informa
healthcare

Original Article

International Journal of Audiology 2006; 45:595–599

Justin A. MacDonald
Paula P. Henry
Tomasz R. Letowski

U.S. Army Research Laboratory,
Human Research and Engineering
Directorate, Aberdeen Proving
Ground, Maryland, USA

Spatial audio through a bone conduction interface

Audición espacial a través de una interfase de conducción ósea

Key Words
Bone conduction
Spatial audio
Head-related transfer function

Abstract

Headphones are the standard presentation device for radio communication in the military. Although bone conduction devices possess several advantages over headphones for some military applications, they are generally considered inappropriate for inclusion in a multi-channel system. The current study tested the feasibility of a multi-channel bone conduction system by measuring the localizability of spatialized auditory stimuli presented through a pair of bone conduction vibrators. Listeners localized a Gaussian noise stimulus spatialized with individualized head-related transfer functions (HRTFs). The sounds were presented from eight virtual locations on the horizontal plane (0, ±45, ±90, ±135, and 180°) through either stereo headphones or a stereo bone conduction system. Localization performance was found to be nearly identical for both audio systems, indicating that bone conduction systems can be effectively used for displaying spatial information.

Sumario

Los auriculares son el instrumento estándar para las comunicaciones por radio en el ejército. A pesar de que los instrumentos de conducción ósea tienen algunas ventajas sobre los auriculares en algunas aplicaciones militares, se consideran generalmente inapropiados para su inclusión en un sistema multicanal. El presente estudio examinó la factibilidad de un sistema multicanal de conducción ósea midiendo la posibilidad de localización de estímulos auditivos distribuidos espacialmente y presentados con un par de vibradores de conducción ósea. Los sujetos localizaron un estímulo de ruido Gaussiano con distribución espacial y funciones individualizadas de transferencia relacionadas con la cabeza (HRTF). Los sonidos se presentaron desde 8 puntos virtuales en un plano horizontal (0, ±45, ±90, ±135, and 180°) a través tanto de auriculares estereofónicos como de un sistema estereofónico de conducción ósea. Se encontró que el rendimiento para la localización fue casi idéntico en los dos sistemas estereofónicos, lo que indica que los de conducción ósea pueden usarse de manera efectiva para presentar información espacial.

The human auditory system is able to perceive sounds received through two pathways: air conduction and bone conduction. Hearing through air conduction involves perception of sounds that arrive to the ears through the ear canals. Hearing through bone conduction involves perception of sounds that arrive to the ears through vibration of the bones of the skull. Communication systems used in the military have traditionally utilized the air conduction pathway, but the increased availability of high-quality bone conduction devices allows for bone conduction communication to be considered as a viable alternative.

Bone conduction offers many potential advantages over air conduction as a means of both transmission and reception of communication in military environments. Headphone-based communication interfaces are incompatible with in-the-ear hearing protection devices. Bone conduction interfaces can circumvent this difficulty by providing effective communication without interfering with hearing protection devices (Henry & Mermagen, 2004; Langford et al, 1989). In quiet environments, the soldier could receive radio communications through bone conduction without obscuring the ears, thereby maintaining full awareness of the surrounding acoustic environment. Alternatively, the use of headphones or in-the-ear devices in quiet environments could reduce the soldier's ability to hear potentially important auditory information. Bone conduction transducers could also be useful to soldiers and security forces in stealth situations: auditory signals can be perceived by the user while remaining inaudible to others, including enemy forces.

However, one of the main disadvantages of current bone conduction communication systems is that they are restricted to single channel operation. Stereo communication interfaces can allow for improved communications through the use of spatialized speech sources (Yost et al, 1996; Drullman & Bronkhorst, 2000; Ericson & McKinley, 1997). In general, bone conduction is thought to be inappropriate for multi-channel communication systems. Presentation of a stereo signal over a pair of bone conduction transducers mounted bilaterally on the skull would result in the output from each transducer arriving at both cochleae. This cross-channel interference would presumably reduce or eliminate any stereo percept.

