CLASSIFYING NON-INDIVIDUAL HEAD-RELATED TRANSFER FUNCTIONS WITH A COMPUTATIONAL AUDITORY MODEL: CALIBRATION AND METRICS

Rapolas Daugintis[⋆] *Roberto Barumerli*[†] *Lorenzo Picinali*[⋆] *Michele Geronazzo*[⋆‡]

[⋆]Audio Experience Design, Dyson School of Design Engineering, Imperial College London, UK
[†]Acoustics Research Institute, Austrian Academy of Sciences, 1040 Vienna, Austria
[‡]Department of Engineering and Management, University of Padova, Italy

ABSTRACT

This study explores the use of a multi-feature Bayesian auditory sound localisation model to classify non-individual head-related transfer functions (HRTFs). Based on predicted sound localisation performance, these are grouped into 'good' and 'bad', and the 'best'/'worst' is selected from each category. Firstly, we present a greedy algorithm for automated individual calibration of the model based on the individual sound localisation data. We then discuss data analysis of predicted directional localisation errors and present an algorithm for categorising the HRTFs based on the localisation error distributions within a limited range of directions in front of the listener. Finally, we discuss the validity of the classification algorithm when using averaged instead of individual model parameters. This analysis of auditory modelling results aims to provide a perceptual foundation for automated HRTF personalisation techniques for an improved experience of binaural spatial audio technologies.

***Index Terms*—** HRTF personalisation, auditory modelling, human sound localisation, Bayesian inference

1. INTRODUCTION

Binaural spatial audio technologies rely on what are known as head-related transfer functions (HRTFs), which are convolved with monophonic sound signals to render virtual spatial sound scenes for headphone use. The HRTFs contain spatial hearing cues arising from the individual morphology of the outer ear, head, and torso, which modify the sound reaching the ear canal. The human auditory system learns to detect these unique spectral differences in the incoming sound (known as monaural hearing cues) as well as differences between the sound coming to both ears (binaural hearing cues) to infer the direction of the incoming sound [1].

For the best rendering results, individual HRTFs can be acoustically measured in a laboratory environment; however, detailed measurements for each individual on a mass scale are labour-intensive and thus unfeasible. To retain a satisfactory spatialisation quality when individual HRTFs are not available, various methods have been proposed [2], some of which are based on selecting a non-individual HRTF. The latter methods rely on a chosen HRTF similarity metric, which can be qualitative (e.g. based on subjective human ratings [3]) or quantitative (e.g. spectral distance [4], or geometric pinnae matching based on notch frequency analysis [5, 6] or computer vision techniques [7]). While the subjective tests are known to produce inconsistent results, especially for naïve listeners [8, 9], the quantitative numerical similarity metrics lack the perceptual foundations to describe the quality of the HRTF fit in a meaningful way (they may be well suited for machine audition aiming for the 'best localisation' but not necessarily a human one).

Machine learning has been used to develop a more perceptually informed HRTF distance metric based on human sound localisation data [10]. Other HRTF personalisation studies used machine learning techniques with psychoacoustic input in stages of, e.g. feature generation [11] or dimensionality reduction [12]. Since the HRTF selection problem is constrained by limited datasets and perceptual complexities of the auditory system, domain knowledge can help develop more efficient and precise automated HRTF fitting procedures.

A few quantitative non-individual HRTF selection studies have used computational auditory models to evaluate the quality of spatial processing without extensive listening tests [13, 14]. The human sound localisation models used in these studies estimate the localisation error, focusing on the sound direction along a sagittal plane, where listener uncertainty is the most impactful [15, 16]. They rely on the assumption that the auditory system has learned the mapping between spatial features and sound direction [17]. This procedure, known as template matching, enables the subject to estimate the source direction from the observed spatial features. To account for the variation across humans in distinguishing the sound localisation cues, the models have a set of parameters, which can be tuned based on individual sound localisation data to accurately predict individual localisation performance. Nevertheless, since collecting the individual localisation data is time-consuming, previous model-aided HRTF selection studies relied on average non-individual model parameters [13, 18]. However, the impact of additional modelling uncertainty due to such model simplification on an HRTF classification task has not yet been addressed.

