



HAL
open science

Impact of HRTF individualization on player performance in a VR shooter game I

David Poirier-Quinot, Brian F. G. Katz

► **To cite this version:**

David Poirier-Quinot, Brian F. G. Katz. Impact of HRTF individualization on player performance in a VR shooter game I. AES Conference on Spatial Reproduction, Aug 2018, Tokyo, Japan. hal-01856505

HAL Id: hal-01856505

<https://hal.science/hal-01856505>

Submitted on 12 Aug 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Audio Engineering Society Conference Paper

Presented at the Conference on
Spatial Reproduction
2018 August 6 – 9, Tokyo, Japan

This conference paper was selected based on a submitted abstract and 750-word precis that have been peer reviewed by at least two qualified anonymous reviewers. The complete manuscript was not peer reviewed. This conference paper has been reproduced from the author's advance manuscript without editing, corrections, or consideration by the Review Board. The AES takes no responsibility for the contents. This paper is available in the AES E-Library (<http://www.aes.org/e-lib>), all rights reserved. Reproduction of this paper, or any portion thereof, is not permitted without direct permission from the Journal of the Audio Engineering Society.

Impact of HRTF individualization on player performance in a VR shooter game I

David Poirier-Quinot and Brian F.G. Katz

Sorbonne Université, CNRS, Institut Jean Le Rond d'Alembert, UMR 7190, F-75005 Paris, France

Correspondence should be addressed to David Poirier-Quinot (david.poirier-quinot@upmc.fr)

ABSTRACT

We present preliminary results of an experiment to assess the impact of individualized binaural rendering on player performance in the context of a VR “shooter game”, as part of a larger project to characterize the impact of binaural rendering quality in various VR applications. Participants played a game in which they were faced with successive enemy targets approaching from random directions on a sphere. Audio-visual cues allowed for target localization. Participants were equipped with an Oculus CV1-HMD, headphones, and two Oculus Touch hand tracked devices as targeting mechanisms. Participants performed two sessions, once using their best and once their worst-match HRTFs from a “perceptually orthogonal” optimized set of 7 HRTFs [1]. Results indicate significant performance improvement (speed and movement efficiency) with best-match HRTF binaural rendering sessions.

1 Introduction

“Binaural hearing” refers to the capability of integrating information from the two ears to perceive a sound in three-dimensional space (azimuth, elevation, and distance). Psychophysical studies have shown that various mechanisms are involved in the human auditory system for sound localization [2]. To infer the angular direction of a sound source, these mechanisms rely on direction-dependent audio cues, resulting from the propagation of an acoustic wave from the source to both ears. Using digital signal processing, these cues can be applied to any audio input to simulate a sound object at a virtual position in a listener’s 3D auditory space (experienced over headphones). The set of these direction-dependent

cues for a given person is typically referred to as a Head Related Transfer Function (HRTF).

HRTFs are individual, directly resulting from the interactions between a person’s morphology and an impinging acoustic wave during its propagation around the head [2]. Individuals listening to a binaural audio scene rendered using their own HRTF will perceive each of its components with more spatial precision than those presented with a random HRTF set [3]. Various methods have been proposed to select a “best-match” HRTF from an existing database [4, 5], as measuring an individual’s actual HRTF is a demanding operation [6]. This process of selection and use of a best-match HRTF for binaural synthesis is here referred to as HRTF individualization.

This study is part of a larger research project aiming to characterize the impact of binaural rendering quality in the context of different Virtual Reality (VR) application contexts. This study focuses on the impact of HRTF individualization on performance in the context of a VR shooter game. While the core of the gameplay is built around an audio-visual localization task, it extends the existing literature [7, 8, 9, 10] in that the final experience is truly a game, where participants are placed under increasingly difficult time and performance constraints. Besides improving localization accuracy, the underlying hypothesis of this study is that using individualized HRTFs will result in participants more readily relying on their reflexes to *e.g.* dodge enemies or sightlessly aim when incoming targets arrive too quickly for the basic sequence: *hear* → *search* → *focus* → *aim* → *shoot*.