Although vibrations from the transducers are likely to interfere with one another, the transcranial attenuation and delay may be sufficient to allow the listener to segregate the overlapping auditory inputs into separate percepts. Vibrations arriving from the contralateral transducer are attenuated relative to the vibrations from the ipsilateral transducer. This attenuation can be as large as 15–20 dB at higher frequencies. For example, Kirikae (1959), Nolan & Lyon (1981), and Stenfelt & Goode (2005) reported a transcranial attenuation of 10–15 dB at

ISSN 1499-2027 print/ISSN 1708-8186 online
DOI: 10.1080/14992020600876519

Accepted:
June 21, 2006

Justin A. MacDonald
U.S. Army Research Laboratory, Human Research and Engineering Directorate, AMSRD-ARL-HR-SD, Aberdeen Proving Ground, Maryland 21005, USA.
E-mail: jmacdonald@arl.army.mil

2000 Hz. In addition to intensity differences, the time of arrival of the contralateral signal will be delayed relative to the ipsilateral signal, although the amount of delay is open to debate. Estimates of the velocity of sound traveling through the skull vary widely depending upon the estimation method used. Franke (1956) applied a signal to the human forehead and measured the delay between a pair of points. His results were dependent on frequency: the velocity of lower frequency sounds was estimated at 80 m/s, while higher frequencies traveled at up to 300 m/s. Placing transducers on each of the mastoids, Zwislocki (1953) and Tonndorf & Jahn (1981) used psychophysical methods to estimate the transcranial velocity at 260 and 330 m/s, respectively. The velocities that have been reported through a living human skull are comparable to the speed of sound though air and are much lower than the speed of sound through a dry skull that was reported to exceed 2000 m/s (Tonndorf & Jahn, 1981). These data indicate that if the source of vibration is located off the median plane, the time of arrival of the contralateral signal will be noticeably delayed relative to that of the ipsilateral signal, possibly further reducing the effects of cross-channel interference.

The perceptual significance of the transcranial attenuation and delay is supported by our own experience with bone conduction interfaces, indicating that under some conditions listeners can segregate a stereo signal into two separate percepts, despite cross-channel interference. There is also some evidence in the literature suggesting that listeners might benefit from using stereo bone conduction systems for spatial orientation. Stenfelt (2005) suggested that presentation of a stereo signal through a pair of bilaterally implanted bone conduction hearing aids should allow for the extraction of some localization cues. These observations led us to explore the possibility that two or more bone conduction transducers might be used to produce a well-defined stereo percept. Placing a bone conduction transducer near each of the ears to maximize the time, intensity, and spectral differences between the signals arriving to the near and far ears could maximize any stereo effect.

In addition, processing the sound through head-related transfer functions (HRTFs) for airborne sounds could enhance the spatial perception of bone conducted sounds. Spatial audio is usually implemented by filtering sounds with HRTFs so that they are perceived to originate in a surrounding three-dimensional space, when in actuality they are presented over stereo headphones. This approach is especially useful for communication systems, as the intelligibility of multiple speech streams is improved with spatialized presentation (Ricard & Meirs, 1994; Yost et al, 1996; MacDonald et al, 2002; Abouchacra et al, 2001; Shilling et al, 2001; Vause et al, 2001).

The purpose of the present study was to assess the feasibility of presenting spatialized auditory stimuli through a stereo bone conduction system. An experiment was conducted to compare localization accuracy when spatialized sounds were presented over headphones versus through a pair of bone conduction transducers wired for stereo sound. Participants were asked to localize spatialized stimuli that originated from one of eight virtual locations around the head. Assuming that the perceptual isolation of the ears is achievable, the participant should be able to make use of the interaural level difference (ILD) and interaural time delay (ITD) cues to localize the spatialized sound. Therefore, similar localization performance with the two transducer types would suggest that users of the bone conduction apparatus are able to segregate the channels to make effective use of the spatial cues. Otherwise, relatively poor localization performance when using the bone conduction apparatus would suggest that cross-channel interference obscured the location cues present in the spatialized sounds.

Method

Participants

Four participants (two male and two female) between the ages of 31 and 40 completed the study. Each participant completed two 60-minute experimental sessions. Each participant had normal hearing thresholds defined as pure-tone air conduction thresholds better than or equal to 20 dB HL at octave audiometric frequencies from 250 to 8000 Hz (ANSI, 2004). The difference between thresholds in each ear was no greater than 10 dB at any test frequency to ensure hearing symmetry.