Emphasising the synergy between human (perceptual) and machine (numerical) approaches [19], we propose a set of automated non-individual HRTF classification criteria based on the sound localisation performances simulated with an individually-calibrated computational auditory model (Sec. 2 and 3). Due to the complexity of sound localisation error distributions over different directions around the listener, we implemented a novel methodology that accounts for the error distributions within a range of directions in front of the listener, rather than aggregated global error values, to classify non-individual HRTFs as 'good' or 'bad' and select one 'best'/'worst', representing the extremes of the proposed metric (Sec. 4). In Sec. 5, we discuss the validity of using the proposed selection methodology with individual and non-individual (averaged) model parameters. Finally, we summarise the work and propose the potential use cases of the methodology in Sec. 6.

This research is part of the SONICOM project (EU Horizon 2020 RIA grant agreement ID: 101017743). RB is also supported by Dynamates (Austrian Science Fund (FWF) project No. ZK 66). Finally, the authors want to thank Piotr Majdak for giving access to the data used in this study.

2. THE COMPUTATIONAL MODEL

The auditory model used in this study (available as *barumerli2022* from the Auditory Modelling Toolbox (AMT) [20]) is a Bayesian model based on the template-matching procedure [21]. When provided with an incoming binaural sound as an input, the model uses cross-correlation to extract interaural time difference (x_{itd}), envelope time-averaging to obtain interaural level difference (x_{ild}), and gammatone filter bank to acquire spectral amplitudes of the directional transfer function (DTF) [1] for the left and the right ears, respectively ($\boldsymbol{x}_{L,DTF}$ and $\boldsymbol{x}_{R,DTF}$)[2]. The obtained input feature vector

$$\boldsymbol{t} = [x_{itd}, x_{ild}, \boldsymbol{x}_{L,DTF}, \boldsymbol{x}_{R,DTF}] + \boldsymbol{\delta} \tag{1}$$

is compared with the stored internal template ($\boldsymbol{T}(\boldsymbol{\varphi})$) of the sound coming from all possible directions around the simulated listener to infer the vertical and horizontal directions ($\boldsymbol{\varphi}$) of the sound source.

The model adds multivariate Gaussian noise $\boldsymbol{\delta} \sim \mathcal{N}(0, \boldsymbol{\Sigma})$ with a diagonal covariance matrix $\boldsymbol{\Sigma}$, controlled via three independent parameters σ_{itd}, σ_{ild}, and σ_{mon}, to the target vector to simulate human performances when inferring the sound source direction. The Maximum a-posteriori estimation (MAP) is then employed by the model to determine the direction of the sound source:

$$\hat{\boldsymbol{\varphi}} = \arg\max_{\boldsymbol{\varphi}} p(\boldsymbol{t}|\boldsymbol{T}(\boldsymbol{\varphi}))\, p(\boldsymbol{\varphi}) + \boldsymbol{m}, \tag{2}$$

where the likelihood $p(\boldsymbol{t}|\boldsymbol{T}(\boldsymbol{\varphi}))$ is weighted by a prior $p(\boldsymbol{\varphi})$, modelled uniformly across the azimuth but normally distributed for the elevation (with zero mean and standard deviation σ_{prior}) to represent human localisation bias towards the eye level [22]. Additionally, the estimate is corrupted by response noise $\boldsymbol{m}$, representing the uncertainty in the action of pointing towards a sound source during a localisation test, which is modelled as a von Mises-Fisher distribution with a standard deviation of σ_{motor}.

The quantities σ_{itd}, σ_{ild}, σ_{mon} (corresponding to $\boldsymbol{x}_{L/R,DTF}$), σ_{prior}, and σ_{motor} are tuned individually to produce accurate human-like localisation responses. To account for the model stochasticity and obtain stable estimates, each target direction is iterated multiple times (300 repetitions were found to produce converging results [21]).

2.1. Data and performance metrics

We used data from the AMT collected as part of two previous studies (6 subjects from [23] and 11 from [24]). The datasets contain subjects' individual HRTFs and their responses from a sound localisation test using short broadband noise bursts in a virtual reality environment. In each simulation of the localisation test, the model was supplied with the individual HRTF as the template and one of the HRTFs as the target, both resampled from 1550 points to a uniform grid of 1500 directions using spherical harmonic interpolation [25].

