DOI: 10.61782/fa.2023.0489

forum acusticum 2023

INITIAL EVALUATION OF AN AUDITORY-MODEL-AIDED SELECTION PROCEDURE FOR NON-INDIVIDUAL HRTFS

Rapolas Daugintis[1*] **Roberto Barumerli**[2]
Michele Geronazzo[1,3] **Lorenzo Picinali**[1]
[1] Audio Experience Design, Dyson School of Design Engineering, Imperial College London, UK
[2] Acoustics Research Institute, Austrian Academy of Sciences, 1040 Vienna, Austria
[3] Department of Engineering and Management, University of Padova, Italy

ABSTRACT

Binaural spatial audio reproduction systems use measured or simulated head-related transfer functions (HRTFs), which encode the effects of the outer ear and body on the incoming sound to recreate a realistic spatial auditory field around the listener. The sound localisation cues embedded in the HRTF are highly personal. Establishing perceptual similarity between different HRTFs in a reliable manner is challenging due to a combination of acoustic and non-acoustic aspects affecting our spatial auditory perception. To account for these factors, we propose an automated procedure to select the 'best' non-individual HRTF dataset from a pool of measured ones. For a group of human participants with their own acoustically measured HRTFs, a multi-feature Bayesian auditory sound localisation model is used to predict individual localisation performance with the other HRTFs from within the group. Then, the model selection of the 'best' and the 'worst' non-individual HRTFs is evaluated via an actual localisation test and a subjective audio quality assessment in comparison with individual HRTFs. A successful model-aided objective selection of the 'best' non-individual HRTF may provide relevant insights for effective and handy binaural spatial audio solutions in virtual/augmented reality (VR/AR) applications.

Keywords: *HRTF personalisation, human sound localisation, computational auditory modelling, listening tests.*

**Corresponding author*: *r.daugintis21@imperial.ac.uk.*

1. INTRODUCTION

A head-related transfer function (HRTF) encodes binaural and monaural spatial hearing cues, which the human auditory system has learnt to use when estimating the direction of an incoming sound source [1]. The individual HRTF depends on the anatomy of the human ears and body. It can be acoustically measured in a laboratory and used for binaural (headphone) spatial audio reproduction. However, these measurements are unfeasible on a mass scale, so widely available binaural spatial audio applications use non-individual HRTFs, typically measured on a mannequin head. This may impact the quality of the spatial audio experience and affect the localisability of sounds [2], or again the sensation that the sound is coming from outside of the listener's head (externalisation) [3].

Various methods have been proposed to personalise HRTFs without acoustically measuring them for an individual [4]. Some attempt to select an HRTF measured on another person that is similar to an individual one. These methods rely on establishing a similarity metric between different HRTFs, which is a multidimensional challenge due to a combination of acoustic and cognitive aspects of spatial hearing. On the one hand, previous studies have used objective numerical metrics, such as spectral differences [5] or matching based on pinna geometry [6], but these metrics are based on empirical simplifications of the human perception [7]. On the other hand, some studies have employed subjective rankings to select the best-fitting non-individual HRTF [8]. However, they require direct human input and often output unreliable and non-repeatable results, especially for naïve listeners [9, 10].

Aiming to combine the numerical and the perceptual worlds, we present a metric based on an auditory sound localisation model as a step towards a perceptually-motivated automated non-individual HRTF selection procedure. The auditory model predicts the human sound localisation performance with a specific HRTF, so the method is numerical but accounts for the human perception of sound in space. In this study, the method, proposed in [11] (summarised in Sec. 2.1), has been applied to a new HRTF dataset. The selection is then evaluated using sound localisation and spatial audio quality tests (detailed in Sec. 2.2). Sec. 3.1 presents the initial HRTF selection results while Sec. 3.2 and Sec. 3.3 show and discuss some preliminary data from the listening tests. The work is concluded and the future outlook is highlighted in Sec. 4.

