Relative threshold curves for implementation of auditory displays on bone-conduction headsets in multiple listening environments.

Raymond M. Stanley and Bruce N. Walker
Georgia Institute of Technology

AUTHOR NOTE
Preliminary discussion of this work, with different analyses, was presented at the 2005 International Conference on Auditory Display. Correspondence concerning this article should be addressed to Bruce Walker, School of Psychology, Georgia Institute of Technology, Atlanta, GA, 30332-0170. Electronic mail may be sent to bruce.walker@psych.gatech.edu.
ABSTRACT

Headphones can be unacceptable because they attenuate ambient sounds. Bone conduction headsets leave the ears uncovered, yet maintain portability and privacy. We studied detection thresholds for “bonephones” with open and plugged ears, in quiet and noisy conditions. Relative thresholds for 150-13500 Hz are plotted, and we discuss the utility of this information for designing auditory displays using bonephones.
Relative threshold curves for implementation of auditory displays on bone-conduction headsets in multiple listening environments.

0 Introduction
Designers of auditory displays are often careful to consider and include the research on auditory perception, stream segregation, and psychoacoustics as they determine what sounds to create. However, we have found that less consideration is paid to the actual hardware used to present the auditory display to the listener. In some cases the oversight can be problematic, especially in listening situations that are somewhat atypical.

For example, although headphones are often used for mobile audio, they are not always the most appropriate hardware. For mobile applications the ability to hear ambient sounds is crucial. Example users would be visually impaired people who rely on environmental audio cues for their primary sense of orientation, but who also have a GPS-enabled guidance system presenting spoken directions to them. Another example would be bicycle couriers who need to listen for cars, while also being able to monitor their dispatcher over the radio. In both of these examples, using headphones deteriorates the detection and localization of external sounds, so an alternative would be preferred [see 1]. In addition to the performance aspects of wearing headphones, there are aesthetic and social issues to consider in some cases: participants using auditory navigation systems may not want to have their ears occluded by headphones or earphones.

On the other hand, there are situations where hearing the ambient sounds is actually not desired. For example, in noisy environments such as on a factory floor or in a military vehicle hearing protection is required, but using headphones to present an auditory display or radio communication does not allow insertion of earplugs for hearing protection. This makes
implementation of auditory displays in a noisy environment difficult.

Unfortunately, when the human factors aspects of the situation dictate an alternative to headphones, there are relatively few options. One approach that has great promise is bone conduction, since the auditory system is also sensitive to pressure waves transmitted through the bones in the skull [2]. Although bone-conducted sound waves are naturally transmitted when listening to one’s own voice and when exposed to loud external sounds, mechanical transducers can also directly transmit sound through the skull. Bone-conduction headsets leave the ear canal and pinna unobstructed, allowing environmental sounds to be accessed or blocked, as appropriate.

Because air conduction and bone conduction are quite different pathways, guidelines for the design of sounds for auditory displays (warnings and speech communication included) cannot necessarily generalize across pathways. Unfortunately, there is little empirical research available about perception of sound through modern consumer-audio bone conduction devices (“bonephones”). The present research uses psychophysical methods to begin to provide information about the perceptual properties of listening through bonephones. These data can then be used to optimize the design of auditory displays that will employ bone conduction output devices.

Bone-conduction headsets have traditionally been used in the field of audiometry, which will be discussed in greater detail below. A new style of bone-conduction headsets differs from traditional bone-conduction transducers used for audiometry: these bonephones can be binaural, come with a standard 1/8” input jack, and are small in size, all of which makes them especially suitable for use with auditory displays in mobile applications (see Figure 1). We begin to study the use of bonephones for mobile auditory displays in this paper by investigating the relative
signal amplitudes that have to be delivered to the bonephones in order for a listener to hear sounds of various frequencies. The particular model of bonephones tested in this study (Temco HG-28) has transducers that contact the mastoid (a raised portion of the temporal bone located directly behind the ear). The mastoid is as close to the inner ear as possible (it encases the inner ear) and is also as lateral as possible on the head, which together maximize the effectiveness of dichotic (stereo) sound presentation through bonephones at that location. Although others have found lower thresholds at the condyle location [see 3], the mastoid may be considered a preferable location in many applications. For example, the mastoid location results in transducer-to-head coupling that is unaffected by jaw movements involved in talking, which is important in radio communication. Placing the transducer on this location is also quite unobtrusive, which generally supports users’ aesthetic preferences.