Apparatus

The experiment was conducted using a computer (IBM compatible) outfitted with a Chaintech AV-710 eight-channel sound card. The sound card was connected through a headphone amplifier to both a pair of AKG K240DF circumaural headphones, and a Temco HG-17 bone conduction system with two vibrators modified to provide stereo output. The placement of the transducers on the Temco HG-17 headset was at the location of the condyle, nearly in front of the wearer's ears. The listening station consisted of a table, chair, monitor, computer, and mouse. All responses were made using the mouse and recorded through proprietary computer software designed for this study.

Stimuli

The stimulus was a train of eight 250-ms Gaussian noise bursts separated by 300-ms intervals. All stimuli presented through the headphone and bone conduction systems were first filtered by the inverse of each transducer's frequency response. The stimuli were equalized to a constant dB SPL (for the headphones) and force level (for the bone conduction transducers) across frequency. The response of the Temco transducer was obtained though a Brüel & Kjær Artificial Mastoid (Type 4930). The response of the AKG headphone was obtained using a Brüel & Kjær Artificial Ear (Type 4153) in conjunction with an ACO (Model 7012) microphone. The frequency response of each transducer was estimated using the MLSSA system (DRA Laboratories). This procedure resulted in the frequency responses illustrated in Figure 1. Inverse filters were created from the measured responses (see MacDonald & Tran, in press) and were applied to the Gaussian noise stimuli. This resulted in a theoretically flat response for both the Temco and AKG transducers. However, the frequency range of the Temco HG-17 bone conduction transducers is quite narrow relative to that of the headphones (see Figure 1). Therefore, all stimuli were bandpass-filtered (300–5000 Hz) in an attempt to eliminate the frequency range of the transducers as a potential confounding variable in the experiment.

The stimuli were then filtered through each participant's individually measured head-related transfer function (HRTF) measured at 45° intervals on the horizontal plane. HRTFs were obtained using the Tucker-Davis Technologies (TDT) RoboArm

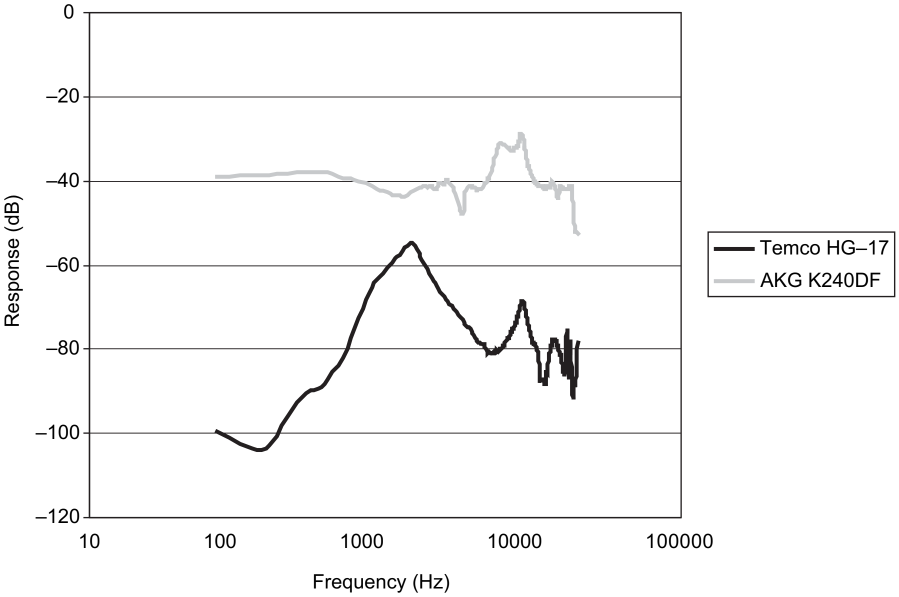

Figure 1. Frequency responses of the Temco HG-17 bone vibrator and the AKG K240DF headphones. The responses are reported as the output level of the transducer relative to a constant input voltage.

360 HRTF measurement system. This system consists of a GF0876 loudspeaker (CUI Inc.) attached to the end of a robotic arm. The distance from the loudspeaker to the center of the head was 1 m. The input to each of the ears was recorded using a pair of EM-125 miniature electret microphones (Primo Microphones Inc.) mounted in foam inserts in the ear canal of the participant. Control of the robotic arm as well as HRTF estimation was conducted in MATLAB.

