EVALUATION OF BONE-CONDUCTION HEADSETS FOR USE IN MULTITALKER COMMUNICATION ENVIRONMENTS

Bruce N. Walker and Raymond M. Stanley
Sonification Lab, School of Psychology
Georgia Institute of Psychology,
Atlanta, GA 30332-0170
bruce.walker@psych.gatech.edu

Nandini Iyer, Brian D. Simpson, and Douglas S. Brungart
Air Force Research Laboratory,
Wright–Patterson Air Force Base, Ohio 45433-7901

Standard audio headphones are useful in many applications, but they cover the ears of the listener and thus may impair the perception of ambient sounds. Bone-conduction headphones offer a possible alternative, but traditionally their use has been limited to monaural applications due to the high propagation speed of sound in the human skull. Here we show that stereo bone-conduction headsets can be used to provide a limited amount of interaural isolation in a dichotic speech perception task. The results suggest that reliable spatial separation is possible with bone-conduction headsets, but that they probably cannot be used to lateralize signals to extreme left or right apparent locations.

INTRODUCTION

Although notable advantages of using headphones to display auditory information include privacy, portability, and the potential for spatialization of sounds, they do have some drawbacks. Most importantly, headphones cover the ears of the listener, which can impair the detection and localization of ambient sounds in the environment.

Headphones and most other audio output hardware deliver auditory stimuli to the cochlea through the medium of air. However, the human auditory system is also sensitive to pressure waves that are transmitted through the bones in the skull (Tonndorff, 1972; von Bekesy, 1960), a mechanism called “bone conduction.” Bone-conduction headsets that deliver sound through direct application of vibrators to the skull allow the privacy and portability that headphones offer. Unlike headphones, however, the ear canal and pinna remain unobstructed. Recently, binaural bone-conduction headsets have become available. Due to their small size and comfort, as well as potential for dichotic stimulation, these “bonephones” are suitable for implementation in auditory displays.

Most psychoacoustics research and all of the human factors research on auditory displays has focused on air-conducted sound. Since guidelines established for air conduction will not necessarily apply to bone conduction, effective auditory display designs need to be re-evaluated for bonephones. Basic psychophysical data such as audibility thresholds (i.e., Walker & Stanley, 2005) pave the way and constrain the problem space for applied work on auditory display designs.

Research Related to Spatial Audio via Bone Conduction

The goal of most previous research has been to establish threshold norms for clinical testing of middle ear disorders. This research is limited in its applicability to the eventual goal of using the bonephones in a spatial auditory display.

Binaural separation is a basic requirement for spatial hearing. Many researchers have assumed that no spatial hearing is possible with bone conduction because there is too much crosstalk between the ears (Studebaker, 1962a; von Bekesy, 1960). Other evidence, however, suggests some interaural attenuation (Blauert, 1983; Tonndorff, 1972), which may permit lateralization.

Nearly all of the research investigating interaural attenuation (IA) of signals is audiology research that measures air-conducted interaural attenuation, which is estimated to be 60 dB, on average. Audiology handbooks indicate that IA for bone conduction (BC) ranges between 0-20 dB, and that audiologists often assume its lower bound estimate of 0 dB (Katz, 2002). It is important to note, however, that there are few empirical investigations of this subject.

In the only published research available that investigated spatial audio with bone conduction, Kaga, Setou, and Nakamura (2001) found self-reports of sound image lateralization that systematically depended on interaural differences delivered through binaural application of bone-conduction vibrators with the ears plugged. They showed sensitivity to interaural time differences (ITDs) and interaural level differences (ILDs) in both children with normal hearing and children with abnormalities of the middle and outer ears. Furthermore, participants with normal hearing had sensitivities that were not significantly different from ITDs and ILDs assessed through air conduction. This research provides convincing evidence that some interaural attenuation occurs in a stereo bone-conducted signal. Although the mechanisms underlying this binaural separation are not clear, Kaga et al. demonstrate that there may be more binaural separation than typically thought, and therefore that spatial audio with bone conduction may be possible.

