PROCEEDINGS of the
23rd International Congress on Acoustics
9 to 13 September 2019 in Aachen, Germany

Tournament Formats as Method for Determining Best-fitting HRTF Profiles

Tray Minh Voong(1), Michael Oehler(2)

(1)University Osnabrück, Germany, tray.minh.voong@uni-osnabrueck.de

(2)University Osnabrück, Germany, michael.oehler@uni-osnabrueck.de

Abstract

An approach is presented how to practically determine best-fitting HRTF profiles for individuals wearing conventional headphones and bone conduction headphones. The latter may be particularly useful for visually impaired people (e.g., for navigation applications), as the outer ear is not covered and the perception of environmental sounds is not affected. For a fast and user-friendly identification of best-fitting HRTFs, different tournament methods are compared. There are several studies that investigate (a) aspects of spatial sound perception with bone conduction headphones and (b) tournament systems as selection procedure for best-fitting HRTFs, but to our knowledge there is no study that analyzes the interdependence of (a) and (b). Compared with other tournament systems the swiss system tournament produces more accurate results than a knock-out tournament in less time than the round-robin tournament system. By matching the preferred HRTF profiles against each other early on, a direct comparison between these profiles can be achieved. In addition profiles disadvantaged by the randomized algorithm in earlier rounds are not excluded but still matched against similar profiles worth considering and still have a chance to compete in later rounds. Therefore the swiss system tournament offers a viable solution in determining fitting HRTF profiles.

Keywords: Tournament System, HRTF, bone conduction Headphones

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 [12, 14]. Spatialized audio cues in AR and MR applications can be used to draw the user's 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 [6, 5]. bone conduction 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. In an exploratory study, we investigated this interaction with a small number of participants [11]. A correlation between fitting HRTF profile and accuracy of the localization test could be shown. In the current study some algorithms have been optimized and the number of participants has been increased.

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. It can be assumed that HRTF profiles have a bigger impact on sound perception when used with bone conduction headphones, since these types of headphones exclude the outer ear. 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. This includes the attempt to shorten the time of the selection procedure as much as possible.

3 METHOD

3.1 Tournament Formats

Basically, two HRTF profiles are compared with each other. Many studies suggest tournament formats, such as a knock-out tournament or round-robin-tournaments [9], or a score system [8] to determine the most fitting HRTF profile. We decided to use a swiss system 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.
The formula for determining the minimum number of rounds for a knock-out tournament where $\mathrm{N_R}$ is the amount of rounds and $\mathrm{N_p}$ is the amount of HRTF profiles equals to

$$N_R = \log_2(N_p) \tag{1}$$

For a round-robin tournament the formula for determining the minimum number of rounds is

$$N_R = N_P - 1 \tag{2}$$

Since a swiss system tournament has no fixed number of rounds, it is necessary to determine the number of top places to prevent more than one HRTF from taking first place. Since the first place is the only relevant position, the factor P (number of singled out ranks) is 1:

$$N_R > 0,2 * N_P + 1,4 * P \tag{3}$$

Eight HRTF profiles correspond to three rounds in a knock-out tournament, seven rounds in a round-robin tournament and a minimum of three rounds for a swiss system tournament. While a round consist of several matches, the total number of matches differs between the tournament variants.
The formula for the number of matches $\mathrm{N_M}$ for a knock-out tournament is

$$N_M = N_P - 1 \tag{4}$$

In this case, there are no iterating rounds: After each round the number of HRTF profiles is halved. The other two tournament formats have similar rounds where there are no preliminary dropouts of profiles. In a round-robin tournament the number of matches is equal to the number of profiles minus one times half of the participants:

$$N_M = (N_P - 1) * (\frac{N_P}{2}) \tag{5}$$

Considering the minimum number of a swiss system tournament to determine a clear winner (three round with eight profiles), six rounds were planned for the study:

$$N_M = (N_R) * (\frac{N_P}{2}) \tag{6}$$

The number of matches would be seven in the knock-out tournament, twenty-eight for the round-robin, and twenty-four for the swiss system tournament (with six rounds), respectively. Therefore the swiss system tournament is chosen as the preferred tournament system because it is not as time consuming as the round-robin tournament, but still allows a higher accuracy than a knock-out tournament system where a single loss excludes the profile from further matches.

In order to increase the accuracy of the swiss system, profiles already assigned to each other in previous rounds are still allowed, while orthodox swiss systems omit already occurred pairings., e.g., when the first ranked already had a match with the second rank, it is assigned to the third rank instead. The resulting iterations of the direct comparison between the HRTF profiles serve as a double check to obtain a clear result between them.