2. METHODOLOGY

2.1 Non-individual HRTF selection

40 HRTFs from the SONICOM dataset [12] were used in this study. To avoid any post-processing artefacts affecting the selection (as well as the listening tests, discussed in the following sections), windowed (5 ms) and not free-field compensated versions of HRTFs were chosen (in line with previous similar work [8]). For each subject from the subset, the 'best' and the 'worst' non-individual HRTFs were selected from the other 39, according to the selection procedure detailed in [11]. The procedure was based on the sound localisation test simulation using a Bayesian auditory sound localisation model [13], available from the Auditory Modelling Toolbox (AMT) [14].

In summary, the model is supplied with the individual HRTF of the listener as a template, and it predicts the localisation errors for given directions when using a non-individual (target) HRTF. It does so by extracting binaural and monaural features from the target HRTFs, corrupting them by internal noise (to account for the limitations of the auditory system), and matching them with the template features using Bayesian inference. The model has five free parameters to account for acoustic and non-acoustic factors of individual human sound localisation performance. However, sound localisation data required to calibrate the model for each individual were unavailable for the given set of subjects, so the parameters were set to the medians of the values obtained from the calibration exercise with 16 subjects from another dataset (reported in [11]).

The selection procedure analyses the predicted local polar root-mean-square (rms) error (PE; calculated over responses that were within 90° from the target position in the polar angle) and quadrant error (QE; percentage of polar errors $> 90°$) distributions within a small range of directions in front of the virtual listener ($\pm 11.5°$ polar and $\pm 30°$ lateral angles in the interaural-polar coordinate system) for each non-individual HRTF (these metrics, initially defined in [15], are described in more detail in [11]). As reported in [11], we expect a normal PE distribution in this region when localising sounds rendered with the individual HRTF, while poorly matched non-individual HRTFs might result in a skewed or multimodal PE distribution, deviating from normality. Thus, our procedure firstly classifies the HRTFs into *good* and *bad* based on the Shapiro–Wilk test, preferring HRTFs that would maintain normality in the simulated PE distribution. It then selects one 'best' HRTF from the good HRTFs and one 'worst' from the bad HRTFs based on the combination of the smallest/highest PE and QE.

2.2 Listening tests

Listening tests were conducted to evaluate the non-individual HRTF classification methodology. The tests were done in a virtual reality (VR) environment using Meta Quest 2. A standalone test application was developed in Unity and built for the VR headset, running an Android operating system. The binaural sound spatialisa-

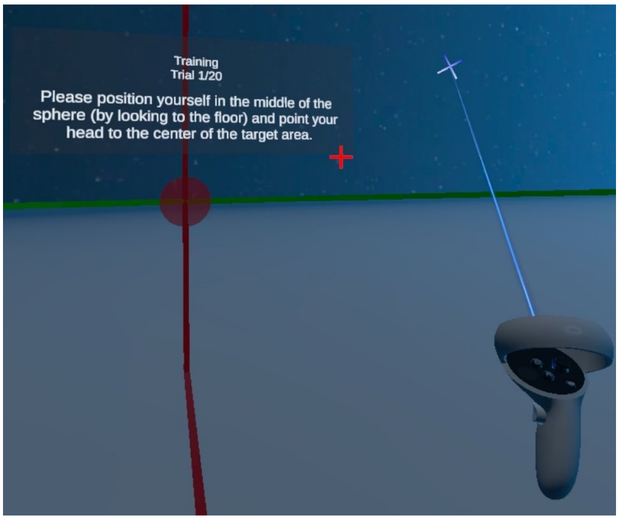

Figure 1. Sound localisation test environment during the alignment process before each trial. The head-anchored reticle (red +) must be aligned to the circular target for the sound to be played. The perceived direction is registered using the controller.

tion with custom HRTFs was handled by the 3D Tune-In Toolkit [16] via its Unity Wrapper [1]. For the listening tests, subjects were sat on a swivel chair (to ease rotations), and the tests were performed in a sound-insulated lab. The sounds were played via Sennheiser HD 599 headphones connected directly to the VR headset. These headphones were used due to their lower impedance compared to other studio-quality open-back headphones, which allowed us to obtain sounds of sufficient loudness from the low-power VR device. No headphone equalisation was performed. However, the headphone response is static and direction-independent, so listeners could adapt to it during the short accommodation phase before the tests.