Figure 1. Photograph of “bonephones.” These Temco HG-28 bone-conduction headsets are designed for use with auditory displays. They are binaural, have a standard 1/8” input jack, are small in size, and contact the mastoid portion of the skull.

The use of bone conduction to present sounds to listeners is by no means new – published investigations of hearing by bone conduction date back at least as far as 1603 [Ingrassia, 1603, as cited in 4]. Much research since then has been devoted to specifying thresholds for the purpose
of identifying the locus of hearing damage in the field of audiometry [e.g., 5, 6]. Specifically, bone conduction can be used to assess the locus of hearing damage by comparing air-conduction thresholds to bone-conduction thresholds [7]. Most other bone-conduction research has focused on understanding some other aspect of treating hearing disorders [8-11], or the basic mechanisms of bone conduction [2, 12-14].

The thresholds for hearing sounds of different frequencies are important to have when designing an auditory display. Clearly, this information can be used to avoid an inaudible sound, which would prevent information from being extracted from the display. Previous research, however, has used clinical bone-conduction headsets that are substantially different in design, and the results have been specified in terms of the physical output of the devices, which is less useful when implementing auditory displays via bonephones.

In an applied design setting, the relative signal levels sent into the bonephones is a more useful specification of thresholds than the level of transducer output. This is because the auditory display designer generally only directly controls the output of the audio generation device (i.e., relative levels in a sound file), and thus the input into the acoustic apparatus. Further, measuring and calibrating the absolute output levels of a bone-conduction device requires substantial equipment (e.g., an artificial mastoid, oscilloscope, etc.) and training beyond that of most auditory display designers. Therefore, thresholds specified at the level of the relative input into the bonephones are of most practical use.

For the reasons discussed to this point, the purpose of this study was to provide information about the required input to hear a sound at various frequencies; this could then be used to guide the selection of the spectral components of a sound being designed. Components that require more energy to be heard through the bonephones can be avoided, and those that require less
energy can be selected.

In cases where the ambient soundscape is important (and available) to the listener, the entire acoustic environment will interact with the sounds presented via the bonephones. Understanding this interaction better will enable a designer to tune their auditory display for maximum comprehension. Since the effect of listening environments relevant to auditory display design has not been investigated—regardless of the type of transducer being used—a secondary goal of this study was to consider bonephone listening in different ambient noise conditions.

0.1 *The Present Study*

The present study examined detection thresholds, as measured by relative bonephone input levels, at various frequencies throughout the audible range. This yielded relative threshold curves, and was completed in several different listening conditions: open ear, open ear with noise, and plugged ear. The open ear condition provided information about the relative threshold curve for a quiet environment with no earplugs inserted. The open ear with noise condition delivered external pink noise that is representative of the spectral content of sounds coming from the everyday environment. The plugged ear condition showed the relative threshold curve with bonephones when earplugs are inserted into the ear canal.

The few previous standardized measurements of bone-conduction thresholds have been defined at the level of transducer output. While the intent is somewhat different, those studies can provide some insight into the general shape of thresholds we might expect to find in the current study. For example, the ANSI standardized bone-conduction threshold measurements have the lowest thresholds between 2000 and 4000 Hz, somewhat higher thresholds from 5000 – 8000 Hz, and the highest thresholds below 1000 Hz [15].
Threshold curves for bone conduction

1 Method

1.1 Participants

Participants were five graduate students at the Georgia Institute of Technology (3 males, 2 females, mean age = 24.6 years, range = 23-26). Each were compensated a total of $30 for their participation. Participants’ audiometric hearing thresholds were tested before each experiment session with a Micro Audiometrics Corporation DSP Pure Tone Audiometer. All participants had audiometrically normal hearing, with pure-tone thresholds less than or equal to 20 dB HL at 250, 500, 750, 1000, 1500, 2000, 3000, 4000, 6000, and 8000 Hz [see 16].

