Auditory spatial perception using bone conduction headphones along with fitted head related transfer functions

Tray Minh Voong* Michael Oehler†

University of Osnabrueck

ABSTRACT

An approach is presented how to practically determine which head-related transfer function (HRTF) profiles fit best for individuals wearing a bone conduction headphone. Such headphones may be particularly useful for visually impaired people (e.g., for navigation applications) as they do not obstruct the outer ear. Hence, it is still possible to perceive environmental sounds without restraints while wearing such headphones. For a fast and user-friendly identification of fitting HRTF profiles, an adapted tournament system is proposed. It could be shown that the results of the tournament method, where participants had to rate overall preference, externalization and envelopment, correlated well with the results of the localization task. The correlation was higher for the conventional headphones condition than for the bone conduction headphones condition.
Analyses of the transmission characteristics show an uneven frequency response of bone conduction headphones compared to conventional headphones or speakers. In future research it will be investigated whether these findings are relevant for the auditory spatial perception at all and to what extent best fitting HRTFs may compensate for these phenomena.

Index Terms: Head-related-transfer-functions—bone-conduction-headphone—tournament methods

1 INTRODUCTION

The usage of spatialized audio in VR and AR applications generates high immersion and helps listeners/individuals navigate through virtual environments more intuitively. Studies have shown that these solutions turn out to be helpful as user guides for auditory displays and similar interfaces [11, 13]. Spatialized audio cues in AR and MR applications can be used to draw the users attention to points of interests [3]. In conjunction with headphones, head-related transfer functions (HRTF) become essential for a natural (auditory) perception in virtual reality. These functions simulate the outer ear of the user which in turn give directional features to the sound signals. As each person has a unique HRTF, the measurement of individual HRTFs delivers the best auditory spatial perception results. However, the measurement of individual HRTFs is difficult to apply in everyday life. Alternatively, several methods are proposed here by which the user selects a best suiting HRTF profile out of many provided profiles. This approach was deemed fit in various studies as offering a compromise between accuracy and technical feasibility [7, 9].
Because headphones conceal the outer ear of the listener, other solutions for audio in AR and MR applications include the use of bone conduction headsets. The main advantage is that the user is able to perceive both sounds from the virtual environment and real life [5, 6]. Bone conducting headphones radiate the sound through the cranial bone via vibrations directly into the inner ear. Visually impaired persons especially need to rely on their full awareness of their environment in order to avoid accidents - for example, traffic hazards. Therefore, the area of auditory displays and navigation systems for blind people is supposed to be an important field of application.
While the relevance of (a) the best-fitting HRTF profiles and (b) the different aspects of sound transmission and perception with bone conduction headphones has already been a topic of research in different studies, to our knowledge, the interaction between both areas has not been addressed so far.

*e-mail: travoong@uni-osnabrueck.de
†e-mail: michael.oehler@uni-osnabrueck.de

2 AIM

The main objective is the development of a method which allows us to determine fitting HRTF profiles out of a pool of various profiles for each user. The profile should especially be reliable and compatible in conjunction with bone conduction headphones. It can be assumed that HRTF profiles have a bigger impact on sound perception when used with these types of headphones. Therefore, necessary sound processing methods will be evaluated.
Another important objective is the application in everyday life. Consequently, a compromise between accuracy, technical feasibility and a user-friendly selection procedure of the HRTF profile is important.

3 METHOD

Basically, two HRTF profiles are compared with each other. Many studies suggest tournament formats, such as a knockout tournament or round-robin-tournaments [9], or a score system [8] to determine the most fitting HRTF profile. We decided to use a Swiss-style tournament format: an advantage of this format is the allocation of matches between winning profiles which in turn allows a direct comparison and helps obtain a clear result.

3.1 Participants

In a within-subject design 5 male und 4 female participants with normal hearing took part in an exploratory experiment (M = 24.5; SD = 2.4). Each participant had to rate the same stimuli (a) presented over conventional headphones and (b) over bone conduction headphones.

3.2 Hard- and Software

The devices and software in this project are limited to inexpensive and easy-to-use solutions in order to achieve a high practicability. Unity is used as a software framework, and the SOFAlizer Plugin is added as a binaural audio engine that allows switching between different HRTF profiles without delay [4]. As a bone conduction headphone and a regular headphone, the Trekz Titanium and Stax SR-507 headphones were chosen, respectively.
The selection of HRTF profiles used in this project consists of a mix of profiles from different databases. Four profiles were chosen from the CIPIC database [1] and another four from the LISTEN database [12]. The overall number of HRTF profiles in this project is eight.

3.3 Stimuli

The focus of the intended application is navigation in urban areas. Therefore, the stimuli were narrowed down to sound files of vehicles and other road users. We chose two different sounds and implemented them in Unity: a driving car and a drone . These sound objects have been positioned and programmed in Unity to move around the object representing the position of the virtual user. The audiolistener-object in Unity represents the user and perceives the sound through his position. The car is positioned in the same height as the user. During Phase Two (see Section 3.4), the stimuli tremble laterally to simulate tiny head movements, which humans do to improve the accuracy of sound localization. In addition four ambient sound objects generating a soundscape are positioned around the user and form a square. These objects also move slightly around the user to simulate head movements.