Two tests were conducted in sequence: a static sound localisation test and a spatial audio quality assessment. Both tests were performed in a similar visual VR environment that contained a ground plane, a starry night skybox, and a coordinate sphere with three different coloured circles to indicate the median, horizontal, and frontal planes with respect to the original listener position. These were included as an aid for head positioning and pointing in an otherwise minimal VR environment. Listeners were given the Quest controller, corresponding to their dominant hand, to control the test. A 3D model of the controller in the VR environment was mapped to the physical controller to provide sensory consistency. The differences between the two test setups are highlighted below.

2.2.1 Localisation test

The localisation test in VR was inspired by the test design first proposed by [17] and later updated in [6] and [18]. The listeners performed a static localisation test to replicate the scenario modelled with the auditory sound localisation model. A screenshot from the test interface is shown in Fig. 1. The lighting of the environment was dimmed to minimise the influence on the localisation from the visual environment while maintaining some visibility of the coordinate axis. Before each trial, subjects were asked to position themselves in the middle of the coordinate sphere and orient their heads towards the original direction (set at the start of the test). This was done with an aid of a reticle which followed the eyesight and a target disk positioned in the original front position. These were red when not aligned with each other and turned green once the alignment was reached (within 10 cm position and 2.5° angle tolerance from the origin). The sound then played after 1 s to ensure the listener was still. The sound used in the test consisted of three consecutive 100 ms Gaussian noise bursts, each windowed using a Hann window (total duration of the signal was 300 ms). The choice to use three bursts instead of one continuous noise was motivated by previous research which showed better localisability of such sounds [19]. The listeners were instructed to stay still while the sound was playing. If they moved, the sound would stop and the listener would be asked to reposition again. After the sound played, the subjects were free to rotate their head. They were asked to point to the perceived position of the sound source using the controller, which had a virtual laser pointer attached to it in the VR environment. The position was registered by pressing a button on the controller. Then the listeners were asked to realign themselves for the next trial (a similar approach was already validated in [20]).

For each of the three conditions (individual HRTF, the 'best' non-individual HRTF, and the 'worst' non-individual HRTF), 40 source positions were tested [2]. The directions were pre-selected, but their order was randomised for each subject and each trial. The selected

[1] https://github.com/3DTune-In/3dti_AudioToolkit_UnityWrapper

[2] The fourth condition, which used individual direct transfer function (DTF), obtained by removing a direction-independent component from the HRTF, was also included but its results are not reported in this study.

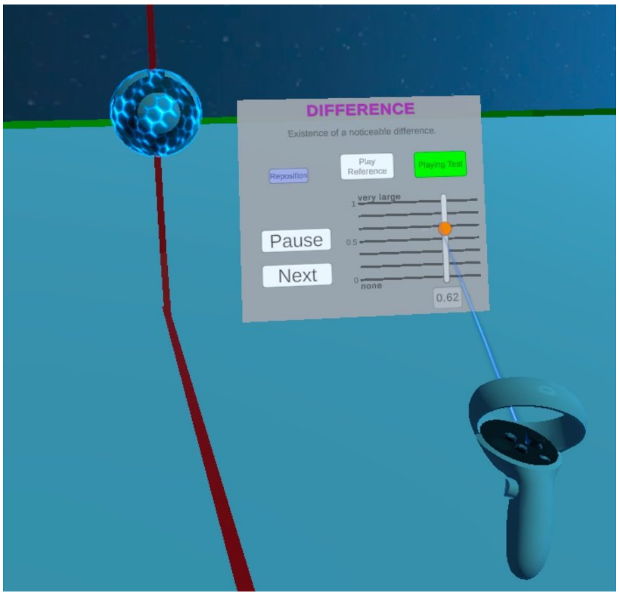

Figure 2. Spatial audio quality test environment.

positions were limited to be within $\pm 30°$ lateral angle (both front and back hemispheres) and $\pm 45°$ polar angle in the front and between $135°$ and $225°$ polar angle at the back of the listener. The positions were selected to sample these areas of the sphere relatively evenly but with a slightly denser grid of 10 positions selected to be within $\pm 11.5°$ polar angle in the front, which corresponds to the area used in the HRTF selection procedure [11]. The presentation order of the conditions was randomised across the participants, but an even spread of orders was ensured across the subjects to minimise biases on the localisation performance due to the running order. Before the main test, a short training session was run with 20 positions and a different non-individual HRTF, measured on a KEMAR mannequin, to familiarise subjects with the test setup. During the first 10 training session trials, the sound source was visible to provide minimal procedural adaptation to the listening environment.

