Toward Adapting Spatial Audio Displays for Use With Bone Conduction: The Cancellation of Bone-conducted and Air-conducted Sound Waves

Raymond M. Stanley and Bruce N. Walker

Sonification Lab, School of Psychology

Georgia Institute of Technology
Toward Adapting Spatial Audio Displays or Use With Bone Conduction: The Cancellation of Bone-conducted and Air-conducted Sound Waves

The application of virtual 3D audio technology in auditory displays has many important uses, such as increasing the detectability of radio communication signals amidst distracters and noise (e.g., Brungart & Simpson, 2002) and providing orientation cues in cases of vision loss (e.g., Walker & Lindsay, 2006). Because virtual 3D audio displays are typically delivered via headphones, the detection and localization of environmental sounds can deteriorate due to the ears being covered. This can be a problem when spatialized audio and sounds in the environment are both important for the user’s task, such as with audio navigation systems like the System for Wearable Audio Navigation (SWAN) (Walker & Lindsay, 2006), or a tactical environment where a soldier needs to hear events occurring around him in combination with radio communications. Aside from performance, the level of comfort that the individual has when their ears are covered may decrease – users of SWAN have reported that they would not use the system outdoors if it covered their ears (Walker, Stanley, & Lindsay, 2005).

These situations would benefit from an alternative to headphones. Because the auditory system is also sensitive to pressure waves transmitted through the bones in the skull (Békésy, 1960; Kelley, 1937; Tonndorff, 1972), bone-conduction headsets may lead to an acceptable solution. Bone-conduction headsets leave the ear canal and pinna uncovered, which may facilitate improvement in the detection and localization of environmental sounds.

Furthermore, bone-conduction headsets allow the display of auditory information even when earplugs are inserted into the ear canal, which cannot be done with headphones. Simultaneous hearing protection and 3D virtual audio displays may be useful in aviation or
artillery environments where simultaneous engine noise and spatialized radio communications could occur.

Though the utility of bone-conduction headsets is justified, there is little research indicating how auditory display design should change when delivering it through bone-conduction headsets rather than headphones. Furthermore, there is no research on how the filters applied to make virtual 3D audio displays could be optimized for delivery through bone-conduction headsets. The purpose of the study presented here is to better understand the feasibility of optimizing HRTFs for bonephones and provide preliminary data for making adjustments to spatial audio filters for bone conduction.

The use of bone-conduction transducers to deliver sound is not new. Bone-conduction transducers have typically been used in clinical audiology settings to assess the locus of hearing damage in patients (e.g., Robinson & Shipton, 1982; Small & Stapells, 2003). There has also been research seeking to understand the mechanisms of bone-conduction hearing (e.g., Békésy, 1960; Stenfelt, Hato, & Good, 2002; Tonndorf, 1966).

Recently, bone-conduction headsets that are better suited for use in auditory displays (“bonephones”) have become available. We use the term “bonephones” to describe any bone-conduction headset that is designed to take the place of headphones. These bonephones are compact, binaural, and have a standard 1/8” stereo phono input jack. Some bonephones also have transducers that rest in an ideal location - the mastoid, which is the raised portion of the temporal bone located directly behind the pinna. This location is closest to the inner ear, is suitable for dichotic stimulation, and is unaffected by jaw movements.

Sensitivity to lateralization cues (interaural level differences and interaural time differences) is a minimum requirement for implementation of virtual 3D audio on bonephones.
The nature of the medium that bone-conducted sounds travel through may seem to prevent these cues from occurring. Specifically, it may seem that the attenuation across the skull is not sufficient for interaural level cues—that is, when one side of the head is stimulated, the contralateral cochlea receives nearly as much energy as the ipsilateral cochlea. It may also seem that the speed of sound travel is much faster than through air, which would prevent interaural time cues from occurring— that is, the cochlea contralateral to stimulation would be activated at nearly the same time as the ipsilateral cochlea.

