DOI: 10.61782/fa.2025.0643

ON COMPARING AUDITORY MODELS AND PERCEPTUAL ASSESSMENT WHEN RATING HEAD-RELATED TRANSFER FUNCTIONS

Rapolas Daugintis[1*] **Michele Geronazzo**[1,2] **Lorenzo Picinali**[1]
[1] Audio Experience Design, Dyson School of Design Engineering, Imperial College London, UK
[2] Department of Engineering and Management, University of Padova, Italy

ABSTRACT

Predicting the perceived quality of binaural audio with different head-related transfer functions (HRTFs) is essential when attempting to automate improvements to spatial audio rendering. To assess the selection accuracy of a numerical HRTF matching algorithm based on computational auditory model estimates, this study compares its results with the findings of a subjective HRTF rating study. In a previously published behavioural experiment, participants rated various HRTFs from the LISTEN database. The procedure was based on noise bursts rendered at different positions along horizontal and vertical trajectories. Possible ratings included 'bad', 'ok', or 'excellent'. In the numerical selection, one 'best' and one 'worst' non-individual HRTFs are chosen from the dataset based on estimated polar and quadrant errors from a modelled localisation experiment with static sound sources. The results indicate an above-chance probability that the HRTF selected as the 'best' using the numerical method would be rated as 'excellent' or at least 'ok' with the behavioural one. However, limitations of the preliminary results can be ascribed to the challenges of repeatability in the subjective listening tests, discrepancies between the two methods (rating based on static vs. dynamic sounds) and differences in metrics (localisation performances vs. subjective ratings).

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

Keywords: computational auditory model, HRTF personalisation, perceptual listening tests.

1. INTRODUCTION

Binaural spatial audio rendering is achieved by convolving monophonic sound sources with head-related transfer function (HRTF) of desired directions [1]. When an individual HRTF is not available, a well-matched non-individual one might still provide a comparable perception of sound location or rendering quality to the listener [2]. However, HRTF matching is challenging due to the multifaceted perceptual nature of the problem [3]. A listening test-based selection might be considered the most faithful approach, but it can be demanding to listeners, while the reliability of the test could vary [4–6]. Instead, numeric methods can provide an efficient and automated way to personalise HRTFs [7], but they require an appropriate perceptually motivated HRTF similarity metric.

This study is part of a more extensive project that aims to establish a relationship between numerical and perceptual HRTF personalisation approaches [8,9]. Here, results are compared between HRTF selection based on a computational auditory model [8] and a perceptual rating [10]. The ability to explain subjective rating results with the use of an auditory model would provide a validation of the latter to be employed for numerical HRTF selection.

2. METHODOLOGY

2.1 Subjective listening test

The procedure for the subjective HRTF rating, based on a listening test, is detailed in [10]. In summary, 45 subjects participated in a test in which they rated 46 HRTFs [1] . The participants listened to a series of repeated noise bursts, rendered in horizontal (every 30°) and vertical (every 15°) trajectories with a corresponding HRTF. They had to rate each HRTF according to how well the rendering corresponded to the trajectories, assigning them one of three available ratings: ‘bad’, ‘ok’, or ‘excellent’. For the purposes of this study, the results of 44 out of 45 subjects and 45 out of 46 HRTFs are used because one of the HRTFs was unavailable.

2.2 Model-based HRTF selection

Numerically, HRTFs were selected using an auditory sound localisation model, described in [11]. The model is based on the template matching procedure: The HRTF of interest (known as the *target*) is processed to extract monaural and binaural features, which are compared to the features of the individual HRTF (the *template*) in a stochastic way, accounting for the limitations of the hearing system. The model is used to estimate the human sound localisation error as if the participant (whose *template* HRTF was used) performed a sound localisation test in which a static broadband noise burst was rendered via the *target* HRTF of a desired direction.

For the purposes of this study, a selection of the ‘best’ and the ‘worst’ non-individual HRTF was made by comparing the predicted local polar error and quadrant confusion rate (based on the definitions in [12]), across different non-individual HRTFs. Initially, the classification methodology described in [8], which only considers errors for sound source directions in front of the listener, was used (referred to as the *original* method). It first categorises the HRTFs as ‘good’ or ‘bad’ based on the normality of the polar error distributions in this area and then selects one ‘best’ HRTF from the ‘good’ ones and one ‘worst’ from the ‘bad’ ones.

Since the subjective test was conducted using the specific directions along the two trajectories, a second version of the model selection was made, where errors were calculated using the specific directions used in the listening test (the *trajectory* method). In this case, error distributions were not analysed, so only the ‘best’ and the ‘worst’ HRTF were selected: The ‘best’ HRTF was selected from the lowest polar and quadrant errors (as described in [8]) and the ‘worst’ was defined as the HRTF which resulted

[1] The set included individual HRTFs, which were indicated to the participants. However, individual HRTF ratings are not considered in this study.

