Audio Engineering Society

Conference Paper

Presented at the 2020 Conference on
Audio for Virtual and Augmented Reality
2020 August 17 – 19, Redmond, WA

Localization of Virtual Sounds in Dynamic Listening Using Sparse HRTFs

Zamir Ben-Hur[1], David Lou Alon[1], Philip W. Robinson[1], and Ravish Mehra[1]

[1] *Facebook Reality Labs, Facebook, 1 Hacker Way, Menlo Park, CA 94025, USA*

Correspondence should be addressed to Zamir Ben-Hur (`zamirbh@fb.com`)

ABSTRACT

Reproduction of virtual sound sources that are perceptually indistinguishable from real-world sounds is impossible without accurate representation of the virtual sound source location. A key component in such a reproduction system is the Head-Related Transfer Function (HRTF), which is different for every individual. In this study, we introduce an experimental setup for accurate evaluation of the localization performance using a spatial sound reproduction system in dynamic listening conditions. The setup offers the possibility of comparing the evaluation results with real-world localization performance, and facilitates testing of different virtual reproduction conditions, such as different HRTFs or different representations and interpolation methods of the HRTFs. Localization experiments are conducted, comparing real-world sound sources with virtual sound sources using high-resolution individual HRTFs, sparse individual HRTFs and a generic HRTF.

1 Introduction

With recent advances in the development of virtual and augmented reality technologies, the demand has emerged for high fidelity spatial sound reproduction through headphones [1–3]. The aim of such spatial sound reproduction systems is to make virtual sounds perceptually indistinguishable from real-world sounds. The localization accuracy of the reproduced virtual sound is important for achieving this goal [4, 5]. A key component for such reproduction systems is the Head-Related Transfer Function (HRTF), a function in both the space and the frequency domains, which is different for every individual [6, 7].

Many efforts have been made to find an optimal representation of the HRTF that will be of low dimensionality and will facilitate real-time spatial interpolation without degrading its accuracy [8–15]. A simple interpolation method is a linear interpolation of the HRTF at a desired direction from its nearby measured directions, which can be performed in either the time or the frequency domains. For a far-field HRTF, this can be done, for example, using bilinear rectangular or triangular interpolations [16]. More complex interpolation methods for sparse HRTFs have been suggested: these include using Principal Components Analysis (PCA) [17, 18], Karhunen–Loeve expansion [19], wavelets [20], and Spherical-Harmonics (SH) [8]. While previous studies investigated the different methods for interpolation of sparse HRTFs by evaluating the overall differences between them [21, 22], the localization performance in dynamic listening environments with sparse HRTFs using linear interpolation has not been fully explored yet.

In this study, we introduce a new experimental setup for accurate evaluation of the localization performance. The localization experimental setup is used to compare between localization performance with real-world sound sources and with virtual sound sources using high-resolution individual HRTFs, sparse individual HRTFs and a generic HRTF. A linear interpolation is used to enable reproduction of virtual sound sources in dynamic listening conditions.

2 Localization Experimental Setup

The localization experimental setup enables the examination of subjects' localization performance when presented with both real-world sound sources (originating from loudspeakers in the room) and virtual sound sources (originating from headphones). The experiment was performed in an acoustically treated room (4m×4m×3m with reverberation time of ∼0.1sec at 1kHz). Figure 1 shows the experimental setup in the room.

To report the perceived direction of the sound source a laser pointer was mounted on the subject's head, and a head-pointing protocol was used. A fabric dome made of a fabric with low acoustic reflectivity, with a radius of 1.2 m, was placed with its center located at the center of the room, which is precisely where the subject is seated, in order to provide a neutral visual surrounding. It also serves as the projection surface for the head-mounted laser pointer.

To evaluate real-world localization, 33 loudspeakers (Mini-DSP SPK-4P) were located at different locations in the room, and were not visible to the subject due to the fabric dome. To evaluate virtual sound localization, a pair of Sennheiser MX475 in-ear headphones were mounted about 1 inch from the subject's ear canal entrance. These *floating headphones* make it possible to perform individual headphone equalization at the beginning of each experiment, and their small form factor is beneficial when comparing real-world sound with virtual sounds [23].