1.2 Stimuli

Participants listened to 1-second long pure tones (.wav format, 44.1 kHz sampling rate, 16-bit depth), presented binaurally via the bonephones. Pure tones were generated for each of the following one-third octave band centers: 150, 250, 350, 450, 570, 700, 840, 1000, 1170, 1370, 1600, 1850, 2150, 2500, 2900, 3400, 4000, 4800, 5800, 7000, 8500, 10500, 13500 [see 17]. The one-third octave band centered at 50 Hz was not tested because the bonephones could not accurately and reliably reproduce this frequency. For each frequency, a set of tones was generated with attenuation levels ranging from 0 to -135 dB, in 3 dB steps. The voltage level of the signal at 0 dB attenuation sent to the bonephones can be seen in the Appendix. The pink noise used in the masking condition was digitally generated, stored as an hour-long audio file (.wav format, 44.1 kHz sampling rate, 16-bit depth), and written onto a CD for playback during the noise condition.

1.3 Apparatus

The presentation of the stimuli was controlled with a program written in Eprime (Psychology Software Tools, Inc.), running on a Dell Dimension 8100 Pentium-4 PC with Windows XP. The digital sound files were output via a Creative Labs Audigy sound card, and
then amplified with a Behringer HA4600 headphone amplifier, to which the bonephones were connected. Participants listened to tones delivered to their mastoid with Temco HG-28 bone-conduction headsets. In the masked listening condition, the pink noise was played from the CD using a Samsung V1000 DVD player, amplified with a Denon DRA-275R stereo receiver, and delivered through Klipsch KSB 1.1 speakers. The speakers were located directly out to each side of the listener (i.e., at –90 and +90 degrees), approximately 42 inches from the center of the listener’s head, and the noise was played at a level of 45 dBA SPL, as measured at the approximate center of the listener’s head by an Extech Instruments 407750 Digital Sound Level Meter. In the plugged listening condition, participants inserted EAR foam earplugs into their ear canals, with instruction from the experimenter on correct insertion technique. Participants completed the procedure in an Industrial Acoustics Company sound-attenuated room, with the computer located outside of the room. Visual information was presented on a 14-inch (35.56 cm) Viewsonic LCD monitor, and participants made their responses using a standard computer keyboard.

1.4 Procedure
Each participant completed three sessions lasting approximately two hours each. Sessions were separated by one day of rest and each involved a different listening condition (open, plugged, or masked). A session consisted of (1) finding a rough estimate of their threshold at each frequency, followed by (2) a systematic assessment of their threshold at each frequency with a staircase procedure (described below). First, the approximate threshold of each participant was estimated by the method of limits performed with a verbal yes/no task. This initial threshold estimate was used to guide the settings for the upper and lower bounds of the staircases. Next, the detailed threshold measurement began with a block of practice trials at 1000 Hz, followed by
23 experimental blocks (one for each frequency). At the beginning of each frequency block, participants heard a sample tone, which was the same intensity as the loudest sound presented in the given staircase. During the plugged and masked listening conditions, participants were subjected to the earplugs or masking noise soon after entering the room, and completed the rest of the session in that state. The threshold measurement task was two-interval forced choice: participants were asked to indicate in which of two time intervals the sound had played. Five hundred milliseconds (msec) after initiating a trial with a press of the space bar, a “1” appeared on the display for one second, followed by an interval of 500 msec with a blank screen, and then a “2” appeared on the display for one second. One of the sound files was always played in either the first or second interval, and no sound was delivered in the other interval. Participants would then indicate which interval they heard the sound in by responding with a “1” or “2” on the numeric pad of the keyboard. Participants then received accuracy feedback before beginning another trial. Threshold was assessed for each frequency with two randomly-interleaved staircases, one ascending and one descending. Each staircase was an up-down transformed staircase (UDTR), following a 1-up, 2-down rule so that it converged on the 70.7% threshold [18]. This method of threshold assessment allows a high amount of efficiency, while maintaining sufficient complexity to avoid participants anticipating stimulus values. The step size was 3 dB of attenuation, and 10 steps could occur before the attenuation would no longer increase or decrease. Each block ended when 7 reversals occurred for both the ascending and descending staircases, or when responses continually drove the stimulus values to the upper or lower bounds of the staircases.