However, in the context of an auditory display, many questions remain. For example, the effect of an open ear canal is not known. In addition, a subjective response does not necessarily indicate the ability to perform tasks on a spatial audio task. Furthermore, the high amount of variability associated with vibrator type and placement (Studebaker, 1962b) may alter sensitivity to binaural cues. The present research is intended to address these issues.

A Dichotic Speech Perception Task

One possible application of spatialized bone-conducted audio might involve a multichannel communication system designed to improve intelligibility by spatially separating the apparent locations of two or more simultaneous talkers. Bolia, Nelson, Ericson, and Simpson (2000) have employed the Coordinate Response Measure (CRM) task to assess the efficacy of using spatial audio to enhance speech intelligibility in multitalker communications environments. The CRM task requires listeners to correctly identify a spoken color name and number imbedded in a carrier phrase. This is only to be done, however, if the phrase is addressed to the listener via the target call sign. There are simultaneous talkers uttering distracter phrases of the same structure. The extent to which a listener can correctly identify color-number combinations addressed to the target call sign 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).

Performance in the CRM task as a function of acoustic manipulations corresponding to changes in spatial location of the sound can be used assess the extent to which spatial separation is possible with a stereo bonephone system. The present study examined the effect of ITD and ILD on CRM performance in three conditions, one with audio presented via bonephones with open ears, one with audio presented via bonephones with plugged ears (to eliminate the possibility of acoustic leakage of airborne sound from the bonephone transducers to the listener's ears), and one with audio signals presented via standard headphones.

METHOD

Participants

Eight listeners (3 males, 5 females) with ages ranging from 21 to 55 years participated in the experiment. All had audiometric thresholds of 20 dB HL or less at octave frequencies between 250 and 8000 Hz, and were native speakers of American English. They were trained listeners and were paid for their participation.

Stimuli & Apparatus

The speech materials used in the experiment were phrases from the CRM corpus. The corpus contains recordings of four male and four female talkers, saying phrases in the form "Ready [call sign], go to [color] [number] now." All possible combinations of eight call signs, four colors, and eight numbers (1-8) result in 256 phrases per talker or 2048 phrases in the corpus. On each trial, three phrases were presented to listeners. Of the three phrases, one was designated the target by the presence of the call sign "Baron," whereas the other two (with two different call signs) served as maskers. All three phrases were spoken by the same gender talkers, and were selected to have different color and number combinations. The maskers were presented with no interaural level or time differences, which should have resulted in their being perceived in the center of the listener's head. The target phrases, however, were presented with a non-zero ITD or an ILD, which should have resulted in their being perceived to the left or right of center. ITD values of 0, 100, 200, 400, 800, and 1600 μsec and ILD values of 0, -2, -4, -8, -12, and -16 dB were tested. Note that the ILDs were generated by attenuating the target signal in one ear. This prevented the listener from gaining any advantage from an increased SNR ratio in one of the two ears.

Signals generated by a Creative Labs Soundblaster Audigy sound card were presented to listeners seated in front of a computer screen in a quiet room. Three headphone conditions were tested: 1) the Temco bone-conduction headset with unoccluded ears; 2) the bone-conduction headset with occluded ears (EAR Classic foam earplugs); and 3) a standard air-conduction headset (Sennheiser HD-520). The transducers of the bone-conduction headset rested on the mastoids behind the ear.

Procedure

The independent variables were the type of headphones worn by the listeners and the interaural differences (ILD or ITD) introduced in the target phrase. Each listener completed blocks of trials with ILD manipulations and ITD manipulations in counterbalanced order. At the beginning of each block of trials, listeners were randomly instructed to wear one of three headphone configurations. Within each block, the interaural difference was presented so that the target shifted either to the right or left relative to the location of the maskers. Each value of ITD or ILD was randomly presented 30 times, for a total of 180 trials per block. Each participant ran 24 blocks of trials (3 headphones x 2 target locations x 2 repetitions x 2 interaural differences).