Procedure

After completion of a hearing test the participant was seated at the listening station and familiarized with the experimental setup and response interface. The participant wore both sets of transducers (headphones and bone vibrators) on the head simultaneously. Proper care was taken that there was no mechanical contact between the sound transmission systems. At the beginning of each session one of the experiment stimuli (located at 0°) was presented alternately though the bone conduction and headphone apparati. The intensity of the stimulus presented though headphones was set to 75 dB(A). The participant was asked to manipulate the amplification of the bone conduction system until the loudness was equivalent to that produced by the headphones.

For the remainder of the experiment session, Gaussian noise was presented through four speakers located at ±45° and ±135° relative to the participant. Each speaker was 2.5 m from the center of the participant's head, resulting in a 55 dB(A) background noise measured at the location of the participant. The noise was included to mask any air-conducted output of the bone conduction transducers. The response interface consisted of a mouse and a computer screen with a large blue circle outlined in yellow. The center of the blue circle represented the participant's position and the vertical axis of the screen (up-down) represented the anteroposterior (front-back) direction. The participant began each trial by clicking within the blue area in the center of the interface. A sound was played from one of eight virtual locations on the horizontal plane (0, ±45, ±90, ±135, and 180°). The task of the participant was to indicate the perceived location of the sound source by clicking on the yellow outline. No feedback on performance was provided. A total of 240 trials were completed in each of two experimental sessions (eight virtual locations × 30 replications), resulting in a total of 480 trials per participant. Stimuli were presented through a single transducer type for each session, either bone conduction or headphones. Participants were not told which transducer type was in use. The order of transducer type was counterbalanced across participants, and the order of all stimuli within each session was completely randomized. The participant was encouraged to take breaks at any time, and all aspects of the experiment were self-paced.

Results

Because the pattern of responses was very similar across participants, experiment results are presented using pooled response data. The responses for the bone conduction and headphone conditions are shown in Figure 2. The patterns of responses in the bone conduction and headphone conditions are nearly indistinguishable from one another, indicating that localization performance was very similar for both types of transducers.

Overall measures of localization performance are detailed in Table 1. Localization error was defined as the angular distance between the perceived and actual source locations in degrees. Reversals occurred when the perceived and actual locations of the sound source were on opposite sides of the interaural axis. Completely chance performance in this task would lead to a mean error of 90 degrees and a reversal rate of 50%. Assuming that the listener correctly assigns the sound to either the left or right hemisphere but otherwise responds randomly, 63 degrees of

