Audio Engineering Society

Conference Paper 17

Presented at the AES 5th International Conference on Audio for Virtual and Augmented Reality
2024 August 19–21, Redmond, WA, USA

This paper was peer-reviewed as a complete manuscript for presentation at this conference. This paper is available in the AES E-Library (http://www.aes.org/e-lib), all rights reserved. Reproduction of this paper, or any portion thereof, is not permitted without direct permission from the Journal of the Audio Engineering Society.

Assessing the Potential Impact of Direction-Dependent HRTF Selection on Sound Localization Accuracy

Sapir Goldring[1], Zamir Ben Hur[2], David Lou Alon[2], and Boaz Rafaely[1]

[1]*School of Electrical and Computer Engineering, Ben Gurion University of the Negev*
[2]*Reality Labs Research, Meta, Redmond, WA 98052, USA*

Correspondence should be addressed to Sapir Goldring (sapirgol@post.bgu.ac.il)

ABSTRACT

This study investigates the approach of direction-dependent selection of Head-Related Transfer Functions (HRTFs) and its impact on sound localization accuracy. For applications such as virtual reality (VR) and teleconferencing, obtaining individualized HRTFs can be beneficial yet challenging, the objective of this work is therefore to assess whether incorporating HRTFs in a direction-dependent manner could improve localization precision without the need to obtain individualized HRTFs. A localization experiment conducted with a VR headset assessed localization errors, comparing an overall best HRTF from a set, against selecting the best HRTF based on average performance in each direction. The results demonstrate a substantial improvement in elevation localization error with the method motivated by direction-dependent HRTF selection, while revealing insignificant differences in azimuth errors.

1 Introduction

The Head Related Transfer Function (HRTF) represents the acoustic response between a distant sound source and the sound pressure at a listener's ear [1, 2]. Since individuals possess a unique anatomical ear structure and dimensions, the HRTF varies between individuals, in addition to changing along direction [3]. For applications such as Virtual Reality (VR), teleconferencing, and hearing aids [4], where accurate sound spatialization is critical, the rendering of the auditory scene relies on these personalized HRTFs [5]. However, obtaining individual HRTFs is often impractical due to the time-intensive measurement process, leading to the utilization of non-individual HRTFs from existing databases [6].

The selection of suitable HRTFs from a database typically involves conducting listening tests where the most appropriate HRTF is assigned to each individual based on perceptual criteria [7, 8, 6]. Various approaches are employed, ranging from participants rating sound trajectory relative to provided descriptions [6][8], to tournament-style selection methods [7]. Some works advocate for concentrating exclusively on source pathways and positions within the horizontal plane, where binaural cues such as the inter-aural time difference (ITD) and inter-aural level difference (ILD) hold prominence [9]. Since they vary less between listeners, effective azimuth localization can often be achieved even when using generic HRTFs [10]. This highlights the need for a selection that is based on a perceptual test that considers a range of directions of arrival, including both elevation and azimuth. Moreover, current HRTF selection methods aim to find the "single most suitable"

HRTF. However, HRTFs are frequency and direction dependent functions, thus a single HRTF may not provide the best fit in all directions and frequencies.

In this study, our objective is to assess the validity of the current HRTF selection approaches by comparing sound localization accuracy when choosing a single HRTF filter for a listener with an alternative approach of selecting the best-fitting HRTF for each specific Direction-of-Arrival (DOA), both in elevation and azimuth. The structure of this paper unfolds as follows: First, we introduce a database reduction method, grounded in the observation of spectral notches within HRTFs, typically appearing above 3 kHz [5][11]. Next, we outline the design of a listening experiment. Finally, we present our evaluation methods and the obtained results. The results of this study show that selecting different HRTFs for different directions, may provide improved individualization in terms of sound localization.

2 Methods

To detect the impact of direction dependent HRTF selection on sound localization accuracy, a localization listening test was conducted. This section, structured into three parts, outlines the methodology employed throughout the test.