An OptiTrack™ system was used to track the subjects' head movements inside the dome, and a headband was used to mount the OptiTrack™ markers, and the head-mounted laser pointer on the subject's head. In addition, the headband makes it possible to control the floating headphones placement (see Fig. 1.c). A projector was used to display the test instructions, and a laser pointer was mounted on each loudspeaker in order to make it possible to indicate on the dome the relative angular position of that loudspeaker, as will be elaborated later.

During the experiment, at the beginning of each trial the subject is instructed to return to a predefined starting position. Then, sound is played from one of the real or virtual sound sources. After the sound has stopped, the subject points with the head-mounted laser to the location on the dome where he/she perceived the sound's origin and presses a button to capture this location.

A spatial sound reproduction system was developed for the virtual localization experiment with dynamic listening conditions. The reproduction software was implemented in C++ and a localization application was developed in Unity. The application enables the selection of the different conditions of the experiment (HRTF, source positions, dry audio, etc.), presents the instruction to the user and captures the localization responses. The audio signals are updated in real-time according to the head-tracking data, where the head rotations are described with 3 degrees of freedom. The system's end-to-end latency, from physical movement to sound playback, is less than 40 ms, which is well below the perceptual threshold of detection in dynamic binaural listening [24, 25].

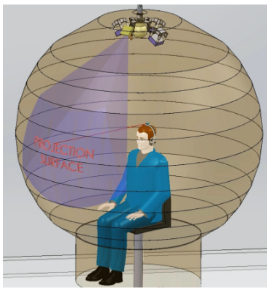
3 Experiment I - Comparison of Real and Virtual Sound Localization Accuracy

To get a baseline for the localization experimental setup, a preliminary session was first conducted with real sound sources, comprising training and evaluation phases. Then, a second session was conducted with virtual sound sources and different HRTFs.

3.1 Session 1 - Ground Truth Localization

In the first session, ground truth localization performance was evaluated using real sound sources. The aim of the ground truth localization experiment was to yield a baseline for the localization of each subject and

(a) The experiment room with the fabric dome folded up.



(b) Illustration of the fabric dome with the subject inside.

(c) KEMAR head with the modular headband.

Fig. 1: Localization experimental setup. The setup includes: 33 loudspeakers with laser pointers, a fabric dome, alignment lasers, subject chair, OptiTrack™ cameras, a projector, and a modular headband with the ability to remove/add the head-mounted laser/ floating headphones/OptiTrack™ markers.

to validate the accuracy of the localization experimental setup by obtaining localization results that are comparable to those published in previous literature [26–28].

In this experiment, the following dynamic listening condition was tested. A 3 sec audio stimulus of white-noise, band-passed filtered between 200 Hz and 16 kHz, was used. The subjects were instructed to freely rotate their head while the audio stimulus was played, but within a limitation of $\pm 15^\circ$ in azimuth and elevation. If the subject exceeded the allowed rotation, the stimulus stops. The aim of the dynamic listening condition is to provide the subject with the ability to use the binaural dynamic cues, while preserving the intended source direction. Removal of the head rotation limitation might lead to an experiment where the most frequently tested source direction is in front of the subject's head.

In the ground truth localization experiment 33 loudspeakers were used as the sound sources. Several steps were taken to minimize the listener's ability to associate coloration and source gain to the location of the sound source in the room. First, the responses from all loudspeakers at the center of the room were measured and a gain was applied to each loudspeaker in order to reduce the deviation caused by the gain differences. Second, during the experiment, a random gain of ± 2 dB was applied, in addition to a random spectral equalization that was implemented using five band-pass filters with random amplification, according to the measured STDs of the loudspeakers responses.

25 subjects with normal hearing participated in the experiment (20 males, 5 females, aged 21-56, median 30 years old). All subjects were naïve, with no previous experience in such a localization experiment. Each trial started with the subject's head pointing to the front of the dome (i.e. 0° azimuth, 0° elevation), the sound was then played, and then the subject undertook a localization judgment by head-pointing to the location on the sphere from where the perceived sound came. The head-mounted laser pointer provides the user with visual feedback.

