Assessing the Impact of Head-Related Transfer Function Individualization on Task Performance: Case of a Virtual Reality Shooter Game

David Poirier-Quinot, Brian F.G. Katz

OPEN ACCESS Freely available online

D. Poirier-Quinot and B. F.G. Katz, "Assessing the Impact of Head-Related Transfer Function Individualization on Task Performance: Case of a Virtual Reality Shooter Game"
J. Audio Eng. Soc., vol. 68, no. 4, pp. 248–260, (2020 April.).
DOI: https://doi.org/10.17743/jaes.2020.0004

Assessing the Impact of Head-Related Transfer Function Individualization on Task Performance: Case of a Virtual Reality Shooter Game

DAVID POIRIER-QUINOT, *AES Member,* **AND BRIAN F.G. KATZ,** *AES Member*
(brian.katz@sorbonne-universite.fr)

Sorbonne Université, CNRS, Institut Jean Le Rond d'Alembert, Lutheries - Acoustique - Musique, Paris, France

This paper presents the results of an extended experiment to assess the impact of individualized binaural rendering on player performance in an ecologically valid use context, specifically that of a VR "shooter game," as part of a larger project to characterize the impact of binaural rendering quality in various VR type applications. Participants played a simple game in which they were faced with successive targets approaching from random directions on a sphere. While audio-visual cues allowed for general target localization, only sections of the game that relied on audio cues were used for analysis. Two HRTF exposure protocols were used, comprising best and worst-match HRTFs from a "perceptually orthogonal" optimized set of HRTFs, during the course of six game sessions. Two groups performed the game sessions exclusively using either their best or worst-match HRTF. Two additional groups performed the game sessions alternating between best and worst-match HRTFs. Results suggest that HRTF quality had minimal general impact on in-game participant performance and improvement rate. However, performance for extreme elevation target positions was affected by the quality of HRTF matching. In addition, a subgroup of participants showed higher sensitivity to HRTF choice than others.

0 INTRODUCTION

The human auditory system relies on direction-dependent audio cues to infer the angular direction of a sound source [1]. The set of these direction-dependent cues for a given person is commonly referred to as the Head Related Transfer Function (HRTF) [2]. Binaural rendering is a signal processing technique that relies on these HRTFs to reproduce spatial hearing over headphones.

Similar to fingerprints, HRTFs are unique to each individual. Per-user HRTF measurement is not a practical option for casual Augmented and Virtual Reality (AR/VR) applications [3]. As such, binaural rendering is often achieved using *non-individual* (generic) HRTFs, selected from existing databases [e.g., 4, 5]. The use of such HRTFs usually results in renderings that warp the perceived auditory space [6–8].

Various methods have been proposed to obtain an HRTF *adapted* to a given subject, i.e., resulting in a binaural rendering they would perceive with more precision than one based on a random or generic HRTF. This process, hereafter referred to as "HRTF individualization," is discussed in more detail in [9] and Sec. 1.

While the benefits of HRTF individualization for completing simple audio localization tasks have been ascertained [10], it is of interest to assess how they extend to more complex and typical tasks of emerging VR games and applications. Previous research has been conducted on the impact of HRTF individualization on task performance in casual VR applications. [11] studied for example the impact of the audio rendering condition on the "Quality of Experience" as judged by participants after short VR scene explorations. Experiencing the scenes with either stereo, "generic" HRTF, or individualized HRTF audio rendering, participants rated their experience in terms of immersion, naturalness, externalization, etc. HRTF individualization was performed based on an anthropomorphic selection method [12] applied on the CIPIC database [4]. Results did not indicate any impact of the audio rendering condition on the rated attributes.

[13] compared the evolution of participant performance during a virtual audio game, playing with either their own HRTF or a set selected from a database using a tournament-style method [14]. While participants clearly improved in localization accuracy over the seven game sessions (30-min games played over two weeks), the reported results suggest

that, in this context, there was no benefit of using one's own HRTF over a selected best-match HRTF.

To the authors' knowledge, no previous research has investigated the objective and subjective impact of HRTF individualization in the context of a full-fledged audio-visual VR application/game. This study proposes to assess participant performance evolution when using a best-match HRTF compared to a worst-case scenario, where they would be given their "worst-match" HRTF, after e.g., a random selection from an existing database. The selected task takes the form of a VR shooter game, where gameplay heavily relies on players' ability to rapidly locate sound sources in their surrounding environment. While the game was designed as a best-effort to highlight the impact of HRTF individualization in an "out of the lab" VR scenario, it was kept as linear[1] as possible to generate reproducible results and facilitate the investigation of correlations between audio parameters and player behavior. Each element of gameplay, discussed in Sec. 2.2.2, has been carefully weighted to keep the overall experiment as close as possible to a classical localization task [7], allowing for the results and observations to be sufficiently generalizable with regards to the role of individualized HRTF.

The proposed protocol allows for an evaluation of how the HRTF profile quality (best versus worst match) impacts participant performance. Two exposure paradigms are considered: the first where participants always use the same HRTF to study the interaction between profile quality and game performance, the second where HRTFs alternate between best and worst match, to assess short-term participant-wise reactions to HRTF quality variations in this context (e.g., the experience of a player switching between VR games using different HRTFs).

Based on previous studies in the literature on the importance of HRTF quality and HRTF rating ability [15], as well as those on HRTF learning ability [16], the following list of hypotheses are proposed for the current study:

H1 : HRTF profile quality has a positive impact on game performance.

H2 : The impact of HRTF profile quality on game performance is more pronounced during the early stages of the game.

H3 : Maintaining a unique HRTF profile across sessions has a positive impact on game performance improvement in the long run (compared to participants alternating between best and worst-match HRTF).

H4 : The use of a best-match HRTF profile reduces cone-of-confusion–related errors during gameplay.

H5.1 : The impact of HRTF profile quality on game performance is more pronounced for participants who are consistent in their selection of a best and worst-match HRTF.

H5.2 : Self-declared "audio experts" are more sensitive to HRTF profile quality (consistent HRTF selection and improved game performance with best-match HRTF).

H5.3 : Participants consistent in their HRTF selection perceive (consciously) the benefits of the HRTF quality during the game.

The remainder of the manuscript is organized as follows. Sec. 1 presents a short discussion on the HRTF selection method used to determine participants' best and worst-match HRTF. Sec. 2 describes the experimental design: participants grouping, HRTF classification, audio stimuli, and localization task gamification. Sec. 3 reports the experiment results. Sec. 4 discusses these results with regard to existing literature. Finally, Sec. 5 summarizes the paper contribution and concludes on the validity of the study's hypotheses.

1 HRTF SELECTION METHOD

Various methods have been proposed for HRTF individualization [9]: generating transfer functions using morphological measurements [17], selecting HRTF from an existing set based on morphological measurements [18] or subjective ratings [10, 19, 20], tuning existing transfer functions using subjective criteria [21, 22], etc.

[23] proposed a selection method in which participants rate the quality of a simple sound source trajectory relative to a description of the trajectory. Two trajectories, using short noise bursts moving along a fixed distance horizontal or median arc paths, were rendered for a select set of HRTFs. Each of the the binaural renderings for each trajectory are rated.

This method was chosen as it is quite fast ($\approx$10 min), it can easily be integrated in a VR application (no extra sensors) or online profile selection (no head tracking required) [24], and it produces a full HRTF ranking data set that can be used as an assessment of subjective rating stability [15]. Full details of the method employed are provided in Sec. 2.1.

This classification of the set of HRTFs, rather than solely selecting a unique best-match candidate, allows for a performance comparison using both ends of the "HRTF quality scale" (for the given set).

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 34 individuals participated in the experiment (11 females, 23 males, mean age 30.8$\pm$11.4 years).

2.1 Part 1: HRTF Classification by Quality

The subset of available HRTFs was assembled from the LISTEN database, defined from a "perceptually orthogonal" optimized HRTF collection [25]. Per-participant

[1] A game with linear gameplay follows a strict path from which players cannot really deviate regardless of their choices and actions.