Despite pre-testing the thresholds, the procedure did not converge to a clear threshold in every single frequency block. In the cases where it did not converge, participants came back and
collected data with a different set of upper and lower bounds. Most of the time these re-
collections were successful. Sometimes an exact threshold was still not determined. This 
occurred when participants’ response in conjunction with the staircase rules caused the stimulus 
values to stay consistently at the loudest or softest value. This occurred for three participants, in a 
total of 5 blocks. When this occurred, the intensity of the softest sound in the staircase was used 
as the threshold value if the participant’s threshold stayed near the softest sound, and the loudest 
sound was used if the participant’s threshold stayed near the loudest sound.

2 Results

The signal intensity determined to represent the 70.7% threshold was averaged across 
participants at each frequency and for each listening condition. A visual depiction of these results 
can be seen in Figure 2. Error bars in the figure represent confidence intervals based on the 
within-subjects error term, as recommended by Loftus and Mason [19] in studies primarily 
concerned with the within-subjects effects. In our case, the result we are most interested in is the 
threshold pattern across frequencies (more than across individuals), making the within-subjects 
error terms more appropriate. A secondary concern is the difference between listening 
conditions, and this is clearly noted in the figure by the different lines representing open, open 
with noise, and plugged conditions.
To create a within-subjects confidence interval (CI), the subject-by-condition interaction mean squares (the “error term” used in ANOVA) is used in the following equation:

\[ CI = M_j \pm \sqrt{(MS_{SC}/n) \times \text{criterion } t(df_{SC})} \]

where \( M_j \) is the mean threshold value at a given frequency in a given condition, \( MS_{SC} \) is the mean sum of squares for the Subject x Frequency interaction, and criterion \( t(df_{SC}) \) is the critical value of the \( t \)-statistic for a 95% interval [see 19]. Note that the confidence intervals shown here

---

**Figure 2:** Relative threshold curves for each listening condition. From the top of the figure: open ears with noise (worst performance overall); open ears with no noise; and plugged ears with no noise (best performance). Error bars are confidence intervals based on the error term for the effect of frequency.
are computed differently from the error bars that we have used in a preliminary discussion of this work [see 20], but we believe them to be more accurate for analyzing the pattern of mean thresholds in this study.

The visual depiction of the threshold curves (see Figure 2) gives the richest description of the data, both in terms of the effect of listening condition, and the effect of frequency within each listening condition. In general, for each condition, thresholds are highest (i.e., performance is worst) at the lowest frequencies, and thresholds decrease rapidly until 1170 Hz, which is approximately where the minimum threshold occurs. After that point, the thresholds increase again, but more slowly than in the low-frequency descent. Figure 2 shows that the open ears with noise listening condition has significantly higher thresholds than the other curves at all frequencies, with the difference diminishing somewhat at the highest frequencies. The plugged ears condition has lower thresholds than the open ears condition at frequencies below about 1850 Hz; the two curves converge as frequency increases beyond that point.

For the open ears with noise condition, above about 700 Hz there were similar thresholds across frequencies. Below 700 Hz, there were consistently increasing thresholds in the open ears with noise listening condition. For the open ears condition, the highest thresholds were below 700 Hz and above 7000 Hz, with the lowest thresholds between these frequencies. For the plugged ear condition, the thresholds were especially high at frequencies greater than or equal to 7000 Hz and less than or equal to 350 Hz. The thresholds were lower between 1600 and 4800 Hz, as well as between 450 and 700 Hz. The thresholds were lowest between 840 and 1170 Hz.

3 Discussion

It is important to keep in mind that in this research, the statements about threshold are not absolute in nature, since they are measured in terms of the energy input into the bonephones,
relative to other listening conditions and frequencies. A lower threshold represents less energy
input (less voltage, or an attenuation in an audio file), and thus, overall, a sound that is easier to
hear. All of the audibility measurements include a combination of the bonephone frequency
response characteristics and the listener sensitivity to the transducer’s output.