The experiment commenced with a training phase, followed by an evaluation phase. In both phases a similar head-pointing protocol is used. As part of the training, each trial is repeated twice. After the first trial of localization judgment a laser pointer mounted on the played loudspeaker is turned on, providing visual feedback. In the second trial the visual feedback stays on and the audio stimulus is played again from the same loudspeaker, providing audio feedback. This double training trial is repeated for a minimum of 20 trials, and continues until the subject's localization performance becomes stable. Localization performance was defined as the angular error between the true source direction and the reported direction, which was calculated for each trial, together with a moving average of the angular errors of the previous 10 trials. The performance was considered to be stable when the difference between the maximum and minimum over the preceding 10 moving averages is below a threshold of 2° and 3°, for azimuth and elevation, respectively. Once the subjects' performance has converged to be within the thresholds, the training stage is completed and the subjects may progress to the evaluation stage. 16 out of the 33 loudspeakers were used for the training stage, while the other 17 directions were used for the evaluation phase, with 5 repetitions for each direction, leading to a total of 85 trials for each subject.

3.2 Session 2 - Localization of Virtual Sound Sources

In the second session, localization performance with virtual sound sources was evaluated using both individual and generic HRTFs. Both HRTF sets were measured with a measurement system similar to the one used in [29], where the generic HRTF was measured for a KEMAR (Knowles Electronics Mannequin for Acoustics Research). The same 25 subjects that participated in the ground truth session participated in this session. The same methodology was implemented, except that the audio stimulus for the evaluation phase was spatialized through the floating headphones instead of the loudspeakers. The training phase was still performed using the real loudspeakers, to eliminate any learning effects of the HRTF that may occur when listening to virtual sounds with visual feedback [30]. The directions of the virtual sound sources were chosen to be similar to those in the ground truth session. To reduce the chance of subjects learning the directions of the sound sources, a random mirroring between right and left directions was added, assuming left-right symmetry in localization performance. Again, 16 directions were used for training and the other 17 directions were used in the evaluation phase with 4 repetitions, resulting in a total of 136 trials for each subject (including both the individual and the generic HRTFs).

The measured HRTF comprises 612 directions over the sphere. For the dynamic listening, real-time interpolation is performed by linear interpolation in the space domain. This interpolation uses the three nearest measured directions and calculates the HRTF for the desired direction as a weighted average of the three measured HRTFs, using Barycentric weights [31]. The interpolation is performed on the magnitude response of the HRTF in the frequency domain, and the phase is computed according to a minimum-phase assumption directly from the magnitude response using the Hilbert transform [32], while the time delays are interpolated separately in a similar manner and added to the HRTF as a linear phase component.

3.3 Results

In this section the localization performance from the ground-truth session are compared with the localization of virtual sound sources with an individual HRTF and a generic HRTF. The localization performance is evaluated using four error metrics: total angular error, defined as the spatial angle between the true sound source direction and the reported direction, its azimuth and elevation components, and the percentage of trials that results in a front-back reversal.

Figure 2 shows a summary of the results for the four localization error metrics, for all the localization conditions (ground-truth, virtual sound source with an individual HRTF and with a generic HRTF). First, the error metrics are calculated for each trial and averaged for each subject. Then, the average and STD across subjects is calculated and presented as a bar plot with error bars for each localization experiment condition and each error metric. A one-factor within-subjects ANOVA with the factor "localization condition" (ground-truth, individual HRTF, generic HRTF) paired with a Tukey-Kramer post hoc test at a confidence level of 95% was used to determine the statistical significance of the results. The main effect of the "localization condition" is significant for all the four error metrics, with $p < 0.001$. As expected, the errors for tests with a generic HRTF are significantly higher compared to those with an individual HRTF, especially in the elevation error, with an average elevation error of 21.4° compared to 10.9° with an individual HRTF ($p < 0.001$). The azimuth error is also significantly higher for the generic HRTF ($p < 0.001$). Furthermore, a higher percentage of front-back reversal is found when using the generic HRTF (9.8% compared to 4.7% ($p = 0.044$)). Higher STDs can also be seen for the results from the tests with a generic HRTF, which is also expected because the KEMAR HRTF that was used may be worse for some people than it is for others.