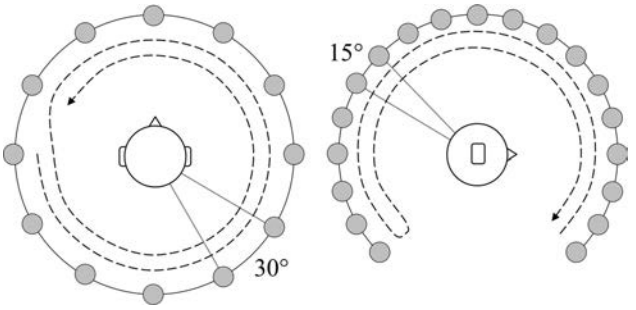

Fig. 1. Trajectory descriptions for HRTF quality ratings: horizontal (left) and median (right) plane trajectories indicating the start/stop position and trajectory direction.

best and worst-match HRTF sets were selected based on the method elaborated in [26], 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°:30°:330°]) and median plane (19 angles [–45°:15°:225°]), as illustrated in Fig. 1.

Each audio trajectory was generated for 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. Following the results of [15], which examined the reliability and repeatability of HRTF judgments by naive and experienced subjects, this rating task was performed 3 times, leading to a total of 6 ratings per subject, counting the two trajectories. An overall judgment rating was taken as the mean of the two trajectory judgments across the three repetitions. Taking the highest and lowest-rated HRTF for each subject results in that subject's perceptually *best* and *worst* HRTF, respectively. None of the participant ratings resulted in a tie between several high or low scores.

Participants completed Part 1 of the study in a listening booth, ambient noise level <30 dBA, using open reference headphones (Sennheiser HD600) connected to an audio interface (RME Fireface UC). Prior to the test, sound levels were calibrated to 80 dBA for all the trajectory (post-HRTF convolution) files, with the headphone placed on a baffled microphone, used as a simple coupler suitable for most types of over-the-ear headphones.

2.2 Part 2: VR Shooter Game

2.2.1 Hardware and Software Architecture

For the game task, participants were equipped with an Oculus CV1 Head Mounted Display (HMD), the CV1 headphones, and a pair of hand tracked devices (Oculus Touch controllers). Evaluations of the HMD by [27] showed sufficient precision and jitter performance (accuracy up to 1 cm and jitter below 0.35 mm) for use of this device in serious gaming studies. [28] showed comparable performance regarding head movement (tracking) to visual scene update latency of 5.8 ± 0.5 ms (stdev).

The game was designed using the Unity v2017.3.0 game engine with modeled assets designed in Blender v2.79. Sounds were spatialized using the Anaglyph binaural audio engine v0.9.1 [29]. The Open Sound Control (OSC) protocol [30] was used for communications between Unity and the Anaglyph engine running as a VST (Virtual Studio Technology) in Cycling'74 Max v7.3. The Anaglyph HRTF update rate, i.e., the time interval between source position update request and new HRTF fade-in completion, was measured at approximately 60 ms for this setup, resulting in an overall audio latency below 70 ms (accounting for tracking latency), judged below the perceptive threshold for the considered scene [31].

Contrary to the trajectory renderings in Part 1, the original Interaural Time Difference (ITD) of the HRTF used in the game was replaced by an individualized ITD, employing an adaptation model based on the participant's head circumference [32]. The same individualized ITD was used for both HRTFs in the case of participant groups using multiple HRTFs. With this individualization, the study compares the two extremes of a "best-effort selection process" scale, improved using a morphological parameter easily accessible to casual VR users, rather than a truly worst possible HRTF of a given database to its counterpart where temporal and spectral cues are poorly matched. This study focuses on the quality of the spectral cues.

Anechoic conditions were employed in the binaural audio engine; no room effect was included to keep the study's focus on HRTF effects. No specific headphone compensation was included, while the different sound effects described below were all designed for playback over the employed headphones. Any such supplemental equalization, as proposed by some studies, being applied globally irrespective of virtual source position, would act simply as an omni-directional source coloration filter and therefore should not affect localization tasks.

2.2.2 Game Description

The game has been designed to closely resemble a classical localization task, so as to facilitate generalization of the results and observations regarding the role of individualized HRTF. The designed gameplay mostly follows the codes of a genre commonly referred to as "survival shooter": players stand against successive waves of enemies, using mid to short-range weapons to defend their position, trying to survive as long as possible as the pace of the game increases. With a full-sphere immersive experience, the ever-increasing rate at which enemies attack players forces them to rely on audio cues for localization as a systematic 360° visual search is too slow a process to survive all but the first few waves.

The game started with each participant being 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. Fig. 2 depicts the game setup and the VR scene.

Enemy targets continually "spawn" one at a time from any of the 29 evenly distributed holes in the structure, flying in straight lines toward the participant until collision, either

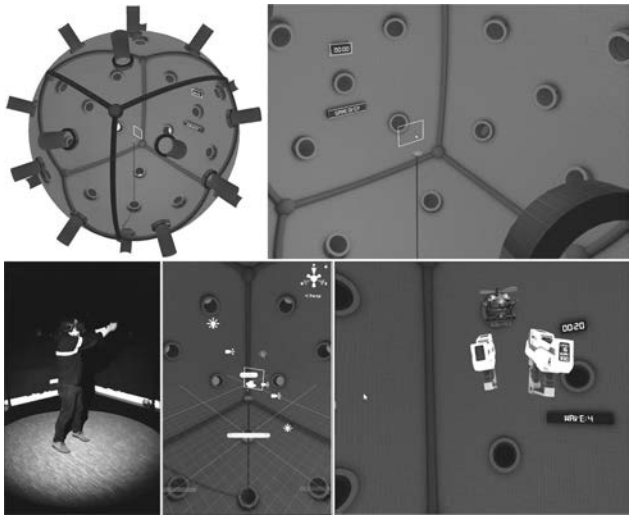

Fig. 2. Game scene overview: (upper left) overall view of the virtual environment, (upper right) focus on the platform atop which participants stand during the game, (lower left) participant in the VR room, (lower middle) virtual environment during gameplay, and (lower right) in-game screenshot.

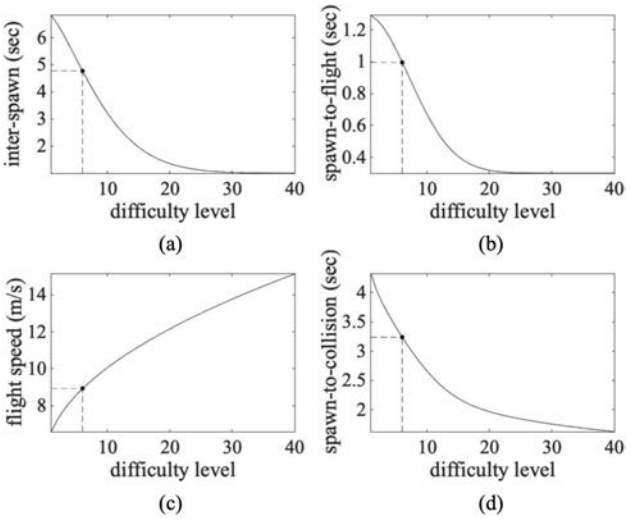

Fig. 3. Game dynamics as a function of game level. (a) Inter spawn interval, the time interval between two target spawn. (b) Spawn to flight interval, the time interval between a target spawn and its flight toward the player. (c) Flight speed, the speed at which a target moved towards the player. (d) Spawn to collision interval, the time interval between a target spawn and its arrival at the center of the sphere, i.e., its collision with the player (= spawn to flight interval + sphere radius / flight speed). (•) indicates initial parameter value (level 6).

with a bullet or the participant. Participants were instructed to shoot at the incoming targets using a pair of hand-held blasters, the avatar representations of the hand tracked devices in the virtual scene. The stated goal for the game was to destroy as many as possible as fast as possible in the given time limit. A video extract showing a game session is available.[2,3]

Game dynamics/difficulty evolved based on three parameters affecting target/enemy behavior: flight speed, inter-spawn interval, and spawn-to-flight interval. The latter defines a small delay between a target's spawn event and the beginning of its flight toward the participant. During this interval, the target remains fixed. Parameter values are a function of the current game difficulty level, increasing one level for every three consecutive kills and decreasing one level for every two consecutive fails (target collides with the participant). Fig. 3 illustrates the relation between game level and these parameters.