Indeed, many researchers have assumed that spatial audio with bone conduction is not possible, because the interaural attenuation, and thus the maximum interaural level difference (ILD), was not considered sufficient (see Blauert, 1983; see Goldstein & Newman, 1994). This is likely due to audiologists’ conservative estimate of bone-conducted interaural attenuation (BC IA) to be zero dB (e.g., Katz & Lezynski, 2002). This conservatively-biased estimate of BC IA is appropriate for clinical purposes where cross-hearing can lead to inaccurate assessments of hearing thresholds. 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) showed sensitivity to ILDs and ITDs delivered through bone-conduction transducers in children with normal hearing, as well as in children with abnormalities of the middle and outer ears. Furthermore, in participants with normal hearing, there was not a statistically significant difference in the sensitivities to these cues between the bone-conducted sound and air-conducted sound. MacDonald, Henry, & Letowski (2006) implemented HRTFs on bone-conduction
vibrators and regular headphones and found similar localization performance. In addition, others have shown sensitivity to spatial audio with Bone-Anchored Hearing Aids (BAHAs) (e.g., Snik, Beyon, van der Pouw, Mylanus, & Cremers, 1998; Snik, Bosman, Mylanus, & Cremers, 2004). Studies of bone conduction hearing in vivo have also shown that the speed of sound through bone-conduction is similar if not slower than air conduction (Boezeman, Bronkhorst, Kapteyn, Houffelaar, & Snel, 1984; Boezeman, Kapteyn, Visser, & Snel, 1983; Franke, 1956; Tonndorf & Jahn, 1981).

Researchers have also shown sensitivity to spatial audio with bonephones. Specifically, Walker, Stanley, Iyer, Simpson, and Brungart (2005) showed that ITDs and ILDs implemented through bonephones can enhance speech intelligibility in multi-talker communications environments with bonephones, presumably due to the spatial separation invoked and Cherry’s “cocktail party effect,” (Cherry, 1953). Furthermore, people have successfully used the SWAN navigation system with bonephones, (Walker & Lindsay, 2005), which guides users to their destination using spatial cues.

Together, these studies suggest that sufficient spatial separation can be induced for implementing virtual 3D audio displays on bonephones. Researchers have already established head-related transfer functions (HRTFs) for air-conduction to produce virtual 3D audio displays (e.g., Wightman & Kistler, 1989). The critical issue remaining is to specify exactly how the sound could be altered to optimize HRTFs for bonephones. There is no way to measure HRTFs for bone-conduction in an analogous way to how they are measured for air conduction, since spatial audio does not occur naturally through bone conduction. It is possible, however, to understand how to alter a given bone-conduction signal so that it sounds like an air-conduction signal. Specifically, the amplitude and phase in different frequency bands could be shifted by an
appropriate amount to match the two signals. The result of understanding the relationship between bone-conduction and air-conduction signals would be a function of shift values that match the two signals. A function of shift values could be used to adapt the HRTFs so that they are suitable for use with bonephones. We call this function a “bone-adjustment function,” or BAF. We call the resultant combination of the HRTF and shift function a “bone-related transfer function,” or BRTF. The BRTF is analogous to an HRTF, but customized for bone-conduction headsets. The present study collects an initial data set of shift values that provides anchor points for a BAF. The purpose of collecting only anchor points was to establish appropriate methodology for finding the shift values and determine factors that affect these shift values. Considering the consistency of BAFs across and within people will also provide prediction about the utility of BAFs, and the stability of the spatial images that would result from implementing BRTFs.

We chose the cancellation method for this research to reveal the amplitude and phase relationships between air-conducted and bone-conducted tones of different frequencies. This method has a participant adjust 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 cancellation method was pioneered by Békésy in 1932 to show that air and bone conduction share hearing mechanisms (Bekesy, 1932 as cited in Békésy, 1960), and has been used by several researchers since then (e.g., Boezeman et al., 1984; Kapteyn, Boezeman, & Snel, 1983; Khanna, Tonndorf, & Queller, 1976; Levitt, 1987). The purpose of those studies varied considerably. The purpose of those studies included determining inner and outer ear components of bone-conduction hearing (Khanna, Tonndorf, & Queller, 1976), assessing the functional gain of a hearing aid system (Levitt, 1987), measurement of extremely high levels of hearing loss
where cross hearing can be a problem (Kapteyn, Boezeman, & Snel, 1983), and understanding the phase relationship between air and bone-conducted signals (Boezeman et al., 1984).

