Directional Hybrid Optimization of HRTFs for Low-Order Spherical Harmonics Binaural Rendering

Rui Zhang[†], Yuxuan Ke[*†], Qunping Ni[‡], Ge Yao[§], Xiaodong Li[†] and Chengshi Zheng[†]
[†] Institute of Acoustics, Chinese Academy of Sciences, Beijing, China
University of Chinese Academy of Sciences, Beijing, China
[‡] TianJin 712 Communication & Broadcasting Co., Ltd., Tianjin, China
[§] Tianjin Tianan Borui Technology Co., Ltd, Tianjin, China
[*]Corresponding author E-mail: keyuxuan@mail.ioa.ac.cn

***Abstract*—Employing only low-order spherical harmonics for binaural reproduction results in significant spectral distortions. Although existing methods like Magnitude Least-Squares (MLS) can reduce overall errors by means of low-order representation of the head-related transfer functions (HRTFs), residual inaccuracies persist in contralateral regions. To address this limitation, we propose a directional hybrid optimization of HRTFs. This method partitions the space into contralateral, ipsilateral, and transition zones, and each zone is processed using least-squares, MLS, and their linear combination, respectively. Objective evaluation results using KU-100 HRTFs demonstrate that the proposed method achieves much higher performance across all objective metrics, suppressing rendering errors in contralateral source while improving spatial cue accuracy at lateral positions when compared with MLS. Subjective MUSHRA tests further confirm the superiority of the proposed method, with perceptual scores exceeding those of MLS by 14% at order $N = 1$ and 4% at $N = 4$. The results of both subjective and objective experiments indicate that the proposed method effectively suppresses spectral distortions of HRTFs in the contralateral regions while preserving critical spatial information for immersive auditory experiences.**

I. Introduction

Binaural reproduction of spatial audio delivers immersive virtual acoustic scenes to listeners via headphones and is a crucial issue in the growing realm of virtual and augmented reality applications [1]. Recently, binaural signals are generally rendered by processing the spherical harmonics representations of the sound field (represented as a plane-wave density function, also known as Ambisonics), as well as the corresponding head-related transfer functions (HRTFs) [2]. In practice, Ambisonic signals are often captured using low-order spherical microphone arrays, necessitating the truncation of HRTF order to match the captured sound field order [3]. Therefore, binaural rendering with only low-order Ambisonic signals can introduce significant spatial and spectral artifacts, including degraded localization cues, timbral distortion, broadened perceived source width, and a pronounced high-frequency roll-off [4], [5].

Various methods have been proposed to mitigate the above-mentioned problems by reducing the effective spherical harmonics order of HRTF. Overall, existing methods can be divided into two broad categories including post-processing and pre-processing. As for post-processing methods, Ben-Hur et al. [4] proposed to apply a global spectral equalization filter on the binaural signals to reduce the coloration existing in order-limited Ambisonic signals. As an extension, Hold et al. [6] introduced spherical harmonics tapering, by windowing the spherical harmonics coefficients, to alleviate the perceived source widening associated with order truncation. Besides these post-processing methods, pre-processing methods have also been proved to be effective [7]. Zaunschirm et al. [8] presented a method for reproducing binaural signals in the spherical harmonics domain by pre-processing HRTFs via frequency-dependent time-alignment, and their extended work [9] refined this concept by introducing the magnitude least-squares (MLS) method, which employs magnitude-only optimization at high frequencies.

Although existing methods can reduce the overall error in binaural reproduction, residual deviations still persist, including inaccurate median plane location [10], errors in interaural level differences (ILDs) and interaural time differences (ITDs) [1], and increased HRTF reconstruction errors, particularly at contralateral directions [5]. To address these persistent limitations, we propose a novel directional hybrid (DH) optimization framework. This framework divides the space into three parts, including contralateral, ipsilateral, and transition zones, and we then apply different tailored processing strategies to each zone.

This paper is structured as follows. In Sec. II, we summarize the least-squares (LS) and MLS methods, and analyze the existence of the residual inaccuracies. The proposed method is presented in detail in Sec. III. The evaluation of the proposed method and a comparison to existing methods via subjective and objective tests are presented in Sec. IV. Finally, the proposed method is summarized in Sec. V.