The localization errors for the individual HRTFs are higher compared to the ground truth results, with an average azimuth error of 7.5° (compared to 5.5°, $p = 0.031$) and elevation error of 10.9° (compared to 6.9°, $p < 0.001$). An explanation for these differences could be the fact that the sound sources in the ground truth session were loudspeakers in a room, while the virtual experiment is with anechoic HRTFs. Another source of errors in the virtual sound source conditions is the imperfection of the spatial sound reproduction system, where errors may be ntroduced by the interpolation, as well as in the headphone equalization stage. Errors in the HRTF measurement may also affect the quality of the spatial sound reproduction and may yield inferior localization performance.

Overall, the ground truth localization errors are comparable to, or even better than, those found in previous literature [26–28]. These results indicate that the experimental setup is reliable in capturing accurate localization performance in dynamic listening conditions, even for naïve subjects. Furthermore, the localization performance using individual HRTFs is much better compared to the performance using the generic (KEMAR) HRTF; although it is higher compared to the ground truth, the results are still comparable to those found in previously published free-field localization literature [23, 33, 34]. This validates the accuracy of the HRTF measurement system and of the spatial sound reproduction system.

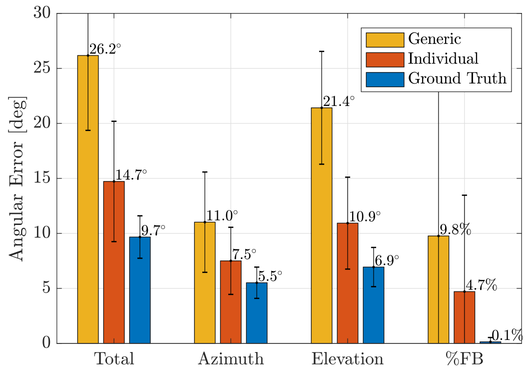


Fig. 2: Localization results for the virtual sound source experiment, for both individual and generic HRTFs compared to the ground truth results. Results are averaged over all subjects, directions and repetitions, in terms of the total angular error and its azimuth and elevation components, as well as for the percentage of trials which resulted in a front-back reversal (%FB). The error bars present the STDs between subjects.

4 Experiment II - Elevation Perception using Sparse Individual HRTFs

This section presents the results of a second localization experiment that aimed to evaluate the effect of sparse HRTFs in dynamic listening. As discussed before, the use of sparse individual HRTFs in dynamic listening requires efficient real-time spatial interpolation. The interpolation method used in this experiment is linear triangle interpolation with Barycentric weights.

4.1 Objective Analysis

To objectively evaluate the accuracy of the interpolated sparse HRTFs, a dense HRTF was measured using a setup similar to that described in [29] for one subject, with 2° resolution in azimuth, resulting in a HRTF with 3060 measured directions. Subsets of different numbers of measurements were taken from this dense set. Figure. 3 presents the subsets that were used, with $Q = 612$, 238, 141, 84 and 36. The $Q = 612$ subset was chosen to be the grid used for measuring the individual HRTFs for the localization experiment, and the other subsets were chosen to be as close as possible to nearly-uniform sampling over the sphere.

As observed in Sec. 3, the main advantage of an individual HRTF over a generic HRTF, is in elevation perception. Therefore, the current evaluation of sparse HRTFs is focused on the effect of reducing the number of HRTF directions on the perception of sound source elevation. The magnitude response of the right ear dense measured HRTF is presented in the leftmost plot of Fig. 4, as a function of frequency and elevation angle, for azimuth direction 315° (45° to the right). The plot highlights the spectral cues that are relevant for elevation perception, such as notches that are frequency and elevation dependent [7]. The figure also shows similar plots for the sparse individual HRTF subsets of $Q = 612$, 84 and 36, as well as for the generic HRTF with $Q = 612$, interpolated to the same directions as in the dense individual HRTF. Comparing the dense