3.2 Participants

In a within-subject design 21 male and 14 female participants with normal hearing took part in the experiment (M = 24.37; SD = 2,75). Each participant had to rate the same stimuli (a) presented over conventional headphones and (b) over bone conduction headphones.

3.3 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 conventional 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 [13]. The overall number of HRTF profiles in this project is eight.

3.4 Stimuli

The focus of the intended application is navigation in urban areas. Therefore, the stimuli were limited to sounds 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 were positioned and programmed in Unity to move around the object representing the virtual user's position. 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 3.5), the stimuli tremble laterally to simulate tiny head movements that 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.

3.5 Test Procedure

The study is conducted with the participant sitting on a chair. A pink noise sample is played back by the conventional headphones and the bone conduction headphone. The participant is asked to adjust the level of the headphones until the perception of the sound is loud enough and satisfactory. No head tracking features will be used in this study. For determining the direction, the frontal direction of the seated test person will be declared as twelve o'clock. The whole study consists of two phases:

First phase: 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

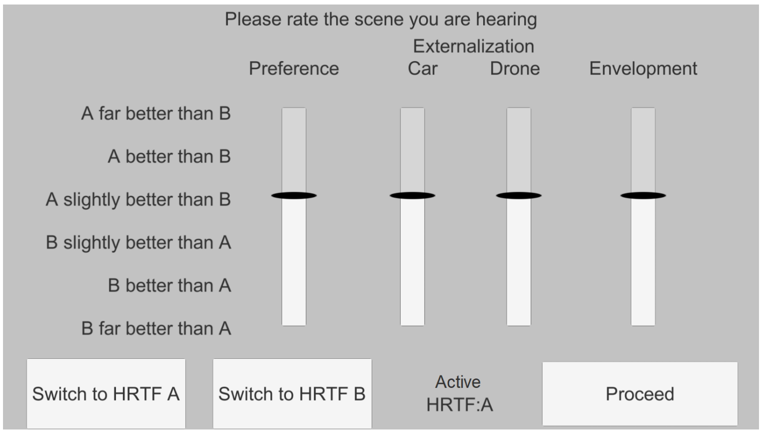

Figure 1. Screenshot of phase one of the study

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 (A and B) during the perception of the sounds. The user is asked to select the preferred profile by means of the criteria preference, externalization, and envelopment.
Whereas preference and envelopment is referencing to the whole scene, the participant is asked to rate externalization for the car and the drone respectively, which results in a total of four evaluable items. Each of these items is represented by a fader which can be moved to different statements ranging from "A far better than B" to "B far better that A". There is no option to rate both as "equally good" as the aim is to force the participant to decide for one preferred HRTF profile (see fig.1).
Except for 'preference', externalization and envelopment is calculated with a coefficient of 1. To prevent ties between the matches, preference has a higher value/weight, namely 1.5. As long as the rating shifts towards either A or B, the HRTF gets a point for the specific item, regardless of how much better the HRTF is compared to the other one. The maximum number a HRTF can get is 4.5, one point each for envelopment, externalization (two stimuli), and 1.5 points for preference. The HRTF profile with the most points (of at least 2.5) is the match winner and gets a tournament point. After all matches have been played, a new ranking is generated. According to the ranking list, new matches are assigned according to the top-down principle: the first rank is matched up against the second rank and so forth. After six iterations the profile with the most tournament points is declared the preferred HRTF profile for the user.

Second phase: 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, see fig 2.). Each tested stimulus will be positioned anew and moves slightly to the left and right (2 degrees in each direction with a modulation frequency f_{mod} = 5 Hz) during the determination of the direction. The purpose of these movements is the simulation of slight head movements which are unconsciously made when trying to localize acoustic signals. Since both the conventional and the bone conduction headphones offer no headtracking, the tiny movements are an attempt to compensate for it. In contrast the ambient sound objects around the user stay active. The randomization algorithm is programmed to prevent subsequent stimuli with identical directions: If these events occur, the stimuli are newly randomized.

