DOI: 10.61782/fa.2025.0434

FORUM ACUSTICUM EURONOISE 2025

HRTF DATABASE MATCHING FOR LOCALIZATION ACCURACY APPLICATIONS

Eleni Tavelidou **Konstantinos Bakogiannis** **Areti Andreopoulou***
Laboratory of Music Acoustics and Technology (LabMAT)
National and Kapodistrian University of Athens, Greece

ABSTRACT

This study investigates HRTF database matching as a means for creating an individualized spatial user experience, optimized for specific application needs. Focusing on localization accuracy, a procedure is designed assisting the effective navigation, evaluation, and selection of the optimal dataset for each user from a large HRTF repository. The navigation within this HRTF collection is achieved by means of clustering and the identification of representative datasets, to be evaluated in a binaural localization test, developed in Unity using the 3DTI toolkit. The winning HRTF in each localization test guides further clustering and refinement, iteratively narrowing the options to a few optimal datasets. The dataset pool was constructed using Barumerli's model and assessed using metrics derived from the localizationerror function in the Auditory Modelling Toolbox. Principal Component Analysis (PCA) reduced redundancy and noise, enabling more coherent clustering with the k-means technique. The effectiveness of the proposed methodology is assessed through a preliminary study featuring an HRTF pool of 84 datasets from five databases, and seven participants. This paper details the methodology, presents findings, and explores the evaluation, limitations, and potential applications of this technique.

Keywords: *binaural hearing, HRTF individualization, HRTF personalization, localization accuracy, HRTF database matching.*

**Corresponding author*: *a.andreopoulou@music.uoa.gr.*

1. INTRODUCTION

Immersion and spatial accuracy are essential for creating a realistic and seamless binaural listening experience, particularly in extended reality (XR) environments. Achieving a personalized auditory experience that closely replicates natural sound field exposure in a physical space is a key objective in spatial audio research. A listener's ability to perceive spatial properties of sound is primarily determined by their Head-Related Transfer Functions (HRTFs)—individualized filters that encode directional and spectral cues based on the shape of the head, ears, and torso. HRTFs are fundamental to binaural rendering and spatial hearing, making their personalization crucial for accurate sound localization and an enhanced sense of presence in virtual and augmented environments.

Nevertheless, obtaining personally measured or individualized HRTFs for each user is challenging, as it is a time-consuming and costly process that requires specialized equipment. To address this limitation two main approaches for acquiring individualized HRTFs have been proposed, as noted in [1]: the first involves numerical simulations using 3D models of the head, pinnae and torso (e.g., [2, 3]); the second relies on transformation of, or selection from, existing HRTF datasets (e.g., [4, 5]). As an alternative, a survey of machine learning techniques for HRTF personalization is presented in [6].

This study focuses on another such approach: database matching [7], an HRTF individualization procedure which can be based on objective [8] or subjective criteria [9]. The work aims to identify the best-matching HRTF for a user from a repository of datasets, constructed through the aggregation of publicly available databases. The proposed technique is based on the premise that an effective HRTF matching should be tailored to the application type, meaning it should consider the importance of spatial perception for specific applications. For example, certain applications may need improved localization accuracy while for others spectral smoothness might be the goal.

Navigation through the repository is achieved by gradually focusing on the region of the collection around which one's best match is located. To achieve this, the HRTF repository is clustered into regions, each represented by the cluster's centroid, and assessed through tailored tasks, based on the application type. Upon selection, the winning cluster is re-clustered into smaller zones and the procedure is repeated iteratively until no further clustering is possible and the best matching HRTF is identified.

This paper focuses on applications that would benefit from individualized binaural renderings to enhance localization accuracy. To achieve this, relative metrics are employed to construct the dataset space (Section 2.1) and an approach is proposed for handling the multi-objective nature of problems involving multiple metrics (Section 2.2). HRTF evaluation is conducted through a Virtual Reality (VR)-based listening test (Section 3), and results of the pilot implementation of this method are presented and discussed (Section 4).