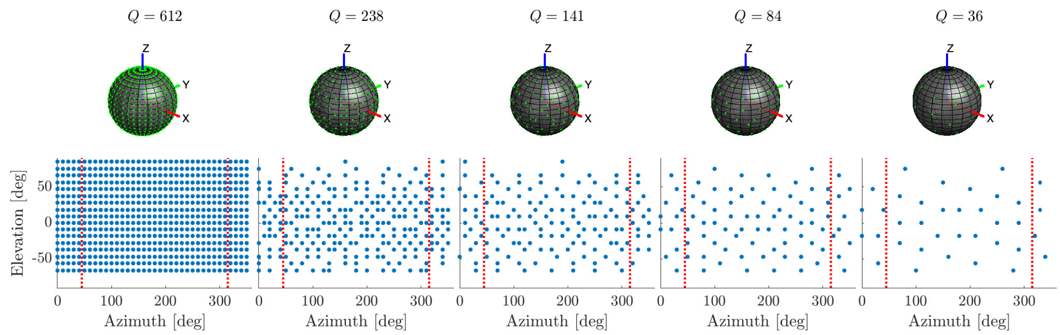


Fig. 3: Directions of the diluted HRTFs used for the objective analysis. The red dashed lines represent the azimuth directions 45° and 315°, which are the azimuth angles that were used for the evaluations in Sec. 4.

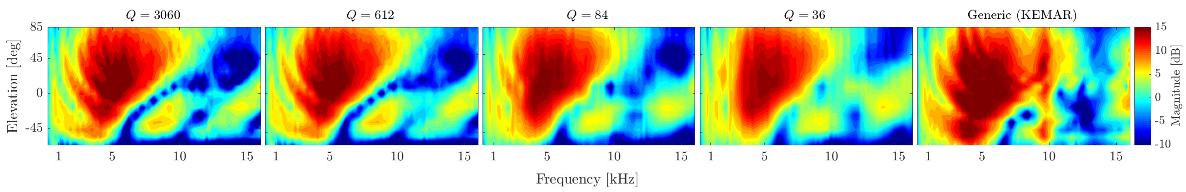


Fig. 4: Right ear HRTF magnitude as a function of frequency and elevation angle, for azimuth 315° (45° to the right), using different subsets from the dense measurement (left plot) and interpolated using the Barycentric method (three central plots). Right plot presents the generic (KEMAR) HRTF.

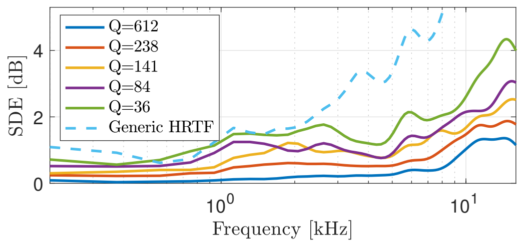


Fig. 5: SDE between sparse HRTFs and a reference dense HRTF. Average across elevation directions for azimuth 315° (45° to the right). Dashed line represents the error from the generic HRTF measurement. The errors are smoothed over $1/3$ octave.

HRTF plot to the interpolated ones from sparse subsets provides a qualitative illustration of the interpolation errors, which can be seen as distortion in the magnitude response. High errors are evident, for example, at high frequencies and the low number of HRTF measurements, $Q = 36$. Comparing the dense individual HRTF to the generic HRTF shows, as expected, large differences at high frequencies.

Figure. 5 shows the Spectral Difference Error (SDE) for the different HRTF subsets of a single representative subject. The errors are computed between the dense measured HRTF, $H(\Omega, f)$, where Ω is the source direction and f is the frequency, and the sparse HRTF after interpolation to the dense HRTF directions, $\widehat{H}(\Omega, f)$, as follows:

$$\mathrm{SDE}(\Omega, f) = \left|20\log_{10}|H(\Omega, f)| - 20\log_{10}|\widehat{H}(\Omega, f)|\right| \quad (1)$$

The errors are computed for the right ear's HRTF and averaged over the elevation angles on the arc of azimuth 315°. The errors between the generic HRTF and the individual HRTF are also presented. From the SDE values it can be seen that, as expected, reducing the number of measurements leads to higher errors.