Thus, those studies were not designed to be applied to developing BAFs. The results from these studies are difficult to aggregate across studies, and have limited generalizability to the purpose of the present research question, due to several important factors. They include the design and location of the transducer, the method of measuring amplitude, and the stimuli tested.

In those studies, there was some report of cancellation instability. Specifically, Kapteyn et al. (1983) noted a high degree of variability in phase adjustments across people, and Khanna et al. (1976) noted a high degree of cancellation sensitivity due to head and jaw movements. This suggests that getting a stable BRTF and spatial image may be difficult, due to unstable phase relationship between the signals.

The present study used the cancellation method with stimuli presented through bonephones mounted on the mastoid and through earbud headphones. The results of this study should yield a description of the amplitude and phase relationships between air-conducted and bone-conducted tones at different frequencies, defining a set of preliminary series of “bone-to-air” shifts that can later be combined with other frequencies to form a complete BAF. The variability of these relationships will also be considered to understand the consistency of the resulting percept, both across time and people.

**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 1.
The schematic of the conditions in Figure 1 shows the left ear as the test ear (TE) and the right ear as the non-test ear (NTE). The bone-to-air condition was administered to describe the amplitude and phase relationships between air-conducted and bone-conducted tones at select frequencies. Figure 1 shows that in the bone-to-air condition, 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. Piloting showed that this was required for cancellation to occur. 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 one of the tones in the TE until it canceled out the other tone in the TE.

Figure 1 also depicts the air-to-air condition, which was administered to calibrate for participants and equipment. As in the bone-to-air condition, bandstop noise was delivered in the
TE to mask the harmonics, and bandpass noise was delivered to the NTE to prevent it from contributing to the response. In the air-to-air condition, 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. The deviation from these values suggests the amount of participants’ error.

**Participants**

There were 10 volunteer participants (six males, four females, mean age = 25.8) from the graduate student community of the Georgia Institute of Technology. They were screened for normal hearing (sensitivity to 20 dB pure tones), using a Micro Audiometrics Corporation DSP Pure Tone Audiometer.

**Stimuli**

Each of the two listening conditions was tested with sinusoidal tones at the following three frequencies: 500, 3150, and 8000 Hz. Table 1 shows the initial energy level at which the tone not being adjusted was delivered (the tone being adjusted began at zero amplitude). These levels were established by preliminary testing, setting the levels at the point at which the tone was first clearly audible.

<table>
<thead>
<tr>
<th>Center (Hz)</th>
<th>Air-conducted$^1$</th>
<th>Bone-conducted$^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>500</td>
<td>72.9</td>
<td>24.8</td>
</tr>
<tr>
<td>3150</td>
<td>50.1</td>
<td>18.2</td>
</tr>
<tr>
<td>8000</td>
<td>48.5</td>
<td>19.6</td>
</tr>
</tbody>
</table>

$^1$dB re 20µPa; i.e., dB SPL  
$^2$dB re 3.16 cm/s$^2$ (acceleration)
The 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. 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 frequency of the tone being tested (ANSI, 2004). Table 2 shows the upper and lower bound of the maskers’ frequency bands. White noise was filtered through 4-pole Butterworth filters to produce the bandpass or bandstop filters. The sound pressure level output of the maskers can also be seen in Table 2.

Table 2

<table>
<thead>
<tr>
<th>Center (Hz)</th>
<th>Lower Limit (Hz)</th>
<th>Upper Limit (Hz)</th>
<th>Stop dB&lt;sub&gt;A&lt;/sub&gt;</th>
<th>Pass dB&lt;sub&gt;A&lt;/sub&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>500</td>
<td>445</td>
<td>561</td>
<td>66.1</td>
<td>58.0</td>
</tr>
<tr>
<td>3000</td>
<td>2670</td>
<td>3370</td>
<td>59.5</td>
<td>57.9</td>
</tr>
<tr>
<td>8000</td>
<td>7130</td>
<td>8980</td>
<td>60.2</td>
<td>58.5</td>
</tr>
</tbody>
</table>