The first part focuses on the description and reduction of the HRTF database. This includes an overview of the database used for the experiment and a detailed account of the reduction method applied to select a manageable number of HRTFs suitable for the localization experiment.

The subsequent sections will delve into the specifics of the experiment setup and the three distinct methods for HRTF selection examined in this paper.

2.1 HRTF Dataset

The study relies on an internal HRTF dataset, which includes HRTFs from 96 individuals measured across 612 directions[12]. Notably, some HRTFs were measured multiple times (3 to 4 repetitions).

To streamline the sound localization experiment, the objective was to reduce the HRTF database to just five selected HRTFs. The primary goal was to ensure perceptual diversity among these chosen HRTFs, an approach aligned with the concept of "perceptually orthogonal" representations [6]. Given the limitations of traditional objective measures [13], an alternative approach was selected, focusing on the first spectral notch within the HRTF.

Table 1: Selected HRTFs' Notch Frequencies and Standard Deviations.

ID	Notch Mean Frequency [kHz]	Notch Std [Hz]
5	9.4118	0
10	8.4706	266.2049
17	8.1568	54.3388
42	7.5764	66.5512
63	8.0940	133.1025

Spectral notches exhibit variations among individuals, resulting in distinct perceptions of sound elevation [14]. This suggests that when aiming to ensure perceptual differentiation among selected HRTFs concerning sound localization, it may be advisable to opt for HRTFs with distinct notches, positioned at a noticeable frequency gap from each other. The first spectral notch of all HRTFs within the database was determined using the methodology outlined in [11]. Subsequently, the mean and standard deviation among repetitions were computed, resulting in the selection of five HRTFs for this experiment. These selected HRTFs exhibited both low standard deviation, ensuring consistency, and a diverse range of mean values in the first notch frequency. Table 1 provides a description of both the mean and standard deviation (std) values for the selected HRTFs, alongside their corresponding subject IDs.

2.2 Localization Experiment

The localization experiment was conducted using a Virtual Reality (VR) headset, with specific details outlined as follows [15]. Participants utilized the Meta Quest 2 headset for visualizations and wore AKG K702 headphones for sound immersion.

During the experiment, participants were presented with a black screen featuring sphere grids to aid in their spatial orientation. This visual context allowed participants to gain a sense of the direction they were pointing.

Auditory stimuli were delivered in the form of 3-second far-field unfiltered white noise, which was simulated

from various DOAs and played in a free field environment.

The reference point for head orientation was established as 0 degrees azimuth and 0 degrees elevation, indicating the front-facing direction. In this system, elevation spans from -90 degrees (down) to 90 degrees (up), while azimuth ranges from 0 degrees (front) to 360 degrees (full circle).

For the investigation, we specifically targeted five distinct elevations: -47, -27, 0, 27, and 65 degrees. These elevations were symmetrically positioned at azimuths of either 40 degrees (left) or 320 degrees (right), representing the left and right directions, respectively. Azimuth values were randomized throughout the experiment.

Participants were exposed to the white noise and instructed to orient their heads in the direction from which they perceived the sound. The Quest 2 built-in head tracker was leveraged to precisely capture each participant's head pointing direction. To enhance data reliability, each combination of DOA and HRTF was repeated four times, resulting in a total of 100 experiment repetitions. Besides monitoring the subjects' pointing direction, the head tracker introduced dynamic elements to the experiment. Participants were able to move their heads freely. If they exceeded a 15-degree deviation from the reference point in any direction, the stimuli would cease. This design enabled enhancing spatial perception, while preventing participants from aligning their heads with the sound source while it was active. Such alignment would lead to the consistent usage of the HRTF associated with the 0,0 direction for all presented directions. Note that Headphone Equalization (HpEq) were not applied due to the dependency of HpEq on the specific HRTFs, which were not available for the HRTFs utilized in the experiment.