2. METHODOLOGY

2.1 Building an HRTF Repository

HRTF datasets are available in various formats, such as free or diffuse-field compensated, full or minimum phase, full-range through low-frequency extension or band limited, etc. To ensure consistency across the HRTF pool, general-purpose free-field equalized HRTFs from the SOFA repository [10] were collected. When multiple formats were available for a single dataset, the least processed version was selected, prioritizing datasets with minimal processing. Simulated datasets were also excluded, leading to a collection of exclusively recorded data obtained from humans and mannequins.

The following databases were selected based on their spatial resolution, number of available datasets, and widespread approval within the research community: SONICOM [11], RIEC [12], ARI [13], HUTUBS [14], and BiLi [15]. To ensure comparability and enable fair metric calculations across databases, a uniform selection criterion for spatial positions was applied. In terms of elevation, positions corresponding to -30°, -20°, -10°, 0°, 10°, 20°, 30°, 40°, and 60°were retained, provided they were available, and the preferred azimuthal resolution was set to 5°. When a specific position was missing, the closest available position was identified, ensuring the angular deviation did not exceed 5°.

2.1.1 *HRTF Post-Processing*

To ensure consistency across datasets, a comprehensive post-processing pipeline was applied, following the steps outlined in [16, 17]. All datasets were converted to Directional Transfer Functions (DTFs) and downsampled to the lowest common sampling rate across the HRTF pool. A custom low-pass cutoff frequency of 8 kHz was selected to retain essential spatial cues, including pinna-related spectral information, while eliminating high-frequency noise introduced by the recording equipment. Additionally, a high-pass filter with a cutoff of 200 Hz was applied to eliminate low-frequency artifacts, before the DC offset was removed.

A rectangular window was applied to truncate the impulse responses (IRs) to a uniform length, based on the shortest IR duration (5.3 ms, corresponding roughly to 234 samples). For computational efficiency, a uniform length of 256 samples was applied across all data. A consistent starting point for truncation was established at 20 samples before the first detected onset, defined as the point where the signal first exceeded 10 dB relative to the peak value.

Variations in stimulus amplitude and loudspeaker distance across setups was accounted for through the application of dataset-specific scaling. The measurement with the highest RMS value for both ears was identified within each dataset, and was used to calculate the median global RMS across datasets. Subsequently, each HRTF dataset was adjusted to match this common median RMS value, ensuring consistent loudness across all data in the repository. As a last processing step, the first and last sample of each impulse response were set to zero.

2.1.2 *HRTF Space Construction*

Upon creation of the standardized HRTF pool, the next step involved the calculation of key metrics for data comparison. Barumerli's Bayesian Spherical Sound Localization model [18] was applied on the data, and its output was fed to the *localizationerror* function within the Auditory Modeling Toolbox (AMT) for the computation of a comprehensive set of metrics. The resulting HRTF space was constructed using the 35 metrics, derived from the output of the said function.

To reduce dimensionality of the output metric data, while preserving essential information, Principal Component Analysis (PCA), a method widely employed in HRTF-related studies (e.g., [19–21]), was applied with a target of 95% variance retention. The PCs served as the basis for k-means clustering of the data. The resulting clusters consisted of HRTFs sharing similar spatialization characteristics. Each cluster was represented by its centroid, identified as the most characteristic set for each group.

2.2 HRTF database matching procedure

Since the goal of this work is the development of an HRTF database matching procedure targeted at localization accuracy applications, the selection criteria should be based on corresponding metrics. The methodology is defined as follows. Users navigate through the standardized HRTF repository by means of a localization test. In each test round users evaluate the most representative dataset of each cluster —that closest to the cluster centroid. To determine the optimal dataset, which guides the selection of the cluster explored in the subsequent round, a selection process based on three key performance metrics was implemented, based on [22] and available through the AMT localization error function:

1. Lateral RMS Error (rmsL): This metric quantifies the average lateral error by computing the root mean square (RMS) of lateral deviations, measured in degrees.
2. Local Polar RMS Error (rmsPmedianlocal): This metric assesses the precision of polar localization while excluding responses with large quadrant errors (i.e., errors exceeding 90°). It is computed as the RMS of polar errors for targets positioned near the median plane (lateral angles within ±30°).
3. Quadrant Error Rate (querrMiddlebrooks): This metric calculates the percentage of gross polar misjudgments, defined as cases where the response falls outside the correct quadrant (i.e., absolute polar errors $> 90°$).