Measurement and Calibration of Acoustic Signals

The air-conduction equipment was calibrated to a reference force value of 20 microPascals. The bone-conduction measurement system was calibrated to a reference acceleration value of 0.0316 m/s². This reference value was chosen as an approximation of the bone-conduction threshold (see Brüel & Kjaer, 1974), much like the traditional value used for air conduction. This was done instead of the 1 microNewton standard for audiometry (see ANSI, 2004) to have a set of logarithmic metrics that were both based on a reference value that approximates the human threshold for hearing sound delivered through that pathway.
Acceleration was used instead of force because it avoids difficulties associated with the mass of other parts of the system used during calibration interfering with the measurement’s accuracy (see Brüel & Kjaer, 1974). Using acceleration also creates a measure independent of the transducer’s mass, unlike force.

A system of sound measurement hardware was used to measure the physical output of the air-conduction and bone-conduction transducers. In the case of air conduction, the earbud was secured to the microphone assembly with electrical tape and a seal between the two was formed with a #8 rubber O-ring. The microphone assembly consisted of a 2cc coupler secured to a Bruel & Kjaer Type 4146 microphone. For the measurement of bone-conducted sounds, the bonephone was clamped into a Bruel & Kjaer Type 4930 Artificial Mastoid. The output of the microphone or artificial mastoid was then connected to a Bruel & Kjaer Type 2610 measurement amplifier, which was connected to a Bruel & Kjaer Type 2260 sound level meter.

The output of both the headphones and bonephones was measured at a variety of levels, so that a function could be plotted between input values (specified by the software) and the corresponding device output values. The software-specified input values that the participant submitted were recorded in a file. Based on the functions obtained during measurement, these input data were processed to compute the resultant output value.
**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 in the computer. The rotary knob sent messages to the computer to increase the amplitude and phase, and the software made these adjustments in set intervals, or “steps”. The software used to generate sounds and provide a user interface was designed with SuperCollider, a sound programming language for real-time audio synthesis running on the Macintosh OS X operating system. The SuperCollider interface visually displayed relative phase and amplitude values that updated as they were adjusted, and allowed participants to submit their final phase and amplitude settings that led to cancellation.

The sound delivery apparatus, set up for delivering the bone-to-air condition, can be seen in Figure 2. The tone and masker in the TE originated from the computer while the bandstop noise in the NTE originated from a CD player that was playing a track of noise.
Figure 2. 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

The experiment consisted of two sessions, each lasting between 45 minutes and two hours. For each session, participants were first screened for normal hearing with the audiometer. Then the participant completed one block of the air-to-air condition, followed by a block of the bone-to-air
condition. Each block began with 10 practice trials at a constant but randomly chosen frequency (500, 3150, or 8000 Hz) and a 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 tone was softest, and then go back and modify amplitude and/or phase until the sound was as quiet as possible. The phase and amplitude each began at zero. 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. Participants were instructed that they could search in the same general area for a cancellation point during the whole air-to-air block, and within a run in the bone-to-air block (each run tested the same frequency tone). This avoided them having to search the entire amplitude–phase space on every trial.

The blocks of trials consisted of five phase/amplitude adjustments at each ear, for three frequencies, yielding a total of 30 phase/amplitude adjustments per block (not including practice trials). The TE was blocked and counterbalanced, and both ears were tested before moving to the next randomly-selected frequency. The air-to-air condition was the same in each session. The bone-to-air condition, however, differed between sessions in the tone that was being adjusted. In the “adjusting air” 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 (see Table 2). In the “adjusting bone” 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 (see Table 2). The order of sessions was counterbalanced between participants.
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. Error was defined in terms of the deviation from the physically correct values for cancellation of two waves passing through the same medium: namely 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. The step error was aggregated across trials by computing the root-mean-square (RMS) step error metric across trials. Practice trials were not considered in the analysis, and data were collapsed across both sessions and all trials, then averaged across all participants.

There were a total of 30 steps on the amplitude slider. The RMS step error for amplitude was on average, 0.70 steps. There was a range of 88 steps on each side (positive and negative) of the phase slider. The RMS step error for amplitude was on average 2.81 steps.

