DOI: 10.61782/fa.2023.1076

forum acusticum 2023

HRTF INDIVIDUALISED AMBISONICS-TO-BINAURAL RENDERING USING MAG-LS AND COMPASS: A PERCEPTUAL EVALUATION OF OVERALL QUALITY

Frank Schultz[1*] **Shaimaa Doma**[2] **Sascha Spors**[1] **Janina Fels**[2]

[1] Institute of Communications Engineering, University of Rostock, Germany

[2] Institute for Hearing Technology and Acoustics, RWTH Aachen University, Germany

ABSTRACT

Headphone-based spatial audio reproduction nowadays often uses Ambisonics-to-Binaural rendering [1]. For optimum beam-forming, non-parametric methods, such as the Magnitude Least Squares (Mag-LS) [2], and parametric methods, such as the coding and multidirectional parametrisation of Ambisonic sound scenes (COMPASS) [3], were recently introduced. The use of individual (rather than non-individual) head-related transfer functions (HRTFs) technically optimises rendering. In a listening experiment, spatial audio-experienced participants were asked to blindly rate the overall perceived quality of six renderings with different HRTF data sets (2x their own HRTFs, 2x random non-individual HRTFs from humans, 2x hat-and-torso simulators (HATSs)), relative to their own internal reference. In 14 panels, different rendering approaches (Mag-LS vs. COMPASS and Ambisonics orders 1,2,3 vs. 7) and two scenes (speech vs. music in dedicated rooms) were varied. Participants were informed that at least one of the six renderings per panel used their own HRTFs. The results of this experiment indicate that renderings with individual HRTFs do not obviously and not necessarily yield highest overall perceived quality ratings. This can be explained by individually different quality concepts and by the intentional test design with no reference and no anchor.

Keywords: *individual HRTF, personalised binaural audio, Ambisonics to binaural rendering, perceived quality*

**Corresponding author*: *frank.schultz@uni-rostock.de.*

1. PRECIS

Headphone-based Higher-Order-Ambisonics-to-Binaural rendering (HOA2Bin) is a profiled approach for spatialisation of audio scenes, for which non-parametric beamformers [1, Fig. 2-4] (cf. Mag-LS) and parametric beamformers [1, Fig. 5] (cf. COMPASS) were introduced. Since the sound field and HRTFs can be acquired separately, HOA2Bin allows for convenient selection of different HRTF datasets and head-tracked dynamic rendering via 3D rotation of the HOA scene.

Due to its inherently most precise antenna directivity characteristic, individual HRTFs – when properly measured or numerically simulated – constitute the optimum selection choice from a purely technical viewpoint. Perceptual studies, concentrating on anechoic spatial audio scenes and on localisation, externalisation and front-to-back confusion indeed suggest, that individual HRTFs can significantly improve binaural renderings compared to HRTFs from HATS or to randomly chosen, non-individual HRTFs, cf. [4, 5]. However, these performances seem to be highly dependent on the experimental conditions, as other studies indicate no statistically significant effect for individual HRTFs alone, cf. [6–8], but rather only in interaction with head-tracking, cf. [9, 10], for example. Hence, the fundamental questions still seem to be, which applications benefit from individual HRTFs, which do not, and why. In [11], the spatial attributes brightness, richness, externalisation and preference were perceptually evaluated for non-head-tracked binaural renderings of a jazz piece in anechoic conditions. SADIE II database's participants rated their individual HRTF vs. non-individual vs. two HATS HRTFs. The results indicate that individual HRTFs do not outperform, but rather a preference tendency for HATS HRTFs. In [12], differences regarding the spatial attributes reverberance, source width/distance/direction and overall quality were rated for a real scenario – a female voice from 3 loudspeaker positions in a seminar room – vs. head-tracked binaural auralisations of this scene including non-individual and HATS HRTFs. The results indicate highly plausible auralisations, and that individual HRTFs do not outperform the HATS HRTFs conditions.