Since the winner determination relies on multiple performance metrics, the problem falls under the multiobjective optimization category, in which no single dataset can simultaneously minimize all metrics. To address this, the Pareto front technique [23] was applied to identify non-dominated solutions —i.e., datasets that cannot be improved in one metric without worsening another. In this context, the Pareto front consists of datasets that exhibit the most favorable trade-offs among the evaluated metrics. From the tested datasets in each round, only those belonging to the Pareto front are considered potential winners. To determine the final winner from this set, the Euclidean distance of each Pareto-optimal dataset from an ideal performance point -defined as zero error across all metrics- is calculated. Since all three metrics in the study are minimization criteria, the dataset with the smallest Euclidean distance to the origin is selected as the winner, hence ensuring the best overall localization performance.

3. EVALUATION

The evaluation procedure of the above implementations involved the localization of invisible sound sources in a VR environment. The overall task included a training session, used to enhance user familiarization with hardware, interface, and VR interactions, as well as the main listening task, and was set to be completed on a single day.

The coordinates considered for the listening test, established as a general reference for the localization task, are shown in table 1 and illustrated in figure 1. However, the actual positions used in the test were slightly adjusted, within reported JND values, to match the specific spatial grid of each dataset. This approach eliminated the need for interpolation of positions that did not exist within the different HRTF grids considered. As a last step, in order to minimize the impact of the non-symmetric test grid on the localization task, two sets of coordinates were utilized in the test with each set being a left-right mirror of the other. The potential impact of each grid on the subjects' test performance and localization accuracy was assessed during the analysis of the responses.

3.1 Test Corpus

To reduce the duration of the localization test, an optimizer was developed to refine the original dataset space of 822 HRTFs, while preserving its representativeness. The reduction process involved outlier removal using the z-score method, followed by PCA for dimensionality reduction, and partitioning the space into grid cells from which representative datasets were selected. The optimizer evaluated multiple configurations by testing different parameters and discarding those that resulted in either uneven dataset representation or excessive test durations. The final configuration included 84 HRTF datasets (41 for SONICOM, 1 for RIEC, 25 for ARI, 5 for BiLi, 12 for

Table 1. General Azimuth (Azi) and elevation (Ele) angles of the 22 positions used for the localization test.

No.	Azi	Ele	**No.**	Azi	Ele
1	0	-30	12	150	0
2	180	-30	13	-150	0
3	20	-30	14	0	30
4	-160	-30	15	180	30
5	0	0	16	30	30
6	180	0	17	-150	30
7	80	0	18	0	60
8	-100	0	19	180	60
9	-20	0	20	-30	60
10	40	0	21	150	60
11	-60	0	22	-40	-30

HUTUBS), ensuring that the average estimated test time remained under one hour.

3.2 Experimental Protocol

The training session consisted of 20 trials of audio-visual stimuli, projected at 20 distinct positions on the surface of a four-meter radius virtual sphere. The structure of the main listening task adhered to the principles of the database matching method as described in section 1. As such, it comprised of a number of Rounds which were divided into Sessions of 44 Trials each. The listening task trials included the presentation of the selected audio stimulus at 22 distinct locations; hence, each position was repeated once within the session. The stimulus used was a 260 ms audio of White Gaussian Noise Bursts, three 60 ms of noise separated by 40 ms of silence.

Seven adults (six female), aged 21 to 43, participated in the localization task. All participants self-reported normal hearing, no visual impairments, as well as no prior experience with VR equipment. Participants were equipped with an Oculus Quest 2 head-mounted display, hand-tracked controllers, and open-back Sennheiser HD 650 headphones for binaural listening. The experiment was carried out on PCs running 64-bit Windows 11 using a Steinberg UR22 USB audio interface. The virtual test environment was developed in Unity v2020.3 integrating the 3D-Tune-In Toolkit for anechoic binaural rendering [24], which allows for the individualization of Interaural Time Differences (ITDs) for each participant.

