Proceedings of ICSA 2019

5th International Conference on Spatial Audio
September 26th to 28th, 2019, Ilmenau, Germany

DOI: 10.22032/dbt.39954

DOI (proceedings): 10.22032/dbt.39936

Full Reviewed Paper at ICSA 2019

Presented by VDT.

Listening Tests with Individual versus Generic Head-Related Transfer Functions in Six-Degrees-of-Freedom Virtual Reality

Olli S. Rummukainen, Thomas Robotham, Axel Plinge, Frank Wefers,
Jürgen Herre, and Emanuël A. P. Habets
International Audio Laboratories Erlangen[1]*, Germany, email: olli.rummukainen@iis.fraunhofer.de*

Abstract

Individual head-related transfer functions (HRTFs) improve localization accuracy and externalization in binaural audio reproduction compared to generic HRTFs. Listening tests are often conducted using generic HRTFs due to the difficulty of obtaining individual HRTFs for all participants. This study explores the ramifications of the choice of HRTFs for critical listening in a six-degrees-of-freedom audio-visual virtual environment, when participants are presented with an overall audio quality evaluation task. The study consists of two sessions using either individual or generic HRTFs. A small effect between the sessions is observed in a condition where elevation cues are impaired. Other conditions are rated similarly between individual and generic HRTFs.

1. Introduction

Our ability to localize sounds in a 3-dimensional space relies on acoustic cues of interaural time difference (ITD), interaural level difference (ILD), interaural cross-correlation (ICC), and the spectral filtering caused by the physiology of the outer ears, head, and torso [14]. To render virtual audio binaurally over headphones, these cues are usually generated by convolving signals with head-related transfer functions (HRTFs). The HRTFs are individual; no two sets are alike.

Measuring or modeling individual HRTFs is a time consuming process requiring specialized hardware and facilities [4,9]. This process is currently unfeasible for the general public and most virtual experiences rely on non-individual, generic, HRTFs measured from a binaural head and torso simulator or averaged over a set of people. However, using generic HRTFs may result in blurry localization and front-back confusions, which in turn reduce the immersiveness of a virtual reality (VR) experience [25]. This study explores the consequences of the choice between individual and generic HRTFs in a six-degrees-of-freedom (6-DoF) virtual environment.

In VR, in contrast to purely auditory research and applications, multiple modalities help us to construct a mental representation of our surroundings [7]. Visual information improves sound localization in real and virtual environments [1]. Freedom of movement in VR further improves our ability to extract and disambiguate sensory information. Sensory-motor coupling has been argued to form a basis for human cognition, where motor actions support sensory information processing [6]. This is clearly demonstrated by the reduction of front-back confusions when head movements are allowed [11, 24]. On the contrary, when we have 6-DoF full-body motion in a VR environment, the auditory localization resolution was found to be degraded when compared to stationary listening in a recent study [20]. This finding hints towards sound localization being less fine grained regardless of whether we are using individual or generic HRTFs during full-body motion.

A number of studies have looked into adaptation to altered

---

[1] A joint institution of the Friedrich-Alexander-University Erlangen-Nürnberg (FAU) and Fraunhofer Institute for Integrated Circuits (IIS).

sound localization cues due to ear molds, hearing aids, or non-individual HRTFs via different training procedures. Active learning with feedback has been found to improve generic HRTF localization accuracy [13]. In VR, audio-visual cross-modal training improves source localization accuracy already after short training periods [3]. A many-to-one mapping mechanism in sound localization has been suggested, where the plasticity in the auditory cortex allows us to learn multiple HRTF sets that lead to the same percept [23]. For additional studies on auditory space adaptations, we refer the reader to [12].

A somewhat different approach is to select perceptually best-matching generic HRTFs from a large pool of HRTFs. Some participants have been shown to benefit from playing a VR-shooter game with their perceptual best-match HRTFs [17]. Related to adaptation, humans may adapt even to their perceptually worst-matched HRTFs through repeated short (12 min) training sessions [22]. However, the adaptation did not happen for everyone, suggesting that there may be limitations to adaptation in the worst match cases. No identifier was found to predict the individual ability to adapt to a new set of HRTFs. Anthropometry-based HRTF matching has been investigated in the context of VR, where no effect on questionnaire results was found between the best match HRTFs and generic HRTFs for a free exploration task in a VR scene [21].