As literature provides evidence that individual HRTFs mostly enhance binaural audio quality, and as crafting true individual HRTFs is still time and tools consuming, strategies for HRTF individualisation [13] appear meaningful for optimum system-to-user adaptation [14]. Models that recommend appropriate HRTFs from a database make use of machine learning advancements, cf. [15–17], and can be trained with perceptual feedback, anthropometric features and raw HRTFs as input data, hence combining and enhancing earlier selection approaches, cf. [18, 19]. The studies [15, 16] involve perceptual tests on HRTF choices and indicate, that model-based selection of an individualised HRTF i) is better than a random HRTF choice and ii) is close in rating to the individual HRTF. Similar results are provided by [17] with special focus on vertical localisation when testing the model performance by technical metrics. This model ranks the examined KEMAR HATS with highest mismatch to recommended individualised and individual HRTFs, which was then shown to be consistent with elevation localisation performances in a listening test. Currently, HRTF individualisation is lively examined for audio-visual VR applications, indicating benefits in certain conditions but – due to multimodal-, interaction- and learning-aspects – no obvious and no mandatory outperformance of individualised HRTFs, cf. [7, 20, 21].

In-depth perceptual evaluation of personalised HOA2Bin deserves our attention. Hence, in this contribution, we discuss an experiment for HOA2Bin using different HRTFs focusing on auditory perception only. The test paradigm shall indicate if using individual HRTFs outperforms personalised HOA2Bin compared to non-individual/HATS HRTFs. Detailed information of the experiment, very briefly reviewed here, can be found in [22, 23], the latter being a preliminary study. For 13 participants and 2 HATS (KEMAR and Aachen), individual HRTF datasets as well as individual headphone transfer functions (HpTFs) of a Sennheiser HD650 were measured in a consistent laboratory setup in Aachen. Two spatial-aliasing free HOA scenes based on shoebox room modelling in RAVEN were created, with a female speaker in an empty seminar room and a jazz standard playing trio in an empty concert hall. Real-time HOA2Bin, head-tracked HOA rotation as well as HpTF equalisation was realised with SPARTA plugins hosted by Reaper, all controlled by a Jupyter notebook based listening test GUI. Twelve spatial-audio and listening-test experienced (except one) subjects were asked to rate overall perceived quality relative to their own internal reference on a continuous (bad / poor / fair / good / excellent labelled) scale for 14 x 6 binaural renderings. Written test instructions provided information on scene setups and comments that localisation, externalisation, room impression, timbre, (technical) preference should enter quality assessment. We could fairly assume that our expert listeners were not biased by personal taste of music and speech style and hence, quality ratings cover only technical aspects, both for music and speech. Without explicitly calling attention, subjects actually performed an HRTF selection procedure – intentionally a test design without anchor/reference, cf. [24] – for 14 panels, each comprising renderings with 6 different HRTFs. Subjects were briefed that at least one stimulus per panel is rendered with their own individual HRTF, which might have raised (individually pronounced) expectations with respect to overall quality. Robust multiple comparisons among dependent groups were evaluated with `rmmcppb()` [25, 26] within each of the 14 panels.

The rating results are shown in Fig. 1 and 2 as boxplots, indicating each of the 14 panels as an individual body. The ratings suggest no obvious outperformance of individual HRTFs. Observed significant differences seem to be specific for conditions, here the HRTF type and the rendering approach, but supposedly not the audio content. For some Mag-LS panels, cf. Fig. 1, individual HRTFs outperform randomly chosen human HRTFs. For COMPASS renderings, cf. Fig. 2, individual HRTFs outperform KEMAR HATS in most of the panels. This might be related to this renderer's usage of unaltered HRTF information for the estimated direct sound components, thus potentially leading to fewer errors in externalisation, localisation and colouration for individual HRTFs. General conclusions should be avoided from this observation though, as KEMAR performance varies, cf. [17, 22] (individual HRTFs mostly outperform) vs. [8, 12] (mostly not outperforming). Overall, our findings are in line with the literature, confirming that benefits of individual HRTFs seem to be application and audio scene dependent. A detailed discussion of the experiment, statistical evaluations and interpretation of the results can be found in [22].