The real-time rendering of spatial audio was achieved through direct Head-Related Impulse Response (HRIR) rendering, with all impulse responses for different directions stored and selected for convolution with the signal in real time, based on the head orientation data from the head tracker. The convolution for the binaural signal at each ear is expressed as $s_{left/right}(t) = s(t) * HRIR_{direction}(t)$, where $s(t)$ is the signal and $HRIR_{direction}(t)$ is the impulse response corresponding to the tracked direction. Barycentric interpolation was employed to generate the dynamic listening experience, allowing for smooth transitions between the HRIRs as participants moved their heads. This method effectively provided continuous spatial sound localization cues.

The study involved eight male participants, ranging in age from 25 to 60. All participants had a background in listening tests and spatial hearing. These participants self-reported normal hearing status.

2.3 Three methods for HRTF Selection

The absolute localization errors in azimuth and elevation were analyzed using three distinct methods:
Random HRTF: Randomly selecting one of the five available HRTFs for each of the eight subjects.
Best single HRTF: Selecting the HRTF that, on average across DOAs and repetitions, yielded the minimal error (in either elevation or azimuth).
Best HRTF in each DOA: Selecting for each subject and for each DOA the HRTF that on average over the repetitions yielded the minimal error.

3 Results

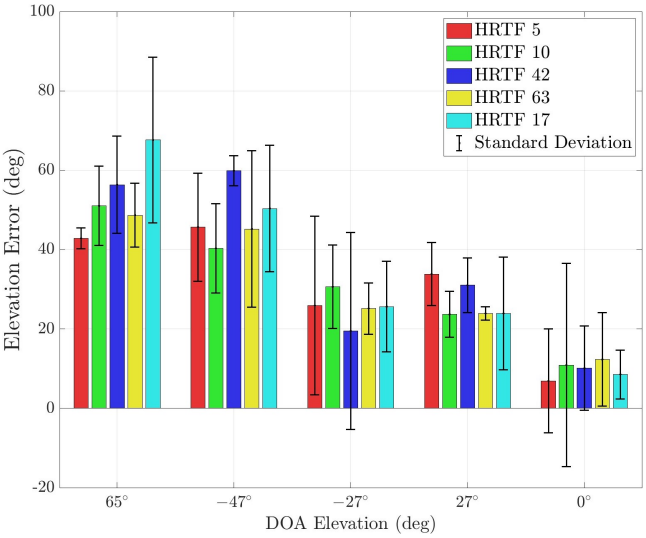


Fig. 1: Elevation errors of subject 1

In the evaluation process, the absolute azimuth and elevation localization errors for each HRTF and participant in the experiment were initially computed. An illustration of the elevation error results for subject 1 is presented in Figure 1. This figure displays the elevation error as a function of the source's Direction of Arrival (DOA), with separate representations for each HRTF. It provides insight into the motivation behind our current