Most studies on HRTF individualization have focused on localization accuracy only. Both timbral and spatial aspects of HRTF preference were considered in [2], where surprisingly a general preference of generic HRTFs over individual HRTFs was found. Potential causes were identified as higher quality of the built-in microphones of the generic binaural head and possible participant movement during HRTF measurement. These findings stress the importance to consider also non-localization-based HRTF quality attributes.

In this study the effect of individual versus generic HRTFs is investigated in the context of 6-DoF audio-visual virtual reality. In contrast to previous studies which have largely focused on localization accuracy with limited freedom of movement, the focus is on overall quality of audio rendering given self-movement cues and a corresponding visual environment. We examine the effect of the HRTF set with similar conditions and same participants in two separate sessions, namely, one with individual and one with generic HRTFs. The conditions include purposefully delayed tracking data and impaired localization cues in addition to a high quality convolution-based rendering and low quality non-spatial anchor conditions. We hypothesize $\mathcal{H}_1$: The individual HRTF session results in lower scores for some impairment conditions compared to the generic HRTF session, $\mathcal{H}_2$: The individual HRTF session results in less variance in the scores compared to the generic HRTF session, and $\mathcal{H}_3$: Individuals whose HRTFs are less like the generic HRTFs show more separation in scores between the sessions.

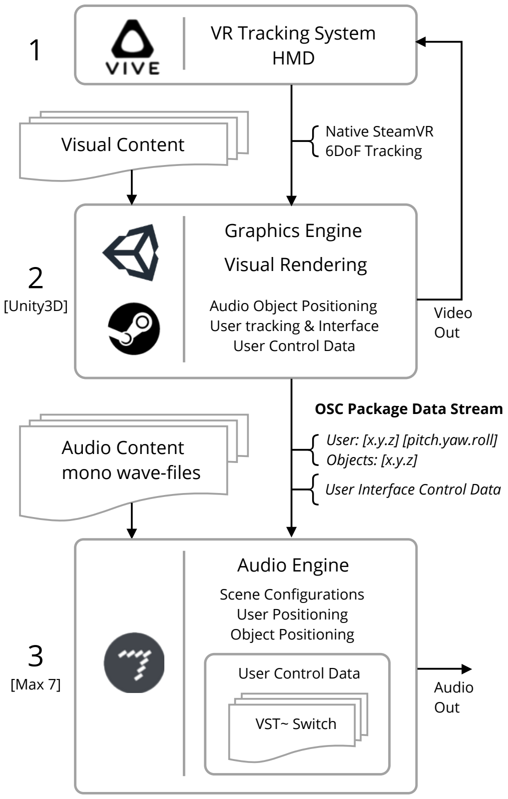

Fig. 1: Overview of the real-time evaluation platform.

2. Method

2.1. Virtual reality environment

A platform for real-time evaluation of binaural renderers in virtual reality was employed. The platform allows participants to switch between conditions, with no interruption to audio-visual sensory input, while exploring a 6-DoF virtual environment. The basic structure is presented in this section; for a thorough walk-through, please see [19]. The platform may be broken up into three components: 1) VR device, 2) Graphical rendering engine, and 3) Audio rendering engine, as depicted in Figure 1.

For Component 1, the HTC VIVE Pro[1] head-mounted display (HMD) is used for positional tracking, visual presentation, and control interface. The tracking accuracy and latency are found suitable for reproducible scientific research [15]. For Component 2, the Unity3D game engine is used for graphical rendering, along with hosting positional information for all audio objects and participant's position and orientation using the SteamVR asset. All relevant positional and rotational data is then sent (via an Open Sound Control (OSC) data package at a 10 ms interval) to Component 3. In Component 3, a binaural renderer is hosted in Max 7 and is fed the positional and rotational information received from Unity3D. All audio content is loaded into Max on a scene by scene basis and triggered to play when the respective scene is loaded inside Unity3D.

---

[1] https://www.vive.com/eu/product/vive-pro/ (Accessed: 29.05.2019)

Current Item: 1
Selected Condition: REF

Fig. 2: Graphical user interface activated within the virtual environment and the interaction device.

As participants' location is non-static, the test control interface is implemented inside the VR environment itself, allowing full freedom of movement while not being forced to return to a specific location to interact with the experiment interface. The interface is designed such that it can be instantiated anywhere in the VR scene. By pressing a button on a hand-held controller, a semi-transparent panel appears at eye level in the participants' field of view. The panel is presented in Figure 2. Pressing the button again hides the panel, allowing the user to fully explore the environment. When instantiated, a virtual laser pointer may be used to operate sliders and buttons on the panel.