Bone-to-Air

Subjective Report of Cancellation

The subjective report of cancellation was solicited from participants during the experiment. Participants indicated that cancellation in the bone-to-air condition was very sensitive to head movements. In general, participants indicated that cancellation comparable to that of the air-to-air condition occurred. There was an occasional run in the experiment 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, the tone got louder (i.e., a local minimum in the loudness space). One participant discontinued participation in the experiment because they could not find this “trough” of loudness at one point in the experiment. That participant’s data were excluded from analysis.

**Amplitude Adjustments**

The physical output of the headphones (“Air dB”) and physical output of the bonephones (“Bone dB”) was computed for each trial based on the scaler-output functions established in measurement. The 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.

A 2x3 within-subjects ANOVA was conducted on the mean amplitude shift across trials, with pathway adjusted (bone or air) 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 3. This analysis revealed statistically significant main effects of pathway and frequency. There was not a statistically significant interaction between pathway and frequency.
Table 3

*Analysis of Variance for Amplitude Shift*

<table>
<thead>
<tr>
<th>Source</th>
<th>$df_{Effect}$</th>
<th>$df_{Error}$</th>
<th>$MSE_{Effect}$</th>
<th>$MSE_{Error}$</th>
<th>$F$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pathway (P)</td>
<td>1</td>
<td>9</td>
<td>42.34</td>
<td>5.42</td>
<td>7.81</td>
<td>.02</td>
</tr>
<tr>
<td>Frequency (F)</td>
<td>2</td>
<td>18</td>
<td>2502.05</td>
<td>28.09</td>
<td>89.07*</td>
<td>&lt; .01</td>
</tr>
<tr>
<td>P X F</td>
<td>2</td>
<td>18</td>
<td>9.21</td>
<td>4.23</td>
<td>2.18</td>
<td>.14</td>
</tr>
</tbody>
</table>

*p < 0.05

The effects of frequency and pathway can be seen in Figure 3 below. The main effect of pathway on amplitude shift was such that the amplitude shift was consistently higher when adjusting air ($M = 32.1$ dB shift, $SE = 0.86$ dB Shift) than when adjusting bone ($M = 30.4$ dB shift, $SE = 0.92$ dB shift).

![Figure 3](image-url)  

*Figure 3.* Main effects of frequency and pathway. The error bars represent one standard error above and below the mean. The minimum y-axis value was chosen to improve graph readability, and does not represent a minimum amplitude shift.
Follow-up pairwise comparisons (with Bonferroni correction) showed that the main effect of frequency was such that there was a statistically significant difference between 500 Hz and 3150 Hz \((p < .01)\), and between 500 Hz and 8000 Hz \((p < .01)\), but no statistically significant difference between 3150 Hz and 8000 Hz \((p > .95)\).

**Time Shift Adjustments**

For the analysis of data in the bone-to-air condition, phase was converted to time shift. This was done to avoid the dependence of phase on frequency. Time shift was then averaged across participants. A 2x3 within-subjects ANOVA was conducted on the median time shift across trials, with pathway adjusted (bone or air) and frequency (500 Hz, 3150 Hz, or 8000 Hz) as within-subjects independent variables. This analysis revealed no statistically significant interaction or main effects (see Table 4). Figure 4 shows the time shifts as a function of frequency and pathway adjusted. This figure shows the high degree of variability in time shift between people, as well as the difference in variability across frequencies.

**Table 4**

*Analysis of Variance for Phase Shift*

<table>
<thead>
<tr>
<th>Source</th>
<th>(d_{\text{f Effect}})</th>
<th>(d_{\text{f Error}})</th>
<th>(MSE_{\text{Effect}})</th>
<th>(MSE_{\text{Error}})</th>
<th>(F)</th>
<th>(p)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pathway (P)</td>
<td>1</td>
<td>9</td>
<td>1066.82</td>
<td>22141.04</td>
<td>0.05</td>
<td>.83</td>
</tr>
<tr>
<td>Frequency (F)</td>
<td>2</td>
<td>18</td>
<td>2227.62</td>
<td>2646.44</td>
<td>0.84</td>
<td>.45</td>
</tr>
<tr>
<td>P X F(^1)</td>
<td>1.1</td>
<td>9.7</td>
<td>2984.54</td>
<td>46237.30</td>
<td>0.07</td>
<td>.82</td>
</tr>
</tbody>
</table>