II. Problem Formulation

The binaural signal, representing the sound pressure observed at the ear of a listener, can be rendered by filtering the plane-wave density function with the HRTFs. In the spherical harmonics domain, this can be formulated as $p^{l/r}(k)$ [11]:

$$p^{l/r}(k) = \sum_{n=0}^{N} \sum_{m=-n}^{n} \left[\tilde{a}_{nm}(k)\right]^* h_{nm}^{l/r}(k), \tag{1}$$

where $a_{nm}(k)$ and $h_{nm}^{l/r}(k)$ are the Nth-order spherical harmonics coefficients of the plane-wave density function

$a(k,\Omega)$ and the left/right ear HRTF, respectively. $\tilde{a}_{nm} = (-1)^m \left[a_{n(-m)}\right]^*$ and $[\cdot]^*$ denotes the complex conjugate. k is the wave number. $\Omega \equiv (\phi, \theta)$ is the spatial angle, where $\phi \in [0, 2\pi]$ is the azimuth angle, measured counterclockwise from the Cartesian positive x-axis in the horizontal plane, and $\theta \in [0, \pi]$ is the elevation angle, measured upwards from the Cartesian positive z-axis. The superscript l/r denotes the left ear or the right ear. For brevity, the superscript l/r and the wave number k are omitted in subsequent expressions when no confusion arises. The spherical harmonics coefficients h_{nm} of the HRTF are typically obtained via a LS formulation [12]:

$$h_{nm}^{\mathrm{LS}} = \arg\min_{h_{nm}} \sum_{\Omega} \left\| \sum_{n=0}^{N} \sum_{m=-n}^{n} h_{nm}\, Y_n^m(\Omega) - h(\Omega) \right\|_2^2, \quad (2)$$

where $h(\Omega)$ represents the true HRTF at direction Ω, and $Y_n^m(\Omega)$ is the spherical harmonic function of order n and degree m. The closed-form solution to (2) is given by:

$$\mathbf{h}_{\mathbf{nm}}^{\mathrm{LS}} = \mathbf{Y}^{\dagger}\mathbf{h} = \mathrm{SFT}_N(\mathbf{h}), \quad (3)$$

where $\mathbf{h} = [h(\Omega_1), h(\Omega_2), \ldots, h(\Omega_Q)]^T$ denotes vector of HRTF measurements at Q directions, $\mathrm{SFT}_N(\mathbf{h})$ represents the Nth-order discrete spherical Fourier transform of $\mathbf{h}$, $\mathbf{Y} = [\mathbf{y}(\Omega_1), \mathbf{y}(\Omega_2), \ldots, \mathbf{y}(\Omega_Q)]^T$, and $()^{\dagger}$ denotes the Moore-Penrose pseudoinverse. $\mathbf{h}_{\mathbf{nm}}^{\mathrm{LS}}$ and $\mathbf{y}(\Omega)$ is given by the $(N+1)^2 \times 1$ vector:

$$\mathbf{h}_{\mathbf{nm}}^{\mathrm{LS}} = [h_{00}^{\mathrm{LS}}, h_{1(-1)}^{\mathrm{LS}}, \ldots, h_{NN}^{\mathrm{LS}}]^T, \quad (4)$$

$$\mathbf{y}(\Omega) = [Y_0^0(\Omega), Y_1^{-1}(\Omega), \ldots, Y_N^N(\Omega)]^T. \quad (5)$$

Although the LS solution matches both the magnitude and phase of the HRTF, its accuracy degrades when using lower spherical harmonics orders N [13]. Consequently, the LS solution is generally recommended only up to a cutoff frequency of:

$$f_{\mathrm{c}} = \frac{c}{2\pi r} N, \quad (6)$$

where r is the head radius of a listener and c is the speed of sound.

According to the duplex theory [14], the perceptual weighting of ITDs decreases with the increase of frequency, while the relative weighting of ILDs increases. Zaunschirm et al. [8] further showed that higher-order spherical harmonics modes primarily capture rapid phase variations occurring at higher frequencies. Inspired by these observations, the MLS method has been developed [9]. This method optimizes the spherical harmonics coefficients solely by matching HRTF magnitudes for the frequencies higher than f_c, thereby effectively reducing the HRTF order with minimal perceptual impact. Formally, for $f \geq f_c$:

$$h_{nm}^{\mathrm{MLS}} = \arg\min_{h_{nm}} \sum_{\Omega} \left\| \sum_{n=0}^{N} \sum_{m=-n}^{n} |h_{nm}\, Y_n^m(\Omega)| - |h(\Omega)| \right\|_2^2. \quad (7)$$

Compared with the LS, the MLS solution improves the perceptual quality of low-order binaural rendering and is considered as the state-of-the-art benchmark for this task.