One might assume, then, that the frequency response of the bonephones could simply be
subtracted from the curve presented here to yield a curve in terms of the physical output of the
bonephones. In some studies that are focusing on the perceptual component of the system, this
sort of approach can be attempted. However, the goal here is to consider the system as a whole,
with the intended user being the auditory display designer who may not necessarily be as
concerned with the purely perceptual phenomenon. Further, the frequency response data on this
device and others is described in terms of the physical force output given a constant input, which
cannot simply be “subtracted out” from a curve of threshold values specified in voltage input
levels.

Further, this research and the recommendations following it are based on testing of a single
set of bonephones. Other designs may be used, and the degree to which these are generalizable to
those designs remains to be determined. Future testing is planned with other varieties of bone-
conduction headsets [e.g., 21].

3.1 Interpretation of Results

The shapes of the relative threshold curves described here have many implications for
choosing spectral content of stimuli for auditory displays implemented on bonephones in
different listening environments. Plots such as Figure 2 can act as a general reference for
choosing that spectral content. We will now discuss some of the more obvious implications from
the curves in terms of inclusion and exclusion of specific frequency components.
The general shape of the thresholds was consistent with the ANSI physical output thresholds, but the particular global minimum was a unique finding. This shows that the response of these devices may differ from those used in the ANSI standard (a RadioEar B71 versus the Temco HG-28 used here), but there are other differences, including the specific pathway involved from transducer to cochlea, slight differences in the frequency makeup of the signal, or the binaural testing used in this study.

The global minimum thresholds in the present study indicate that stimuli with the frequency components around 1200 Hz will be easiest to hear across listening environments. As frequency components lower than that are chosen for stimuli, their audibility will consistently decrease as the frequency decreases, regardless of the listening condition. As frequency components increase, audibility will also decline, but at a less rapid rate. Therefore, if the minimum cannot be used, it is better to increase frequency than decrease frequency.

The relative thresholds between listening conditions revealed how different acoustic environments can effect how easy it is to detect a given frequency component, with a given input into the bonephones. Not surprisingly, introducing ambient noise will require the target sound to be louder for it to be heard through the bonephones, especially at the lower frequencies (see the top line in Figure 2). At frequencies above 4800 Hz, there will be less impact of noise. This is partially due to the reduced high-frequency spectral energy of the masking stimulus, but it is important to remember that pink noise is representative of noise that naturally occurs in the environment.

The relative positions of the open and plugged curves indicates that plugging the ears will actually increase the audibility of the bone conducted stimuli at frequencies below 1850 Hz; at higher frequencies plugging the ears or leaving them open makes little difference in audibility.
These results suggest that if a bonephone auditory display may be used with plugged ears, no audibility will be lost. If lower frequencies are used, using earplugs can even increase the detectability of the stimuli beyond what is possible with open ears. This is consistent with the “occlusion effect” found in the literature [2, 22], in which an increased pressure in the ear canal produces a reduced threshold for bone-conducted sound at lower frequencies. This finding also bodes well for the use of bone conduction audio in tactical situations where ear plugs are required.

The shape of the open ear with noise threshold curve suggests that any frequency component above 700 Hz will have similar audibility, but going below this point will decrease audibility. Based on this curve, it is recommended to avoid using sounds below 700 Hz for bonephone auditory displays implemented in noisy environments.

The shape of the open ear threshold curve suggests that the most audible frequency components to use are between 1000 and 4800 Hz. Designing stimuli to include these frequency components will provide the greatest audibility. If this range of frequencies is not available for use, then using frequencies below 350 Hz should be avoided.

The shape of the plugged listening condition threshold curve suggests that auditory display stimuli will be most likely heard when frequency components between 840 and 1170 Hz are included. If those frequency components cannot be used, then at least frequencies above 7000 Hz and below 350 Hz should be avoided.