Targets emitted specific event-based sounds for spawning, launching, flight, and collision. Sound design emphasized "localizable" signals, with attention paid to rate of attack, spectral content, and spectral masking [33]. Fig. 4 shows the spectral content of the different sounds (prior to binaural filtering) and their temporal envelopes.

A short training session (≈3 min) introduced the controls and difficulty level mechanism, highlighting that overall game dynamics increased as the game progressed and participant skill improved. Participants then played a series of six sessions of the 5-min game. To avoid fatigue for this rather demanding task (see video extract), participants played sessions S1–2 directly following the HRTF classification task, followed by a week pause interval, then sessions S3–6.

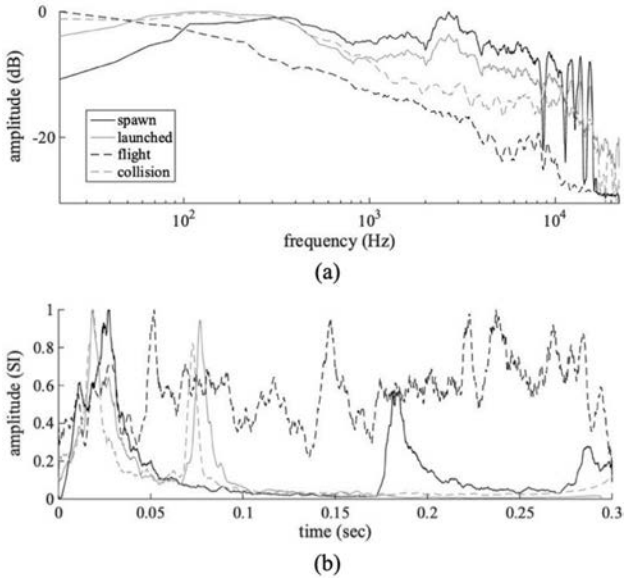

Fig. 4. Normalized (a) frequency spectrum and (b) temporal envelope (first 300 ms) of the four game sounds emitted by the target.

2.2.3 Participants Grouping

Four groups were constructed to test the hypothesis concerning various HRTF conditions: two groups were constructed to compare the overall impact of HRTF quality on game performance and two groups were constructed to

[2] youtu.be/q6muds1qW-w.
[3] www.youtube.com/c/LAMSorbonneUniversite.

Table 1. Participant group definitions.

Name	Description	Num. of part.
BB	always used best-match HRTF	7
WW	always used worst-match HRTF	7
BW	alternating, started with best-match	10
WB	alternating, started with worst match	10

assess the short-term impact of HRTF quality on a per-participant basis. The first two groups were composed of seven participants each, playing solely with either their identified best or worst-match HRTFs: groups **BB** and **WW**, respectively. Ten participants comprised each of the last two groups, alternating between their identified best and worst-match HRTF between sessions. Best and worst-match presentation order was evenly balanced, resulting in two groups, **BW** and **WB**, identifying the initial HRTF pair tested.[4] Table 1 summarizes participant grouping.

2.2.4 Questionnaire

After sessions S1–S2, participants were asked to assess whether they perceived any difference in the audio rendering between the two sessions. They rated S1 versus S2 audio renderings according to which was more "natural," "coherent with visuals," and "efficient" (regarding the target localization task) on a five-point scale. Participants also rated their prior "video game," "virtual reality," and "spatial audio" expertise on a five-point scale.

3 RESULTS

3.1 Statistical Analysis Tools

Statistical significance between game metric distributions was assessed using a Wilcoxon non-parametric test (p-value threshold of 0.05), as all compared distributions proved to follow non-normal (skewed) distributions, assessed using a Lilliefors test (calculated using the `lillietest` function). The signed rank version of the test (respectively rank-sum test) was used for paired/dependent (resp. unpaired/independent) samples comparison. Inter-variable dependence was assessed using linear correlation (p-value threshold of 0.05). Statistical significance and correlation estimation were calculated using Matlab `signrank`, `ranksum`, and `corrcoef` functions, respectively. The notation $p < \varepsilon$ is adopted to indicate p-values below 10^{-3}. Significant differences between fits pertaining to the analysis of learning rate across sessions in Sec. 3.4.3 were assessed based on the comparison of their coefficients (exponential fit parameters). Significant difference is discussed when at least one of the coefficients of two fits differ beyond 50% of their estimate's 95% Confidence Interval (CI) [36].

[4]Preliminary results for alternating HRTF conditions (groups **WB** and **BW** only) with 30 participants for only sessions S1–2 were presented in [34]. Extended preliminary results for groups **WB** and **BW** to sessions S1–6 with 20 participants were presented in [35].

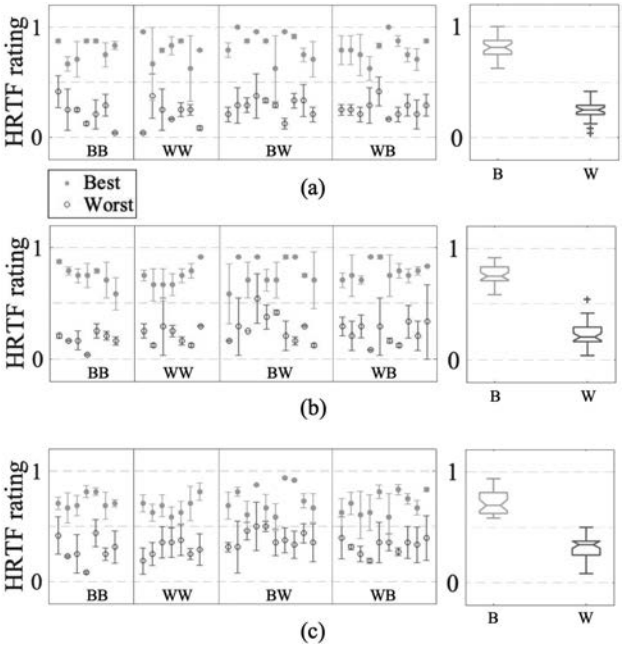

Fig. 5. Scores of the HRTF classification task for all participants for (a) horizontal and (b) median trajectories. The reported rating values correspond to the average normalized rank given by each participant to their later determined best and worst-match HRTF. A value of 0 (resp. 1) indicates that the HRTF was consistently rated as the least (resp. most) representative of the described trajectory across the 3 rating repetitions. (c) Combined trajectory mean score results for the *determined* best and worst-match HRTF. Error bars indicate the variance of participant ratings for best and worst HRTF. Right-hand plots illustrate overall participant ratings. Groups are separated by vertical lines; participants are otherwise unsorted.

Likewise, a significant difference between a pair of ordinate scales (HRTF ratings, questionnaire ratings, etc.) was established when they differ beyond 50% of their estimate's 95% CI.

3.2 HRTF Classification Results

Results of the HRTF ratings of Part 1 are summarized in Fig. 5, focusing on scores corresponding to each participant's best and worst HRTF match for both trajectories. Participants are consistent in their classification with regards to these extrema as previously observed [26, 16, 37]. Most participants were able to clearly distinguish between best and worst-match for both horizontal and median trajectories.

As audio sources in Part 2 of the experiment were to arrive from all directions, an average best and worst-match HRTF across the two trajectories was calculated for each participant. The rating statistics for selected best and worst HRTF are shown in Fig. 5(c). As observed in [26], participant HRTF ratings for the horizontal and median trajectories were not correlated. This explains the reduced separation of mean rating values for the combined trajectory results [Fig. 5(c)] as compared to the individual horizontal and median plane mean trajectory ratings [resp. Fig. 5(a) and Fig. 5(b)]. For almost all participants, average-best and

Table 2. Independent and dependent variables of the VR game experimental protocol. The 29 spawn positions have been grouped into 6 "elevation regions."