2 Experimental design

The experiment consisted of two sequential parts. The objective of Part 1 was to identify the *best* and *worst* match HRTFs from a subset of 7 for each participant. Part 2 was the VR shooter game. A total of 30 participants undertook the experiment (13 female, mean age 33.2 ± 11.4 years).

2.1 HRTF classification

The HRTF subset database was assembled from the LISTEN database, defined from a “perceptually orthogonal” optimized HRTF collection [1]. Per-participant best and worst match HRTF sets were selected based on the method elaborated in [11], establishing a classification based on perceptual-space distance between a spatialized audio trajectory and a described reference. Two trajectories were presented: horizontal plane (12 angles $[0^\circ:30^\circ:330^\circ]$) and median plane (19 angles $[-45^\circ:15^\circ:225^\circ]$), as illustrated in Fig. 1.

Each audio trajectory was generated with the 7 HRTFs from the subset. Participants were instructed to rate the 7 resulting versions of each trajectory on a fixed 9-point scale. They were encouraged to distribute their notations on that scale, and required to indicate at least one best (9) and one worst (1) match. Both median and horizontal rating sessions were repeated 3 times, to take into account HRTF rating consistency [12]. Participants completed the study in a listening booth, ambient noise level < 30 dBA using Sennheiser HD600 headphones and RME Fireface UC audio interface.

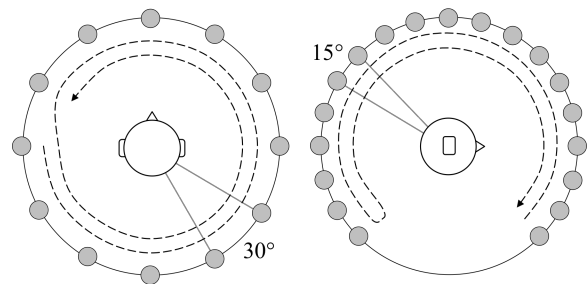


Fig. 1: Descriptions for HRTF quality ratings: horizontal (left) and median (right) plane trajectories indicating the start/stop position and trajectory direction (— →).

2.2 VR Shooter Game

During the VR shooter game, participants were equipped with an Oculus CV1 Head Mounted Display (HMD), a pair of headphones (those of the CV1), and a pair of hand tracked devices (Oculus Touch controllers). The game started with participants immersed in a virtual scene, standing on a 0.5 m radius platform mounted on a pole at the center of a 20 m radius spherical structure. Enemy targets could “spawn” from any of the 29 evenly distributed holes in the structure, flying in straight lines towards the participant until collision, either with a bullet or the participant. Enemy targets emitted specific event-based sounds for: spawning, launching, flight, and collision. All sounds were spatialized using the freely available Anaglyph binaural audio engine v0.9 [13]. Anechoic conditions were employed, no room effect was included to keep the study’s focus on HRTF effects. Participants were instructed to shoot at the incoming targets using the pair of hand-held blasters, the avatar representations of the hand tracked devices in the virtual scene, destroying as many as possible, as fast as possible, in the given time limit. If unable to shoot at an incoming target (*e.g.* when the game accelerated), participants were advised to try and dodge them to avoid collisions.

The game was designed using the Unity v2017.3.0 game engine with modelled assets designed in Blender v2.79. The OSC (Open Sound Control) protocol [14] was used for communications between Unity and the Anaglyph engine running as a VST in Cycling’74 Max v7.3. Figure 2 depicts the game setup and the VR scene. A video extract showing a game session is available¹.