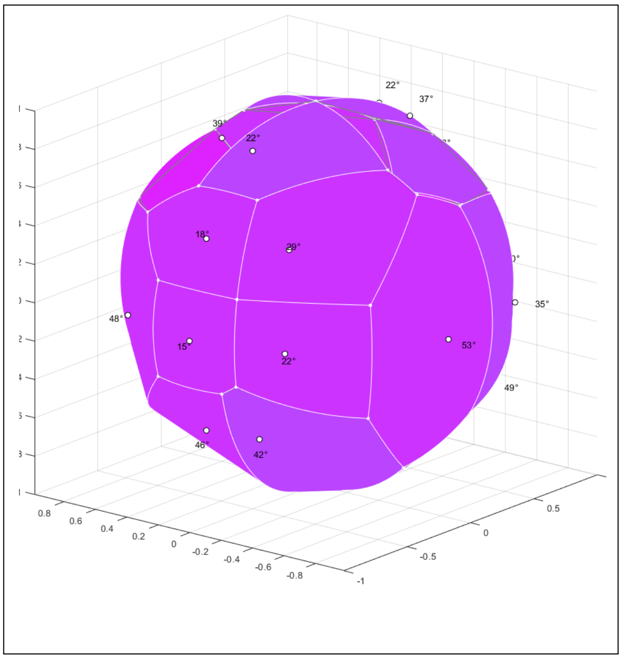

Figure 1. Visualization of the 22 positions used in the localization task.

Participant performance using each HRTF, was assessed at the end of each Round using the Auditory Modeling Toolbox (AMT) in Matlab. The winning HRTF was used to further refine the clustering of the HRTF universe, gradually refining the selection until the user-specific optimal HRTF was identified (see Sec. 2.2).

More specifically, following the assessment of the participant's performance using the initial HRTFs characterizing the clusters of the first database matching Round, the process advances to the next Round, which is determined by the winning HRTF. From the second round onward, the set of HRTFs under evaluation includes not only the HRTF candidates from a more refined region of the HRTF universe but also an additional 'escape' HRTF. This escape dataset represents a cluster of the previous Round and is included in the process to detect potential inconsistencies in the participant's responses. The number of HRTFs presented in each Round is predetermined by the clustering process. At this stage, based on performance evaluation using the AMT, the procedure either progresses to the third Round — further narrowing the selection toward the optimal dataset — or returns to the previous Round if the escape HRTF yields better performance. The process is repeated until the Final HRTF Winner is ide-

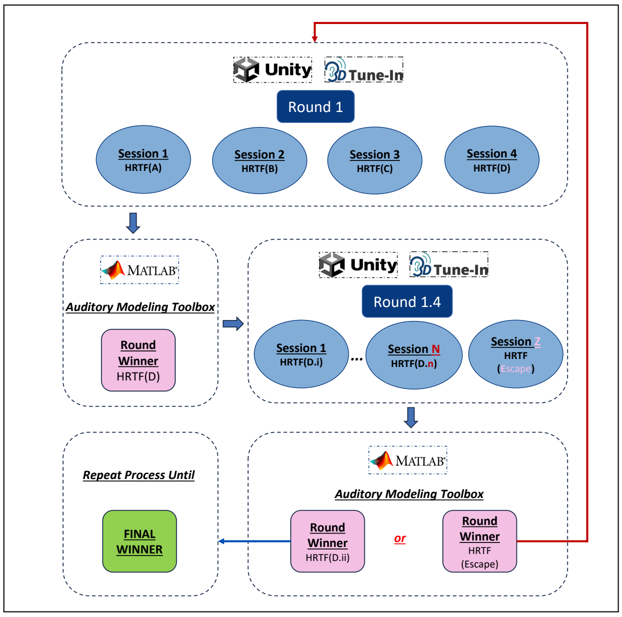

Figure 2. Overview of the database matching protocol.

tified. A detailed representation of the experimental sequence can be found in figure 2.