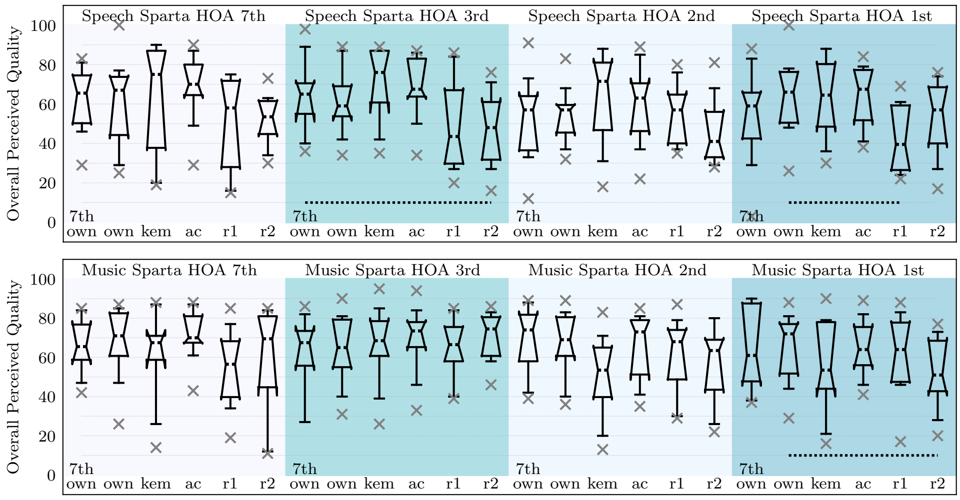

Figure 1. Results as box-plots (median: dotted, whiskers: [5, 95] % percentiles, rectangular box shape: interquartile range (IQR), notch shape: 95 % confidence interval (CI) of the median, outliers: grey $\times$) for speech scene (top) and music scene (bottom) rendered with **Sparta** plugin, i.e. **Mag-LS** algorithm. HRTF: `own` individual, `kem` KEMAR HATS, `ac` Aachen HATS, `r1`/`r2` two different random humans included in this study's HRTF database. Densely dotted, horizontal lines indicate statistical significant differences w.r.t. individual HRTFs vs. non-individual or HATS HRTFs (by comparing dependent groups using `rmmcppb()` [26] v1.1.0 defaults, percentile bootstrap, 20 % trimmed mean, $\alpha_{\text{cumulated}} = 0.05$ per panel.)

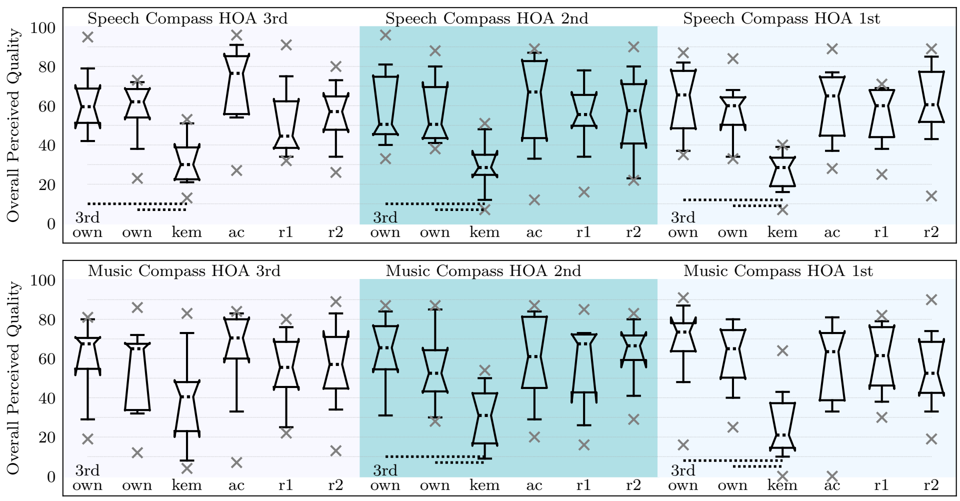

Figure 2. Results for speech scene (top) and music scene (bottom) rendered with **COMPASS** plugin. Same depiction strategy as in Fig. 1 above. Here, non-overlapping median CIs are highly consistent with corresponding `rmmcppb()`-based statistical significant group comparisons indicated by the horizontal lines.

2. REFERENCES

[1] B. Rafaely *et al.*, "Spatial audio signal processing for binaural reproduction of recorded acoustic scenes - review and challenges," *Acta Acust.*, vol. 6, p. 47, 2022.

[2] C. Schörkhuber *et al.*, "Binaural Rendering of Ambisonic Signals via Magnitude Least Squares," in *Proc. 44th German Annu. Conf. Acoust. (DAGA)*, 2018.

[3] A. Politis *et al.*, "COMPASS: Coding and Multidirectional Parameterization of Ambisonic Sound Scenes," in *Proc. IEEE Intl. Conf. Acoustics, Speech and Signal Process. (ICASSP)*, 2018.

[4] H. Møller *et al.*, "Binaural Technique: Do We Need Individual Recordings?," *J. Audio Eng. Soc*, vol. 44, no. 6, 1996.