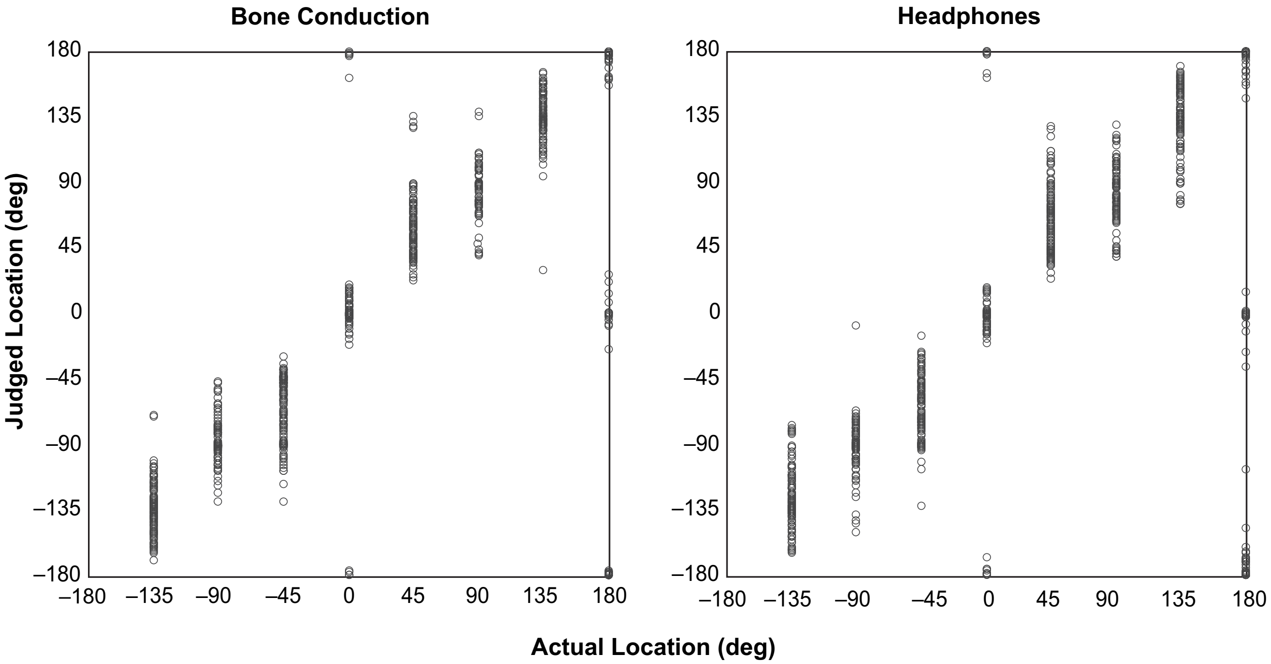

Figure 2. Scatter plots of actual source location versus judged source location for the bone conduction and headphone conditions. Actual locations ranged from −135° to 180° in 45° increments.

absolute error can be expected. If the listener lateralizes rather than localizes the sound (responds at −90 degrees for all sounds localized in the left hemisphere, and at +90 degrees for all sounds located in the right hemisphere), the expected absolute error is 45 degrees. The observed error and reversal rates indicate much better than chance performance: the overall absolute error rate was 17° for the bone conduction condition and 22° for the headphone condition. Reversals were relatively rare at 7% and 12% for the bone conduction and headphone conditions, respectively.

Discussion

Presenting spatialized stimuli through a stereo bone conduction interface could potentially lead to confusing and unnatural sounds that do not generate the desired percept. The interaural time delay (ITD) and interaural level difference (ILD) spatial cues artificially inserted during spatialization may be partially or completely obscured by cross-channel interference. However, the bone conduction apparatus used in the present study allowed for very similar localization performance to that afforded by headphones, indicating that a stereo bone conduction apparatus can provide an effective spatial audio interface.

The wearing of the circumaural headphones during the bone conduction condition introduced some degree of occlusion (Dirks & Swindeman, 1967; Fagelson & Martin, 1998; Watson, 1938; Watson & Gales, 1943). The effect of occlusion on sound localization remains unknown, however. Occlusion results in amplification of the low frequency components of a signal (mainly below 500 Hz). This is likely to result in an increased perception of the ITD cues present in the spatialized stimuli. At the same time, amplifying the low frequencies should lead to increased masking of the high-frequency components, thereby obscuring the ILD cues associated with those frequencies.

Table 1. Summary measures of localization performance by actual location and transducer type.

	Mean absolute error		*% reversals*	
Location (degrees)	*Bone conduction*	*Headphones*	*Bone conduction*	*Headphones*
−135	13.1	15.7	1.7	5.8
−90	9.4	9.4	–	–
−45	21.8	23.0	12.5	10.0
0	21.2	32.8	10	16.7
45	18.7	21.4	3.3	12.5
90	10.5	15.0	–	–
135	11.7	16.8	0.8	7.5
180	29.6	39.1	15.8	20.0
Overall	17.0	21.7	7.4	12.1

Whatever the result, occlusion should distort the spatialized stimulus, and therefore reduce localization performance. A further empirical study specifically addressing the issues of occlusion should serve to answer this question.

In conclusion, our results demonstrate the feasibility of a spatial audio interface implemented with two bone conduction transducers wired for stereo sound. Apparently, all four listeners were able to segregate the channels based solely upon the relatively small attenuation and delay introduced as the vibration travels through the cranium toward the opposite ear. Indeed, our own observations and the participants' comments confirmed that only a small qualitative difference existed between the spatialized stimuli presented through the different apparati. Based on the results of this study, bone conduction systems seem to be effective interfaces for inclusion in communication and warning systems requiring spatial representation of the signals. The accurate localization of sound sources observed with the bone conduction interface is all the more surprising given the reduced bandwidth of the stimulus. The future development of bone conduction transducers with an expanded frequency range can only enhance the utility of the approach. The merging of bone conduction communication with enhanced spatialized sound should also lead to a greater understanding of how a listener assembles disparate and conflicting spatial cues into a coherent percept of sound source location.