4. RESULTS

General observations from the pilot localization test are first presented. All participants successfully completed the test in three rounds, with no instances of an escape winner (i.e., no participant reverted to a previous round). The average test duration was 64 min (std: 19 min), including breaks. A comparison between the average performance in the first and final rounds shows a clear trend of improvement across all three evaluated metrics. For the rmsL metric, the average error of the winning HRTF decreased by 5°, with a reduction in standard deviation from 4°to 3°, indicating both improved accuracy and consistency. Similarly, for the querrMiddlebrooks metric, the average error dropped by 4% and the corresponding standard deviation reduced was reduced by 2%. While the average winning performance for rmsPmedianLocal remained steady, the decrease in variability (std from 6°to 5°) suggests increased stability. Notably, in the final round, the selected HRTFs outperformed the master escape dataset across all three metrics by 2°in rmsL, 3°in rmsPmedianLocal, and 7% in querrMiddlebrooks, further supporting the effectiveness of the selection process.

Participant responses are presented with a focus on the rmsL metric as a representative example, due to space constraints. The remaining two metrics used for cluster selection and test evaluation —rmsPmedianlocal and querrMiddlebrooks— exhibited similar trends. Figure 3 summarizes the average localization performance across all participants with respect to the rmsL error metric. For each round, the average localization performance of the winning HRTF is shown alongside the average performance across all HRTFs in that round (excluding the localization results of the escape HRTF). The average rmsL error of the master escape HRTF —defined as the dataset most distant from the initial winning cluster— is also included as a control reference to evaluate localization performance changes.

As can be seen, across all rounds the average localization error of the non-winning HRTFs is consistently higher than that yielded by the winning set. When focusing on the error distribution between the three experiment rounds the following pattern emerges. Responses of the first round exhibit higher data variability (wider boxplots) compared to those in the other two rounds whose boxplots are more compact, reflecting a higher concentration of values around the median, and hence more comparable localization performance between HRTFs. This result is consistent with expectations, as the initial test-round included HRTFs of very different spatial qualities, occasionally resulting in incidental good and bad matches for users. In contrast, the second round operates on a more constrained dataset space, which may have introduced suboptimal HRTFs prior to further refinement. By the third and final test-round, performance shows consistent improvement (lower rmsL error), indicating that the selected winning HRTFs yielded higher localization accuracy than those in earlier rounds. Furthermore, responses to the final round's winning HRTFs demonstrate a greater performance gap relative to the master escape, underscoring the effectiveness of the iterative selection process. Similar trends were observed across all evaluated performance metrics.

The overall performance of each winning HRTF was evaluated relative to the remaining datasets within the same round by computing the signed difference between the round's average performance (excluding the localization results of the escape HRTF) and the performance of the winning dataset (Figure 4). A positive difference indicates that the winner outperformed the remaining sets, leading to lower localization error. As can be seen, in the

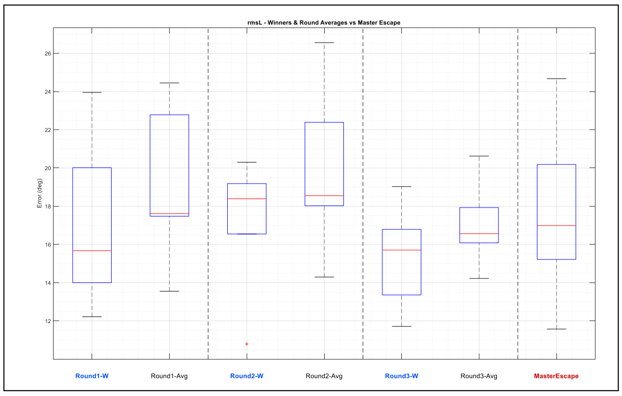

Figure 3. Average localization error (rmsL) across all participants for each round's winning HRTF, alongside the round's average (excluding escape HRTFs). The Master Escape HRTF is also shown as a control reference.

majority of cases, positive values are observed, supporting the effectiveness of the selection process in enhancing localization performance with respect to that metric. Similar observations were made across all three evaluated metrics.