¹<https://youtu.be/q6muds1qW-w>

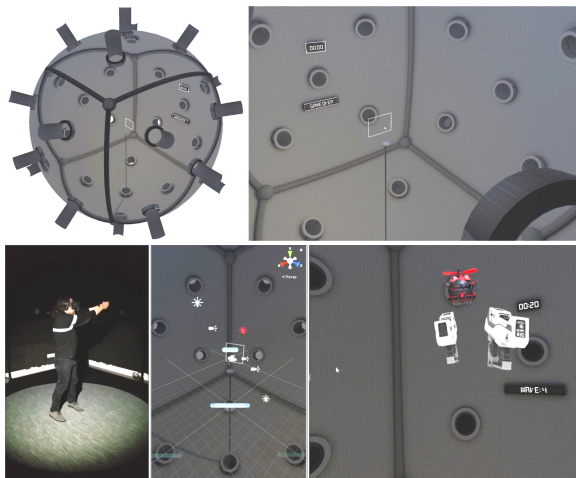


Fig. 2: Game scene overview: (upper-left) overall view of the virtual environment, (upper-right) focus on the platform atop which participants stand during the game, (lower-left) participant in the VR room, (lower-middle) virtual environment during gameplay, and (lower-right) in-game screenshot,

A short training session introduced the controls and the difficulty level mechanism, implemented so that the overall game dynamic (enemy spawn interval, flight speed, etc.) increased as the game progressed and participant's skill ("in-game level" below) improved. The 5 min game was played in two sessions, once with participant's best match HRTF, once with the worst. Best/worst match presentation order was evenly balanced among participants, resulting in two groups. The game was kept short to avoid fatigue on this rather demanding task (see video extract).

3 Preliminary Results

Performance assessment was based on 3 metrics, calculated from sessions logs: in-game level, spawn-spot reaction time and spawn-spot travelled angular distance. The in-game level was related to the number of enemies destroyed versus those that hit participants: increasing one unit for every three consecutive kills, decreasing one unit for every two consecutive fails.

The spawn-spot reaction time corresponded to the time interval between the spawn of a target and its entering the visual field of view, defined as a 50° cone centered around the current forward view axis. The event of

Table 1: Independent and dependent variables of the experimental protocol.

| <i>Independent variables</i> | | |
|------------------------------|--------------|-----------------|
| Participant ID | 30 | random variable |
| HRTF ID | 2 | best, worst |
| Session ID | 2 | first, second |
| <i>Dependent variables</i> | | |
| angular distance | event-wise | raw and norm |
| time | event-wise | raw and norm |
| mean level | session-wise | raw and norm |

seeing the target, rather than destroying it, was chosen so as to remove the impact of skill at aiming and destroying targets from the analysis (being independent of the acoustic rendering quality). Targets spawned in the current field of view were discarded from spawn-spot reaction time analysis. For targets that never entered the field of view, the spawn-to-collision time was used as the spawn-spot reaction time. For events where participants destroyed the target without seeing it (representing 0.3% of the total number of target destroyed), the spawn-shoot time was used as the spawn-spot reaction time.

The associated spawn-spot travelled angular distance metric corresponds to the angular distance traversed by the participant's head from spawn to spot events. This last metric represented movement efficiency, and served to differentiate between participants using binaural cues to localize targets and those randomly looking around [15]. Table 1 summarizes the independent and dependent variables of the experimental protocol.

Result significance was assessed using a Wilcoxon signed rank test (p -value threshold of 0.05) as all compared paired-sample distributions proved to follow a non-normal (skewed) distribution.

3.1 HRTF Classification

Results of the HRTF ratings of Part 1 are summarized in Fig. 3, focusing on the scores obtained by each participant's best and worst HRTF match for both trajectories. Participants were consistent in their classification with regards to these extrema. As audio sources in Part 2