However, when the sound source is located contralateral to the considered ear, the MLS solution exhibits much larger magnitude errors shifted towards these contralateral regions when compared with the LS solution. Psychoacoustic experiments suggest that the high-frequency information arriving from contralateral directions plays a diminished role in sound localization [15]. Although this results implies that spectral distortions in contralateral regions may be perceptually less critical, specifically for localization, these distortions may potentially affect other perceptual attributes, such as timbre or externalization.

III. Proposed Method

To address the increased rendering errors observed in contralateral directions under the MLS method, we propose a novel DH optimization framework. In this framework, we divide the space into three zones at first, i.e., contralateral, ipsilateral, and transition zones, and then treat each zone with different strategies.

In the contralateral region (denoted by Ω_{contra}), the HRTF is reconstructed from the LS optimization considering both magnitude and phase. In the ipsilateral region (denoted by Ω_{ipsi}), the MLS optimization is retained. To ensure the spatial continuity, in the transition region (denoted by Ω_{trans}), the HRTF is a combination of the LS and MLS optimization. Thus, the directionally hybrid HRTF of the three regions can be written in a unified form:

$$h_{\mathrm{DH}}(\Omega) = \lambda(\Omega) h_{\mathrm{MLS}}(\Omega) + [1 - \lambda(\Omega)] h_{\mathrm{LS}}(\Omega), \quad (8)$$

where $\Omega \in (\Omega_{\text{contra}}, \Omega_{\text{ipsi}}, \Omega_{\text{trans}})$, $h_{\mathrm{LS}}(\Omega)$ and $h_{\mathrm{MLS}}(\Omega)$ are obtained by conducting discrete inverse spherical Fourier transform on h_{nm}^{LS} and h_{nm}^{MLS}, respectively. $\lambda(\Omega)$ is the region-dependent weight function.

These three regions are divided according to the normalized magnitude error (NME) [16], which is denoted as ϵ_{NME} in this paper. The division strategy is motivated by the conclusion made in [17], that in the high-frequency region, accuracy in magnitude may be more important perceptually than accuracy in phase. To the end, in order to achieve lower ϵ_{NME}, the weight function $\lambda(\Omega)$ is defined as:

$$\lambda(\Omega) = \begin{cases} 0, & \Omega \in \Omega_{\text{contra}} \\ \beta \frac{D(\Omega)}{\max_{\Omega} D(\Omega)}, & \Omega \in \Omega_{\text{trans}} \\ 1, & \Omega \in \Omega_{\text{ipsi}} \end{cases} \quad (9)$$

where $D(\Omega) = \epsilon_{\mathrm{NME}}^{\mathrm{MLS}}(\Omega) - \epsilon_{\mathrm{NME}}^{\mathrm{LS}}(\Omega)$ is the difference function. The spatial regions are then assigned based on $D(\Omega)$ and a threshold parameter β:

$$\Omega \in \begin{cases} \Omega_{\text{ipsi}}, & D(\Omega) < 0 \\ \Omega_{\text{trans}}, & 0 \leq D(\Omega) < \frac{\max_{\Omega} D(\Omega)}{\beta} \\ \Omega_{\text{contra}}. & D(\Omega) \geq \frac{\max_{\Omega} D(\Omega)}{\beta} \end{cases} \quad (10)$$