References

Abouchacra, K., Breitenbach, J., Mermagen, T. & Letowski, T. 2001. Binaural helmet: Speech recognition performance using spatial sound. *Human Factors*, 43, 584–594.

American National Standards Institute. 1996. *Specification for Audiometers*. New York ANSI S3.6 (R 2004).

Dirks, D. & Swindeman, J. 1967. The variability of occluded and unoccluded bone-conduction thresholds. *J Sp Hear Res*, 10, 232–249.

Drullman, R. & Bronkhorst, A.W. 2000. Multichannel speech intelligibility and talker recognition using monaural, binaural, and three-dimensional auditory presentation. *J Acoust Soc Am*, 107, 2224–2235.

Ericson, M.A. & McKinley, R.L. 1997. The intelligibility of multiple talkers separated spatialy in noise. In R.H. Gilkey & T.R. Anderson (eds.) *Binaural and spatial hearing in real and virtual environments*. Mahwah, NJ: Erlbaum, pp. 701–724.

Fagelson, M. & Martin, F. 1998. The occlusion effect and ear canal sound pressure level. *Am J Audio*, 7, 50–54.

Franke, E.K. 1956. Response of the human skull to mechanical vibrations. *J Acoust Soc Am*, 28, 1277–1284.

Henry P. & Mermagen T. 2004. Bone conduction communication in a military vehicle. Proceedings of the NATO Vehicle Habitability Conference, Prague (Czech Republic): October 4–6. NATO, pp. 14–1 to 14–10.

Kirikae, I. 1959. An experimental study of the fundamental mechanism of bone conduction. *Acta Otolaryngol*, Supp. 145, 1–110.

Langford T., Mozo B. & Patterson J. 1989. Evaluation of speech intelligibility through a bone conduction stimulator. USAARL Technical Report #89–13.

MacDonald, J.A., Balakrishnan, J.D., Orosz, M.D. & Karplus, W.J. 2002. Intelligibility of speech in a virtual 3-D environment. *Hum Factors*, 44, 272–286.

MacDonald J.A. & Tran P.K. In press. Loud speaker equalization for auditory research. *Behavioral Research Methods, Instruments, & Computers*.

Nolan, M. & Lyon, D.J. 1981. Transcranial attenuation in bone conduction audiometry. *J Laryngo Otol*, 95, 597–608.

Ricard, G.L. & Meirs, S.L. 1994. Intelligibility and localization of speech from virtual directions. *Hum Factors*, 36, 120–128.

Shilling R., Storms R. & Letowski T. 2001. Applications of spatial audio displays for use within attack rotary wing aircraft. Proceedings on CD of the 22nd Army Science Conference, [Baltimore (MD): December 11–13, 2000], Baltimore (MD): ASA, pp. 823–827.

Stenfelt, S. 2005. Bilateral fitting of BAHAs and BAHA fitted in unilateral deaf persons: Acoustical aspects. *Int J Audiol*, 44, 178–189.

Stenfelt, S. & Goode, R. 2005. Transmission properties of bone conducted sound: Measurements in cadaver heads. *J Acoust Soc Am*, 118, 2373–2391.

Tonndorf, J. & Jahn, A.F. 1981. Velocity of propagation of bone-conducted sound in a human head. *J Acoust Soc Am*, 70, 1294–1297.

Vause N., Abouchacra K., Letowski T. & Resta E. 2001. Spatial (3-D) audio displays improve speech communication in the command and control vehicle (C^2V). Proceedings on CD of the 22nd Army Science Conference, [Baltimore (MD): December 11–13, 2000]. Baltimore (MD): OASA, pp. 834–841.

Watson, N. 1938. Limits of audition for bone conduction. *J Acoust Soc Am*, 9, 294–300.

Watson, N. & Gales, R. 1943. Bone-conduction threshold measurements: Effects of occlusion, enclosures, and masking devices. *J Acoust Soc Am*, 14, 207–215.

Yost, W.A., Dye, R.H. & Sheft, S. 1996. A simulated 'cocktail party' with up to three sound sources. *Percept Psychophys*, 58, 1026–36.

Zwislocki, J. 1953. Acoustic attenuation between the ears. *J Acoust Soc Am*, 25, 752–759.