Interestingly, even with Q as low as 36, the interpolation error is substantially lower than the difference between the generic HRTF and the evaluated individual HRTF (except for frequencies $f \in (700, 900)$ Hz which are not expected to affect elevation perception). Even smaller errors are achieved over all frequencies with $Q = 84$ and above, where the SDE is less than 2 dB up to $\sim$10 kHz. These results present the effects of the interpolation errors on the magnitude responses of the HRTFs when using sparse HRTFs; however, in order to better understand the effects of these errors on elevation perception of the reproduced binaural signals, a listening experiment is required.

4.2 Experiment Method

The experiment is similar to the virtual sound sources localization experiment presented in Sec. 3.2. 11 subjects who participated in the previous experiment participated in this experiment. The measured individual HRTFs were used for the spatial sound reproduction, using the full set with $Q = 612$ and using the subsets of $Q = 238$, 141, 84 and 36 (as seen in Fig. 3). 13 sound source directions were tested, located at elevation angles from -60° to 60° with 10° resolution and with azimuth angles ±45° (both to the left and to the right). A total of 195 trials were performed, comprising 5 HRTFs × 13 directions × 3 repetitions. A training session was conducted at the beginning of each evaluation session, with the same method as presented in Sec. 3.2. A 3 second recording of white-noise, band-passed filtered between 200 Hz and 16 kHz, was used as the stimulus. The subjects were instructed to freely move their head while the sound is playing, with the limitation of ±15° in azimuth and elevation.

4.3 Results

Figure 6 shows the results of the experiment. The figure presents the average elevation errors for all subsets, compared to the localization performance in elevation using the generic HRTF (as presented in Sec. 3.2). Surprisingly, even when using a sparse individual HRTF with only 36 measurement directions the average elevation error (11.3°) is remarkably lower compared to the average elevation error with the generic HRTF (21.4°). A one-factor within-subjects ANOVA with the factor "subset" ($Q = 612$, 238, 141, 84 and 36) paired with a Tukey-Kramer post hoc test at a confidence level

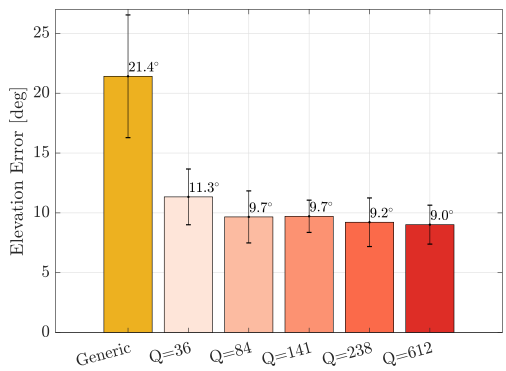


Fig. 6: Elevation localization errors for the virtual sound source experiment with individual sparse HRTFs compared to generic HRTF. Results are averaged over all subjects, directions and repetitions, in terms of the elevation error. The error bars present the STDs between subjects.

of 95% was used to determine the statistical significance of the results. The main effect of the "subset" factor is significant ($F_{(4,40)} = 8.47, p < 0.001$). Only when reducing the number of HRTF measurements to $Q = 36$ the elevation error becomes significantly higher compared to the $Q = 612$ condition ($p < 0.01$). Interestingly, the elevation errors for all other subsets, $Q = 84$, 14, 238, are not significantly different from the errors with the full HRTF with $Q = 612$ ($p > 0.5$ for all pairwise comparisons).

