

BINAURAL SPATIALISATION OVER A BONE CONDUCTION HEADSET: ELEVATION PERCEPTION

Amit Barde, Gun Lee

HIT Lab NZ,
University of Canterbury,
Private Bag 4800,
Christchurch 8140, New Zealand.
amit.barde@pg.canterbury.ac.nz,
gun.lee@canterbury.ac.nz

William S. Helton

Department of Psychology,
University of Canterbury,
Private Bag 4800,
Christchurch 8140, New Zealand.
deak.helton@canterbury.ac.nz

Mark Billinghamurst

School of ITMS,
University of South Australia,
Mawson Lakes, SA 5095, Australia.
mark.billinghurst@unisa.edu.au

ABSTRACT

Preliminary results from an on-going experiment exploring the localisation accuracy of a binaurally processed source displayed via a bone conduction headset are described. These results appear to point to decreased localisation accuracy in the horizontal plane when the vertical component is introduced. There also appears to be a significant compression in the area directly in front of the observer $\pm 15^\circ$ in elevation from 0° . This suggests that participants tended to localise stimuli presented at elevations greater than and less than $\pm 30^\circ$ within a 30° 'window' extending 15° vertically either above or below the horizontal plane defined by the 0° azimuth. The results gathered until now suggest that binaural spatialisation over a bone conduction headset can also reproduce the perception of an elevated source to an acceptable degree of accuracy.

1. INTRODUCTION

Current forms of mobile spatial auditory displays almost all rely on delivery of sound through headphones or earphones. These mediums of sound delivery isolate the ears from the ambient acoustic environment [1]. Besides this, most spatial auditory displays are only capable of providing very basic information such as the arrival of a message, email etc. This is achieved by the use of basic tones or sounds. Inevitably, the user of such a display is forced to engage in a visual interaction with the mobile device in order to retrieve information he/she has been conveyed has arrived [2].

With the widespread availability of information 'on-the-go', the need for spatial auditory displays has become greater. It is now necessary for mobile auditory displays to provide more than just alerts for incoming messages and emails. The mobile auditory display of the future must be seamlessly integrated into a wearable computing system capable of delivering useful and actionable information. For example, a wearable auditory display must not only inform a person of incoming message, but also be able to provide navigation information via a binaurally spatialised auditory beacon. Such

functionality incorporated into an auditory display will be able to reduce the cognitive load that current visually demanding mobile displays exert on their users [2] [3].

In addition to the problem of too much information attempting to be displayed on ever shrinking screens of wearable interfaces, there also exist safety concerns. Mobile devices that constantly attempt to engage the visual faculty may end up being a distraction and divert the user's attention away from the primary task. If that task is an attention critical one such as driving or navigation in hazardous environments, the risk posed to the user is great. The lack of attention to the primary task could prove to be fatal in either of these situations. While spatial auditory displays have been developed in response to the challenges posed by visually demanding displays, they suffer from the issue of sensory deprivation. Most spatial auditory displays involve the use of headphones or earphones to deliver auditory information to the user. The use of these mediums to deliver the sound isolates the user from the ambient acoustic environment by covering the ears or blocking the ear canals [1]. This isolation from the acoustic environment is undesirable, since a lot of our information about the environment outside of our visual field is gathered via the auditory faculty. There is a need to develop auditory displays that allow us to retain our natural acoustic perception of the surroundings while simultaneously being able to provide synthesized auditory cues for information presentation and retrieval. The bone conduction headset (BCH) makes for an ideal candidate for such an auditory display. Its relatively small size and the fact that it does not obstruct the pinnae or the ear canals are design aspects that work in its favour.

We are currently carrying out an experiment as part of larger study to explore the feasibility of the BCH as an auditory display device. In the following sections we will cover the research that has been carried out in to the use of a BCH as auditory display device, the design and execution of our study and preliminary results from the study and what they appear to suggest about auditory perception over a BCH.

2. RELATED RESEARCH

While bone conduction technology has been around for a long time, it is only now gaining some ground, both in the research and commercial fields. There is relatively little known about auditory perception of binaurally spatialised sound sources over a BCH. With a few notable exceptions [1] [4] [5], there are few studies that have closely evaluated binaural spatialisation and localisation performance over a BCH. Our previous study has shown that binaural spatialisation over a BCH induces an acceptable level of externalisation and that localisation performance is within parameters acceptable for a such an auditory display [6].

Up until now research related to the use of the BCH has primarily been restricted to its use as a navigation aid for the visually challenged [1] [7]. Few researchers have explored the use of a BCH as part of an auditory display device for AR or VR environments [8] [9]. A large part of the existing literature also concentrates on the use of individualised HRTFs with BCH based reproduction [4] [10]. McDonald et al's results [4] tend to suggest that the use of individualised HRTFs for BCH based reproduction is able to reproduce spatial resolution that is comparable to or better than that achieved over headphones. Studies with non-individualised HRTFs also have been shown to achieve good results [1] [5]. All this research suggests that the BCH has a great potential for being used as a spatial auditory display device as part of a wearable interface incorporating auditory and visual cues.