2.2.2 *Spatial audio quality test*

After the sound localisation test, listeners were asked to perform a qualitative spatial audio assessment (see Fig. 2). The test design was based on the *Spatial Audio Quality Inventory* (SAQI) [21]. For each session, the listener was asked to rate the difference between a reference and a test sound according to one of the attributes presented. The session control panel had buttons to play the reference, the test, or pause the sound, a slider to rate the difference based on the quality indicated at the top of the panel, and a button to go to the next session. The rating scale on the panel was adapted to the specific quality with corresponding scale labels. The rating, made using the slider, was reflected in a number box below. The slider could only be moved once both sounds were played, and the listener could only go to the next session after moving the slider.

Following the SAQI protocol, the listeners were first asked to rate the overall difference between the two sounds. If they indicated no difference, no further questions would be asked. However, if the difference was reported, they were asked to rate the same set of test/reference sounds in the difference of ‘Tone colour bright-dark’, ‘Externalisation’, and ‘Naturalness’ (SAQI descriptions of each quality were also presented). Finally, the listeners were asked if they wanted to report any other difference, which they could describe to the invigilator. The sound source was visible throughout the qualitative assessment and could be moved to another place on the coordinate sphere (without changing the distance). The participants were asked to explore the sound scene by either rotating their heads or moving the source before making the final rating.

The test consisted of six sessions: three with a continuous noise train (similar to the one used for the localisation task) and three with an unintelligible speech sample (*International Speech Test Signal* (ISTS) [22]). The reference sound was always rendered with the individual HRTF. The test sound was rendered using either the individual HRTF (the same as the reference), the ‘best’ non-individual, or the ‘worst’ non-individual HRTFs. Before the actual test, the subjects were presented with a set of instructions within the VR test environment which explained the control panel and helped subjects familiarise with the test procedure.

3. RESULTS AND DISCUSSION

3.1 HRTF selection results

The model-based selection results for the 40 subjects are presented in Fig. 3. The figure shows the classification of each non-individual HRTF for each subject and highlights the ‘best’ and the ‘worst’ HRTF from each group. The

Figure 3. Classification of non-individual HRTFs based on the auditory sound localisation model prediction for 40 subjects. The ‘best’ and the ‘worst’ HRTFs were used in the perceptual validation test.

latter two HRTFs were used in the perceptual evaluation.

The selection methodology appears to consistently select the same HRTF as the 'worst' for most subjects (no. 33), which may indicate an HRTF with especially poor spectral cues or an instance of a flawed HRTF measurement. On the other hand, the selection of the 'best' HRTF is more varied but favours some HRTFs (especially no. 10), implying that some HRTFs may allow for better sound source localisation performances thanks to their higher spectral variance. Moreover, the HRTF selection is not reciprocal between subjects, which is a result already shown in the procedure calibration study on another HRTF dataset [11], in other HRTF selection studies employing a different computational auditory model [23], as well as previous perceptual studies [8, 24].

3.2 Localisation test results

In this section, we present some preliminary results of the sound localisation test with a few subjects whose HRTFs were included in the selection procedure.

At the moment of reporting, 17 subjects have performed the listening tests, including a mix of naïve and experienced listeners. Fig. 4 shows boxplots representing PE and QE across the participants for three conditions. Four participants who had PE $\geq$ 40° or QE $\geq$ 50% with the individual HRTFs were excluded from the plots because their errors were deemed to be too big to show meaningful results (approaching or exceeding chance level). Therefore, the plots include data from 13 subjects. The error in the lateral direction is not reported because it is typically lower and less dependent on the HRTF than the polar and quadrant errors [25].