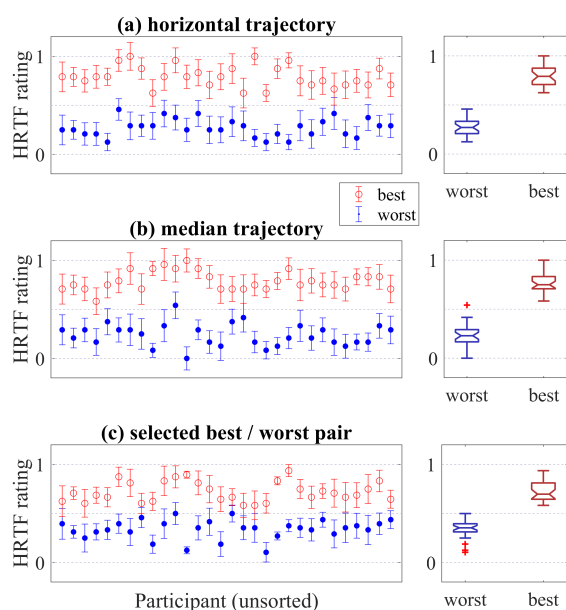


Fig. 3: Results of the HRTF classification task for all participants. The reported rating value corresponds to the average normalized rank given by participants to their best/worst HRTF. A value of 0 (resp. 1) indicates that the HRTF was always rated as the least (resp. most) representative of the described trajectory across the 3 rating repetitions. Scores for the best and worst HRTF matches for the (a) horizontal and (b) median trajectory. (c) Combined mean score results for the *selected* best and worst match HRTF.

of the experiment were to arrive from all directions, an average best- and worst-HRTF match across trajectories was established for each participant. The rating statistics for selected best- and worst-HRTF matches are shown in Fig. 3. Participants proved consistent in their ratings to clearly distinguish between best and worst match for both horizontal and median trajectories. As reported in [11], participant HRTF ratings for the horizontal and the median trajectories were not correlated. This explains the observed decrease in rating values for the trajectory mean results as compared to the individual horizontal and median plane trajectory ratings. For almost all participants, average-best and average-worst HRTF scores remained sufficiently distinct to distinguish both populations (Fig. 3c).

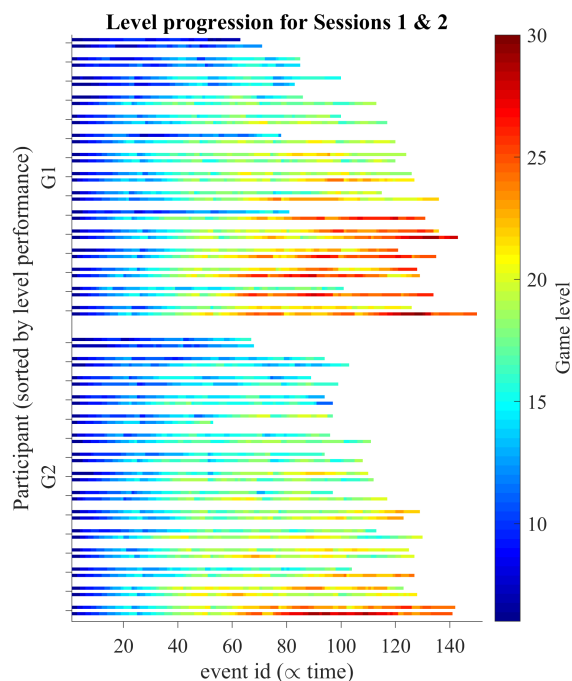


Fig. 4: Evolution of each participants' level throughout the game (from 6 to approx. 30). Each pair of bars represents Session 1 & 2 of a participant. Participants are separated according to group, sorted (top to bottom) based on individual maximum session-wise mean level.

3.2 VR Shooter Game

Figure 4 illustrates the evolution of game level during the course of each session. As can be seen, large inter-participant differences can be observed for the VR shooter game results.

3.2.1 Statistical analysis across participants

Analyzing the average statistical metrics per participant between the two sessions across both groups, mean spawn-spot reaction time significantly decreased between Sessions 1 and 2 (1.40 s vs. 1.35 s). Mean spawn-spot angular distance traversed (averaged per participant's session) also significantly decreased between both session (155° vs. 144°). No significant differences in mean level were observed between sessions. These results highlight the game task learning effect independent of the HRTF condition.

Repeating the same analysis as a function of HRTF across both groups showed no significant impact of the HRTF on these metrics. These results would suggest at first glance that there was no benefit to using individual HRTFs in the VR shooter game.

3.2.2 Statistical analysis by event

Combining the results of each event (from target spawn to target destruction) for all participants, the event-wise analysis also reflects the impact of the game learning effect on performance for spawn-spot angular distance traversed (154.7° vs. 143.4°) and reaction time (1.38 s vs 1.34 s). Contrary to mean participant results (see Section 3.2.1), event-wise statistics show that the impact of HRTF quality on reaction time (1.39 s vs. 1.34 s) and angular distance traversed (151° vs. 147°) is significant.