Each listener responded to the color and number combination spoken by the target talker (designated by the call sign "Baron"), which was displayed on the computer screen as a four-by-eight matrix of colors and numbers. Feedback was provided on each trial and percentage correct feedback was provided at the end of each block of trials.

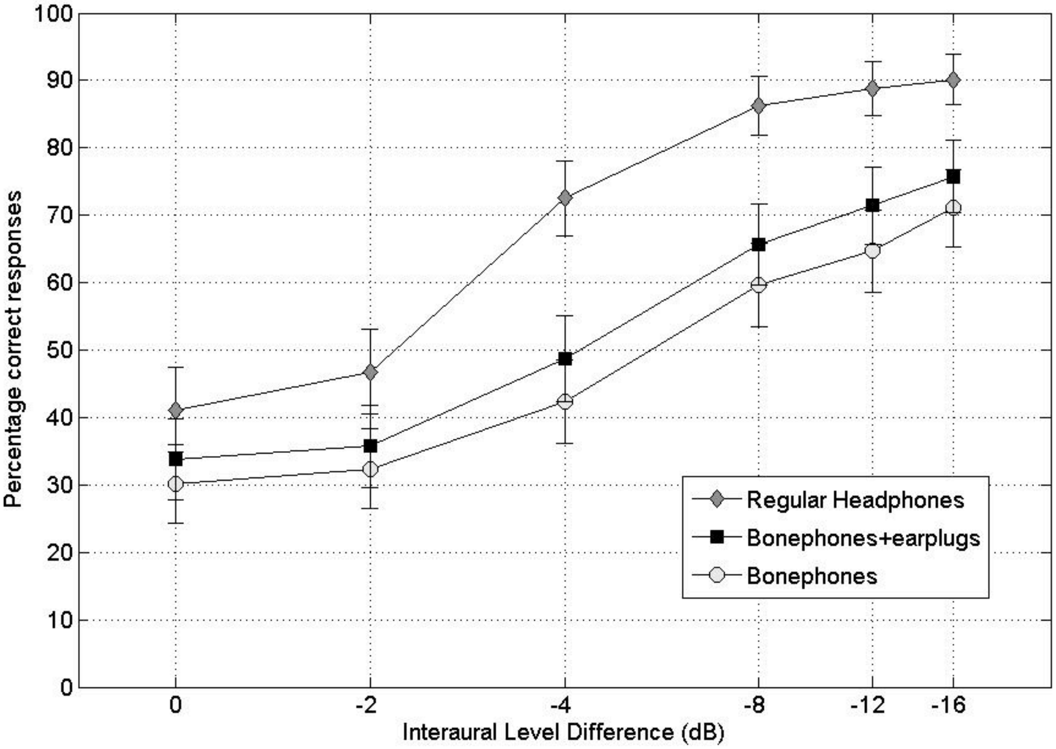

Figure 1. Percentage of correct color and *number identifications as a function of ILD for each of the three headphone conditions tested. The error bars represent the 95% confidence intervals around each data point.*