To separate the directions, dominated by the binaural vs monaural cues, an interaural-polar coordinate system was used for localisation data analysis [26]. It decouples the location on an imaginary sphere around a listener $\boldsymbol{\varphi} = (\theta, \phi)$ into a lateral angle θ, which represents the horizontal angle between the median plane and the sagittal plane of interest, and a polar angle ϕ, defined anticlockwise from the front along the sagittal plane. A graphical representation can be found in [24].

[1] The DTF is obtained by removing the directionally independent part of the HRTF, known as the Common Transfer Function (CTF) [4].

[2] Among the different feature vectors proposed by [21], we selected the ones more consistent with past literature [17].

Algorithm 1 Greedy *barumerli2022* calibration algorithm

```
σ_ild = 1, σ_mon = 5, σ_motor = 10                ▷ Initialise parameters
Calculate ε_LE, ε_PE, ε_QE(RAU)                   ▷ Quick model setup
while ε_QE(RAU) ≥ 0.2 do:
    σ_mon −= sign ε_QE(RAU)
    Calculate ε_LE, ε_PE, ε_QE(RAU)               ▷ Quick setup
    if σ_mon ≤ 2 or σ_mon ≥ 8 then break
while ε_PE ≥ 0.15 do:
    σ_motor −= sign ε_PE
    Calculate ε_LE, ε_PE, ε_QE(RAU)               ▷ Quick setup
    if σ_motor ≤ 5 or σ_motor ≥ 25 then break
if ε_LE ≥ 0.1 then:
    Check if ε_LE reduces with σ_ild = 0.5        ▷ Quick setup
    while ε_LE ≥ 0.1 do:
        σ_motor −= sign ε_LE
        Calculate ε_LE, ε_PE, ε_QE(RAU)           ▷ Quick setup
        if σ_motor ≤ 5 or σ_motor ≥ 25 then break
Calculate ε_LE, ε_PE, ε_QE(RAU) to validate       ▷ Full setup
```

In a localisation test with N trials, each target source direction $\boldsymbol{\varphi}_i$ has an associated response direction $\tilde{\boldsymbol{\varphi}}_i$, where $i = 1, 2, \ldots, N$. Following the description of [26], we define a set of local responses $\mathcal{A} = \left\{ i : \text{wrap}\left|\tilde{\phi}_i - \phi_i\right| < 90° \right\}$ and three errors in the interaural-polar coordinate system:

$$PE = \sqrt{\frac{\sum_{i \in \mathcal{A}} \left(\text{wrap}\left(\tilde{\phi}_i - \phi_i\right)\right)^2}{|\mathcal{A}|}}, \tag{3}$$

$$QE = \left(1 - \frac{|\mathcal{A}|}{N}\right) \times 100\%, \tag{4}$$

$$LE = \sqrt{\frac{\sum_{i=1}^{N} \left(\text{wrap}\left(\tilde{\theta}_i - \theta_i\right)\right)^2}{N}}, \tag{5}$$

where root-mean-square (rms) local polar error (PE) is the aggregated error in the polar dimension for responses within 90° from the target, quadrant error rate (QE) corresponds to the percentage of polar errors larger than 90° and accounts for top-down and front-back confusions, and LE is the rms lateral error. Angle differences are wrapped to $[-180°, 180°]$ range. To avoid highly distorted polar errors on the far left and right sides of the listener, PE and QE are only defined for targets within lateral angle $|\theta| \leq 30°$, while LE is defined within $|\theta| \leq 60°$ to avoid a ceiling effect [26].

3. MODEL CALIBRATION

The free model parameters were tuned based on the individual data from the real localisation tests. The model parameters for the first five subjects had already been reported in [21]. Following that calibration procedure, the model was fitted to additional eleven subjects using the greedy algorithm, formalised in Alg. 1, which was empirically developed to produce the most stable results [21]. For each subject, the model was supplied with individual HRTFs as both the template and the target, and relative differences between aggregated real and simulated errors, ϵ_{PE}, ϵ_{LE}, and $\epsilon_{QE}(RAU)$, were computed. For the calibration procedure only, QE percentage values were transformed to a scale comparable to the other two angle-based metrics using a rationalised arcsine transform (RAU) [27].