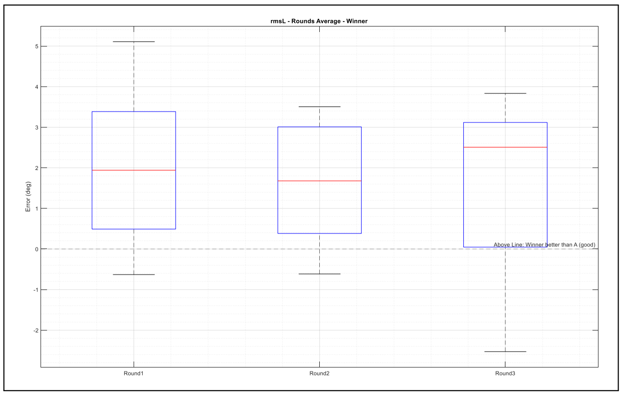

Figure 4. Signed difference between each round's average localization error and the winning HRTF's error (rmsL metric).

5. CONCLUSIONS AND FUTURE WORK

This work investigated HRTF database matching as a method for creating individualized spatial audio experiences optimized for specific application needs. A methodology was developed to support the selection and evaluation of the matched HRTFs through iterative clustering and localization testing using a Unity-based VR environment. The proposed method was validated in a pilot study involving 84 HRTF datasets from five databases and seven participants.

The key findings from the pilot localization test revealed a general improvement trend, according to which the final round consistently outperformed the Master Escape condition. Moreover, the winning HRTFs generally outperformed the other datasets in the same round across participants. The incorporation of evaluation dimensions such as those proposed in [25] may enhance the robustness and relevance of the assessment framework.

Future work will focus on the robust re-design of the experimental protocol. The previously discussed repository of 84 HRTFs constrains the range of possible outcomes and the depth of analysis across test rounds. The expansion of the HRTF data collection is expected to facilitate trend identification. Similarly, an increase in the number of test participants will strengthen the statistical power and generalizability of the findings. To support this expansion, improvements in test efficiency are required. Efforts will focus on developing a faster and more streamlined testing procedure that reduces overall duration without compromising the effectiveness and quality of the match.

6. ACKNOWLEDGMENTS

The research work was supported by SONICOM (www.sonicom.eu), a project that has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No. 101017743.

7. REFERENCES

[1] L. Picinali and B. F. Katz, "System-to-user and user-to-system adaptations in binaural audio," in *Sonic interactions in virtual environments*, pp. 115–143, Springer International Publishing Cham, 2022.

[2] B. F. Katz, "Boundary element method calculation of individual head-related transfer function. i. rigid model calculation," *The Journal of the Acoustical Society of America*, vol. 110, no. 5, pp. 2440–2448, 2001.

[3] H. Ziegelwanger, W. Kreuzer, and P. Majdak, "Mesh2hrtf: Open-source software package for the

numerical calculation of head-related transfer functions," in *22nd international congress on sound and vibration*, 2015.

[4] K. H. Shin and Y. Park, "Enhanced vertical perception through head-related impulse response customization based on pinna response tuning in the median plane," *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, vol. 91, no. 1, pp. 345–356, 2008.

[5] D. Zotkin, J. Hwang, R. Duraiswaini, and L. S. Davis, "Hrtf personalization using anthropometric measurements," in *2003 IEEE workshop on applications of signal processing to audio and acoustics (IEEE Cat. No. 03TH8684)*, pp. 157–160, Ieee, 2003.

[6] D. Fantini, M. Geronazzo, F. Avanzini, and S. Ntalampiras, "A survey on machine learning techniques for head-related transfer function individualization," *IEEE Open Journal of Signal Processing*, 2025.

[7] A. Andreopoulou and A. Roginska, "Database matching of sparsely measured head-related transfer functions," *Journal of the Audio Engineering Society*, vol. 65, no. 7/8, pp. 552–561, 2017.