3.4 Test Procedure

The study is conducted with the participant sitting on a chair. A calibration at the beginning is necessary with a speaker positioned in front of the participant at a distance of one meter. A pink noise sample is played back by the speaker and the bone conduction headphone. The test person is asked to adjust the level of the bone conduction headphone until both noise stimuli are perceived as equally loud. No headtracking features will be used in this study. For determining the direction, the frontal direction of the seated test person will be declared as twelve oclock. The whole study consists of two phases:

- In the first phase, a fitting HRTF profile for the user is determined. The Swiss-system tournament format is used to match preferred profiles against each other until six rounds are completed. The first stimulus (a car sound) moves around the user in the horizontal plane and enables him to perceive sounds from all directions (in that plane). The second stimulus is a drone sound that moves in an arc from front-left to right in order to include the perception of the elevation. It is possible to switch between two different profiles during the perception of the sounds. The user is asked to select the preferred profile by means of the criteria preference, externalization, and envelopment. The profile with the most wins is considered the preferred HRTF profile for the user.

- In the second phase, each HRTF profile will be assigned to the user once. For each profile, the user is asked to determine the direction of ten randomized stimuli. The participant is requested to use the angles made by the hands of a clock (half-hours included). Each tested stimulus will be positioned anew and moves left and right (2 degrees in each direction with a modulation frequency $f_{mod} = 5$ Hz) during the determination of the direction. The ambient sound objects around the user stay active.

- The two phases are repeated with the other type of headphones (bone conducting vs. conventional headphones) subsequently. The order of type of headphones is randomized.

4 Results

Preliminary analyses reveal that the first or second winner HRTF profile in the tournament task of the first phase correlates with the localization accuracy in the second phase: (a) Five of nine participants showed best localization results for the first or second winner HRTF and (b) all participants showed best or second-best localization results for the first or second winner HRTF in the conventional headphones condition. In the bone conduction headphones condition four of nine participants showed best localization results for the first or second winner HRTF and (b) five of nine participants showed best or second-best localization results for the first or second winner HRTF.

5 Conclusion

The exploratory study showed that a tournament method, where participants had to rate overall preference, externalization and envelopment, correlated well with the results in the following localization task. The correlation was higher for the conventional headphones condition than for the bone conduction headphones condition. A reason for this might be the transmission characteristics of the bone conduction headphone that show a slight attenuation of the lowest frequencies up to 300 Hz. Furthermore, the attenuated representation of the lower frequencies might be problematic and should be reduced. It appears that this does not affect the perception of the stimuli in the tournament task but decreases the localization accuracy. Attempts to amplify the lower frequencies by using compressor or equalizer functions cause noticeable vibrations and affects perception. In the frequency range between 1kHz and 4kHz, crosstalk artifacts emerged. Such artifacts become noticeable when there is a high correlation between the left and right audio channel (e.g., if the stimulus is positioned directly in front of the virtual user). To reduce the crosstalk artifacts or the time to get used to frequency responses of bone conducting headphones the use of various filters or phase shifters should be addressed in future research.

References

[1] V. R. Algazi, R. O. Duda, D. M. Thompson, and C. Avendano. The cipic hrtf database. In *Proceedings of the 2001 IEEE Workshop on the Applications of Signal Processing to Audio and Acoustics (Cat. No.01TH8575)*, pp. 99–102. IEEE, 2001. doi: 10.1109/ASPAA.2001.969552

[2] C. Armstrong, G. Kearney, and L. Thresh. Sadie II database — Binaural and anthropomorphic measurements for virtual loudspeaker rendering, June 2018. https://www.york.ac.uk/sadie-project/database.html

[3] J. R. Blum, M. Bouchard, and J. R. Cooperstock. What's around me? Spatialized audio augmented reality for blind users with a smartphone. In *International Conference on Mobile and Ubiquitous Systems: Computing, Networking, and Services*, pp. 49–62, ICST, Springer, Ber, 2011.

[4] C. Jenny, P. Majdak, and C. Reuter, eds. *SOFA Native Spatializer Plugin for Unity—Exchangeable HRTFs in Virtual Reality*, AES 144th Convention, 2018, AES, Milan.

[5] R. W. Lindeman, H. Noma, and P. G. de Barros. An empirical study of hear-through augmented reality: Using bone conduction to deliver spatialized audio. In M. C. Lin, ed., *IEEE Virtual Reality Conference, 2008*, pp. 35–42. IEEE, Piscataway, NJ, 2008.

[6] J. A. MacDonald, P. P. Henry, and T. R. Letowski. Spatial audio through a bone conduction interface. *International journal of audiology*, 45(10):595–599, 2006. doi: 10.1080/14992020600876519

[7] A. Roginska, G. H. Wakefield, and T. S. Santoro. User selected hrtfs: Reduced complexity and improved perception. In *Undersea Human System Integration Symposium, Providence, RI*, pp. 1–14, 2010.

[8] B. U. Seeber and H. Fastl. *Subjective selection of non-individual head-related transfer functions*. Proceedings of the 2003 International Conference on Auditory Display, 2003, Boston, MA, USA.

[9] R. Shukla, R. Stewart, A. Roginska, and M. Sandler, eds. *User Selection of Optiman HRTF Sets via Holistic Comparative Evaluation*, International Conference on Audio for Virtual and Augmented Reality 2018, Redmond, WA, USA.

[10] L. Simon, SR, N. Zacharov, and B. F. G. Katz. Perceptual attributes for the comparison of head-related transfer functions. *The Journal of the Acoustical Society of America*, 140(5):3623–3632, 2016.

[11] B. N. Walker and J. Lindsay. *Navigation performance in a virtual environment with bonephones*. Proceedings of ICAD 05-Eleventh Meeting of the International Conference on Auditory Display, 2005, Limerick, Ireland.

[12] O. Warusfel. Listen HRTF Database, May 2003. http://recherche.ircam.fr/equipes/salles/listen/index.html

[13] E. M. Wenzel, M. Godfroy-Cooper, and J. D. Miller. *Spatial auditory displays: Substitution and complementarity to visual displays*. 20th International Conference on Auditory Display. 2014, New York, USA.