The study in to the localisation performance achievable for binaurally presented sources over a BCH is an attempt to explore the limits of the device. Knowledge of the operational limits of the BCH will help with the design and development of an auditory display device that is capable of providing useful information to its user.

3. METHOD

3.1. Apparatus

An ecologically valid approach has been adopted for this study. We have chosen to use inexpensive 'off the shelf' hardware and software components that are representative of those used at the developers' and consumers' ends. The experiment was developed in and run using the Unity3D engine [11]. Binaural spatialisation was achieved using the 3Deception [12] plugin for Unity. The binaural engine was selected after an exhaustive pilot study comparing three popular binaural engines available in the market at the time the experiment was conceived. For the BCH we have used the Aftershokz Sport3 [13]. This combination of hardware and software is a step away from the traditional form virtual auditory display studies that incorporate the use of individualised HRTFs [4] [10] or non-individualised HRTFs (HRTF databases) [1] [5] [14].

The experiment was developed in the Unity3D environment. Sources were created using the binaural engine and the 'global listener' of the engine was slaved to the main camera in Unity to render a first person perspective to the auditory events in the scene (see figure 1). Sound was delivered to the BCH via a Zoom UAC-2 audio interface. A Dell Inspiron Laptop (Windows 8, 2.2 GHz Intel Core i7) was

used to run the study. The experiment was carried out in a sound proof booth that conformed to the HTML 2045, ISO 8253 and ISO BS EN6189 standards.

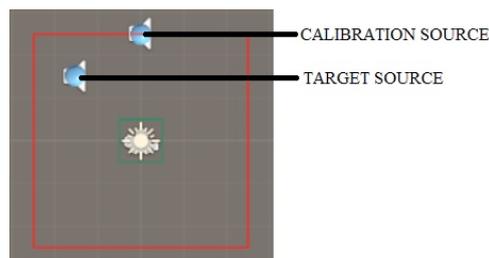


Figure 1: Experimental setup in Unity3D

3.2. Stimuli and Calibration

A single 1 second burst of pink noise (25 ms onset and offset time) was used. Pink noise was chosen since it has been shown that broadband stimulus is easier to localise than one with a restricted spectral range [1] [15] [16] [17]. Calibration of the headset was achieved by asking the participants to adjust the level on the BCH until they felt that it matched the level set on a loudspeaker placed 1m away. During the process participants were asked to look directly at the loudspeaker (PhonicEar AT578-S) and align their heads with it in a manner such that their ears were approximately at the same level as the loudspeaker (see figures 2 and 3). The level of the calibration source played over the loudspeaker was set to approximately 70 dBA measured at 1m from the speaker. The duration and level of the stimulus was chosen to represent that used by previous researchers [1] [15] [18].

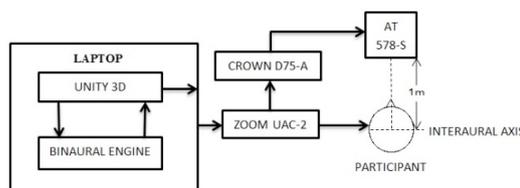


Figure 2: Block diagram of experimental setup

3.3. Participants

We've had 6 participants (1 female, 5 males) between the ages of 19 and 29 (Mean: 23.3, SD: 3.3) take part in the experiment until now. All participants reported normal hearing. No audiometric screening was performed to check for normal hearing function.

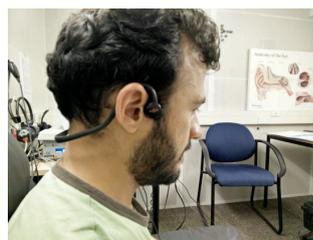


Figure 3: Participant wearing the bone conduction headset

3.4. Procedure

Participants were presented with a single one second burst of pink noise at elevation ranging from -45° to +45° in steps of



15°. This resulted in a total of seven elevation angles. The vertical range within which the stimulus was spatialised was limited since pilot testing showed severe degradation in azimuth perception for vertical angles greater than ±45°. This is likely due to the fact that azimuth, like longitudes, are compressed as they approach the poles of the imaginary sphere surrounding the participants' heads within which the source is spatialised.

In the horizontal plane (azimuth), locations varied from 0° - 90° i.e. single quadrant size. Step sizes for the azimuth was also 15° resulting in a total of 7 azimuth locations per quadrant. Each elevation was reproduced twice for every azimuth giving us a total of 14 trials per elevation. 98 trials were conducted for every quadrant (14 trails/elevation x 7 azimuth) giving us a total of 392 trials per participants encompassing a complete 360° range in the azimuthal plane.