5 Conclusion

The paper presents a flexible localization experimental setup that facilitates the use of different virtual sound conditions, such as different HRTFs, interpolation methods and listening conditions. The system provides accurate evaluation of the localization performance of both real and virtual sound sources. A comparison was made between the localization of real sounds and virtual sounds with generic and individual HRTFs, validating the reliability of the setup. Furthermore, comparison between sparse individual HRTFs was made. Interestingly, even with a sparse individual HRTF that was measured over only 36 directions, the localization performance was not substantially different from when using a high-resolution individual HRTF, but remarkably better than using a high-resolution generic HRTF, which further highlights the advantage of individual HRTFs over a generic HRTF. This result is important as, although a simple linear interpolation was used, relatively high localization accuracy was obtained, which may imply that a linear interpolation method can be used as a baseline for future evaluation of more complex and (presumably) more accurate interpolation methods.

Acknowledgment

The authors would like to thank our colleagues from Facebook Reality Labs, Kevin Scheumann, Alex Gustafson, Scott Colburn, Xuzhou (Mike) Ye, David Kalamen, Manoel Soares, Samuel Clapp, Pablo F. Hoffmann and Vamsi Krishna Ithapu, who provided insight and shared their expertise, which greatly assisted in the research and improved the manuscript.

References

[1] Begault, D. R. and Trejo, L. J., "3–D sound for virtual reality and multimedia," *NASA, Ames Research Center, Moffett Field, California*, pp. 132–136, 2000.

[2] Vorländer, M., *Auralization: fundamentals of acoustics, modelling, simulation, algorithms and acoustic virtual reality*, Springer Science & Business Media, 2007.

[3] Blauert, J., *The technology of binaural listening*, Springer, 2013.

[4] Møller, H., "Fundamentals of binaural technology," *Applied acoustics*, 36(3), pp. 171–218, 1992.

[5] Kleiner, M., Dalenbäck, B.-I., and Svensson, P., "Auralization-an overview," *Journal of the Audio Engineering Society*, 41(11), pp. 861–875, 1993.

[6] Blauert, J., *Spatial hearing: the psychophysics of human sound localization*, MIT press, 1997.

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

[8] Evans, M. J., Angus, J. A., and Tew, A. I., "Analyzing head-related transfer function measurements using surface spherical harmonics," *The Journal of the Acoustical Society of America*, 104(4), pp. 2400–2411, 1998.

[9] Hugeng, H., Anggara, J., and Gunawan, D., "Implementation of 3D HRTF interpolation in synthesizing virtual 3D moving sound," *Int. J. of Technology (IJTech)*, 8(1), pp. 186–195, 2017.

[10] Ramos, G. and Cobos, M., "Parametric head-related transfer function modeling and interpolation for cost-efficient binaural sound applications," *The Journal of the Acoustical Society of America*, 134(3), pp. 1735–1738, 2013.

[11] Duraiswaini, R., Zotkin, D. N., and Gumerov, N. A., "Interpolation and range extrapolation of HRTFs [head related transfer functions]," in *2004 IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 4, pp. iv–iv, IEEE, 2004.

[12] Brinkmann, F. and Weinzierl, S., "Comparison of Head-Related Transfer Functions Pre-Processing Techniques for Spherical Harmonics Decomposition," in *Audio Engineering Society Conference: 2018 AES International Conference on Audio for Virtual and Augmented Reality*, Audio Engineering Society, 2018.

[13] Ben-Hur, Z., Alon, D. L., Mehra, R., and Rafaely, B., "Efficient Representation and Sparse Sampling of Head-Related Transfer Functions Using Phase-Correction Based on Ear Alignment," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 27(12), pp. 2249–2262, 2019, doi:10.1109/TASLP.2019.2945479.

[14] Porschmann, C., Arend, J. M., and Brinkmann, F., "Directional Equalization of Sparse Head-Related Transfer Function Sets for Spatial Upsampling," *IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP)*, 27(6), pp. 1060–1071, 2019.

[15] Zaunschirm, M., Schörkhuber, C., and Höldrich, R., "Binaural rendering of Ambisonic signals by head-related impulse response time alignment and a diffuseness constraint," *The Journal of the Acoustical Society of America*, 143(6), pp. 3616–3627, 2018.

[16] Freeland, F. P., Biscainho, L. W., and Diniz, P. S., "Interpositional transfer function for 3D-sound generation," *Journal of the Audio Engineering Society*, 52(9), pp. 915–930, 2004.