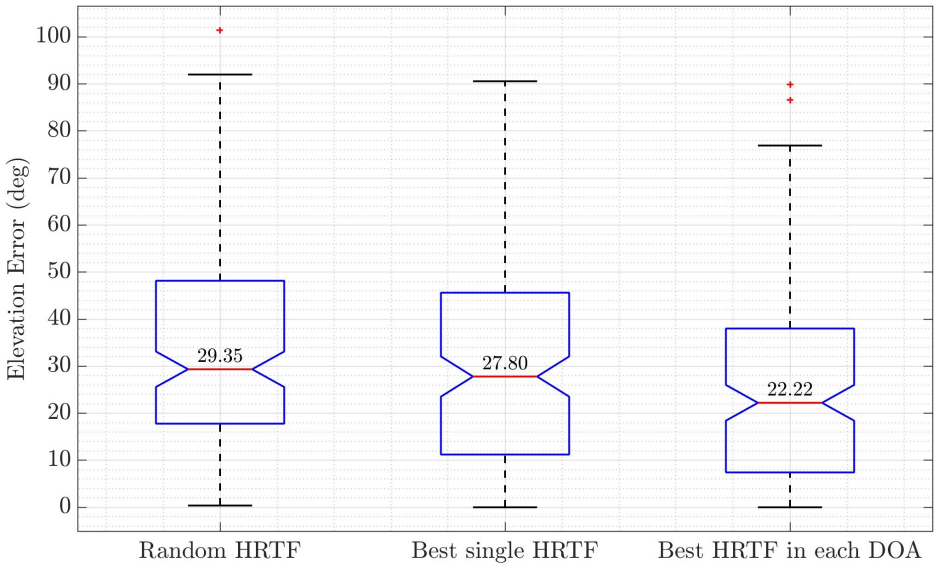


Fig. 2: Elevation error in degrees for the three HRTF selection methods

research, which aims to determine whether the HRTF that results in the smallest localization errors varies across different DOAs. For instance, at DOA elevation 0, the HRTF of subject number 5 yielded the smallest error on average, while at DOA -27 degrees,the HRTF of subject number 42 exhibited the smallest errors. These initial findings motivated the investigation of direction-dependent HRTF selection. As previously outlined, three approaches for HRTF selection were performed. For each of the methods, analysis of the elevation and azimuth localization errors was performed. The outcomes for the analysis of elevation error are shown in Figure 2, while the results for azimuth errors, adjusted to project all subject localization evaluations onto the front hemisphere, are displayed in Figure 3. It should be noted that azimuth error analysis focuses exclusively on two symmetric azimuth directions, specifically 40 degrees and 320 degrees (left and right).

3.1 Statistical Analysis

A Kruskal-Wallis non-parametric test was conducted to compare the efficacy of two of the methods described above: the selection of the best-fitting HRTF on average, versus the approach of individually selecting the best-fit HRTF for each DOA. For elevation errors, A very close to statistically significant difference with a p-value of 0.056 was observed, which was associated with a 5.6-degree difference in the medians (27.8 degrees for the method of selecting a single HRTF compared to 22.2 degrees for the best-fit in each DOA). Additionally, for azimuth errors, an insignificant difference with a p-value of 0.12 was noted in conjunction with a 2.4-degree difference in the medians (18.13 degrees for the method of selecting a single HRTF compared to 15.74 degrees for the best-fit in each DOA). Moreover, an insignificant 0.87 -degree difference in the medians was observed between the method of randomly choosing an HRTF and selecting the averaged best one for the elevation errors, with a p-value of 0.96.

3.2 Directional Variability of Selected HRTFs

This chapter analyzes the distribution patterns of HRTFs identified as the best fit for elevation sound localization. Figure 4 illustrates the averaged, ordered distribution of these HRTFs, highlighting how frequently each was determined to be the best fit across various directions. For example, the HRTF most commonly identified as providing the best fit was found to be the best for an average of approximately 2.5 out of 5 possible directions. In contrast, in a scenario where one HRTF is the best fit in all directions, the distribution would be expected to feature a single column at the full count of 5. However, it is observed from our data that the distribution neither peaks at a single point nor is uniformly flat.