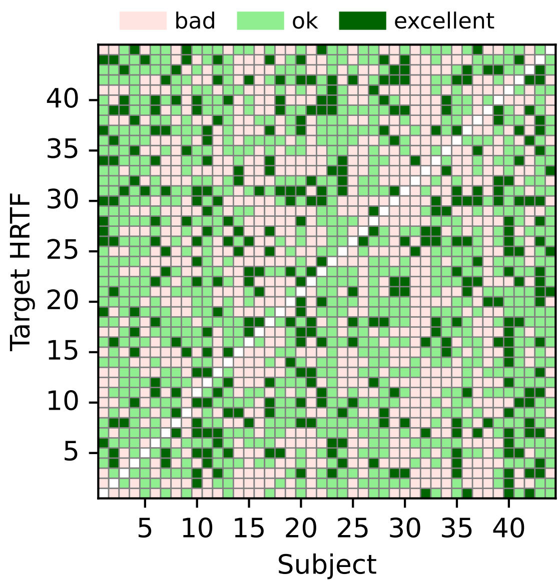

Figure 1. Subjective non-individual HRTF ratings. Figure reproduced from [10].

in the highest quadrant error rate estimate. In both cases, only non-individual HRTFs were used in the selection procedure, because including individual HRTFs would have resulted in them being selected as the ‘best’ HRTFs.

3. RESULTS

Fig. 1 reproduces the subjective rating results, presented in [10, Fig. 1] for non-individual HRTFs, considered in this study. In total, 17.9% of the time the HRTFs were rated as ‘excellent’, 42.0% as ‘ok’ (making 59.9% of subjective ratings at least ‘ok’), and 40.1% as ‘bad’. These percentages indicate the chance level for selecting an HRTF to be in one of the categories.

Fig. 2 shows model-based selection, following the *original* method. Comparing the model-based selection with the perceptual ratings, the ‘best’ HRTFs according to the model were rated as ‘excellent’ 15.9% of the time while 59.1% of the time they were rated to be at least ‘ok’. These levels are similar to chance, indicating poor agreement between the model and the subjective test. On the other hand, the ’worst’ HRTFs, according to the model-based selection, were rated as ‘bad’ in the perceptual test at 52.3%, indicating an above chance level of correspondence between the two methods.

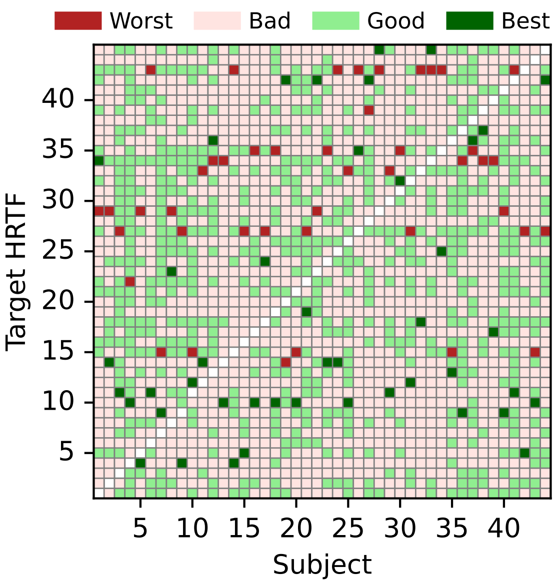

Figure 2. Auditory model-based HRTF selection (*original* method).

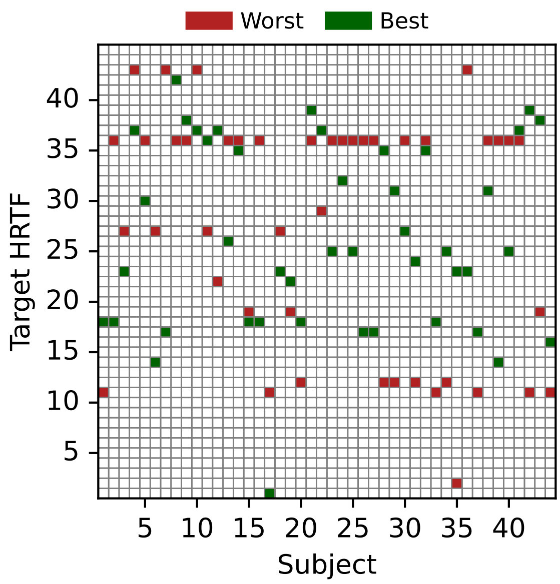

Figure 3. Auditory model-based HRTF selection (*trajectory* method).