Trials were divided in to three blocks consisting of a trial block not exceeding 5 minutes and 2 main blocks of trials separated by a 10 minute break. Participants were told to use the two response charts provided to localise the stimulus (see figure 4). There was no compulsion to look or point at the chart to give the response. Positions were to be called out using the signed angles protocol displayed on the response charts. This method of judgement estimation in localisation studies has been validated by previous studies [19] [20] [14] [15]. Participants were asked to face forward during the experiment, and try and keep their head in line with the loudspeaker used for calibration (see figure 3). This wasn't strictly enforced though. No chin brace was employed either to keep the head in a fixed position.

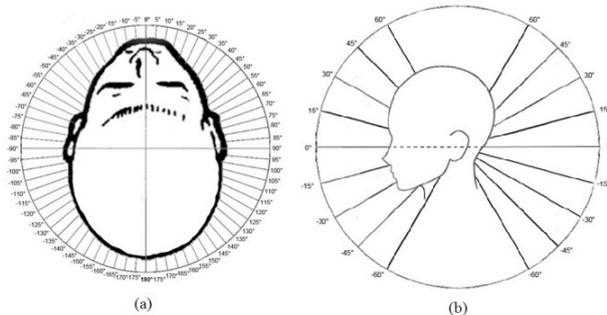


Figure 4: Response charts: azimuth (a) and elevation (b)

Participants were also asked to rate externalization at the end of the experiment. These ratings were based on a method previously employed by Stanley [10] and Gardner [15] in their experiments.

4. RESULTS

Several standard components of localisation such as angular deviation from the source, front-back confusions etc. were measured. The standard front-back and up-down division of the listening space around the user was applied. The division was based of the interaural axis passing through the ear (see figure 5). Participants displayed established phenomena of reversals in the azimuth and elevation (see figure 6). Approximately 82% of the trials in the front resulted in the stimulus being localised to the rear. For trials in which stimulus was presented to rear, only 4% of these were localised to the front. These results are similar to the ones demonstrated in [6], with the exception that the back-front reversals appear to be almost 10% lower in this study.

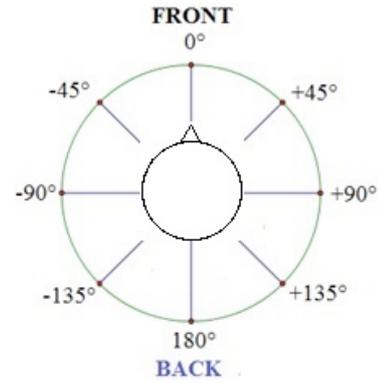


Figure 5: Front-Back division of the auditory space. Line joining -90° and +90° is the interaural axis

Up-down and down-up confusions were also observed during this experiment. Up-down confusions refer to the misrepresentation of a sound source as being below the interaural axis when it is in actuality above it (see figure 6). This is generally caused when an inaccurate representation of the sound source, particularly in the spectral domain, is rendered. This is generally known to occur with non-individualised HRTFs. Down-up confusions can be looked up as phenomena that are exactly opposite of up-down confusions. In this study up-down confusions in the front, -90° to +90°, were approximately 24%. At the rear this rate dropped to about 22%. Down-up confusions on the other hand were relatively low. In front they occurred in approximately 5% of the trials for the front and 7% of the trials for the back.

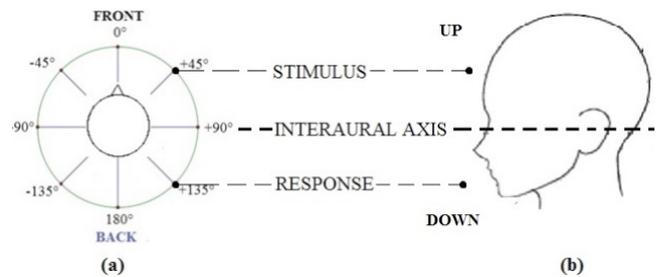


Figure 6: Confusions: Front-Back (a) Up-Down (b)

Angular deviation in the horizontal plane appeared to suffer significantly in comparison with [6]. An average angular deviation from the target of 44.9° was observed for the front. That rose to 51.6° for the rear. These results though appear inflated due one participant who we consider a bad localiser. If this participant's results are excluded, we get an average deviation of 38.9° for the front and 47° for the rear. In the vertical plane, an average error of 21° for front and 20° for the rear was observed. Angular deviation for the horizontal and vertical planes was calculated after resolving confusions. These preliminary results also appear to demonstrate a 'compression' for elevation estimations between -30° and +30°. A large number of trials across all elevations appear to consistently be localised within 15°, top and bottom, of the interaural axis. A more in-depth analysis also shows that early angular deviation results appear to be well correlated with those obtained by Wenzel et al. [14] for high and low elevation across the front, side and back for the headphone condition. Four off of the six participants reported externalization. Two of these four participants reported the stimulus to be located at a distance of 1m or more from the surface of the head.