Independent variables	
participant ID	random variable
HRTF quality	best, worst
group ID (HRTF pres. order)	**BB**, **WW**, **BW**, **WB**
session ID	S1, S2, . . ., S6
spawn elevation region	R1, R2, . . ., R6
Dependent variables	
mean game level	session-wise
spawn-spot reaction time	event-wise
spawn-spot traveled angular distance	event-wise

average-worst HRTF scores remained sufficiently distinct to distinguish both populations.

3.3 VR Shooter Game Results Analysis Preamble

Result analysis is subdivided into session-wise and event-wise analysis. Session-wise analysis concerns participant mean results across sessions. Event-wise analysis decomposes each participant session into events, with an event being defined as a target {spawn, launch, flight, collision} sequence. The following analysis concerns $34 \times 6 = 204$ sessions for a total of 24,764 events: average of 121.4 ± 18.8 events per session, as game dynamics varied with game level and hence participant. Session-wise analysis is based on paired-samples comparison (paired by participant ID) and event-wise based on unpaired-samples comparison due to the uneven number of per-session events for any given participant.

3.3.1 Metrics Definition

Performance assessment is based on three metrics, calculated from sessions logs: mean game level (see Sec. 2.2 for description of game mechanics), spawn-spot reaction time, and spawn-spot traveled angular distance. Table 2 summarizes the independent and dependent variables of the VR game experimental protocol.

Participant's *mean game level* represents their overall performance during a session of the game. *Spawn-spot reaction time* corresponds to the time interval between the spawn of a target and its entering the visual Field Of View (FOV) of the participant, defined as a 50° cone centered around the current forward view axis of the HMD. The 50° value represents the HMD's FOV [27]. The event of 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 (the task of visual *targeting* being judged as independent of the acoustic rendering quality). The associated *spawn-spot traveled angular distance* corresponds to the angular distance traversed by the participant's head during that time. Compared to reaction time, traveled angular distance represents movement efficiency and serves to differentiate between participants using binaural cues to localize targets and those randomly looking around [38].

Targets spawned within a participant's FOV are discarded from the analysis. For targets that never enter participant FOV (colliding with the participant or a stray bullet), the spawn-to-collision time and angular distance were used. Targets shot without ever entering participants FOV represented less than 3% of the total number of targets spawned.

3.3.2 Data Normalization

Three different types of normalization were used in the analysis: per-group, per-participant, and per-event normalization. *Per-group* normalization is a centering of all participants' results around their group mean:

$$x^i_{G\text{norm}} = x^i - \text{mean}(\mathbf{x}_G),\ \mathbf{x}_G \text{ values in } G \tag{1}$$

where x^i represents the i^{th} participant's raw metric and $x^i_{G\text{norm}}$ its normalized counterpart. $\mathbf{x}_G$ is the ensemble of values of x for group G. *Per-participant* normalization uses Eq. (1), replacing $\mathbf{x}_G$ by the ensemble of values of x for participant i. *Per-event* normalization was only applied to angular distance. The normalized angular distance is a ratio between the distance traversed by a participant's head during the event and the minimum angular distance "required" for this event, expressed for the i^{th} event as:

$$\theta^i_{E\text{norm}} = \int_{t=spawn}^{t=collision} \frac{\delta\theta_t}{\Delta\theta^i_{(\overrightarrow{head},\overrightarrow{target})}} \tag{2}$$

where $\delta\theta_t$ represents the raw traversed angular distance for a time step. $\Delta\theta^i$ is the i^{th} angle between the forward head orientation upon spawn and the target spawn position. The resulting metric, in $[1, \infty[$, indicates event-wise participant efficiency; the smaller the more efficient.

3.4 VR Shooter Game Results

3.4.1 Overall Game Statistics

All groups combined, the mean game level significantly increased between sessions, reaching a plateau from S5 on, from 14.7 for S1 to 18.7 for S5–6 ($p < \varepsilon$ but for $p_{S4-S5} = 0.012$). Spawn-spot angular distance and response time likewise decreased across sessions, from 145° and 1.35 s for S1 to 126° and 1.26 s for S6. This decrease is significant for both metrics between S1 and S2 (145° versus 134° and 1.35 s versus 1.30 s, $p < \varepsilon$), and for angular distance between S3 and S4 (134° versus 129°, $p = 0.01$).

Table 3 reports the non-normalized results of each group across all sessions. Overall, group **BW** performance was significantly below that of the other groups for all metrics listed in Table 2 ($p < \varepsilon$ for all pair-wise metric comparison between **BW** and the other groups). As a direct result of the difference in mean level, group **BW** participants faced, on average, fewer targets during their sessions. The evolution of the four groups' mean game level across the six sessions is illustrated in Fig. 6. *Per-participant* normalization was applied to allow for a relative comparison of group results, independent of inequality in participant performance distribution.

Table 3. Summary of mean performance metrics across groups. Mean game level is averaged over sessions, reaction time, and angular distance over events and sessions. Group **BW** performance is significantly below that of the other groups ($p < \varepsilon$ for all pair-wise metric comparison between **BW** and the other groups).

	BB	WW	BW	WB
mean game level	17.5	17.9	16.3	17.7
reaction time (sec)	1.28	1.27	1.34	1.26
angular distance (deg)	129	131	136	131
inter spawn interval (sec)	1.7	1.6	1.8	1.7
spawn-collision interval (sec)	2.1	2.1	2.1	2.1
targets per subject	728	748	702	741

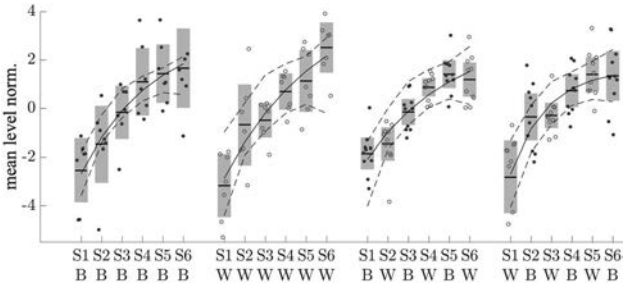

Fig. 6. Mean value, confidence interval (95%), and scattered representation of mean game level normalized (per participant) for the four groups. **B** and **W** indicate best and worst HRTF, respectively. Black continuous and dotted lines represent per-group nonlinear regression fit and fit prediction 95% CI, respectively, based on the exponential decay form of Table 5, computed with Matlab `nlinfit` and `nlpredci` functions.

As can be inferred from the results reported in Table 3, there is a certain interdependence between the experimental metrics (defined in Table 2). To assess whether this interdependence was low enough to justify a separate analysis, or if any two metrics simply represented the exact same information and should therefore be aggregated, paired correlation values were calculated, compounding results across groups and sessions:

- mean game level versus angular distance (mean per session): $r = -0.30$ ($p < \varepsilon$),
- mean game level versus reaction time (mean per session): $r = -0.70$ ($p < \varepsilon$),
- angular distance versus reaction time: $r = 0.79$ ($p < \varepsilon$).

While clearly suggesting a certain degree of interdependence, none of the above correlation coefficients are high enough to justify metric aggregation at this point.

3.4.2 Overall Impact of the HRTF Condition

No significant effect was observed for the HRTF condition on participant performance when comparing the aggregated results of the four groups. Direct comparison between groups **BB** and **WW**'s performances showed no overall impact of the HRTF condition, nor did a comparison between groups **WB** and **BW**'s aggregated best versus worst HRTF sessions.