3.2 Conclusions

The results presented here give a detailed guide to designers who have to decide on the frequency components of a stimuli within an auditory display implemented on bonephones, in several listening conditions. With this empirical data, frequency components which require less
energy to be heard can be selected to enhance audibility, and frequency components which require more energy to be heard can be avoided. Furthermore, the effect of plugging the ears or delivering the auditory display in a noisy environment can be taken into consideration in the design of stimuli that will be displayed in these environments. With specification in terms of input, the designer is freed from having to conduct extensive device calibration and testing. With this knowledge of relative audibility, stimuli that are most likely to be heard can be chosen for auditory displays implemented on bonephones. This is an important first step in the cognition of auditory displays: if a user cannot hear the sound, then she or he cannot interpret the sound and perform the desired task with the auditory display.

Rather than guiding the optimal frequency component selection to hear a sound, the curves could also be interpreted in terms of the relative boost in power that needs to be delivered for the sounds to be audible. Thus, the curves could also be used as an approximation of the zero-phon equal-loudness curve. It is also important to note that any boost specification would work best for a just-audible sound. The non-linearity of equal loudness curves [23] suggests that deviation from this audibility level would cause inaccurate equalization.

Other recent investigations have considered perceptual issues related to the use of bone-conducted transducers in an applied environment. These issues include spatial audio [24, 25], equal-loudness investigations [26], and speech intelligibility [27]. Together with the data presented here, a better understanding of perceptual issues relevant to use of bone-conduction headsets in auditory displays is emerging, thereby increasing the effectiveness and practical utility of such displays.
4 References


## Appendix

Table A1. Voltage output of system at 0 dB Attenuation (Baseline)

<table>
<thead>
<tr>
<th>Tone Frequency (Hz)</th>
<th>Voltage into Bonephones (V)</th>
</tr>
</thead>
<tbody>
<tr>
<td>150</td>
<td>2.118</td>
</tr>
<tr>
<td>250</td>
<td>2.211</td>
</tr>
<tr>
<td>350</td>
<td>2.293</td>
</tr>
<tr>
<td>450</td>
<td>2.370</td>
</tr>
<tr>
<td>570</td>
<td>2.442</td>
</tr>
<tr>
<td>700</td>
<td>2.509</td>
</tr>
<tr>
<td>840</td>
<td>2.562</td>
</tr>
<tr>
<td>1000</td>
<td>2.597</td>
</tr>
<tr>
<td>1170</td>
<td>2.622</td>
</tr>
<tr>
<td>1370</td>
<td>2.635</td>
</tr>
<tr>
<td>1600</td>
<td>2.508</td>
</tr>
<tr>
<td>1850</td>
<td>2.627</td>
</tr>
<tr>
<td>2150</td>
<td>2.604</td>
</tr>
<tr>
<td>2500</td>
<td>2.567</td>
</tr>
<tr>
<td>2900</td>
<td>2.351</td>
</tr>
<tr>
<td>3400</td>
<td>2.283</td>
</tr>
<tr>
<td>4000</td>
<td>2.198</td>
</tr>
<tr>
<td>4800</td>
<td>2.249</td>
</tr>
<tr>
<td>5800</td>
<td>2.107</td>
</tr>
<tr>
<td>7000</td>
<td>1.961</td>
</tr>
<tr>
<td>8500</td>
<td>1.802</td>
</tr>
<tr>
<td>10500</td>
<td>1.641</td>
</tr>
<tr>
<td>13500</td>
<td>1.467</td>
</tr>
<tr>
<td>No Tone*</td>
<td>0.003</td>
</tr>
</tbody>
</table>

* This was the voltage output when no tone was played. This represents the noise floor of the system.

Note: Measurements of output of headphone amplifier / input to bonephones were taken with a Fluke 75 Series II Multimeter.
Figure Captions

Figure 1. Photograph of “bonephones.” These Temco HG-28 bone-conduction headsets are designed for use with auditory displays. They are binaural, have a standard 1/8” input jack, are small in size, and contact the mastoid portion of the skull.

Figure 2: Relative threshold curves for each listening condition. From the top of the figure: open ears with noise (worst performance overall); open ears with no noise; and plugged ears with no noise (best performance). Error bars are confidence intervals based on the error term for the effect of frequency.