The participants have a possibility to teleport in the virtual scene, which means they are not bound by the tracked lab area. They have a 2 m × 3 m floor area for free walking, but this area may be re-positioned in the virtual world via the teleport function.

2.2. Participants

In total 10 people (2 female, 8 male) participated in the test. Their average age is 36.6 years (SD = 9.2). All the participants are doctoral students or employees at Fraunhofer IIS working in audio, and all of them have prior experience of VR systems. The participants may be considered as experts in audio quality evaluation, but none had experience on quality evaluation in VR context and they received no information about the conditions under test. Some of the participants had prior listening experience with the generic HRTFs used in this study. None reported any known hearing impairments.

2.3. Stimuli

Scenes There were four scenes with different characteristics used in the study. The scenes and the audio objects in the scenes are summarized in Table 1. The scenes were always evaluated in the same order as follows: *Restaurant* scene with three audio sources close to the horizontal plane, *Living room* with a fireplace audio object on the floor and a piano and wind chimes objects positioned at the same horizontal position but separated in elevation, *Outdoor* scene with bird sounds, tree cutting, and an airplane at different elevations, and *Fountain music* with a piano and a water fountain audio object.

Tab. 1: Visual scene and audio object descriptions.

Scene	Features
Restaurant	Three audio objects: Guitar, conversation, and bottle opening
Living room	Three audio objects: Piano, wind chimes, and fireplace
Outdoor	Four audio objects: Tree cutting, ducks, airplane, and birds
Fountain music	Two audio objects: Piano and a fountain

The audio objects in the three first scenes were presented visually as yellow spheres at the location of the audio event and there was no semantic congruency otherwise. In the fourth scene, *Fountain music*, the sound producing objects were visually modeled according to their real world counterparts as a piano and a fountain. All audio objects were initially within sight of the participant at the start of a specific scene. The participants could then teleport or walk closer to the objects to examine different aspects of the scene more carefully.

The audio samples were constructed of about one minute long segments that could be looped infinitely. They were recorded at 48 kbps with 24 bits. The acoustics of the virtual space was not modeled in the rendering stage. Some of the audio samples contained a small amount of reverberation from the recording location, but, due to the lack of virtual acoustics, the direct-to-reverberant ratio cue was not modeled. Thus, distance rendering relied only on the intensity cue realized by applying the inverse square law with a maximum level reached at 0.1 m from the sound object. Auditory near-field effects were not modeled. The audio was played back through Beyerdynamic DT770 Pro headphones and there was no individual headphone equalization applied. The headphones are diffuse-field equalized by the manufacturer.

HRTFs Individual HRTFs were measured in a semi-anechoic chamber at 1.2 m distance with a procedure described in [18]. The measurement setup is depicted in Figure 3, where the loudspeaker arc used for measurement signals is shown behind a participant who is standing on a rotating platform. The resolution was 5° in azimuth and 2.5° in elevation resulting in 4608 source positions with a filter length of 386 samples. The measurements were made at the entrance to the blocked ear canal. To remove any non-directional characteristics, the HRTFs were diffuse-field equalized for this experiment by calculating the average magnitude over all directions, inverting it, and creating a minimum-phase filter, which was then convolved with the HRTFs. Diffuse-field equalization through the recording stage to reproduction has been found to result in consistent playback for different listeners and to reduce the need for individual headphone equalization [10].

The generic binaural head (Neumann KU100) HRTFs were obtained from the spatial audio for domestic interactive en-

Fig. 3: HRTF measurement setup with a participant.

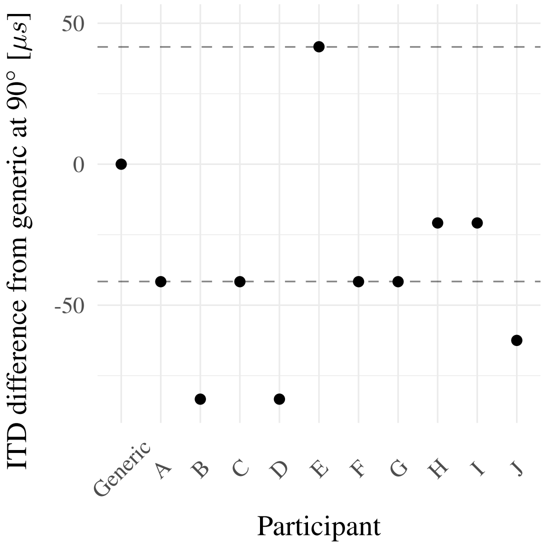

Fig. 4: ITD differences between the generic and the 10 individuals estimated from the HRTFs of the generic head (KU100) and the participants at azimuth = 90°, elevation = 0°. The dashed lines at 42 μs denote the boundaries of *low* and *high* similarity to the generic ITD groups, which are employed in the data analysis.

