices used in the experiment described in this document. The measurements described below pertain to the tone that the participant adjusted in the study. The equations on the figures indicate the logarithmic function that was used to calculate output. The equation matches with the function that it is closest to. The R^2 values indicate the fit between the equation and the measurement data collected.
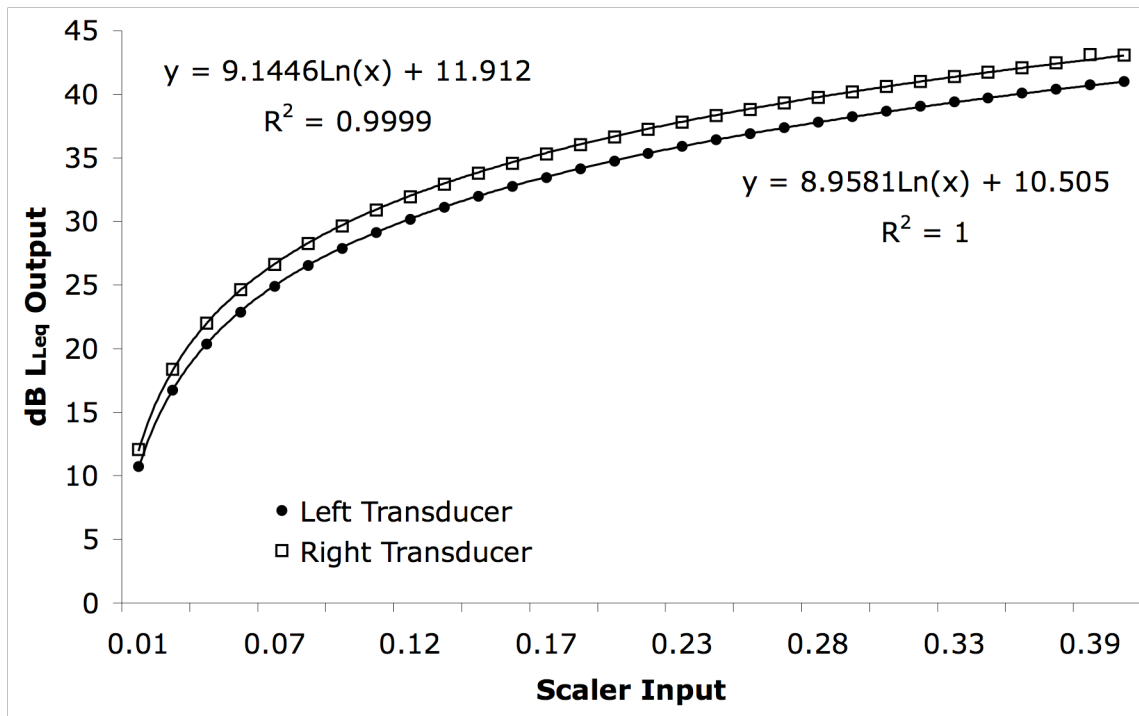

Figure 20. Bonephone Measurements, 500 Hz.

Figure 21. Bonephone Measurements, 3150 Hz.

Figure 22. Bonephone Measurements, 8000 Hz.

Figure 23. *Headphone Measurements, 500 Hz.*

Figure 24. *Headphone Measurements, 3150 Hz.*

Figure 25. Headphone Measurements, 8000 Hz

Variability in Measurement

For measurement of all stimuli, three measurement sessions were completed. There was some variability in the measurements across these sessions. This variability is quantified for measurements of the adjustment tones, static tones, and maskers in Table 15. The variability for the measurement of the adjustment tones is the standard deviation of the measurement sessions, averaged across scaler input values and then across ears. The variability for the measurement of the static tones and maskers is the standard deviation averaged across ears.

Table 15

Variability in Stimuli Measurements

	SD: 500 Hz	SD: 3150 Hz	SD: 8000 Hz
Adjusting Tones			
Bone	1.2	0.9	1.4
Air	3.0	1.9	0.3
Static Tones			
Bone	1.2	0.9	1.4
Air	3.4	2.0	0.3
Maskers			
Bandstop	2.2	2.1	2.0
Bandpass	3.0	2.7	0.3

APPENDIX C

INDIVIDUAL DATA

Tables 16 through 18, below, contain amplitude and phase shifts of individual participants in the bone-to-air condition. The data are further broken down by ear, frequency, and pathway adjusted ("pathway"). The data were aggregated across trials (n = 5).

Table 16

Individual data for amplitude shift required for cancellation.