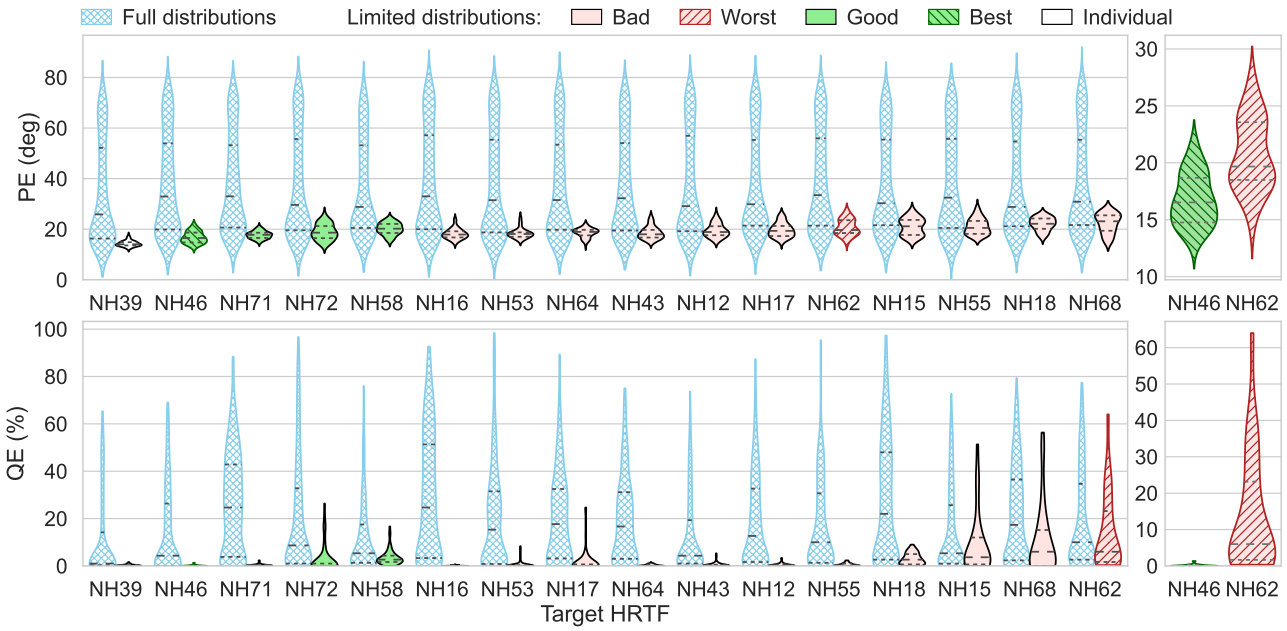

Fig. 1. Violin plots of modelled error distributions for subject NH39 with different target HRTFs. Dashed lines indicate quartiles. Full distributions across the listener sphere are plotted next to the distributions within a limited range of $|\phi| \leq 11.5°$ in the front. The non-individual HRTFs are categorised and arranged from left to right based on PE and QE values, respectively, according to the proposed procedure (Sec. 4). The limited distributions for the 'best' and 'worst' HRTFs are plotted separately on the right.

The values of $\sigma_{itd} = 0.569$ and $\sigma_{prior} = 11.5°$ were set based on derivations from previous psycho-acoustics experiments [28, 22]. σ_{mon}, σ_{motor} and σ_{ild} were tuned in the specified order to reduce ϵ's below previously reported thresholds [21]. A quick model setup with 5 repetitions was used during the adjustment process to reduce computational time, while a full setup (300 repetitions) was run at the end to validate the calibration. Termination conditions were added to keep the parameters within stable bounds, but none of the subjects exceeded them. Once the automated calibration was finished, the parameters were fine-tuned to one decimal place to reduce the ϵ's further by running the model manually (a more precise automated calibtation procedure is subject to future research).