tertainment (SADIE) database[2]. The measurements comprise 1550 source positions measured at 1.5 m distance and the resulting HRTFs are diffuse-field equalized. Filter length is 256 samples. Notably, the KU100 binaural head does not include a torso, removing the shoulder reflection effect from the generic HRTFs. The generic and individual HRTF sets were level aligned to 68 dB_A at 1 m distance using pink noise and a binaural head.

To characterize the HRTFs objectively, the ITD was estimated from the HRTF sets for a source position directly to the side and at ear level. The estimation was based on finding the lag value corresponding to the maximum peak in the cross-correlation between the left and right ear head-related impulse responses. This value correlates with the individual head size. The resulting estimated differences between the generic HRTFs and the 10 participating individuals are displayed in Figure 4. Additionally, Figure 5 displays the diffuse-field equalized frequency responses of the individual HRTFs together with the generic HRTF for a source in the median plane at 40° elevation. The generic HRTF is observed to display one major notch between 7 kHz and 8 kHz, whereas the individual HRTFs have possibly multiple sharp notches in the range between 7 kHz and 14 kHz. Overall, the generic HRTF shows less detail in the frequency domain compared to the individual HRTFs, which results most likely from the shorter filter length (256 vs. 386 samples) and the simplified and torso-less geometry of the KU100 binaural head.

Conditions There were five audio rendering conditions in the test. The first, *Convolution*, convolved the nearest pair of HRTFs for each audio object's direction of arrival with the source signal. There was no interpolation between positions, i.e., the nearest HRTFs would always be selected. This condition functions as a basis for all the subsequent conditions, which were constructed by manipulating the tracking data or the scene setup. The *No elevation* condition modified the scene setup by placing all the audio objects' elevation to the ear level based on the tracking data of each participant. Here, the visual scene remained unaltered, but the corresponding binaural rendering emanated from the ear level at the correct azimuthal direction. Similarly, the *Angle offset* condition biased the horizontal tracking data to shift all audio rendering by 15° counter-clockwise. In this condition the elevations were rendered correctly. These two conditions resulted in a constant audio-visual mismatch in location, that became increasingly evident the closer one is to the audio object. The *Delay 500 ms* condition added a delay of 500 ms to all tracking data for the audio rendering. Effectively, this condition resulted in a sound scene that reacted slowly to listener movements in position and rotation. Finally, the *Stereo mix* was not reactive to the listener movements, rather all audio objects were rendered in a static manner and without any HRTF processing. This condition resulted in a sound scene that is mostly localized within the head. The conditions are summarized in Table 2.

It was assumed that the degradation in elevation cues would most likely lead to differences between the sessions. The generic HRTFs are known to result in poor localization in elevation, which, in turn, could lead to a stronger audio-visual integration effect. Thus, there could be less reduction in overall quality scores because the visual target would capture the weakly localized auditory percept. The individual HRTFs would lead to stronger auditory localization and the audio-visual mismatch would be more easily perceived.

---

[2] https://www.york.ac.uk/sadie-project/index.html (Accessed: 29.05.2019)