\(^1\)Mauchly’s test of sphericity showed that sphericity could not be assumed, so Geisser-Greenhouse corrected degrees of freedom were used.
Further analyses were done to better understand the lack of main effects and interaction of the independent variables for time shift. The large error bars indicate that there was a large degree of variability between participants, which suggests that the null effects were not due to the time shift being the same across participants. Knowing this, there are two remaining explanations for the null effects of the independent variables on time shift. One is that there was a consistent lack of dependence of time shift on the factors investigated here, but that this constant time shift was different for each person. The other is that the pattern of means generated by invoking the independent variables was not consistent across people. To investigate these two possibilities, the data were plotted individually in Figure 5, panels A and B. Figure 10 shows the median time shift averaged across ears, for each frequency. Panel A shows these data for when
the bone pathway was being adjusted, and panel B shows these data for when the air pathway was being adjusted. Each separate line indicates a different subject for both panels of Figure 5, and the same symbols for each subject were used in each panel. This figure suggests that a difference in the pattern of mean phase between people across the independent variables prevented the null hypotheses from being rejected. [Bruce – I was a little unsure about how much detail to go into here – I hope that I summarized without being inaccurate. The most detailed way of investigating the null effects would be to consider each of the null main effects and the null interaction individually…here I tried to paint in broad brushstrokes all three.]
Figure 5, Panels A and B. Each panel shows the median phase shift (expressed in microseconds) at different frequencies tested. Each line indicates a different subject. Panel A shows when the bone pathway was being adjusted, and Panel B shows when the air pathway was being adjusted. The same symbols for each subject were used in each panel.

Figure 5 shows between-subject (inter-individual) differences in time shift patterns. It is also interesting to look at how stable the time shifts were for a participant within a given condition, and thus consider intra-individual differences. The range of time shift values across trials (“intra-individual time shift range”) was computed for each condition and participant. Then, for each participant, this value was averaged across conditions. Across participants, the mean intra-individual time shift range was 42.7 microseconds, and the standard deviation of the
mean intra-individual time shift range was 23.6 microseconds. For comparison, the range of median time shift values across participants for a given condition was also computed, and then averaged across conditions (“inter-individual time shift range”). Across participants, the mean inter-individual time shift range was 350.4 microseconds, and the standard deviation of the inter-individual time shift range was 346.6 microseconds.

**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. Participants were typically less than one step, out of 30, away from the correct amplitude shift. For phase, they were typically less than three steps out of 88 away from the correct value. These data suggest 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 similarly small amount of error.

**Bone-to-Air**

**Subjective report of cancellation**

The subjective reports, combined with the calibration block, suggest that the participants in this study were able to achieve the perception of cancellation by manipulating the relationship between air and bone-conducted sounds, as has been documented in previous studies (e.g., Boezeman et al., 1984; Boezeman et al., 1983; Kapteyn et al., 1983; Khanna et al., 1976; Levitt, 1987). This study tested the mastoid location with a transducer designed for auditory displays at a variety of frequencies, and physically measured the output of the transducer corresponding to
when cancellation occurred, all of which make this more suitable for the goal of adjusting HRTFs for bonephones than previous research.

The dependency of cancellation on head movements confirms previous findings (e.g., Khanna et al., 1976), and is likely due to the coupling between the bonephones and the skull changing as the head moves, which in turn alters the phase and amplitude of the signal at the level of the cochlea. This sensitivity to head movement, in combination with the occasional difficulty with achieving cancellation in the bone-to-air condition, suggest that there may be difficulty with producing stabilized spatial images. 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.

**Amplitude**

The amplitude shift required for cancellation depended on which tone was being adjusted: the amplitude shift that participants adjusted was greater when adjusting bone than when adjusting air. These differences are likely due to loudness differences in the static tone between when the bone-conducted and air-conducted tone was being adjusted.

The dependence of amplitude shift on frequency shows that establishing a set of shifts, rather than a single overall shift, is important for producing the appropriate amplitude shift for adjusting HRTFs. The follow-up pairwise comparisons suggested that amplitude adjustment of HRTF filters should be similar at 3150 and 8000 Hz, and higher for 500 Hz. Future work should consider more frequencies between these points and generate a full BAF (Bone-adjustment function).