4. HRTF CLASSIFICATION PROCEDURE

Localisation errors were modelled for every subject using each of the 16 HRTFs (1 individual and 15 non-individuals). Contrary to the calibration procedure using global errors, the errors were aggregated only over the 300 repetitions for each simulated direction, resulting in 16 distributions per subject per error. Full sphere PE distributions appeared to be multimodal for different areas around the listener, so we limited the directions of interest to a section in front of the listener within $\pm 11.5°$ polar angle, equivalent to σ_{prior}. In this range, we expected a more direction-independent localisation performance according to [22]. We hypothesised that a 'good' non-individual HRTF would correspond to a normally distributed PE within this frontal area. At the same time, skewed or multimodal distribution would indicate that a modelled listener cannot effectively use the localisation cues embedded in the HRTF and might rely too much on prior beliefs. Since LE shows less variability among non-individual HRTFs [29], we focused on PE and QE. The classification was performed in the following steps **for each subject**, separately:

1. divide target HRTFs into 'good' ($p > 0.05$) and 'bad' ($p < 0.05$) based on the Shapiro-Wilk test on the PE distributions of the 15 simulated localisation tests;
2. **'good' HRTFs**: order PE distributions of 'good' HRTFs by rms of the distribution and their QE distributions by the 3rd quartile, respectively;
3. **'bad' HRTFs**: order QE distributions by the median.

The 'best' and the 'worst' non-individual HRTFs for each subject were selected from the 15 available using the following criteria:

1. **The 'best' HRTF**: Look for a common HRTF in top $n = 1$ 'good' PE and QE distributions. Increase n until the intersection is found.
2. **The 'worst' HRTF**: select the highest median QE; If multiple HRTFs have the same highest median QE then base the selection on the 3rd quartile, or rms PE value if QE distributions are identical.

Finally, we repeated the classification procedure for each subject using non-individual (median) model parameters to investigate how the results are affected when individual model parameters are not used.

5. RESULTS AND DISCUSSION

Table 1 reports individual model parameters for 11 of the subjects used in the study, obtained from the calibration procedure detailed in Sec. 3. Median parameter values across all 16 subjects (including 5 from the previous study), used to test the selection methodology with non-individual model parameters, are also presented.

Fig. 1 shows an example of modelled PE and QE distributions across different target HRTFs for subject NH39, using individual model parameters. The figure presents error distributions for all directions (within $|\theta| \leq 30°$ as per error definitions) and the limited

Table 1. Model parameters for 11 individual subjects, and median parameter values across 16 subjects.

Subject	Parameters		
	σ_{ild} (dB)	σ_{mon} (deg)	σ_{motor} (deg)
NH12, 15-18	*available from [21]*		
NH39	1	5.4	11.8
NH43	1	2.2	20.5
NH46	1	3.2	13.8
NH53	1	2.4	11.1
NH55	0.5	5.1	17.5
NH58	0.5	3.1	9.2
NH62	1	7.2	13.5
NH64	0.5	4.7	13.2
NH68	0.5	7.5	13.4
NH71	0.5	4.5	13.6
NH72	1	6.9	11.1
Median	**0.75**	**4.3**	**13.45**

ones in the front, colour-coded based on the PE normality criterion. It also includes distributions from the individual HRTF which, as expected, are generally smaller than from the non-individual HRTFs.

Generally, medians of direction-limited PE and QE distributions for 'good' non-individual HRTFs are one of the smallest ones across the simulated HRTFs. The imposed normality criteria may sometimes place HRTFs with median PE, similar to the 'good' HRTFs, in the 'bad' category (e.g. NH16, NH53, or NH64). However, when the full-sphere error distributions are considered, the 'bad' HRTFs appear to be associated with higher overall QE rates. This could suggest that the strong evidence of bimodality of the direction-limited PE distribution may act as a good indicator for a higher QE in a full localisation test with the particular target HRTF.

The error distributions of the 'best' and the 'worst' target HRTFs for subject NH39 are also plotted separately in Fig. 1. The bimodal nature of the PE distribution for NH62 is clearly visible in the violin plot, while the evidence for the bimodality of the NH46 is not strong enough for the normality hypothesis to be rejected. NH46 is also associated with negligibly small QE, compared to NH62.