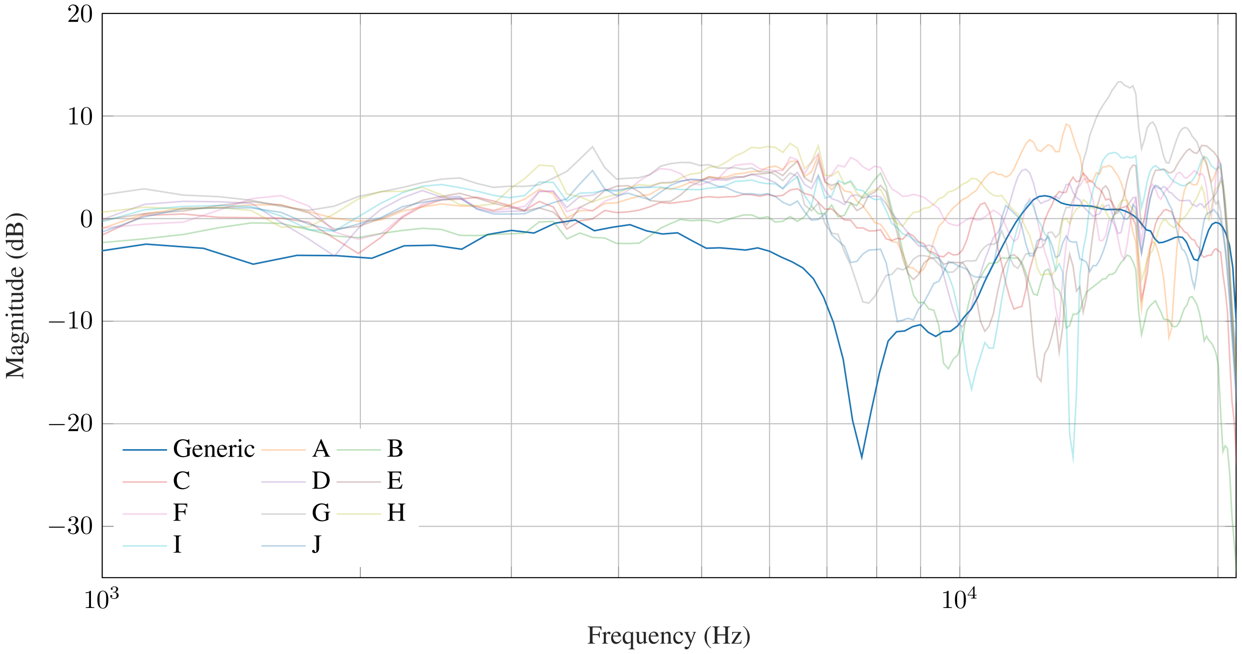

Fig. 5: Diffuse-field equalized frequency responses of the generic and individual head-related transfer functions measured at the left ear with azimuth = 0°, elevation = 40°.

Tab. 2: Audio rendering conditions under study.

Condition	Description
Convolution	HRTF-convolution-based renderer
No elevation	All audio objects are re-positioned to ear level
Angle offset	15° offset to the azimuth tracking data
Delay 500 ms	500 ms delay added to all tracking data
Stereo mix	Static stereo mix of all audio objects in a scene

2.4. Procedure

The participants were instructed to rate the overall quality of audio in the VR scene on a 100-point continuous scale in a multi stimulus test with hidden reference and anchor (MUSHRA) -like paradigm [8]. There were verbal labels marking the regions of the scale as bad, poor, fair, good, and excellent in 20-point intervals. It was made clear that there is no audio reference, but their self-motion and the visual presentations should be understood as creating the reference for expected auditory stimulation. Quality degradations were assumed to result from mismatch between the expectations and the perceived auditory stimulus. When unable to judge the overall quality, the participants were instructed to pay attention to spatial impression, i.e., how well the auditory stimulus is co-located with the visual sensation. The interface (Figure 2) allowed the participants to switch between the five conditions via buttons (A-E) as many times as required. There was no possibility to set loop points to inspect specific segments of the audio samples.

The participants were first familiarized with the VR system and interface in a special scene before beginning with the actual experiment. They were instructed on how to operate the controller to bring up and hide the control panel. They were also able to go through a dummy rating test without audio to get used to the interface buttons. The four experiment scenes were evaluated after the familiarization scene. There were two sessions conducted on separate days: the *individual* or *generic* HRTFs sessions. The conditions and scenes were identical in both sessions. The order of the session was counterbalanced between the participants so that half started with their own HRTFs and the other half with the generic HRTFs to remove the effect of learning. The average time to setup, familiarize, and complete evaluation of the four scenes in one session was 20 min.

3. Results

The study was structured around three independent variables in a within-participants design: *Scene*, *Condition*, and *Session*. The main effects on the dependent variable *Score* were analyzed by a three-way repeated measures analysis of variance (ANOVA). To check for ANOVA's assumptions, Mauchly's test for sphericity was performed on the data and indicated that the assumption of sphericity had not been violated for any independent variable. No main effect of the *Scene* ($F_{(3,27)} = 0.266$, $p = 0.850$, $\eta^2_G = 0.000$) nor any significant interactions with other variables were found. Thus, the data from all scenes were pooled together.

The results presented in the following stem from a two-way repeated measures ANOVA. The significant main effects and post-hoc analysis are presented in Table 3. Post-hoc comparisons were performed with the Tukey's method. Effect sizes are reported as the generalized eta-squared values, where

$\eta^2_G = 0.01$ is considered a small effect, $\eta^2_G = 0.06$ a medium effect, and $\eta^2_G = 0.14$ a large effect [5, 16].