Overall, the sound localisation errors obtained from the tests are relatively high but somewhat comparable to the ones from previous studies (e.g. [17, 20]), especially considering that several naïve subjects were included in the group. Participants generally felt that the localisation task was very difficult. Many of them reported a sensation that the sounds were mainly coming from their back. This is reflected in the figure which shows QE rates approaching 50%, a chance level. Importantly, the errors across the subject group do not appear to reflect the specific choice of the HRTF, including the individual HRTF condition (which was expected to produce lower localisation errors). The lack of relationship between the HRTF and the localisation errors may indicate that the sound source direction was too difficult for the listeners to discern, which led to uncertain and highly-variable answers, irrespective of the HRTF used. One possible explanation of these results is that listeners relied on the most salient interaural cues when indicating the direction, paying less attention and thus not benefiting from better-matched monaural cues. Reducing the test space to a sagittal plane might help investigate the validity of this hypothesis. Furthermore, the subjects' lack of experience with VR might have additionally corrupted the results. A potential improvement to the localisation task would be to provide more extensive

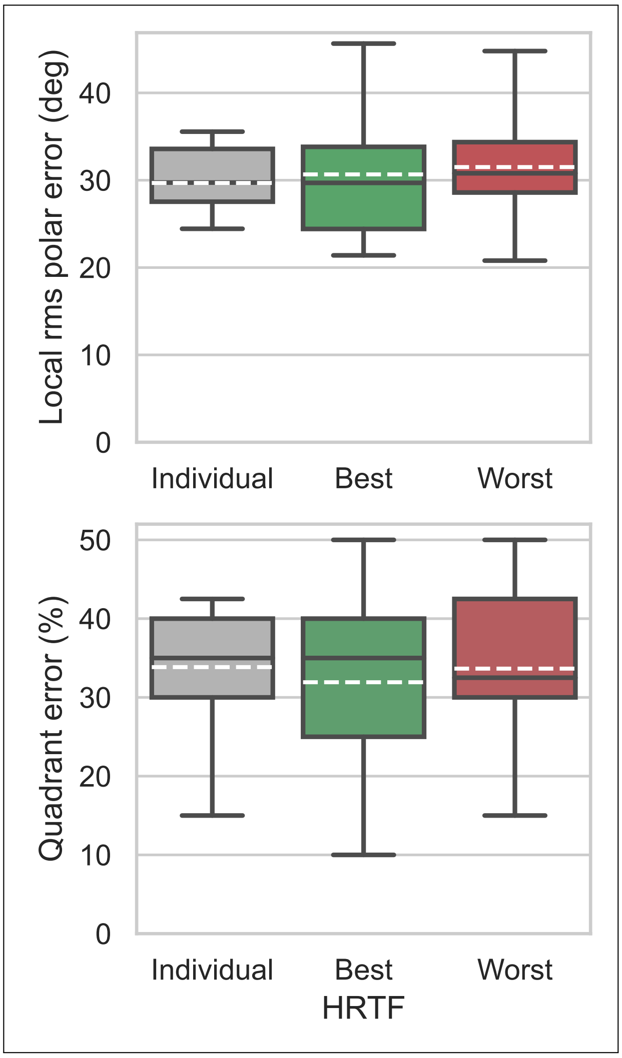

Figure 4. Sound localisation errors across 13 subjects. Boxes show distribution quartiles and whiskers present the full distribution extent. Dashed lines represent distribution means.

training on the VR pointing procedure (similar to [17]) and investigate HRTF processing algorithms [26]. Finally, a high level of front-back reversal might have been caused by the visual environment dominating the perception when the non-individual auditory cues were vague: lack of visual source in the front of the listener led participants to believe that the sound must have come from their back (even if the subjects were informed that the sounds might come from any direction, both front and back). A potential mitigation strategy would be to obscure the visual stimuli even more, using visual effects like fog (similarly to [27]).

3.3 Spatial audio quality test results

Fig. 5 reports early results from the spatial audio quality assessment across the 13 subjects, whose sound localisation data were shown previously. Following the SAQI test result analysis, similar to [28], the plots show the rating medians and their 95% bootstrap confidence intervals (CIs). The top row represents the ratings made for the overall difference between the reference (individual HRTF condition) and different test HRTF conditions (x-axis). The bottom row presents the comparison of perceived sound externalisation. Results from tests with noise bursts are on the left column while the right column shows the results when using the speech sample. The first condition on the x-axis is the control condition where the test and the reference are the same signals, rendered using the individual HRTFs. Although most subjects correctly indicated no difference between the two when noise was used, there were a few outliers who indicated a perceived minor difference (the subjects were told that there may be no difference between the test and the reference in some conditions before the test started). The results were less consistent for the speech sample, where more individuals reported differences between identical conditions. Taking into account the feedback received from participants, it was concluded that the specific speech sample (ISTS) was not a fitting audio signal for the test; different speech samples glued together in the audio sample varied too much over time, and this variation was perceived as caused by the change in condition when subjects were switching between the reference and the test signals.