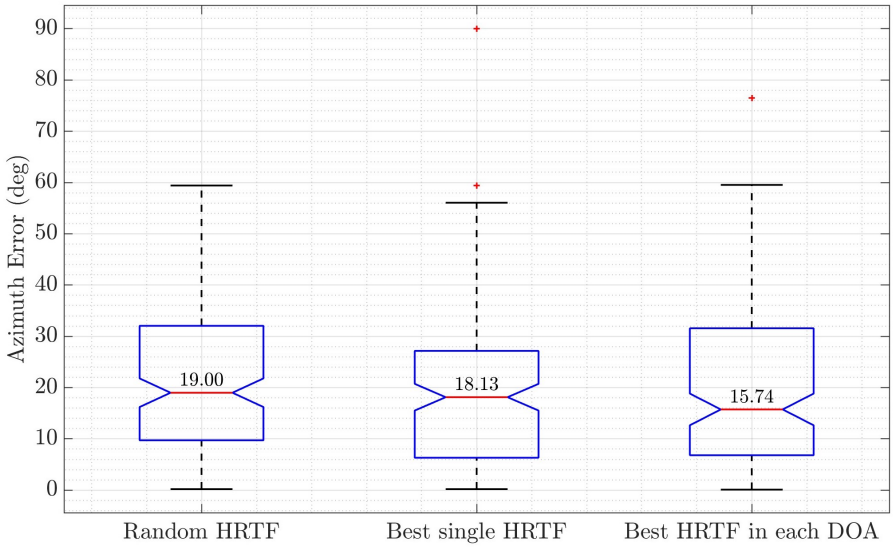


Fig. 3: Azimuth error in degrees for the three HRTF selection methods

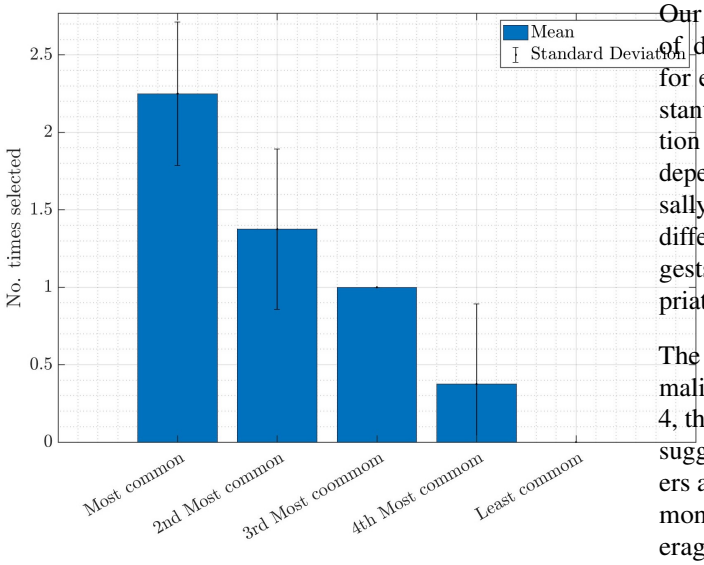


Fig. 4: Averaged, ordered distribution of HRTFs based on their selection as the optimal fit across different directions. The x-axis represents the rank order of HRTFs by their frequency of being chosen as the best fit, from the most to the least common.

4 Discussion and Conclusion

Our research findings strongly support the concept of direction-dependent HRTF selecion, particularly for elevation sound localization. We observed a substantial 5.6-degree difference in the medians of elevation localization errors between employing a direction-dependent HRTF application and using a single, universally optimal HRTF across all directions. This notable difference, underscored by a low p-value of 0.056, suggests a substantial benefit in selecting the most appropriate HRTF for each elevation.

The distribution analysis of HRTFs' directional optimality further reinforces this conclusion. As depicted in 4, the variability in HRTF optimality across directions suggests no single HRTF consistently outperforms others across all tested elevations. Instead, the most commonly selected HRTF provided the best fit for an average of approximately 2.5 out of 5 directions. This indicates a nuanced preference for certain HRTFs in specific spatial orientations, necessitating the use of multiple HRTFs to achieve optimal localization accuracy.

Conversely, our study revealed a slightly different scenario when examining azimuth errors. Selecting the optimal HRTF for each azimuth direction had less impact than on elevation errors, with a 2.4-degree difference

in the medians and a p-value of 0.12. This outcome was anticipated, given that azimuth localization relies not only on monaural HRTF cues but also on binaural cues, such as the Interaural Time Difference (ITD) and Interaural Level Difference (ILD) [9]. These binaural cues exhibit less variability among individuals compared to monaural cues, which are heavily influenced by the distinctive morphology of the pinnae and have a greater effect on elevation perception [10]. Therefore, while our study varied the HRTFs to assess their impact on localization errors, the inherent characteristics of azimuth localization—primarily governed by these robust binaural cues—meant that changes in HRTF selection method had a smaller effect on azimuth errors.