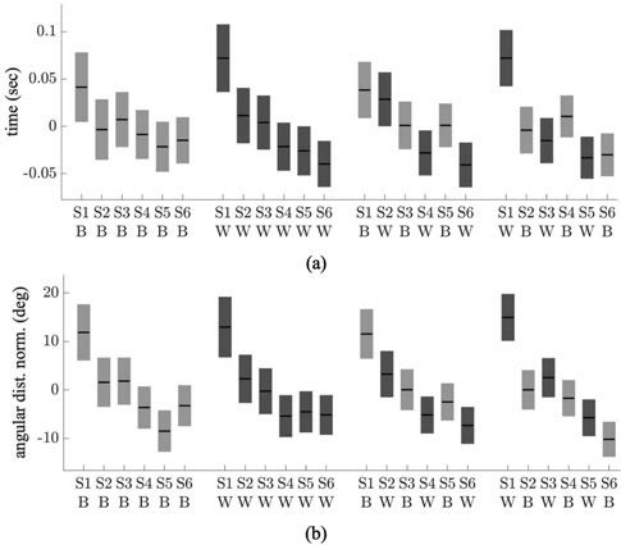

Fig. 7. Aggregated mean and 95% CI of (a) normalized (per-group) reaction time, and (b) normalized (per-group) angular distance across sessions and HRTF conditions for the four groups.

Table 4. Event-wise participant performance comparison between the first two sessions. [*] indicates significant difference with $p < \varepsilon$ between S1 and S2 results.

	Mean level		React. time (s)		Ang. dist. (°/Δ°)	
	S1	S2	S1	S2	S1	S2
group **BB**	15.0	16.0	1.32	1.28	1.41	1.30
group **WW**	14.7	17.2	1.34	1.28	1.46	1.33
group **BW**	14.4	14.8	1.38	1.37	1.49	1.40
group **WB**	14.9	17.3	**1.33***	**1.25***	**1.47***	**1.30***

Analysis focused on Groups **WB** and **BW** suggests that HRTF presentation order had an impact on participant performance. Group **WB**'s best HRTF sessions were significantly better than worst-HRTF sessions regarding mean level (17.1 for best versus 18.3 for worst, $p = 0.018$) and angular distance (134° versus 127°, $p < \varepsilon$). Conversely, Group **BW**'s worst-HRTF sessions were significantly better than best-HRTF sessions regarding angular distance (133° for best versus 139° for worst, $p = 0.003$) and reaction time (1.32 s versus 1.35 s, $p = 0.012$). These results indicate a possible influence of HRTF presentation order.

3.4.3 HRTF Condition × Session Interaction

The effect of HRTF presentation order observed in the previous section was limited to the early stage of the game. No significant effect of the HRTF condition was observed on Group **WB**'s performance when considering S3–6, the same being true for Group **BW** when considering S5–6.

Fig. 7 illustrates the evolution of participant angular distance and reaction time across sessions for each group. No significant effect of the HRTF condition was observed on participant performance for S1, i.e., between aggregated groups **BB** + **BW** and **WW** + **WB**. As reported in Table 4, Group **WB** is the only group for which performance signif-

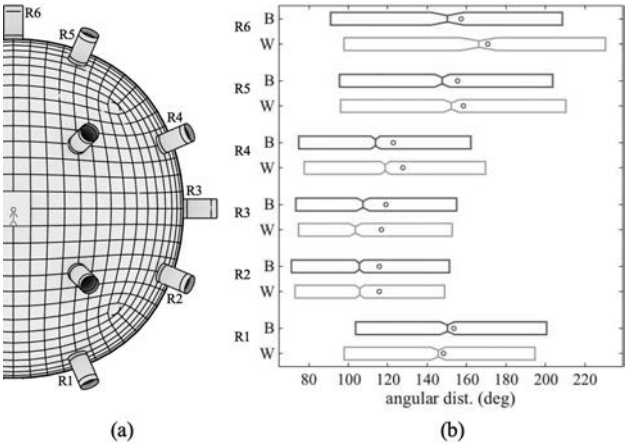

Fig. 8. (a) Definition of R_i spawn elevation regions, distributed on the sphere surrounding the participant during the game. (b) Angular distance traversed across spawn region for aggregated Groups **WB** and **BW**, separating best and worst-match HRTF. Boxplot represents cluster median and 95% CI, black circle its mean.

Table 6. Effect of target spawn elevation region on aggregated group performance. Fig. 8 regions are clustered and sorted based on increasing "difficulty." [*] indicates significant difference from neighboring region aggregates, all with $p < \varepsilon$ but for R1 versus R5–6 angular distance normalized (per event) where $p = 0.031$.

	R2–3	R4	R1	R5–6
reaction time (s)	**1.14***	**1.24***	**1.39***	**1.57***
angular distance norm (°/Δ°)	**1.07***	**1.22***	**1.64***	**1.67***

icantly improved between S1 and S2, switching from worst to best HRTF.

To assess the impact of HRTF *condition* on participant improvement over time, a nonlinear regression fit on group **BB** versus **WW**'s metrics evolution across sessions was performed, against the exponential decay form suggested in [39]. A similar regression was performed on the aggregated group **BB** + **WW** versus **BW** + **WB** to assess the impact of HRTF *consistency* on performance improvement over time. No significant difference was observed between the resulting coefficients, reported in Table 5. Regression curves are plotted against participant performance in Fig. 6.

3.4.4 Impact of Spawn Origin Position

Participant performance analysis suggested that they had difficulty locating targets spawned in the lower and upper regions of the sphere surrounding them, i.e., R1, R5, and R6 *spawn elevation regions* in Fig. 8. Among these regions, events related to R5–6 spawns (upper regions) resulted in significantly higher angular distance traversed and reaction times than R1 (lower region). Among the "easier" mid-regions, targets spawned from R2–3 (mid-center and mid-low) took significantly less time and necessitated less traversed angular distance than those spawned from R4 (mid-up). These results are summarized in Table 6.

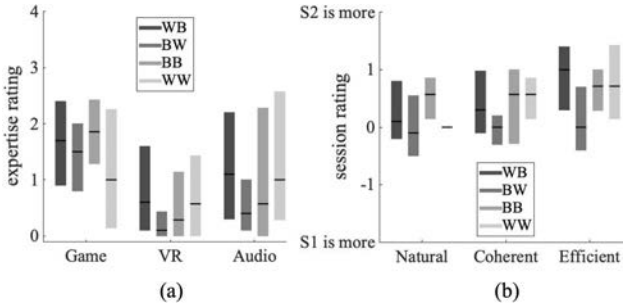

Fig. 9. Mean and 95% CI of (a) participant ratings on their video game, VR, and audio expertise on a [1:5] scale (higher values for higher expertise), and (b) participants' preference of S1 versus S2 audio rendering in terms of naturalness, coherence, etc. on a [−2:2] scale (0 indicates no preference).

3.4.5 HRTF Condition × Spawn Origin Position Interaction

No significant difference was observed between groups **BB** and **WW**'s performances regarding target spawn position. Overall, group **BW** and **WB** participants were more efficient at locating targets spawned from R6 when using their best HRTF. As shown in Fig. 8, R6 traversed angular distance significantly increases for both groups, comparing aggregated best versus aggregated worst HRTF sessions, from 158° to 176° ($p = 0.009$). On average, participants from these two groups only managed to locate and destroy half of the targets spawned from this region (54% with best versus 44% with worst-match HRTF out of ≈280 spawns). This result only holds for the early stage of the game and was no longer observed for S5–6.

3.5 Correlation Between HRTF Classification Results and Game Performances

An analysis was conducted to assess if consistency in rating HRTFs during the classification task was correlated to sensitivity to the HRTF condition during the game. Groups **BW** and **WB**'s *relative* performance, subtracting the results of worst from best HRTF sessions, was compared against HRTF classification scores, reported in Fig. 5. There was no clear correlation between relative performance and consistency in selecting a best or worst-match HRTF. The highest correlation of this analysis was between consistency (variance) for median trajectory ratings and relative efficiency for locating targets during the game (per-event normalized angular distance), with $r = 0.40$ ($p = 0.077$).

3.6 Questionnaire

3.6.1 Metric Correlation

Participant self-rated expertise and self-evaluation of S1 versus S2 audio renderings are reported in Fig. 9. A significant correlation was observed between the 3 attributes proposed to rate subjective preference of S1 and S2 [Fig. 9(b)]:

Regarding expertise ratings, there was a significant correlation between expertise with VR and with spatial audio ($r = 0.76$, $p < \varepsilon$).