Frequency		500	500	3150	3150	8000	8000
Ear		L	R	L	R	L	R
Participant	Pathway						
1	Bone	34.8	32.4	22.5	16.5	25.8	25.2
1	Air	38.7	35.4	23.6	18.5	31.6	18.1
2	Bone	44.6	38.5	20.5	21.3	17.9	30.5
2	Air	48.9	44.8	22.4	22.0	23.2	33.0
3	Bone	52.2	51.3	22.1	26.2	25.8	22.1
3	Air	58.0	53.5	24.8	27.9	18.4	25.0
4	Bone	48.6	44.3	25.4	23.3	21.7	28.8
4	Air	49.8	41.8	26.7	29.9	14.1	20.8
5	Bone	45.9	44.7	33.0	33.8	23.6	31.6
5	Air	53.3	47.1	28.9	31.8	25.5	34.2
6	Bone	50.3	45.0	23.1	22.7	25.9	27.8
6	Air	53.5	47.3	21.7	24.4	24.4	25.7
7	Bone	43.6	42.6	24.7	24.0	17.1	21.6
7	Air	36.4	43.6	22.4	26.3	17.4	26.6
8	Bone	37.7	43.8	24.1	21.8	25.6	33.0
8	Air	44.2	41.4	30.8	19.9	25.7	36.3
9	Bone	46.2	35.1	21.0	16.8	18.8	25.8
9	Air	49.1	41.8	26.6	22.0	21.7	23.9
10	Bone	39.9	33.6	17.0	29.3	20.9	30.0
10	Air	38.7	43.3	33.7	26.9	21.7	35.3

Table 17

Individual data for phase shift (degrees) required for cancellation.

Frequency		500	500	3150	3150	8000	8000
Ear		L	R	L	R	L	R
Participant	Pathway						
1	Bone	24.68	24.68	94.63	72.00	-260.54	243.37
1	Air	-24.68	-35.48	-72.14	-75.89	290.59	-190.06
2	Bone	24.68	24.68	-41.14	-92.57	-246.86	-152.23
2	Air	-20.83	-10.80	37.03	47.31	-115.20	176.91
3	Bone	12.34	12.34	-43.20	-6.17	144.00	-148.12
3	Air	-21.60	-15.43	10.28	0.00	246.86	137.83
4	Bone	12.34	12.34	47.31	2.06	-53.49	181.47
4	Air	7.71	-3.09	16.46	14.40	-90.52	-111.09
5	Bone	6.17	6.17	-28.80	14.40	-213.94	-218.06
5	Air	-13.89	-4.63	-67.88	10.28	-135.77	-123.43
6	Bone	-6.17	-6.17	-30.86	-16.46	135.77	109.57
6	Air	-3.09	0.00	-14.40	30.86	-141.58	-142.30
7	Bone	27.77	27.77	69.94	-12.34	-197.49	-230.40
7	Air	18.51	12.34	26.75	12.34	-123.43	-131.66
8	Bone	41.83	41.83	69.94	121.42	131.66	123.43
8	Air	-91.94	60.17	-100.80	-84.35	263.41	-117.62
9	Bone	-11.17	-11.17	19.01	9.79	180.05	-159.12
9	Air	-13.89	-18.51	2.06	26.75	152.23	135.77
10	Bone	-23.92	-23.92	50.85	16.23	101.34	88.64
10	Air	135.00	89.12	-59.79	-31.26	222.80	-78.17

Table 18

Individual data for phase shift (microseconds) required for cancellation.

Frequency		500	500	3150	3150	8000	8000
Ear		L	R	L	R	L	R
Participant	Pathway						
1	Bone	2.39	2.44	1.46	1.11	-1.58	1.47
1	Air	-2.39	-3.44	-1.11	-1.17	1.76	-1.15
2	Bone	2.39	0.60	-0.63	-1.42	-1.50	-0.92
2	Air	-2.02	-1.05	0.57	0.73	-0.70	1.07
3	Bone	1.20	1.95	-0.66	-0.09	0.87	-0.90
3	Air	-2.09	-1.50	0.16	0.00	1.50	0.84
4	Bone	1.20	-2.24	0.73	0.03	-0.32	1.10
4	Air	0.75	-0.30	0.25	0.22	-0.55	-0.67
5	Bone	0.60	1.95	-0.44	0.22	-1.30	-1.32
5	Air	-1.35	-0.45	-1.04	0.16	-0.82	-0.75
6	Bone	-0.60	2.09	-0.47	-0.25	0.82	0.66
6	Air	-0.30	0.00	-0.22	0.47	-0.86	-0.86
7	Bone	2.69	-0.90	1.08	-0.19	-1.20	-1.40
7	Air	1.80	1.20	0.41	0.19	-0.75	-0.80
8	Bone	4.06	0.00	1.08	1.87	0.80	0.75
8	Air	-8.91	5.83	-1.55	-1.30	1.60	-0.71
9	Bone	-1.08	1.58	0.29	0.15	1.09	-0.96
9	Air	-1.35	-1.80	0.03	0.41	0.92	0.82
10	Bone	-2.32	-10.42	0.78	0.25	0.61	0.54
10	Air	13.09	8.64	-0.92	-0.48	1.35	-0.47