Fig. 2 summarises the classification results for all 16 subjects using both individual and non-individual (median) model parameters. The number of HRTFs classified as 'good' varies considerably across subjects. For instance, when considering the individualised predictions, 11 out of 15 non-individual HRTFs are classified as 'good' for subject NH18, while none passed the normality criterion for subject NH72. Although the results might be affected by the limited HRTF selection pool, some subjects appear to be responsive to a wider range of sound localisation cues than others. However, the phenomenon is not reciprocal, i.e. HRTF of a responsive subject is not generally categorised as 'good' for other subjects. In fact, NH18 is marked as the 'worst' for 3 of the subjects. The addition of noise across various stages of the auditory model makes the classification non-commutative. This asymmetry is in line with the outcomes from previous perceptual studies [3, 30]. On the other hand, some HRTFs are often selected as the 'worst' (e.g. NH39 and NH62). The repeated classification suggests that those HRTFs may have less distinguishable spectral cues compared to the rest of the dataset and could be discarded in a process of database optimisation for non-individual HRTF selection (similarly to [13]).

When the non-individual model parameters are used, the resulting HRTF selection is affected, as indicated by the colour mismatch

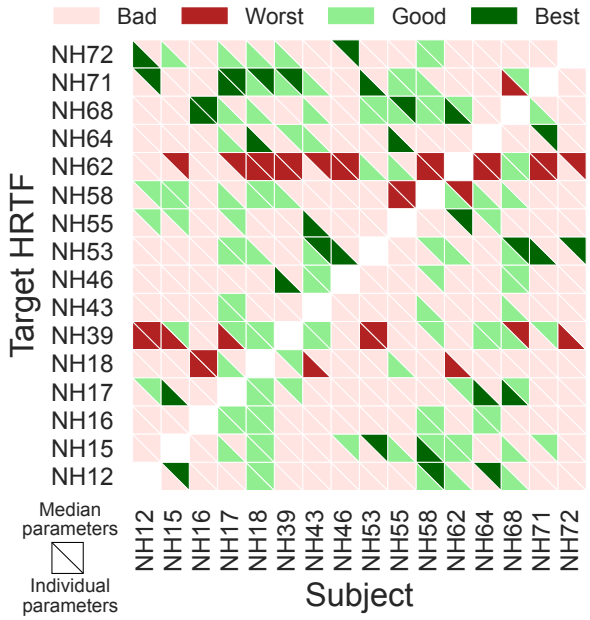

Fig. 2. Non-individual HRTF selection using individual (bottom left half of the cells) and median (top right half) model parameters.

within each cell of Fig. 2. Generally, the selection methodology for the 'worst' HRTF appears to be more robust to model parameter changes than the one for the 'best', indicating a relationship between the parameters and the shape of the PE distributions. The HRTF categorisation is also more robust to parameter changes for some subjects. For example, the selection for NH16 does not depend on the model parameters, while 8 HRTFs are misclassified for NH68, including the 'worst' HRTF, which is considered 'good' if individual model parameters are used. A more systematic analysis of the effect of the parameter change on the HRTF selection is required in the future. Nevertheless, the results demonstrate that one must take care when interpreting results from sound localisation models without individually calibrated parameters.

6. CONCLUSION

Using the individually calibrated Bayesian sound localisation model, this study presented non-individual HRTF classification criteria based on distributions of predicted polar and quadrant localisation errors within a limited directional range in front of the listener ($|\theta| \leq 30°$ and $|\phi| \leq 11.5°$). For 16 subjects, each of the 15 non-individual HRTFs was classified into 'good' and 'bad' categories and the 'best'/'worst' was selected from each group. Misclassification of non-individual HRTFs when the same procedure was performed with non-individual model parameters revealed the sensitivity of the auditory model to parameter changes. Evaluations of the selection methodology with listening tests are currently ongoing. Once perceptually validated, this method could be used as a metric in combination with other approaches, e.g. geometric pinnae matching or HRTF selection based on a sparse, 'low-quality' HRTF, measured from a few positions in uncontrolled conditions, allowing to consistently and repeatedly select a perceptually well-matched laboratory-measured HRTF.

7. REFERENCES

[1] J. Blauert, *Spatial Hearing: The Psychophysics of Human Sound Localization*, 2nd ed. Cambridge, MA: The MIT Press, 1996.