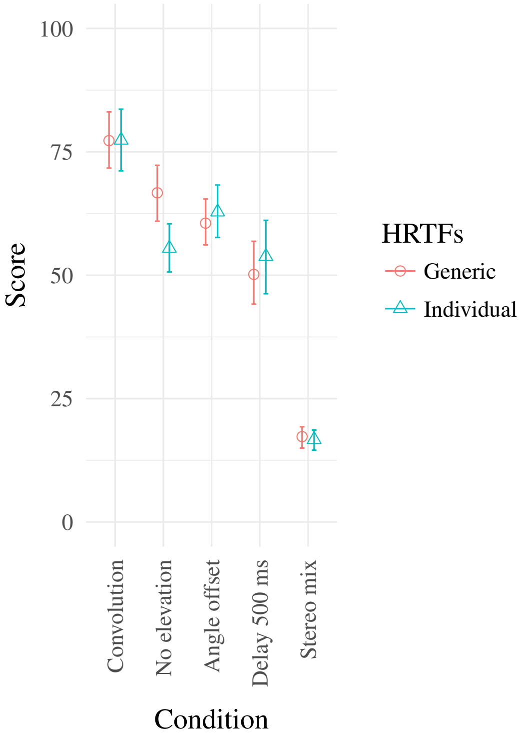

Fig. 6: Mean scores for conditions and sessions. The whiskers denote the bootstrapped 95 % confidence intervals of the mean.

Figure 6 displays the main results in a graphical form. Most importantly for our hypotheses, a significant interaction of *Session* and *Condition* is observed: in the *Generic* HRTF session the *No elevation* condition was scored significantly higher than in the *Individual* condition. Otherwise, both sessions result in similar scoring of the conditions. The unimpaired *Convolution* condition is always rated the highest and the *Stereo mix* the lowest. The *Delay 500 ms* condition receives the second lowest scores in both sessions followed slightly higher by the *Angle offset* and *No elevation*.

Further analysis on the variance between the sessions in each of the conditions was conducted by the Levene's test of homogeneity of variance. The data was split according to the condition and a separate analysis on the effect of session was done on each sub-group. The Levene's test revealed the homogeneity of variance assumption was met in every sub-group and thus no significant difference in the variances between sessions could be concluded based on our data. The analysis is in agreement with visual inspection of the distributions in Figure 6, where the 95 % confidence intervals are approximately similar for the two session in all conditions.

To test the hypothesis that the similarity of HRTFs to the generic ones has an effect, the participants were evenly divided into two groups based on their absolute ITD difference from the generic ITD. The split was done with ITD difference $\leq 42\ \mu s$ defined as the *high similarity* and ITD difference $> 42\ \mu s$ as the *low similarity* groups. Participants at the border were randomly assigned to either group to obtain even groups. By visually inspecting the scores grouped by the ITD difference, no difference between the groups could be observed. The ITD difference was further added as a factor to a linear model to explain variation in the dependent *Score* variable, but no effect was found.

4. Discussion

The *individual* HRTF session was found to be scored lower in one impairment condition, *No elevation*, which lends support to our first hypothesis that the conditions would receive differing scoring between HRTF sessions. No differences were observed in any other condition between the sessions. The other impairment types were not as critical for accurate HRTFs, since they involved errors in horizontal localization, which is robust against spectral differences in the HRTFs, or delayed tracking data, which affects all localization.

The *No elevation* had a rather large impairment in the elevation cues especially in the *Outdoor* scene, where there were multiple elevated audio objects such as tree clipping, birds, and an airplane. In other scenes the participants could have moved close to an audio object and crouched below it to inspect the elevation rendering. This, however, depended on the participants since they received no special instruction what to do in the virtual environment. Similarly, it was easy to miss the 15° angular offset, thanks to the audio-visual integration.

In a 6 DoF VR, the participants may ultimately dictate the audio content by their movements. While the fundamental audio content within a virtual scene is the same, participants' varying position and orientation means that audio in one participant's experience will be different from another. This raises the question whether there should be a participant training program for VR audio quality evaluations and what would that program entail. On the one hand, there is value in evaluations done with naive participants, as they represent the average end-user of a VR service, but on the other hand, they may miss some obvious shortcomings in rendering that become evident by chance to someone else.