For tests with continuous noise bursts, participants tended to rate the 'best' HRTF as being less different from the individual than the 'worst' HRTF. Since the Shapiro-Wilk test didn't reject the null normality hypothesis for both distributions, a paired t-test was used and a statistically significant effect was found between the 'best' and the 'worst' conditions ($t(12) = -3.05$, $p = 0.01$). However, the CIs for both conditions do not extend to zero (which corresponds to no difference to the individ-

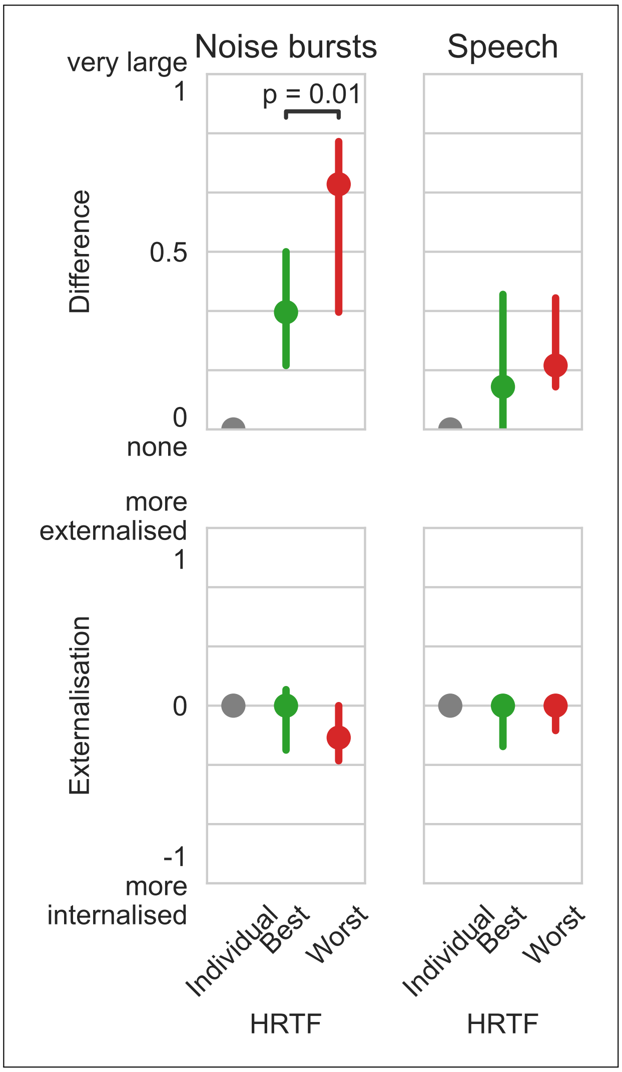

Figure 5. Median ratings across 13 subjects for 'Difference' (top) and 'Externalisation' (bottom) quality attributes compared to the individual HRTF condition as a reference when using noise bursts (left) and unintelligible speech (right) as a signal. Lines represent 95% bootstrap CIs and the p-value indicates a statistically significant difference between the conditions.

ual condition) indicating that, in general, listeners found the sound to be different with non-individual HRTFs as compared to the individual one. This is less of a case when using a speech sample, where the lower CI for the 'best' condition extends beyond zero, suggesting that the perceived difference between the 'best' and the individual HRTFs in speech might not be significant. Although the median rating and the lower CI for the 'worst' condition are higher than for the 'best' condition, the Wilcoxon test (used because the two distributions are not normal) reveals no significant differences between the two conditions.