[5] C. Jenny *et al.*, "Usability of Individualized Head-Related Transfer Functions in Virtual Reality: Empirical Study With Perceptual Attributes in Sagittal Plane Sound Localization," *JMIR Serious Games*, vol. 8, no. 3, 2020.

[6] A. Roginska *et al.*, "Stimulus-Dependent HRTF Preference," in *Proc. 129th Audio Eng. Soc. Conv.*, 2010.

[7] O. S. Rummukainen *et al.*, "Listening Tests with Individual versus Generic Head-Related Transfer Functions in Six-Degrees-of-Freedom Virtual Reality," in *Proc. 5th Intl. Conf. Spatial Audio (ICSA)*, 2019.

[8] M. Fallahi *et al.*, "Dynamic Binaural Rendering: The Advantage of Virtual Artificial Heads over Conventional Ones for Localization with Speech Signals," *Appl. Sci.*, vol. 11, no. 15, 2021.

[9] D. R. Begault *et al.*, "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, 2001.

[10] J. Oberem *et al.*, "Experiments on localization accuracy with non-individual and individual HRTFs comparing static and dynamic reproduction methods," *bioRxiv*, 2020.

[11] C. Armstrong *et al.*, "A Perceptual Evaluation of Individual and Non-Individual HRTFs: A Case Study of the SADIE II Database," *Appl. Sci.*, vol. 8, no. 11, 2018.

[12] M. Blau *et al.*, "Toward realistic binaural auralizations - perceptual comparison between measurement and simulation-based auralizations and the real room for a classroom scenario," *Acta Acust.*, vol. 5, 2021.

[13] C. Guezenoc *et al.*, "HRTF Individualization: A Survey," in *Proc. 145th Audio Eng. Soc. Conv.*, 2018.

[14] L. Picinali *et al.*, *Sonic Interactions in Virtual Environments (Eds.: M. Geronazzo, S. Serafin)*, ch. System-to-User and User-to-System Adaptations in Binaural Audio. Springer, 2023.

[15] A. Andreopoulou *et al.*, "Database Matching of Sparsely Measured Head-Related Transfer Functions," *J. Audio Eng. Soc*, vol. 65, no. 7/8, 2017.

[16] R. Pelzer *et al.*, "Head-related transfer function recommendation based on perceptual similarities and anthropometric features," *J. Acoust. Soc. Am.*, vol. 148, no. 6, 2020.

[17] M. Geronazzo *et al.*, "Applying a Single-Notch Metric to Image-Guided Head-Related Transfer Function Selection for Improved Vertical Localization," *J. Audio Eng. Soc*, vol. 67, no. 6, 2019.

[18] K. McMullen *et al.*, "Subjective Selection of Head-Related Transfer Functions (HRTF) Based on Spectral Coloration and Interaural Time Differences (ITD) Cues," in *Proc. 133rd Audio Eng. Soc. Conv.*, 2012.

[19] B. F. G. Katz *et al.*, "Perceptually based head-related transfer function database optimization," *J. Acoust. Soc. Am.*, vol. 131, no. 2, 2012.

[20] D. Poirier-Quinot *et al.*, "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, 2020.

[21] J. S. Andersen *et al.*, "Evaluation of Individualized HRTFs in a 3D Shooter Game," in *Proc. Intl. Conf. on Immersive and 3D Audio (I3DA)*, 2021.

[22] F. Schultz *et al.*, "Overall Perceived Quality for Mag-LS And Compass Ambisonics-To-Binaural Rendering Using Individualised vs. Non-Individualised HRTFs," *journal article TBA*, 2023.

[23] F. Schultz *et al.*, "HRTF Individualised Mag-LS and COMPASS Ambisonics-To-Binaural Rendering: Overall Perceived Quality for Pre-Conditioned Listeners," in *Proc. 154th Audio Eng. Soc. Conv.*, 2023.

[24] A. Andreopoulou *et al.*, "Subjective HRTF evaluations for obtaining global similarity metrics of assessors and assessees," *J. Multimodal User Interfaces*, vol. 10, no. 3, 2016.

[25] R. R. Wilcox, *Introduction to Robust Estimation and Hypothesis Testing*. Academic Press, 5th ed., 2022.

[26] A. Campopiano *et al.*, "Hypothesize: Robust Statistics for Python," *J. Open Source Softw.*, vol. 5, no. 50, 2020.