[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. Cham: Springer International Publishing, 2023, pp. 115–143.

[3] 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.

[4] 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.

[5] M. Geronazzo, S. Spagnol, A. Bedin, and F. Avanzini, "Enhancing vertical localization with image-guided selection of non-individual head-related transfer functions," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, 2014, pp. 4463–4467.

[6] 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. Acoust. Speech Signal Process. (ICASSP)*, 2020, pp. 411–415.

[7] B. Zhi, D. N. Zotkin, and R. Duraiswami, "Towards fast and convenient end-to-end HRTF personalization," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, 2022, pp. 441–445.

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

[9] 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.

[10] 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, 2021, 044401.

[11] X. Qi and L. Wang, "Parameter-transfer learning for low-resource individualization of head-related transfer functions," in *Interspeech*, 2019, pp. 3865–3869.

[12] F. Grijalva, L. Martini, D. Florencio, and S. Goldenstein, "A manifold learning approach for personalizing HRTFs from anthropometric features," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 24, no. 3, pp. 559–570, 2016.

[13] 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 Lang. Process.*, vol. 26, no. 7, pp. 1247–1260, 2018.

[14] M. Warnecke, S. Jamison, S. Prepeliţă, P. Calamia, and V. K. Ithapu, "HRTF personalization based on ear morphology," in *Proc. Audio Eng. Soc. (AES) Conf. on Audio for Virtual & Augmented Reality*, 2022.

[15] R. Baumgartner, P. Majdak, and B. Laback, "Assessment of sagittal-plane sound localization performance in spatial-audio applications," in *The Technology of Binaural Listening*, J. Blauert, Ed. Berlin, Heidelberg: Springer, 2013, pp. 93–119.

[16] R. Baumgartner, P. Majdak, and B. Laback, "Modeling sound-source localization in sagittal planes for human listeners," *J. Acoust. Soc. Am.*, vol. 136, no. 2, pp. 791–802, 2014.

[17] J. C. Middlebrooks, "Narrow-band sound localization related to external ear acoustics," *J. Acoust. Soc. Am.*, vol. 92, no. 5, pp. 2607–2624, 1992.

[18] R. Barumerli, M. Geronazzo, and F. Avanzini, "Localization in elevation with non-individual head-related transfer functions: Comparing predictions of two auditory models," in *Proc. Eur. Signal Process. Conf. (EUSIPCO)*, 2018, pp. 2539–2543.

[19] L. M. Heller, B. Elizalde, B. Raj, and S. Deshmukh, "Synergy between human and machine approaches to sound/scene recognition and processing: An overview of ICASSP special session," *arXiv:2302.09719*, 2023.

[20] 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.

[21] R. Barumerli, P. Majdak, M. Geronazzo, F. Avanzini, D. Meijer, and R. Baumgartner, "A Bayesian model for human directional localization of broadband and static sound sources," *bioRxiv:2022.10.25.513770*, 2022.

[22] R. Ege, A. J. V. Opstal, and M. M. Van Wanrooij, "Accuracy-precision trade-off in human sound localisation," *Sci. Rep.*, vol. 8, no. 16399, 2018.

[23] 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. Psychophys.*, vol. 72, no. 2, pp. 454–469, 2010.

[24] P. Majdak, T. Walder, and B. Laback, "Effect of long-term training on sound localization performance with spectrally warped and band-limited head-related transfer functions," *J. Acoust. Soc. Am.*, vol. 134, no. 3, pp. 2148–2159, 2013.

[25] D. N. Zotkin, R. Duraiswami, and N. A. Gumerov, "Regularized HRTF fitting using spherical harmonics," in *Proc. IEEE Workshop Appl. Signal Process. Audio Acoust. (WASPAA)*, 2009, pp. 257–260.

[26] 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.

[27] G. A. Studebaker, "A "rationalized" arcsine transform," *J. Speech Lang. Hear. Res.*, vol. 28, no. 3, pp. 455–462, 1985.

[28] J. Reijniers, D. Vanderelst, C. Jin, S. Carlile, and H. Peremans, "An ideal-observer model of human sound localization," *Biol. Cybern.*, vol. 108, no. 2, pp. 169–181, 2014.

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

[30] 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.