For the differences in perceived externalisation, the median ratings (or at least their CIs) are around zero, indicating no significant effect of the HRTF condition on externalisation. And although the median rating value for the 'worst' HRTF condition when using noise bursts is negative, the Wilcoxon test shows no significant difference between the 'best' and the 'worst' conditions. The results suggest that assessment of externalisation with different HRTFs is difficult in such a setup, in line with conflicting findings from previous externalisation studies [3].

To obtain more conclusive results, the test procedure must be reviewed and more participants need to take part in the experiment. Furthermore, a more comprehensive statistical analysis is ongoing, including data with other SAQI attributes ('Tone colour' and 'Naturalness').

4. CONCLUSION

The study presented the use of a previously reported non-individual HRTF selection procedure on the new SONICOM HRTF dataset [12]. The selected 'best' and 'worst' non-individual HRTFs were used in a perceptual validation using a sound localisation test and a spatial audio quality assessment. Preliminary test results reveal that procedurally selected 'best' non-individual HRTF is also perceived to be less different from an individual HRTF than the 'worst' one. However, the sound localisation data is inconclusive but clearer results might arise from improvements to the test design and HRTF processing, data including more subjects (including a separate analysis on expert and naïve listeners), and a more comprehensive statistical analysis. Overall, once validated, the selection methodology could be used to improve binaural audio technologies based on our perceptually-informed metric in combination with other methods, e.g., by using a sparse set of HRTF measurements done at home as an individual template to be matched to a high-quality dense non-individual HRTF set from a database.

5. ACKNOWLEDGMENTS

This research is supported by the SONICOM project (EU Horizon 2020 RIA grant agreement ID: 101017743). Roberto Barumerli is also supported by the Dynamates project (Austrian Science Fund (FWF) project No. ZK 66).

6. REFERENCES

[1] J. Blauert, *Spatial Hearing: The Psychophysics of Human Sound Localization*. The MIT Press, 2nd ed., 1997.

[2] 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.

[3] V. Best, R. Baumgartner, M. Lavandier, P. Majdak, and N. Kopčo, "Sound externalization: A review of recent research," *Trends Hear.*, vol. 24, 2020.

[4] L. Picinali and B. F. G. Katz, "System-to-user and user-to-system adaptations in binaural audio," in *Sonic Interactions in Virtual Environments* (M. Geronazzo and S. Serafin, eds.), pp. 115–143, Cham: Springer International Publishing, 2023.

[5] J. C. Middlebrooks, "Individual differences in external-ear transfer functions reduced by scaling in frequency," *J. Acoust. Soc. Am.*, vol. 106, no. 3, pp. 1480–1492, 1999.

[6] M. Geronazzo, E. Peruch, F. Prandoni, and F. Avanzini, "Applying a single-notch metric to image-guided head-related transfer function selection for improved vertical localization," *J. Audio Eng. Soc. (AES)*, vol. 67, no. 6, pp. 414–428, 2019.

[7] I. Ananthabhotla, V. K. Ithapu, and W. O. Brimijoin, "A framework for designing head-related transfer function distance metrics that capture localization perception," *JASA Express lett. (JASA-EL)*, vol. 1, no. 4, p. 044401, 2021.

[8] 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.

[9] A. Andreopoulou and B. F. G. Katz, "Investigation on subjective HRTF rating repeatability," in *Proc. Audio Eng. Soc. (AES) Conv.*, p. 9597, 2016.

[10] C. Kim, V. Lim, and L. Picinali, "Investigation into consistency of subjective and objective perceptual selection of non-individual head-related transfer functions," *J. Audio Eng. Soc. (AES)*, vol. 68, no. 11, pp. 819–831, 2020.

[11] R. Daugintis, R. Barumerli, L. Picinali, and M. Geronazzo, "Classifying Non-Individual Head-Related Transfer Functions with A Computational Auditory Model: Calibration And Metrics," in *Proc. IEEE Int. Conf. on Acoust., Speech and Signal Process. (ICASSP)*, June 2023.

[12] I. Engel, R. Daugintis, T. Vicente, A. O. T. Hogg, J. Pauwels, A. J. Tournier, and L. Picinali, "The SONICOM HRTF dataset," *J. Audio Eng. Soc. (AES)*, vol. 71, no. 5, pp. 241–253, 2023.