[17] Kistler, D. J. and Wightman, F. L., "A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction," *The Journal of the Acoustical Society of America*, 91(3), pp. 1637–1647, 1992.

[18] Xie, B.-S., "Recovery of individual head-related transfer functions from a small set of measurements," *The Journal of the Acoustical Society of America*, 132(1), pp. 282–294, 2012.

[19] Chen, J., Van Veen, B. D., and Hecox, K. E., "A spatial feature extraction and regularization model for the head-related transfer function," *The Journal of the Acoustical Society of America*, 97(1), pp. 439–452, 1995.

[20] Torres, J. C. and Petraglia, M. R., "HRTF interpolation in the wavelet transform domain," in *2009 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pp. 293–296, IEEE, 2009.

[21] Runkle, P., Blommer, M., and Wakefield, G., "A comparison of head related transfer function interpolation methods," in *Proceedings of 1995 Workshop on Applications of Signal Processing to Audio and Accoustics*, pp. 88–91, IEEE, 1995.

[22] Hartung, K., Braasch, J., and Sterbing, S. J., "Comparison of different methods for the interpolation of head-related transfer functions," in *Audio Engineering Society Conference: 16th International Conference: Spatial Sound Reproduction*, Audio Engineering Society, 1999.

[23] Romigh, G. D., Brungart, D. S., and Simpson, B. D., "Free-Field Localization Performance With a Head-Tracked Virtual Auditory Display," *IEEE Journal of Selected Topics in Signal Processing*, 9(5), pp. 943–954, 2015.

[24] Yairi, S., Iwaya, Y., and Suzuki, Y., "Influence of large system latency of virtual auditory display on behavior of head movement in sound localization task," *Acta Acustica united with Acustica*, 94(6), pp. 1016–1023, 2008.

[25] Sankaran, N., Hillis, J., Zannoli, M., and Mehra, R., "Perceptual thresholds of spatial audio update latency in virtual auditory and audiovisual environments," *The Journal of the Acoustical Society of America*, 140(4), pp. 3008–3008, 2016.

[26] Oldfield, S. R. and Parker, S. P., "Acuity of sound localisation: a topography of auditory space. I. Normal hearing conditions," *Perception*, 13(5), pp. 581–600, 1984.

[27] Carlile, S., Leong, P., and Hyams, S., "The nature and distribution of errors in sound localization by human listeners," *Hearing research*, 114(1-2), pp. 179–196, 1997.

[28] Best, V., Brungart, D., Carlile, S., Jin, C., Macpherson, E., Martin, R., McAnally, K., Sabin, A., and Simpson, B., "A meta-analysis of localization errors made in the anechoic free field," in *Principles and applications of spatial hearing*, pp. 14–23, World Scientific, 2011.

[29] Cuevas-Rodriguez, M., David Lou, A., Clapp, S. W., Robinson, P. W., and Mehra, R., "Evaluation of the effect of head-mounted display on individualized head-related transfer functions," in *The 23rd International Congress on Acoustics (ICA)*, pp. 2635–2642, 2019.

[30] Mendonça, C., Campos, G., Dias, P., Vieira, J., Ferreira, J. P., and Santos, J. A., "On the improvement of localization accuracy with non-individualized HRTF-based sounds," *Journal of the Audio Engineering Society*, 60(10), pp. 821–830, 2012.

[31] Sundareswara, R. and Schrater, P., "Extensible point location algorithm," in *2003 International Conference on Geometric Modeling and Graphics, 2003. Proceedings*, pp. 84–89, IEEE, 2003.

[32] Alan, V. O., Ronald, W. S., and John, R., "Discrete-time signal processing," *New Jersey, Printice Hall Inc*, 1989.

[33] Wightman, F. L. and Kistler, D. J., "Headphone simulation of free-field listening. II: Psychophysical validation," *The Journal of the Acoustical Society of America*, 85(2), pp. 868–878, 1989.

[34] Bronkhorst, A. W., "Localization of real and virtual sound sources," *The Journal of the Acoustical Society of America*, 98(5), pp. 2542–2553, 1995.