Table 5. Coefficients of the nonlinear regression fit of group performance evolution across sessions, against the exponential decay form $y = y_0 e^{-t/\tau} + c$ proposed in [39]. As in Fig. 6, per-participant normalization was applied on each metric to allow for relative comparisons between group performance. ± value is the 95% CI, MSE is the estimate mean square error of the variance of the error term. Regression, CI, and MSE have been computed with Matlab `nlinfit` and `nlparci` functions.

	Relative acute performance, y_0		Improvement time constant, τ		Long-term performance, c		MSE	
	BB	**WW**	**BB**	**WW**	**BB**	**WW**	**BB**	**WW**
mean game level	−7.9±2.5	−9.5±10.4	3.5±4.4	4.7±13.4	20.7±3.9	22.7±13.5	0.09	0.33
reaction time	0.2±0.2	0.2±0.1	1.2±2.1	1.7±1.5	1.3±0.0	1.2±0.0	ε	ε
angular distance	36.2±38.8	41.4±23.0	1.5±2.8	1.3±1.0	123.0±10.7	125.6±4.0	8.7	1.8
	BW	**WB**	**BW**	**WB**	**BW**	**WB**	**BW**	**WB**
mean game level	−6.6±6.1	−7.6±4.6	4.2±11.7	1.8±2.3	19.4±8.5	19.3±1.9	0.20	0.19
reaction time	0.2±0.6	0.6±2.4	5.9±44.0	0.6±1.3	1.2±0.7	1.2±0.0	ε	ε
angular distance	34.6±20.1	36.8±25.7	1.9±2.7	2.9±8.8	128.5±10.6	118.6±37.9	4.6	15.3

naturalness × coherence	$r = 0.30, p < \varepsilon$
naturalness × efficiency	$r = 0.56, p < \varepsilon$
coherence × efficiency	$r = 0.60, p < \varepsilon$

3.6.2 Expertise Distribution and Interactions

No significant differences were observed between participant expertise distributions between the 4 groups. An analysis was conducted on expertise ratings to assess if there was a correlation between expertise and performance during the game or consistency in HRTF classification. There was no clear correlation between expertise ratings and game performance regarding mean game level, reaction time, or angular distance traversed. For the HRTF classification task, a significant correlation was observed between median trajectory rating consistency (variance) and VR ($r = -0.53$, $p < \varepsilon$) and spatial audio ($r = -0.42$, $p = 0.014$) expertise.

3.6.3 Perceived Difference in Audio Rendering Between Sessions

As seen in Fig. 9(b), ratings suggest a slight preference of S2 over S1 for all 3 attributes: aggregated means of 0.12±0.68, 0.32±0.67, and 0.59±0.91 for natural, coherent, and efficient resp. **BW** was the only group not to rate S2 audio rendering as more efficient than S1, indicating a noticeable degradation in quality when changing from best to worst HRTF. A similar trend, though less pronounced, can be observed on **BW** ratings of S1 versus S2 naturalness and coherency.

Spatial audio and VR experts' ratings of S1 versus S2 did not differ from other participants, not showing any preference for best-match HRTF session. Similarly, no clear correlation was observed between consistency in the HRTF classification task and S1 versus S2 ratings.

3.7 Interviews

During the brief interviews that followed S1–2, most participants acknowledged that their attention was more focused on game dynamics than audio rendering during the first session, S1. They reported starting to really pay attention to subtle audio cues during S2, aware that they were not efficient in differentiating targets spawned from the lower and upper regions (R1 and R6 in Fig. 8). To compensate, a few participants mentioned the use of strategies based on micro head-movement (boxer's stance–like motion, head moving side to side) upon target spawns to improve location ability. Most were not familiar with this technique as reported in binaural rendering literature [40].

4 DISCUSSION

Overall, results indicated that the benefits of HRTF individualization were limited to the early stage of the game. As can be expected, the game learning effect that takes place at this point interferes with the analysis of the impact of the HRTF condition. As suggested from subjective appreciation of S1 versus S2 audio renderings, corroborated by the interviews, participants were in a different mindset between these two sessions, more focused on understanding the game's mechanics during S1 and the audio rendering during S2. The impact of the HRTF presentation order observed on responses to the HRTF condition during the game is likely a direct result of that mindset. The difference in performance observed in Table 4 between S1 and S2 for **BW** and **WB** could then be interpreted as the result of an interaction between the HRTF condition and the game learning effect, where switching from best to worst-match HRTF negates the benefit that should result from training while, in contrast, training and HRTF quality combine when switching from worst to best match. Fig. 10 further examines the interpretation, summarizing **BW** versus **WB** performance analysis. The clustering of participant performance, based on HRTF presentation order (highlighted by enveloping ellipses), further suggests that the HRTF presented in session S2 was more likely to be "efficient" than the one presented during S1.

Given the studied task, half-way between a game and an audio localization task, it was not clear if the HRTF condition would impact performance evolution as in a classic learning test [16]. The similarity between Groups **BB** and

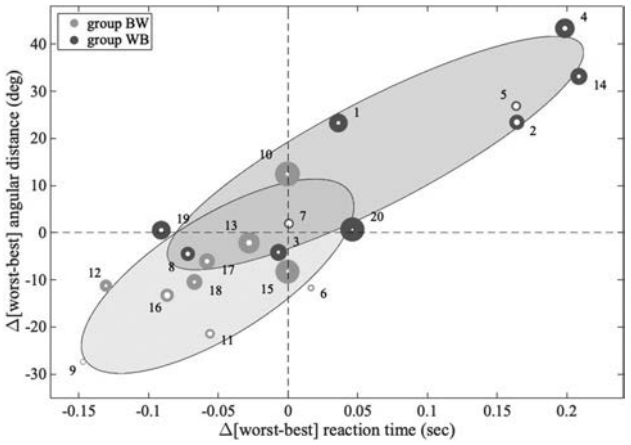

Fig. 10. Cluster analysis of groups **BW** and **WB** as a function of the difference in event-wise performance using best and worst-match HRTF for combined spawn elevation regions R1 and R5–6. Each point represents the performance of a given participant, as a function of differences in mean response time and angular distance between best and worst-match HRTF sessions, respectively. Points in the top-right section (resp. bottom left) represent participants who performed better with their best- (resp. worst) match HRTF. To investigate any interaction between Part 2 performance results and Part 1 HRTF ranking, the radius of each circle is proportional to the difference between best and worst-match HRTF scores obtained during the rating task. As an indicator of rating repeatability, the radius of concentric white circles is proportional to the mean variance of best and worst-match HRTF ratings (see Fig. 5). Numbers indicate participant game performance rank, based on the highest mean level across all sessions (groups **BW** and **WB** only). Coloured patches are ellipse fits around **BW** and **WB** clusters using a least squares criterion.

WW's performance evolution (see Sec. 3.4.3) suggests that the game's mechanics interfered with such straightforward observation, much as the results reported in [13] comparing the benefits of own measured HRTF against an individualized best-match HRTF on participant performance during an audio game built around a localization task. A similar observation can be made regarding HRTF stability and the observation that Groups **BB** and **WW** did not benefit from a consistent use of a given HRTF compared to Groups **BW** and **WB**, who alternated HRTFs. Additionally, Groups **BW** and **WB** did not perceive the change in HRTF between S1 and S2, a result similar to that reported in [11]. Overall, these results do not provide a strong argument in favor of the need for HRTF spectral cue individualization for the given task at hand (i.e., an immersive audio-visual interaction game in which audio rapidly steers attention outside of FOV events).

Complementing these considerations, the absence of an observed effect of HRTF individualization on the very first session raises questions with regards to the HRTF selection method. It should therefore be noted that studies on HRTF learning employing the same method have shown a significant and positive effect on performance during a simple auditory localization task with the best-match HRTF compared to the worst-match HRTF [23, 16]. Given the rough localization task on which the present game is based, addressing directional attention [38], it is likely that performance was more impacted by "up/down" and "front/back" confusions [23] than by fine localization errors, even considering the minor benefits of a best-match HRTF on the R6 region targets reported in Fig. 8. Further work is needed to assess if an individualization method specifically focusing on reducing these confusions would improve game performance.