The two phases are repeated with the other type of headphones (bone conduction headphones vs. conventional headphones) subsequently. The order of type of headphones is randomized and evenly distributed among the

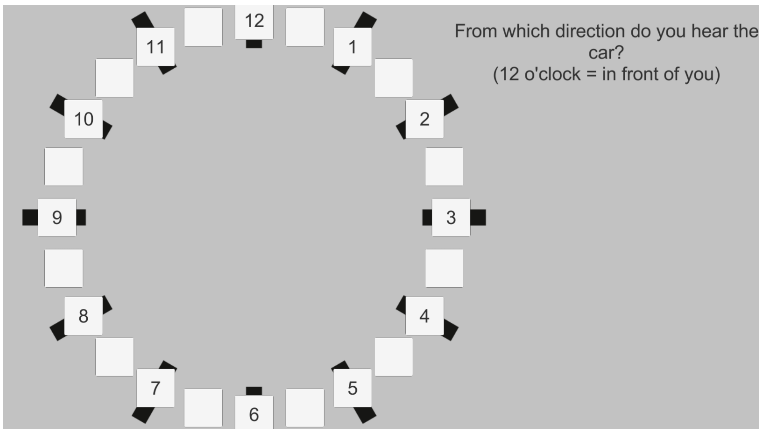

Figure 2. Screenshot of phase two of the study

participants. In this case the starting order changes, i.e., the first participant starts with the conventional headphones, and the second with the bone conduction headphones.

4 Results

The results reveal that there is some correlation between the conventional headphones and BCH: Three participants have exactly the same ranking for the conventional headphones and the BCH. In 13 cases there is at least one identical HRTF in the same rank for conventional headphones and BCH and in another 13 cases a HRTF profile is either ranked first or second for both conventional headphones and BCH, respectively. Six participants do not show overlapping profile rankings for the first or second place.
In both conditions, i.e., stimuli via conventional headphones and via BCH, 46% of the participants had the best localization result in phase two with the same HRTF profile (rank 1 or 2), which emerged from the tournament task in phase 1 as best-fitting profile (rank 1 or 2). A McNemar test showed no significant difference between the two conditions ($p = .598$; one-sided).

Table 1. Correlations between the headphones in phase one and between phase one and phase two

	Correlation between conventional headphones and BCH in phase one	Correlation between phase one and phase two (for both conventional headphones and BCH)
One identical HRTF in identical ranks	13	20
One identical HRTF in different ranks	13	12
Identical HRTFs in both ranks	3	-
No Similarities	6	10
Number of participants with two similarities	-	7
Total	35	35

As for the comparison between phase one and phase two there is no case where the first and second ranked HRTF profiles in phase one is also in first and second place in phase two. Twenty participants have at least one identical HRTF in identical ranks, twelve participants have at least one identical HRTF in a different rank, corresponding to a total number of 32 cases where a HRTF profile ranked first or second is also ranked first or second for the other headphone type. There are seven overlapping cases where participants have two similarities. Ten participants show no similarities between phase one and phase two, i.e., the profiles ranked first or second are different between the conventional headphones and the BCH.

5 Conclusion

The study showed that a tournament method, where participants had to rate overall preference, externalization and envelopment in a simulated scene, 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.
While bone conduction headphones have the advantage of not obstructing the outer ear, the transmission of sound via vibrations still has artifacts and might alter the perception. Conventional headphones use the air as sound carrier. On the contrary bone conduction headphones rely on solid matter to transmit the sound into the inner ear. Since the head of each person differs in various characteristics, e.g., size and shape, the density of the head varies significantly between different people. This further complicates a possible standardization for any person using these types of headphones. Nevertheless, bone conduction headphones still present a viable solution for audio in AR and MR environments.
The rating system in phase one still has room for improvement, as the algorithm only sums up the points of each criteria (preference, externalization and envelopment). Therefore a grading between the HRTF profiles might improve the selection process of the tournament system. In addition, all elements except the preference are equally weighted.. Depending on which items are more important and therefore should have more weight on the result, the matches may have a different winner and lead to a different result. A retrospective adjustment of the grading is not possible because the process of the swiss system tournament changes accordingly to the results of the matches. A possible work around might be a simulated swiss system tournament with the collected data of a Round-robin tournament. While it can work with data on any direct comparison of HRTFs, it can also falsify the overall result and is therefore not recommended.
While phase two has a satisfactory grading system, where the overall deviation is taken into account, that takes into account the overall variance, there are some cases that have a negative impact on the outcome., e.g., in-head-localization and front-back confusion. Especially the latter has a negative impact on the accuracy of the localization test. How these perceptual artifacts can be taken into account needs to be investigated.

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] T. M. Voong, M. Oehler. *Auditory spatial perception using bone conduction headphones along with fitted head related transfer functions*, 26th IEEE Conference on Virtual Reality and 3D User Interfaces 2019, Osaka, Japan.

[12] 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.

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

[14] 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.