Our second hypothesis about the reduced variance in response scores for the *individual* HRTF session is not supported by the data. Inspecting the score distributions visually in Figure 6 the bootstrapped confidence intervals appear similar for both sessions. Furthermore, Levene's test for homogeneity of variance did not find differences in the between sessions variances for any condition. Similarly, our third hypothesis concerning the differing scores based on the likeness of individual HRTFs to the generic HRTFs is rejected by our data. For these hypothesis our sample size (N=10) is probably too limited, since there is not enough variation in the HRTFs. The HRTF similarity comparison was based on ITD differences only, which may not be a descriptive or meaningful enough a metric.

In summary, our findings lend only weak support for the need of individual HRTFs for critical listening in 6-DoF VR. Only one effect between individual and generic HRTFs was observed with a small effect size ($\eta^2_G = 0.03$), signaling

Tab. 3: Main effects, interactions, generalized eta-squared (η_G^2) effect sizes, and post-hoc comparisons with p $<$ 0.05.

Effect	**F-value**	**p-value**	**Effect size**	**Post-hoc** ($p < 0.05$)
Condition	$F_{(4,36)} = 34.17$	$p < 0.001$	$\eta_G^2 = 0.63$	
Session	$F_{(1,9)} = 0.61$	$p = 0.46$	$\eta_G^2 = 0.00$	
Session × Condition	$F_{(4,36)} = 3.30$	$p = 0.02$	$\eta_G^2 = 0.03$	
Session|No elevation	$F_{(1,9)} = 6.58$	$p = 0.03$		Individual < Generic
Condition|Generic	$F_{(4,36)} = 34.22$	$p < 0.001$		Stereo mix < Delay < Angle offset < Convolution; Stereo mix < Delay < No elevation < Convolution
Condition|Individual	$F_{(4,36)} = 25.82$	$p < 0.001$		Stereo mix < Delay < Angle offset < Convolution; Stereo mix < No elevation < Convolution

that errors in elevation rendering may go unnoticed with generic HRTFs. Further studies with larger sample sizes are needed to be able to draw a clearer image of the question. Furthermore, here the HRTF sets had a large difference in the number of source locations with 1550 (generic) versus 4608 (individual), which could be assumed to results in larger perceptual effects. In 6-DoF VR the visuals and self-movement potentially largely mask the reduced spatial resolution of the generic HRTFs through audio-visual-proprioceptive integration. However, our results do not mean that generic HRTFs will result in higher quality in 6-DoF VR compared to individual HRTFs. The individual and generic HRTFs were not directly compared against each other in this study, and most likely different results would emerge from such a comparison.

Finally, our observations point towards a need for a training session to inform participants of the different nature of impairments in 6-DoF audio. In informal discussions many participants commented having learned what to listen for only after the first session or during the second session. Taking previous literature on auditory adaptation into account, a training session in 6-DoF may also serve as a familiarization phase to the generic HRTFs further reducing the effect of individual versus generic HRTFs.

5. Conclusions

In this study the effect of individual versus generic HRTFs for critical listening was investigated in 6-DoF audio-visual virtual reality. The use of individual HRTFs was found to enhance the participants' perception of errors in elevation. However, the effect size was small and most of the conditions showed no difference between the HRTF sets. Future research directions are envisioned to include participant training in six-degrees-of-freedom VR audio evaluation and adaptation to generic HRTFs in 6-DoF VR.

6. References

[1] Ahrens, A., Lund, K. D., Marschall, M., and Dau, T. Sound source localization with varying amount of visual information in virtual reality. *PLoS One 14*, 3 (2019), 1–19.

[2] Armstrong, C., Thresh, L., Murphy, D., and Kearney, G. A perceptual evaluation of individual and non-individual HRTFs : a case study of the SADIE II database. *Applied Sciences 8*, 2029 (2018), 1–21.

[3] Berger, C. C., Gonzalez-Franco, M., Tajadura-Jiménez, A., Florencio, D., and Zhang, Z. Generic HRTFs may be good enough in virtual reality. Improving source localization through cross-modal plasticity. *Frontiers in Neuroscience 12*, February (2018).

[4] Carpentier, T., Bahu, H., Noisternig, M., and Warusfel, O. Measurement of a head-related transfer function database with high spatial resolution. In *Forum Acusticum* (Krakow, Poland, 2014), pp. 1–6.

[5] Cohen, J. *Statistical power analysis for the behavioral sciences*, 2nd ed. Routledge, New York (NY), USA, 1988.