A closer look at individual performance distributions in Fig. 10 suggests that some participants were more sensitive to the HRTF condition. In an attempt to derive a metric to judge whether a participant would benefit from an HRTF individualization beforehand (i.e., based on HRTF classification results only), a post-hoc analysis was conducted on individual performance. The analysis revealed that out of 20 participants, HRTF quality proved to have a significant impact for four participants regarding the angular distance metric during the whole game. Three out of these four performed best with their best-match HRTF, one with their worst match; all of them performed reasonably well during the game (ranked #$\{1, 4, 14\}$ and $\{6\}$ resp. out of 20, based on overall mean level). Such results reflect similar findings on the individual nature of HRTF learning, being limited to a subset of participants [16]. Finally, no clear correlation was observed between expertise, consistency at rating HRTFs (see colored circles in Fig. 10), or subjective appreciation of best versus worst-match HRTF sessions.

In light of these results, the following conclusion can be made on the hypotheses of this study. Overall, HRTF quality had minimum impact on game performance (**H1**) except for the minor reduction in up-down confusions observed in Sec. 3.4.5. Therefore, **H1** is only supported in some instances. This observation on up-down confusions reduction, while limited to participants alternating between HRTFs, directly supports **H4**. Most results show that any potential impact the HRTF quality had was limited to the early stage of the game (supporting **H2**). Neither the regression slopes of Sec. 3.4.3 nor any between-session analysis indicated that alternating between HRTF degraded long-term performance evolution (refuting **H3**). There was no evidence of a correlation between ability at HRTF classification and sensitivity to HRTF quality during the game (refuting **H5.1**). Likewise, results did not suggest that self-declared audio experts could further benefit from HRTF quality during the game nor outperform other participants regarding HRTF classification consistency (refuting **H5.2**). Finally, no correlation could be established between participants' perception and actual performance when using a best-match HRTF, i.e., between those noticing an improvement and those benefiting from it (refuting **H5.3**).

5 CONCLUSION

This paper presented the results of an experiment designed to assess the impact of individualized binaural rendering on player performance in the context of a localization task transformed via gamification into a multimodal interactive VR "shooter game." During the game, participants had to quickly locate and shoot at successive targets ap-

proaching from random directions in a sphere. Designed around a classical localization task, the objective of this game was to both create an engaging experience, resembling the average VR shooter game while accentuating the impact 3D audio could have on player achievement, and produce performance metrics that would give a generalizable insight on the impact of HRTF quality in this context.

Participants performed six game sessions. Two groups used either only their best or worst-match HRTF. Two additional groups alternately used their best and worst-match HRTF across sessions. Best and worst-match HRTFs were determined based on participant ratings of a subset of perceptually orthogonal HRTFs [25] taken from the LISTEN database [5].

Results indicated that the use of a best-match HRTF did not improve overall participant performance with regards to the time needed to localize targets, nor the angular distance they traveled in doing so. No significant difference was observed in the overall learning rate between participants alternating between best and worst-match HRTF and those using a unique HRTF. An interaction was observed between game learning effect and HRTF presentation order for participants alternating between best and worst-match HRTF, leading some participants to "favor" the HRTF used during the second session of the game, being allegedly more focused on the audio rendering at that point.

Detailed analysis focusing on source origin regions revealed that angular distance traversed significantly decreased when participants from groups alternating HRTF quality used their best-match HRTF for the most elevated region. Targets spawned from this region were also more often spotted before collision by participants using their best-match HRTF (54% versus 44%), indicating a benefit of individualized HRTFs when elevated source positions are considered. This benefit of the best HRTF was limited to the early stage of the game (no lasting effect when considering the last two sessions) and was not observed when comparing the performance of participants using a unique HRTF. This last result is likely due to an interaction between game learning and HRTF learning, where participants are no longer subject to HRTF quality (HRTF learning) when they are familiar enough with the game mechanics (game learning) for said quality to make a difference.

Overall, the results of this study suggest that HRTF quality had minimal impact on general in-game participant performance and improvement rate. It is noted that no quantification of the similarity between the selected best-match HRTF as compared to the participant's own HRTF was possible. Despite that, coupled with the perceptually orthogonal set of HRTFs used [25], the authors believe that the selection process employed fairly represents what can be achieved in terms of a perceptual HRTF selection procedure from a database in a reasonable amount of time. Further work is needed to assess if an alternative HRTF quality evaluation and selection method, focused specifically on minimizing front-back and up-down confusions, could prove beneficial for task performance during a full-fledged VR situation when reliable quadrant-wise localization is required before precise auditory localization.

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

REFERENCES

[1] J. Blauert, *Spatial Hearing: The Psychophysics of Human Sound Localization* (MIT Press, 1997).

[2] C. I. Cheng and G. H. Wakefield, "Introduction to Head-Related Transfer Functions (HRTFs): Representations of HRTFs in Time, Frequency, and Space," presented at the *107th Convention of the Audio Engineering Society* (1999 Sep.), convention paper 5026.

[3] T. Carpentier, H. Bahu, M. Noisternig, and O. Warusfel, "Measurement of a Head-Related Transfer Function Database With High Spatial Resolution," *Forum Acust.*, pp. 1–6 (2014).

[4] V. R. Algazi, R. O. Duda, D. M. Thompson, and C. Avendano, "The CIPIC HRTF Database," *App. Sig. Proc. Audio Acoust.*, pp. 99–102 (2001).

[5] O. Warusfel, "Listen HRTF Database" (2003), accessed: 2019-01-25, url: http://hrtf.ircam.fr/

[6] D. R. Begault, E. M. Wenzel, and M. R. Anderson, "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. Audio Eng. Soc.*, vol. 49, no. 10, pp. 904–916 (2001).

[7] E. M. Wenzel, M. Arruda, D. J. Kistler, and F. L. Wightman, "Localization Using Nonindividualized Head-Related Transfer Functions," *J. Acoust. Soc. Am.*, vol. 94, no. 1, pp. 111–123 (1993), doi:10.1121/1.407089.

[8] R. Mehra, A. Nicholls, D. Begault, and M. Zannoli, "Comparison of Localization Performance With Individualized and Non-individualized Head-Related Transfer Functions for Dynamic Listeners," *J. Acoust. Soc. Am.*, vol. 140, no. 4, pp. 2956–2957 (2016), doi:10.1121/1.4969129.

[9] S. Xu, Z. Li, and G. Salvendy, "Individualization of Head-Related Transfer Function for Three-Dimensional Virtual Auditory Display: A Review," *Intl. Conf. Virtual Reality*, pp. 397–407 (2007), doi:10.1007/978-3-540-73335-5_44.

[10] B. U. Seeber and H. Fastl, "Subjective Selection of Non-individual Head-Related Transfer Functions," *Intl. Conf. Audit. Disp.*, pp. 259–262 (2003).

[11] M. Geronazzo, E. Sikström, J. Kleimola, F. Avanzini, A. De Götzen, and S. Serafin, "The Impact of an Accurate Vertical Localization With HRTFs on Short Explorations of Immersive Virtual Reality Scenarios," *Intl. Symp. Mixed Augment. Reality*, pp. 90–97 (2018).

[12] M. Geronazzo, E. Peruch, F. Prandoni, and F. Avanzini, "Improving Elevation Perception With a Tool for Image-Guided Head-Related Transfer Function Selection," *Proc. Intl. Conf. Digital Audio Eff.*, pp. 397–404 (2017).

[13] A. Honda, H. Shibata, J. Gyoba, K. Saitou, Y. Iwaya, and Y. Suzuki, "Transfer Effects on Sound Localization

Performances From Playing a Virtual Three-Dimensional Auditory Game," *Appl. Acoust.*, vol. 68, no. 8, pp. 885–896 (2007).