Fig. 3 shows the results of the selection when the directions along the trajectories used in the subjective test were considered. In this case, 29.5% of the 'best' HRTFs were rated as 'excellent' in the listening test and 77.3% were rated to be at least 'ok'. These values are above the estimated chance level, indicating some level of correlation between the modelled and the subjective results. However, 45.5% of the 'worst' HRTFs were rated as 'bad' in the subjective test, which is only slightly above chance level and lower than with the *original* method.

4. DISCUSSION

The limited success of the modelling results can be attributed both to the challenges of consistency and repeatability of the subjective listening tests (the task of rating 46 HRTFs one after another in a consistent fashion is challenging, especially for less experienced participants [5]) and to the limitations of the model-based selection procedure. The auditory model was specifically designed to replicate a static sound localisation task. However, the subjective test used predefined trajectories to present the HRTFs, so the participants were performing a comparison between the expected sound location and the binaural rendering instead of a static blind localisation task. Studies comparing similar HRTF assessment methods and metrics indicated that the correlation between different methods depended on the specific metrics used [13] and the listeners who performed the tests [6].

The metrics used for the model-based selection might have to be reconsidered to better correspond to the perceptual task. For example, quadrant ambiguity might play a smaller role when listening to a sound source from a known direction, which would explain the poor agreement between the 'worst' selected HRTFs and their subjective ratings. Instead, a polar error might be a better metric for the selection of the 'worst' HRTF. Furthermore, the model, which uses multiple noise parameters to predict individual localisation responses based on real sound localisation performance, was not calibrated to each subject due to the unavailability of the required localisation data. Therefore, individual-level predictions might be inaccurate. Finally, the interaural time differences (ITDs) of the HRTFs used for the subjective test were altered to match each subject's ITD, whereas the HRTFs used in the model selection had unaltered ITDs. Model-based selection only considered errors along the confusion cones, mainly dependent on spectral cue mismatch [14], while the perceptual judgement of horizontal trajectory rendering could have been dominated by binaural cue matching, not taken into account by the chosen estimated metrics.

5. CONCLUSION

Comparison of the results of subjective HRTF evaluation and a non-individual HRTF selection based on a computational auditory model indicates little correlation between the numerical results using the *original* method, proposed in [8] and the perceptual evaluation, presented in [10]. When the *original* method is adjusted to account for the specific directions along the trajectories, used in the subjective evaluation, the correlation between some of the ratings improves to the above-chance level. This result indicates that the predicted localisation error could be a partial indicator of perceived HRTF rendering quality. However, some limitations of the results can be attributed to differences between the model's intended scenario and the exact setup of the subjective test and repeatability challenges of a subjective listening test. More detailed result analysis is needed to find a metric that best correlates with the perceptual results.

6. ACKNOWLEDGMENTS

This research has been supported by the SONICOM project (`www.sonicom.eu`), funded by the Horizon 2020 research and innovation programme of the European Union under grant agreement no. 101017743. The authors thank Brian F. G. Katz for providing the data from the original study.

7. REFERENCES

[1] X.-l. Zhong and B.-s. Xie, "Head-Related Transfer Functions and Virtual Auditory Display," in *Soundscape Semiotics - Localisation and Categorisation* (H. Glotin, ed.), InTech, Mar. 2014.

[2] 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, 2023.

[3] L. S. R. Simon, N. Zacharov, and B. F. G. Katz, "Perceptual attributes for the comparison of head-related transfer functions," *J. Acoust. Soc. Am.*, vol. 140, no. 5, pp. 3623–3632, 2016.

[4] A. Andreopoulou and B. F. G. Katz, "Subjective HRTF evaluations for obtaining global similarity metrics of assessors and assessees," *J. Multimodal User Interfaces*, vol. 10, pp. 259–271, Sept. 2016.

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

[6] 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.

[7] D. Fantini, M. Geronazzo, F. Avanzini, and S. Ntalampiras, "A survey on machine learning techniques for head-related transfer function individualization," *IEEE Open J. Signal Process.*, vol. 6, pp. 30–56, 2025.

[8] 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. Acoust. Speech Signal Process. (ICASSP)*, June 2023.

[9] R. Daugintis, R. Barumerli, M. Geronazzo, and L. Picinali, "Initial Evaluation of an Auditory-Model-Aided Selection Procedure for Non-Individual HRTFs," in *Proc. Conv. EAA Forum Acusticum*, Sept. 2023.

[10] 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.

[11] 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.

[12] 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.

[13] F. Zagala, M. Noisternig, and B. F. G. Katz, "Comparison of direct and indirect perceptual head-related transfer function selection methods," *J. Acoust. Soc. Am.*, vol. 147, no. 5, pp. 3376–3389, 2020.

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