[6] Engel, A. K., Maye, A., Kurthen, M., and König, P. Where's the action? The pragmatic turn in cognitive science. *Trends in Cognitive Sciences 17*, 5 (2013), 202–209.

[7] Ernst, M. O., and Bülthoff, H. H. Merging the senses into a robust percept. *Trends in Cognitive Sciences 8*, 4 (2004), 162–9.

[8] International Telecommunication Union. Recommendation ITU-R BS.1534-3 Method for the subjective assessment of intermediate quality level of audio systems, 2015.

[9] Katz, B. F. G. Boundary element method calculation of individual head-related transfer function. I. Rigid model calculation. *The Journal of the Acoustical Society of America 110*, 5 (2001), 2440–2448.

[10] Larcher, V., Jot, J.-M., and Vandernoot, G. Equalization methods in binaural technology. In *Audio Engineering Society 105th Convention* (San Francisco (CA), USA, 1998), pp. 1–28.

[11] McAnally, K. I., and Martin, R. L. Sound localization with head movement: Implications for 3-d audio displays. *Frontiers in Neuroscience 8* (2014), 1–6.

[12] MENDONÇA, C. A review on auditory space adaptations to altered head-related cues. *Frontiers in Neuroscience 8* (2014), 1–14.

[13] MENDONÇA, C., CAMPOS, G., DIAS, P., VIEIRA, J., FERREIRA, J. P., AND SANTOS, J. A. On the improvement of localization accuracy with non-individualized HRTF-based sounds. *Journal of the Audio Engineering Society 60*, 10 (2012), 821–830.

[14] MØLLER, H., SØRENSEN, M. F., HAMMERSHØI, D., AND JENSEN, C. B. Head related transfer functions of human subjects. *Journal of the Audio Engineering Society 43*, 5 (1995), 300–321.

[15] NIEHORSTER, D. C., LI, L., AND LAPPE, M. The accuracy and precision of position and orientation tracking in the HTC Vive virtual reality system for scientific research. *i-Perception 8*, 3 (2017), 1–23.

[16] OLEJNIK, S., AND ALGINA, J. Generalized eta and omega squared statistics: measures of effect size for some common research designs. *Psychological Methods 8*, 4 (2003), 434–447.

[17] POIRIER-QUINOT, D., AND KATZ, B. F. G. Impact of HRTF individualization on player performance in a VR shooter game II. In *Audio Engineering Society Conference on Audio for Virtual and Augmented Reality* (Redmond (WA), USA, 2018), pp. 1–8.

[18] RICHTER, J.-G., BEHLER, G., AND FELS, J. Evaluation of a fast HRTF measurement system. In *Audio Engineering Society 140th Convention* (Paris, France, 2016), pp. 1–7.

[19] ROBOTHAM, T., RUMMUKAINEN, O., AND HABETS, E. A. P. Evaluation of binaural renderers in virtual reality environments: platform and examples. In *Audio Engineering Society 145th Convention* (New York (NY), USA, 2018), pp. 1–5.

[20] RUMMUKAINEN, O. S., SCHLECHT, S. J., AND HABETS, E. A. P. Self-translation induced minimum audible angle. *The Journal of the Acoustical Society of America 144*, 4 (2018), EL340–EL345.

[21] SIKSTRÖM, E., GERONAZZO, M., KLEIMOLA, J., AVANZINI, F., DE GÖTZEN, A., AND SERAFIN, S. Virtual reality exploration with different head-related transfer functions. In *15th Sound and Music Computing Conference* (Limassol, Cyprus, 2018), pp. 85–92.

[22] STITT, P., PICINALI, L., AND KATZ, B. F. Auditory accommodation to poorly matched non-individual spectral localization cues through active learning. *Scientific Reports 9*, 1 (2019), 1–14.

[23] TRAPEAU, R., AUBRAIS, V., AND SCHÖNWIESNER, M. Fast and persistent adaptation to new spectral cues for sound localization suggests a many-to-one mapping mechanism. *The Journal of the Acoustical Society of America 140*, 2 (2016), 879–890.

[24] WALLACH, H. The role of head movements and vestibular and visual cues in sound localization. *Journal of Experimental Psychology 27*, 4 (1940), 339–368.

[25] WENZEL, E. M., ARRUDA, M., KISTLER, D. J., AND WIGHTMAN, F. L. Localization using nonindividualized head-related transfer functions. *The Journal of the Acoustical Society of America 94*, 1 (1993), 111–123.