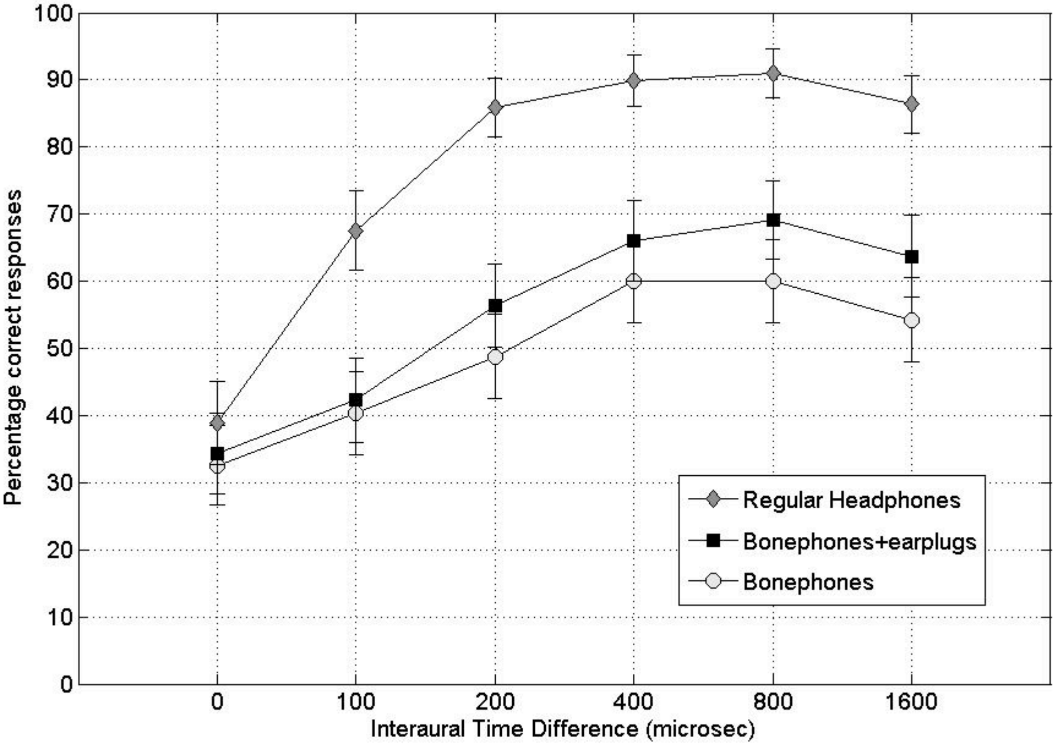

Figure 2. Percentage of correct color and *number identifications as a function of ITD for each of the three conditions tested. The error bars represent the 95% confidence intervals around each data point.*

RESULTS

Inspection of Figure 1 reveals that performance on the task systematically varied with the interaural level difference presented via both standard headphones and bone-conduction headsets. Overall, performance on the task was considerably worse for bonephones than it was for standard headphones. Performance increased more gradually with the bonephones than it did with standard headphones, and performance at the largest ILD value tested (16 dB) was approximately equivalent to the level achieved with an ILD of only 4 dB with the standard air-conducted headset. This may suggest an upper bound on the amount of spatial separation that can be achieved with a bone-conduction headset. Plugging the ear canals with foam earplugs makes no difference in performance with the bonephones.

Inspection of Figure 2 reveals that performance on the task also improved with increasing interaural time delay for all three types of headphones. As in the ILD condition, performance was consistently better with standard headphones than with the bonephones in the ITD condition. In fact, performance with the bonephones never exceeded the performance level achieved using headphones and the smallest ITD. Again, plugging the ears made little or no difference in performance with the bonephones.

DISCUSSION

Despite performance with normal headphones that was consistently superior to performance with bonephones, the results demonstrate that reliable segregation can be produced with bonephones. However, the degradation relative to standard headphones suggests that it may be difficult to produce large enough interaural differences to simulate sound sources at extreme lateral locations. The results also suggest that, for the bonephones, ILDs are more effective at producing this spatial separation than ITDs.

This research confirmed and extended the implications of Kaga et al. (2001), showing that binaural separation was possible with a non-clinical bone-conduction headset and an open ear canal. Furthermore, by showing an increase in performance on a task that is known to improve with spatial separation, it confirmed subjective reports with a more objective assessment of spatial hearing. Although the differences in bone conduction headsets, task, and measurement make direct comparisons to the ITD and ILD thresholds found by Kaga et al. difficult, the essential finding is the implication that sensitivity to ITDs and ILDs persists with a task and device that are designed for application in auditory displays. In this way, it has been demonstrated that bonephones are a promising alternative to headphones that may be suitable for displays that require spatial separation, such as multitalker communication displays (Brungart & Simpson, 2002). In other applications where spatialized audio is important, such as in navigational aides for the blind, bonephones have also shown performance that is acceptable, but not at the level of headphones (Walker & Lindsay, in press). In some cases, the effectiveness of bone conduction headsets may be improved with signal processing, since there are factors other than ITDs and ILDs that can affect performance, such as the frequency response of the devices. Other recent studies (i.e., Walker & Stanley, 2005) have mapped out an equalization curve for the bonephones, under a variety of listening conditions, and these curves could be incorporated into the signal to further improve performance.