[14] K. Saito, Y. Iwaya, and Y. Suzuki, "Sound Localization With Individualized HRTFs Selected by Tournament Matches," *Forum Inf. Technol.*, pp. 381–383 (2005).

[15] A. Andreopoulou and B. Katz, "Investigation on Subjective HRTF Rating Repeatability," presented at the *140th Convention of the Audio Engineering Society* (2016 May), convention paper 9597.

[16] P. Stitt, L. Picinali, and B. F. Katz, "Auditory Accommodation to Poorly Matched Non-individual Spectral Localization Cues Through Active Learning," *Sci. Rep.*, vol. 9, pp. 1–14 (2019), doi:10.1038/s41598-018-37873-0.

[17] R. O. Duda, V. R. Algazi, and D. M. Thompson, "The Use of Head-and-Torso Models for Improved Spatial Sound Synthesis," presented at the *113th Convention of the Audio Engineering Society* (2002 Oct.), convention paper 5712.

[18] D. Schönstein and B. F. Katz, "HRTF Selection for Binaural Synthesis From a Database Using Morphological Parameters," presented at the *International Congress on Acoustics* (2010).

[19] D. Zotkin, J. Hwang, R. Duraiswaini, and L. S. Davis, "HRTF Personalization Using Anthropometric Measurements," *Workshop Appl. Sig. Proc. Audio Acoust.*, pp. 157–160 (2003), doi:10.1109/ASPAA.2003.1285855.

[20] Y. Iwaya, "Individualization of Head-Related Transfer Functions With Tournament-Style Listening Test: Listening With Other's Ears," *Acoust. Sci. Technol.*, vol. 27, no. 6, pp. 340–343 (2006), doi:10.1250/ast.27.340.

[21] J. C. Middlebrooks, E. A. Macpherson, and Z. A. Onsan, "Psychophysical Customization of Directional Transfer Functions for Virtual Sound Localization," *J. Acoust. Soc. Am.*, vol. 108, no. 6, pp. 3088–3091 (2000).

[22] A. Silzle, "Selection and Tuning of HRTFs," presented at the *112th Convention of the Audio Engineering Society* (2002 Apr.), convention paper 5595.

[23] G. Parseihian and B. F. Katz, "Rapid Head-Related Transfer Function Adaptation Using a Virtual Auditory Environment," *J. Acoust. Soc. Am.*, vol. 131, no. 4, pp. 2948–2957, (2012), doi:10.1121/1.3687448.

[24] H. Dejardin and E. Ronciere, "NouvOson Website: How a Public Radio Broadcaster Makes Immersive Audio Accessible to the General Public," presented at the *AES 57th International Conference: The Future of Audio Entertainment Technology – Cinema, Television and the Internet* (2015 Mar.), conference paper 6-2.

[25] B. F. G. Katz and G. Parseihian, "Perceptually Based Head-Related Transfer Function Database Optimization," *J. Acoust. Soc. Am.*, vol. 131, no. 2, pp. EL99–EL105 (2012), doi:10.1121/1.3672641.

[26] A. Andreopoulou and B. F. Katz, "Subjective HRTF Evaluations for Obtaining Global Similarity Metrics of Assessors and Assessees," *J. Multimodal User Interf.*, pp. 1–13 (2016), doi:10.1007/s12193-016-0214-y.

[27] A. Borrego, J. Latorre, M. Alcañiz, and R. Llorens, "Comparison of Oculus Rift and HTC Vive: Feasibility for Virtual Reality-Based Exploration, Navigation, Exergaming, and Rehabilitation," *Games Health J.*, vol. 7, no. 3, pp. 151–156 (2018), doi:10.1089/g4h.2017.0114.

[28] A. Becher, J. Angerer, and T. Grauschopf, "Novel Approach to Measure Motion-To-Photon and Mouth-To-Ear Latency in Distributed Virtual Reality Systems," presented at the *GI VR/AR Workshop* (2018).

[29] D. Poirier-Quinot and B. F. Katz, "The Anaglyph Binaural Audio Engine," presented at the *144th Convention of the Audio Engineering Society* (2018 May), convention paper 431.

[30] M. Wright, "Open Sound Control: An Enabling Technology for Musical Networking," *Organ. Sound*, vol. 10, no. 3, pp. 193–200 (2005), doi:10.1017/S1355771805000932.

[31] P. Stitt, E. Hendrickx, J. -C. Messonnier, and B. F. Katz, "The Influence of Head Tracking Latency on Binaural Rendering in Simple and Complex Sound Scenes," presented at the *140th Convention of the Audio Engineering Society* (2016 May), convention paper 9591.

[32] M. Aussal, F. Alouges, and B. Katz, "HRTF Interpolation and ITD Personalization for Binaural Synthesis Using Spherical Harmonics," presented at the *AES 25th UK Conference: Spatial Audio in Today's 3D World* (2012 Mar.), conference paper 04.

[33] G. Parseihian and B. F. G. Katz, "Morphocons: A New Sonification Concept Based on Morphological Earcons," *J. Audio Eng. Soc.*, vol. 60, no. 6, pp. 409–418 (2012).

[34] D. Poirier-Quinot and B. F. Katz, "Impact of HRTF Individualization on Player Performance in a VR Shooter Game I," presented at the *2018 AES International Conference on Spatial Reproduction - Aesthetics and Science* (2018 July), conference paper P5-4.

[35] D. Poirier-Quinot and B. F. Katz, "Impact of HRTF Individualization on Player Performance in a VR Shooter Game II," presented at the *2018 AES International Conference on Audio for Virtual and Augmented Reality* (2018 Aug.), conference paper P4-1.

[36] G. Cumming, "The New Statistics: Why and How," *Psychol. Sci.*, vol. 25, no. 1, pp. 7–29 (2014), doi:10.1177/0956797613504966.

[37] B. Katz and R. Nicol, "Binaural Spatial Reproduction," in *Sensory Evaluation of Sound*, pp. 349–388 (CRC Press, Boca Raton, 2019), ISBN 978-1-4987-5136-0.

[38] B. Katz, A. Tarault, P. Bourdot, and J.-M. Vézien, "The Use of 3D-Audio in a Multi-modal Teleoperation Platform for Remote Driving/Supervision," presented at the *AES 30th International Conference: Intelligent Audio Environments* (2007 Mar.), conference paper 23.

[39] P. Majdak, T. Walder, and B. Laback, "Effect of Long-term Training on Sound Localization Performance With Spectrally Warped and Band-Limited Head-Related Transfer Functions," *J. Acoust. Soc. Am.*, vol. 134, no. 3, pp. 2148–2159 (2013), doi:10.1121/1.4816543.

[40] S. Perrett and W. Noble, "The Effect of Head Rotations on Vertical Plane Sound Localization," *J. Acoust. Soc. Am.*, vol. 102, no. 4, pp. 2325–2332 (1997), doi:10.1121/1.419642.

THE AUTHORS

David Poirier-Quinot

Brian FG Katz

David Poirier-Quinot is a researcher, presently working on sound spatialization, perception, and room acoustics simulation for virtual and augmented realities. He studied these fields along with signal processing and computer sciences at the ∂'Alembert Institute, the Imperial College London, IRCAM, LIMSI-CNRS, and ETIS labs. With a background in Mathematics, Physics, and Chemistry, he obtained a Master's degree in signal processing and telecommunications from the ENSEA graduate school of Electrical Engineering (France) in 2011 and received a Ph.D. degree in acoustics, signal processing, and computer science from Sorbonne University (Paris VI, France) in May 2015.

•

Brian F.G. Katz is a CNRS Research Director at the ∂'Alembert Institute, Sorbonne Université, CNRS, and coordinator of the Sound & Space research theme. His fields of interest include spatial 3D audio rendering and perception and room acoustics. With a background in physics and philosophy, he obtained his Ph.D. in Acoustics from Penn State in 1998 and his HDR in Engineering Sciences from UPMC in 2011. Before joining CNRS he worked for various acoustic consulting firms, including Artec Consultants Inc., ARUP & Partners, and Kahle Acoustics. He has also worked at LIMSI-CNRS and IRCAM.