Additionally, it is noteworthy that azimuth localization errors were relatively low compared to elevation errors. This reduction in errors was achieved by projecting all measurements onto the front hemisphere, which effectively minimized the impact of front-back confusion—a common issue with non-individualized HRTFs [16]. The smaller error magnitude can also be attributed to the robust binaural cues, which significantly aid azimuth localization.

The comparison of individualized HRTFs with synthesized HRTFs derived by the method studied in this work, is proposed for future research. Further exploration of the challenges involved in generating and effectively using 'mixed HRTFs', where for each selected direction HRTF from different individuals may be used, will be integral to understanding the real-world implications of our findings. These challenges may include finding the appropriate method for interpolating all of these HRTFs, the efficiency of the calculations, and the performance of the listening test.

5 Acknowledgments

This work was partially supported by Reality Labs @Meta.

References

[1] Rafaely, B. and Avni, A., "Interaural cross correlation in a sound field represented by spherical harmonics," *The Journal of the Acoustical Society of America*, 127(2), pp. 823–828, 2010.

[2] Xie, B., *Head-related transfer function and virtual auditory display*, J. Ross Publishing, 2013.

[3] Majdak, P., Balazs, P., and Laback, B., "Multiple exponential sweep method for fast measurement of head-related transfer functions," *Journal of the Audio Engineering Society*, 55(7/8), pp. 623–637, 2007.

[4] Madmoni, L. and Rafaely, B., "Direction of arrival estimation for reverberant speech based on enhanced decomposition of the direct sound," *IEEE Journal of Selected Topics in Signal Processing*, 13(1), pp. 131–142, 2018.

[5] Poirier-Quinot, D. and Katz, B. F., "Assessing the impact of head-related transfer function individualization on task performance: Case of a virtual reality shooter game," *Journal of the Audio Engineering Society*, 68(4), pp. 248–260, 2020.

[6] Katz, B. F. and Parseihian, G., "Perceptually based head-related transfer function database optimization," *The Journal of the Acoustical Society of America*, 131(2), pp. EL99–EL105, 2012.

[7] Iwaya, Y., "Individualization of head-related transfer functions with tournament-style listening test: Listening with other's ears," *Acoustical science and technology*, 27(6), pp. 340–343, 2006.

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

[9] Seeber, B. U. and Fastl, H., "Subjective selection of non-individual head-related transfer functions," in *Proc. 9th Int. Conf. on Aud. Display, ICAD03*, 2003.

[10] Romigh, G. D. and Simpson, B. D., "Do you hear where I hear?: Isolating the individualized sound localization cues," *Frontiers in Neuroscience*, 8, p. 370, 2014.

[11] Iida, K., Ishii, Y., and Nishioka, S., "Personalization of head-related transfer functions in the median plane based on the anthropometry of the listener's pinnae," *The Journal of the Acoustical Society of America*, 136(1), pp. 317–333, 2014.

[12] Cuevas-Rodriguez, M., Alon, D. L., Clapp, S., Robinson, P. W., and Mehra, R., *Evaluation of the effect of head-mounted display on individualized*

head-related transfer functions, Universitätsbibliothek der RWTH Aachen, 2019.

[13] Wightman, F. and Kistler, D., "Multidimensional scaling analysis of head-related transfer functions," in *Proceedings of IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pp. 98–101, IEEE, 1993.

[14] Xuezhi, Y., Iida, K., and Ishii, Y., "anthropOmetric of pinnae between HRTF databases," 2014.

[15] Warnecke, M., Jamison, S., Prepelita, S., Calamia, P., and Ithapu, V. K., "HRTF personalization based on ear morphology," in *Audio Engineering Society Conference: AES 2022 International Audio for Virtual and Augmented Reality Conference*, Audio Engineering Society, 2022.

[16] Hu, H., Zhou, L., Ma, H., and Wu, Z., "HRTF personalization based on artificial neural network in individual virtual auditory space," *Applied Acoustics*, 69(2), pp. 163–172, 2008.