There are certainly other issues that remain to be studied. For example, if bonephones were to be implemented in a spatial auditory task where ambient sounds also need to be processed (e.g., multitalker radio communication with simultaneous monitoring of spoken commands heard directly, not over the radio), it would help to understand better how the perception of bone-conducted and ambient sounds interact. Because air and bone conduction share the same mechanisms after the cochlea, the cognitive and attentional aspects of processing the multiple sound sources should be no different than what decades of previous research on auditory attention has investigated (Brungart, 2001; Cherry, 1953). However, the difference in pathways that air and bone-conducted sound travel to the cochlea suggest that perceptual aspects of this interaction may be quite unique. Future research would benefit from considering these more advanced perceptual interactions with more complex stimuli and tasks.

Now that an ability to use some spatial cues has been demonstrated, a next step is to make direct measurements of these thresholds and other binaural aspects of hearing. Since lateralization is a precursor to spatialization, another step to be taken is the investigation of making bone-conducted audio sound like it is externalized.

REFERENCES

Blauert, J. (1983). *Spatial Hearing: The Psychophysics Of Human Sound Localization* (J. S. Allen, Trans.). Cambridge, Ma: The Mit Press.

Bolia, R. S., Nelson, W. T., Ericson, M. A., & Simpson, B. D. (2000). A Speech Corpus For Multitalker Communications Research. *Journal Of The Acoustical Society Of America*, 107(2).

Brungart, D. S. (2001). Informational And Energetic Masking Effects In The Perception Of Two Simultaneous Talkers. *Journal Of The Acoustical Society Of America*, 109, 1101-1109.

Brungart, D. S., & Simpson, B. D. (2002). The Effects Of Spatial Separation In Distance On The Informational And Energetic Masking Of A Nearby Speech Signal. *Journal Of The Acoustical Society Of America*, 112(2), 664-676.

Cherry, E. C. (1953). Some Experiments On The Recognition Of Speech, With One And Two Ears. *Journal Of The Acoustical Society Of America*, 25, 975-979.

Kaga, K., Setou, M., & Nakmura, M. (2001). Bone-Conducted Sound Lateralization Of Interaural Time Difference And Interaural Intensity Difference In Children And A Young Adult With Bilateral Microtia And Atresia Of The Ears. *Acta Oto-Laryngologica*, 121(2), 274-277.

Katz, J. (2002). Clinical Masking. In J. Katz (Ed.), *Handbook Of Clinical Audiology* (5 Ed., Pp. 124-141). Philadelphia: Williams & Wilkins.

Studebaker, G. A. (1962a). On Masking In Bone-Conduction Testing. *Journal Of Speech & Hearing Research*, 5(3), 215-227.

Studebaker, G. A. (1962b). Placement Of Vibrator In Bone-Conduction Testing. *Journal Of Speech & Hearing Research*, 5(4), 321-331.

Tonndorff, J. (1972). Bone Conduction. In J. V. T. (Ed.) (Ed.), *Foundations Of Modern Auditory Theory* (Pp. 195-237). New York: Academic Press.

Von Bekesy, G. (1960). *Experiments In Hearing* (E. G. Wever, Trans.). New York: Mcgraw-Hill Book Company, Inc.

Walker, B. N., & Lindsay, J. (In Press). Navigation Performance With A Virtual Auditory Display: Effects Of Beacon Sound, Capture Radius, And Practice. *Human Factors*.

Walker, B. N., & Stanley, R. M. (2005). Thresholds Of Audibility For Bone-Conduction Headsets. *Proceedings Of The International Community For Auditory Display*.