5. DISCUSSION

While the results indicated here are just a preliminary evaluation of an on-going study, it seems that the addition of a vertical component appears to result in poorer localisation when compared to [6]. The ‘compression’ of the localisation within a relatively small area in front appears to be an interesting phenomenon, possibly driven by an evolutionary adaptation. The level of externalisation reported by participants though is encouraging. Based on these preliminary results we could possibly recommend that the element of height not be incorporated in to binaural spatialisation over a BCH. This is because the addition of a vertical component to a task which previously had only requested azimuth ratings appears to result in poorer measured localisation performance when compared to [6]. However, a full analysis of the results and comparisons with existing BCH and headphone studies needs to be carried out to judge the efficacy of the BCH in being able to reproduce a convincing percept of elevation.

6. REFERENCES

- [1] Walker, B.N. and J. Lindsay, *Navigation Performance In A Virtual Environment With Bonephones*, in *International Conference On Auditory Displays*2005: Limerick, Ireland.
- [2] Brewster, S. and A. Walker, *Non-Visual Interfaces For Wearable Computers*. COLLOQUIUM DIGEST-IEE, 1999.
- [3] Walker, A. and S. Brewster, *Spatial Audio In Small Screen Device Displays*. Personal Technologies, 2000. **4**(2-3): p. 144-154.
- [4] MacDonald, J.A., P.P. Henry, and T.R. Letowski, *Spatial Audio Through A Bone Conduction Interface*. International Journal Of Audiology, 2006. **45**(10): p. 595 - 599.
- [5] Lindeman, R.W., H. Noma, and P.G.d. Barros, *Hear-Through and Mic-Through Augmented Reality: Using Bone Conduction To Display Spatialized Audio*. 2007.
- [6] Barde, A., et al., *Binaural Spatialisation over a Bone Conduction Headset: Minimum Discernable Angular Difference*, in *Submitted to the 140th Convention of the AES*2016: Paris, France.
- [7] Walker, B.N. and J. Lindsay, *Navigation Performance With A Virtual Auditory Display: Effects Of Beacon Sound, Capture Radius and Practice*. Human Factors: The Journal of the Human Factors and Ergonomics Society, 2006. **48**(2): p. 265 - 278.
- [8] Valjamae, A., et al., *Binaural Bone Conducted Sound In Virtual Environments: Evaluation Of A Portable, Multimodal Motion Simulator Prototype*. Acoustical Science And Technology, 2008. **29**(2): p. 149 - 155.
- [9] Villegas, J. and M. Cohen, *GABRIEL: Geo-Aware Broadcasting For In-Vehicle Entertainment And Localizability*, in *AES 40th International Conference*2010, AES: Tokyo, Japan.
- [10] Stanley, R.M., *Measurement And Validation Of Bone Conduction Adjustment Functions In Virtual 3D Audio Displays*, in *School of Psychology*2009, Georgia Institute of Technology.
- [11] Helgason, D., N. Francis, and J. Ante, *Unity3D*, 2005, Unity Technologies: Copenhagen, Denmark.
- [12] Thakur, A. and V. Nair, *3Deception*, 2015, Two Big Ears: Edinburg, Scotland.
- [13] Aftershokz. *Sportz3*. [cited 2015 2nd June]; Available from: <http://aftershokz.com/collections/wired/products/sportz-3>.
- [14] Wenzel, E.M., et al., *Localization using nonindividualized head-related transfer functions*. The Journal of the Acoustical Society of America, 1993. **94**(1): p. 111-123.
- [15] Gardner, W.G., *3D Audio Using Loudspeakers*, in *School of Architecture and Planning*1997, Massachusetts Institute of Technology.
- [16] Stevens, S.S. and E.B. Newman, *The Localization of Actual Sources of Sound*. The American Journal of Psychology, 1936. **48**(2): p. 297 - 306.
- [17] Weinrich, S.G., *Horizontal Plane Localization Ability and Response Time as a Function of Signal Bandwidth*, in *98th Convention of the AES*1995: Paris, France.
- [18] Wersenyi, G., *Localization In A HRTF Based Minimum Audible Angle Listening Test On a 2D Sound Screen For GUIB Applications*, in *115th Convention of the Audio Engineering Society* 2003.
- [19] Wightman, F.L. and D.J. Kistler, *Headphone Simulation Of Free Field Listening. II: Psychophysical Validation*. The Journal of the Acoustical Society of America, 1989. **85**(2): p. 868 - 878.
- [20] Wenzel, E.M., F.L. Wightman, and D.J. Kistler, *Localization with non-individualized virtual acoustic display cues*, in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*1991, ACM: New Orleans, Louisiana, USA. p. 351-359.