[8] Z. Guo, Y. Lu, H. Zhou, Z. Li, Y. Fan, and G. Yu, "Anthropometric-based clustering of pinnae and its application in personalizing hrtfs," *International Journal of Industrial Ergonomics*, vol. 81, pp. 1–9, 2021.

[9] R. Shukla, R. Stewart, A. Roginska, and M. Sandler, "User selection of optimal hrtf sets via holistic comparative evaluation," in *Audio Engineering Society Conference: 2018 AES International Conference on Audio for Virtual and Augmented Reality*, Audio Engineering Society, 2018.

[10] P. Majdak, F. Brinkmann, J. De Muynke, M. Mihocic, and M. Noisternig, "Spatially oriented format for acoustics 2.1: Introduction and recent advances," *Journal of the Audio Engineering Society*, vol. 70, pp. 565–584, 2022.

[11] I. Engel, R. Daugintis, T. Vicente, A. O. Hogg, J. Pauwels, A. J. Tournier, and L. Picinali, "The sonicom hrtf dataset," *journal of the audio engineering society*, vol. 71, no. 5, pp. 241–253, 2023.

[12] K. Watanabe, Y. Iwaya, Y. Suzuki, S. Takane, and S. Sato, "Dataset of head-related transfer functions measured with a circular loudspeaker array," *Acoustical science and technology*, vol. 35, no. 3, pp. 159–165, 2014.

[13] P. Majdak, M. J. Goupell, and B. Laback, "3-d localization of virtual sound sources: Effects of visual environment, pointing method, and training," *Attention, perception, & psychophysics*, vol. 72, no. 2, pp. 454–469, 2010.

[14] F. Brinkmann, M. Dinakaran, R. Pelzer, P. Grosche, D. Voss, and S. Weinzierl, "A cross-evaluated database of measured and simulated hrtfs including 3d head meshes, anthropometric features, and headphone impulse responses," *Journal of the Audio Engineering Society*, vol. 67, no. 9, pp. 705–718, 2019.

[15] T. Carpentier, H. Bahu, M. Noisternig, and O. Warusfel, "Measurement of a head-related transfer function database with high spatial resolution," in *7th forum acusticum (EAA)*, 2014.

[16] A. Andreopoulou, D. R. Begault, and B. F. Katz, "Inter-laboratory round robin hrtf measurement comparison," *IEEE Journal of selected topics in signal processing*, vol. 9, no. 5, pp. 895–906, 2015.

[17] J. Pauwels and L. Picinali, "On the relevance of the differences between hrtf measurement setups for machine learning," in *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1–5, IEEE, 2023.

[18] 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 Acustica*, vol. 7, p. 12, 2023.

[19] G. Marentakis and J. Hôlzl, "Compression efficiency and signal distortion of common pca bases for hrtf modelling," in *Proceedings of the 18th Sound and Music Computing Conference*, Axea sas/SMC Network, 2021.

[20] F. Grijalva, B. Escobar, B. A. A. Acurio, and R. Álvarez, "Analysis and synthesis of hrtfs using principal component analysis," in *2019 IEEE Fourth Ecuador Technical Chapters Meeting (ETCM)*, pp. 1–6, IEEE, 2019.

[21] M. Zhang, Z. Ge, T. Liu, X. Wu, and T. Qu, "Modeling of individual hrtfs based on spatial principal component analysis," *IEEE/ACM Transactions on Audio,*

Speech, and Language Processing, vol. 28, pp. 785–797, 2020.

[22] J. C. Middlebrooks, "Virtual localization improved by scaling nonindividualized external-ear transfer functions in frequency," *The Journal of the Acoustical Society of America*, vol. 106, no. 3, pp. 1493–1510, 1999.

[23] S. Petchrompo, D. W. Coit, A. Brintrup, A. Wannakrairot, and A. K. Parlikad, "A review of pareto pruning methods for multi-objective optimization," *Computers & Industrial Engineering*, vol. 167, p. 108022, 2022.

[24] 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.

[25] D. Poirier-Quinot, M. S. Lawless, P. Stitt, and B. F. Katz, "Chapter hrtf performance evaluation: Methodology and metrics for localisation accuracy and learning assessment," 2022.