[13] R. Barumerli, P. Majdak, M. Geronazzo, D. Meijer, F. Avanzini, and R. Baumgartner, "A Bayesian model for human directional localization of broadband static sound sources," *Acta Acust.*, vol. 7, p. 12, 2023.

[14] P. Majdak, C. Hollomey, and R. Baumgartner, "AMT 1.x: A toolbox for reproducible research in auditory modeling," *Acta Acust.*, vol. 6, no. 19, 2022.

[15] J. C. Middlebrooks, "Virtual localization improved by scaling nonindividualized external-ear transfer functions in frequency," *J. Acoust. Soc. Am.*, vol. 106, no. 3, pp. 1493–1510, 1999.

[16] M. Cuevas-Rodríguez, L. Picinali, D. González-Toledo, C. Garre, E. de la Rubia-Cuestas, L. Molina-Tanco, and A. Reyes-Lecuona, "3D Tune-In Toolkit: An open-source library for real-time binaural spatialisation," *PLoS ONE*, vol. 14, no. 3, p. e0211899, 2019.

[17] P. Majdak, M. J. Goupell, and B. Laback, "3-D localization of virtual sound sources: Effects of visual environment, pointing method, and training," *Atten. Percept. and Psychophys.*, vol. 72, no. 2, pp. 454–469, 2010.

[18] M. Geronazzo, J. Y. Tissieres, and S. Serafin, "A minimal personalization of dynamic binaural synthesis with mixed structural modeling and scattering delay networks," in *Proc. IEEE Int. Conf. on Acoust., Speech and Signal Process. (ICASSP)*, pp. 411–415, 2020.

[19] M. J.-M. Macé, F. Dramas, and C. Jouffrais, "Reaching to sound accuracy in the peri-personal space of blind and sighted humans," in *Computers Helping People with Special Needs, ICCHP 2012*, vol. 7383, pp. 636–643, 2012.

[20] M. A. Steadman, C. Kim, J.-H. Lestang, D. F. M. Goodman, and L. Picinali, "Short-term effects of sound localization training in virtual reality," *Sci. Rep.*, vol. 9, p. 18284, Dec. 2019.

[21] A. Lindau, V. Erbes, S. Lepa, H.-J. Maempel, F. Brinkman, and S. Weinzierl, "A spatial audio quality inventory (SAQI)," *Acta Acust. united with Acust.*, vol. 100, no. 5, pp. 984–994, 2014.

[22] I. Holube, S. Fredelake, M. Vlaming, and B. Kollmeier, "Development and analysis of an International Speech Test Signal (ISTS)," *Int. J. of Audiology*, vol. 49, no. 12, pp. 891–903, 2010.

[23] M. Geronazzo, S. Spagnol, and F. Avanzini, "Do we need individual head-related transfer functions for vertical localization? The case study of a spectral notch distance metric," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 26, no. 7, pp. 1247–1260, 2018.

[24] M. Cuevas-Rodriguez, D. Gonzalez-Toledo, A. Reyes-Lecuona, and L. Picinali, "Impact of non-individualised head related transfer functions on speech-in-noise performances within a synthesised virtual environment," *J. Acoust. Soc. Am.*, vol. 149, no. 4, pp. 2573–2586, 2021.

[25] G. D. Romigh and B. D. Simpson, "Do you hear where I hear?: Isolating the individualized sound localization cues," *Front. in Neurosci.*, vol. 8, 2014.

[26] A. Andreopoulou and B. F. G. Katz, "Comparing the effect of HRTF processing techniques on perceptual quality ratings," in *Proc. Audio Eng. Soc. (AES) Conv.*, p. 9920, 2018.

[27] D. Poirier-Quinot and B. F. G. Katz, "On the improvement of accommodation to non-individual HRTFs via VR active learning and inclusion of a 3D room response," *Acta Acust.*, vol. 5, no. 25, 2021.

[28] F. Brinkmann, L. Aspöck, D. Ackermann, S. Lepa, M. Vorländer, and S. Weinzierl, "A round robin on room acoustical simulation and auralization," *J. Acoust. Soc. Am.*, vol. 145, no. 4, pp. 2746–2760, 2019.