3.2.3 HRTF \times Session interaction analysis

Subsequent analysis examines the interaction effect of HRTF order. The event-wise mean and 95% confidence intervals for reaction time and angular distance traversed are shown in Fig. 5. These results show a clear interaction effect, with an advantage to using best-match HRTF in Session 2. Analysis of event-wise reaction time and angular distance traversed by HRTF \times session shows that:

- Session 1: Participants starting with their best match HRTF showed performances comparable to those of participants who started with their worst match.
- Session 2: Participants using their best-match HRTF exhibit a clear improvement in performance.
- Session 2: Participants using their worst-match HRTF showed no improvement.

These results suggest that Group 1 (who commenced the game with their worst-match HRTF) was able to benefit from their best-match HRTF after an initial game training period in Session 1. Commencing the game with a best-match HRTF (Group 2) did not improve participants performance immediately, as there was an apparent game task learning effect which appears to supersede that of spatial audio rendering quality. This leads to a potential conclusion that a worst match HRTF negates the benefit that should result from training. Further tests are currently underway to examine this question.

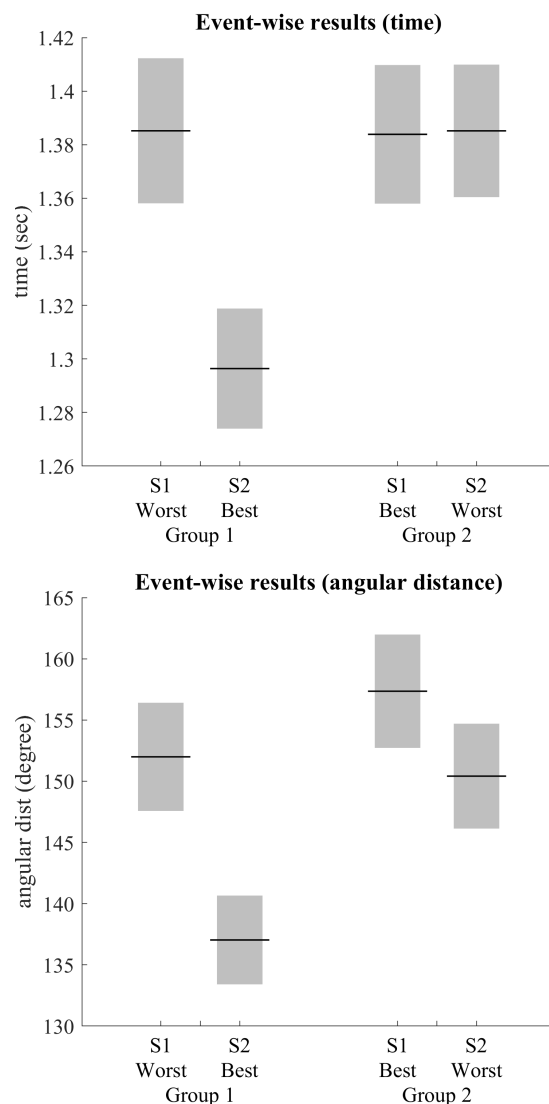


Fig. 5: Reaction time and traversed angular distance (mean and 95% CI) across session and HRTF for each event.

4 Conclusion

We have presented preliminary results of an experiment designed to assess the impact of individualized binaural rendering on player performance in the context of a VR “shooter game”. Participants performed two game sessions, using once their best-match HRTF and once using their worst-match HRTF. During the game, participants had to locate and shoot at successive enemy

targets approaching from random directions within a sphere. Results indicate that the use of a best-match HRTF improved participants performance regarding the time they needed to localise the targets and the angular distance they travelled before doing so (average of 1.34 s and 147° with best-HRTF vs. 1.39 s and 151° with worst).

Preliminary participant-based analysis has shown promising results concerning the creation of an overall metric to rate participants “affinity” with binaural hearing. Using this metric, the authors are investigating how participants who focused on the auditory scene (Part 1) and were comfortable with VR game mechanisms (Part 2) can be identified from the participant pool in order to examine the benefit of HRTF individualization for this sub-group. The hypothesis of this ongoing analysis is that the benefits of HRTF individualization for this sub-group of “aware listeners” exceed those of the average participant.