The empirically-based amplitude shift values specified here provide a starting point for a function of amplitude shifts that could be implemented in BAFs, which can be applied to HRTFs.
to form BRTFs. Amplitude shift values can be used to leverage the already widely-researched air-conduction spatial audio filters so that the same percept can result for bone-conduction. As an example, the 500 Hz amplitude shift data point found in this study could be applied to an HRTF by altering the amplitude component of the filter so that the output of the air-conduction headphones is 45 dB more than the output of the bonephones, using the measurement standard used in this study.

**Time Shift**

The results of the time shift adjustments did not produce as clear recommendations for BAFs as the amplitude adjustments did. Specifically, inferential statistics showed no difference in time shift as a function of pathway or frequency, and no interaction between the two. Individual data plotting showed that the null effects were due to a difference between people in how the independent variables affected time shift. There was also a particularly high amount of variability associated with the lowest frequency tested, 500 Hz. This suggests that lower frequencies are more sensitive to differences between people’s head that result in differential coupling and sound travel. The high variability confirms findings by Kapteyn et al. (1983).

The individual data plots and the inter-individual time shift range show that the time shift that needs to occur for a given frequency within a given person is not predictable based on aggregate data. This lack of consistency between people in the perception of sounds delivered through bonephones brings into question the feasibility of developing generalized BRTFs. The large inter-individual differences suggest that manipulating the time-dependent aspects of generalized BRTFs may not be effective. These large differences are 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. To the extent that BRTFs are necessary for accurate localization, this data also suggests that there would be lack of consistency in the perceived spatial location across people.

However, the intra-individual time shift range was quite small and consistent across people, which shows potential for individualized BRTFs. To give an idea of how the magnitude of intra-individual time shift range corresponds to perceived spatial location, it may be useful to consider the relationship between interaural time difference and perceived lateralization for an air conducted stimulus, which has been plotted by Feddersen (1957). According to their graph, a difference of 50 µsec would make an approximate difference of 10 degrees in perceived azimuth location for a broadband stimulus. It is interesting to note the contrast between the relative stability within a person of the time shift adjustments for cancellation and the instability of the perception of cancellation. This contrast suggests that although there is a high degree of intra-individual consistency, there could be some degree of instability within a person.

**General Implications in the Application of Adjustment Functions**

This study provided an initial description of how to alter sounds to simulate spatial audio through bonephones, by describing amplitude components of a BRTF. The time shift data demonstrated possible problems with the feasibility of making generalized BRTFs, bringing into question the consistency in perceived spatial location of a sound across people. The consistency within a person, however, indicated that an individualized BRTF could be useful, and that sound images within a person could be stable. The amplitude data shows the utility of this paradigm for finding recommended bone-to-air shifts, and provides initial anchor points for a BAF that can be used to form a BRTF.
The implied lack of consistency in perceived spatial location across people provided by these data may seem to be in disagreement with a study that found localization performance to be similar for HRTFs implemented on bone-conduction transducers and headphones (MacDonald et al., 2006). However, there were some important aspects of this study that could explain these differences. The stimulus was broadband, which would allow use of ILDs on the higher frequencies. Also, the noise presented in that study could have degraded the listener’s sensitivity to differences between the conditions in localization performance.

Future work should further consider the stability of temporal-related adjustments in spatial audio across people and within a person and the impact that this has on HRTF adjustment functions, as well as considering how localization performance is altered by using BRTFs on bonephones, as compared to regular HRTFs. It may also be beneficial to consider alternate methods for obtaining BAFs, such as using otoacoustic emissions (e.g., Purcell, Kunov, & Cleghorn, 2003). Current work with physical models of sound traveling through the skull is underway in our lab in collaboration with others (Przekwas et al., 2007; Walker et al., 2007). This physical modeling may lead to a better understanding of differences between and within people in the temporal nature of bone-conducted signals traveling through the human head. In general, any improvements in understanding how to make a bone-conducted signal match an air-conducted signal could improve the implementation of auditory displays with bone-conduction transducers. Specifically, it would allow the application of the large amount of previous research on relevant concepts in air-conducted auditory displays to be applied in bone-conducted auditory displays.
References