The impact of the learning effect along with the research for a metric to evaluate the participants receptiveness to binaural hearing individualization point towards an extension of this study. The extended study will consider roughly the same number of participants but exposing them longer / on repeated occasions to the VR game until the uncontrolled impact of the learning effect can be discarded. A control group will be added to this second study to fully understand the impact of the learning effect during the early stages of the game. A more thorough statistical analysis on participants results will then be conducted, absent in this preliminary presentation.

Acknowledgements

This work was funded in part through a fundamental research collaboration partnership between Sorbonne Université, CNRS, Institut ∂' Alembert and Oculus VR, LLC.

References

- [1] Katz, B. F. G. and Parseihian, G., “Perceptually based head-related transfer function database optimization,” *J Acous Soc of Am*, 131(2), pp. EL99–EL105, 2012, doi:10.1121/1.3672641.
- [2] Blauert, J., *Spatial Hearing: The psychophysics of human sound localization*, MIT press, 1997.
- [3] Begault, D. R., Wenzel, E. M., and Anderson, M. R., “Direct comparison of the impact of head tracking, reverberation, and individualized head-related transfer functions on the spatial perception of a virtual speech source,” *J Aud Eng Soc*, 49(10), pp. 904–916, 2001.
- [4] Seeber, B. U. and Fastl, H., “Subjective selection of non-individual head-related transfer functions,” in *Intl Conf on Auditory Display*, pp. 259–262, 2003.
- [5] Zotkin, D., Hwang, J., Duraiswaini, R., and Davis, L. S., “HRTF personalization using anthropometric measurements,” in *IEEE Workshop on Applications of Sig Proc to Audio and Acoustics*, pp. 157–160, 2003.
- [6] Carpentier, T., Bahu, H., Noisternig, M., and Warusfel, O., “Measurement of a head-related transfer function database with high spatial resolution,” in *Forum Acusticum (EAA)*, 2014.
- [7] Wenzel, E. M., Arruda, M., Kistler, D. J., and Wightman, F. L., “Localization using nonindividualized head-related transfer functions,” *J Acous Soc of Am*, 94(1), pp. 111–123, 1993.
- [8] Xu, S., Li, Z., and Salvendy, G., “Individualization of head-related transfer function for three-dimensional virtual auditory display: a review,” in *Intl Conf on Virtual Reality*, pp. 397–407, Springer, 2007.
- [9] Härmä, A., van Dinther, R., Svedström, T., Park, M., and Koppens, J., “Personalization of headphone spatialization based on the relative localization error in an auditory gaming interface,” in *Aud Eng Soc Conv 132*, 2012.
- [10] Mehra, R., Nicholls, A., Begault, D., and Zannoli, M., “Comparison of localization performance with individualized and non-individualized head-related transfer functions for dynamic listeners,” *J Acous Soc of Am*, 140(4), pp. 2956–2957, 2016.
- [11] Andreopoulou, A. and Katz, B. F., “Subjective HRTF evaluations for obtaining global similarity metrics of assessors and assessees,” *JMUI*, (SI: Auditory Display), pp. 1–13, 2016.
- [12] Andreopoulou, A. and Katz, B., “Investigation on Subjective HRTF Rating Repeatability,” in *Aud Eng Soc Conv 140*, pp. 9597:1–10, 2016.

- [13] Poirier-Quinot, D. and Katz, B. F., “The Anaglyph binaural audio engine,” in *Aud Eng Soc Conv 144*, 2018.
- [14] Wright, M., “Open Sound Control: an enabling technology for musical networking,” *Organised Sound*, 10(3), pp. 193–200, 2005.
- [15] Katz, B., Tarault, A., Bourdot, P., and Vézien, J.-M., “The use of 3D-audio in a multi-modal teleoperation platform for remote driving/supervision,” in *Aud Eng Soc Conf: Intelligent Audio Environments*, pp. 1–9, Saariselkä, 2007.