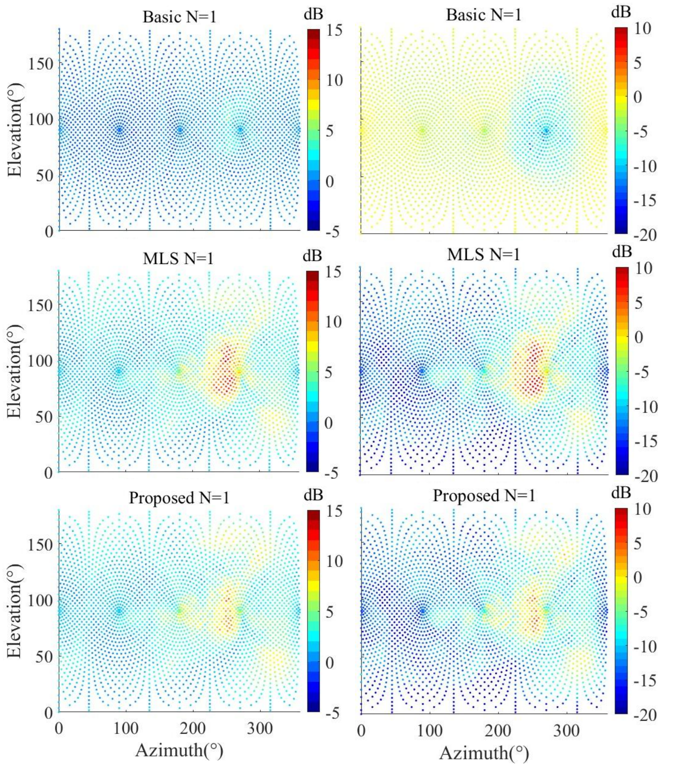

Fig. 1. NMSE (left), $\epsilon_{\text{NMSE}}(\Omega)$, computed as in (11) and NME (right), $\epsilon_{\text{NME}}(\Omega)$, computed as in (12) at the left ear for point sound sources propagating from 434 directions sampled on a Lebedev scheme.

The parameter β was selected empirically as 2.

Finally, we compute the optimized spherical harmonic coefficients h_{nm}^{DH} that can represent the directionally hybrid HRTF $h_{\text{DH}}(\Omega)$ across all directions through (3). This step integrates the hybrid reconstruction into a single set of optimized spherical harmonics coefficients suitable for efficient binaural rendering.

IV. Evaluation

A. Objective Performance Measures

1) Experimental Setup: The KU-100 HRTF dataset [18] was employed for evaluation. The HRTFs were processed using three methods, including the basic method, the MLS method, and the proposed method with an order $N = 1$. The dataset provides HRTF measurements over $Q = 2702$ directions sampled on a Lebedev grid. The reference signal p_{ref} was computed using (1) with $N = 41$, and the evaluated signal p was computed with an order $N = 1$. Four objective metrics were chosen: normalized mean square error (NMSE), NME, ITD, and ILD. The first two measures, NMSE and NME, were evaluated at the left ear ($\phi = 90°$, $\theta = 0°$) for point sound sources propagating from 434 directions sampled on a Lebedev scheme, and the last two metrics, ITD and ILD, were evaluated for point sound sources from 500 directions in uniform sampling on the horizontal plane.

The NMSE metric measures overall physical accuracy and

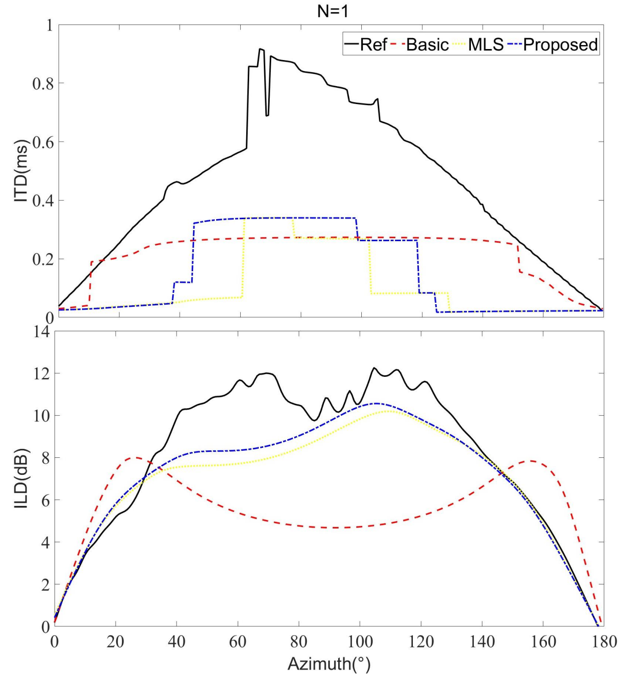

Fig. 2. ITD (top), $\epsilon_{\text{ITD}}(\Omega)$, computed as in (13) and ILD (bottom), $\epsilon_{\text{ILD}}(\Omega)$, computed as in (14) averaged across 500 directions in uniform sampling on the horizontal plane.

is sensitive to both magnitude and phase errors, defined as:

$$\epsilon_{\text{NMSE}}(\Omega) = 10\log_{10}\frac{|p_{\text{ref}}(\Omega) - p(\Omega)|^2}{|p_{\text{ref}}(\Omega)|^2}. \tag{11}$$

Then, the NME measures magnitude errors and is critical for coloration and high-frequency localization, defined as:

$$\epsilon_{\text{NME}}(\Omega) = 10\log_{10}\frac{||p_{\text{ref}}(\Omega)| - |p(\Omega)||^2}{|p_{\text{ref}}(\Omega)|^2}. \tag{12}$$

The ITD is estimated using the peak of the interaural cross-correlation function computed in the frequency domain. The mean absolute ITD error over all evaluated directions is:

$$\epsilon_{\text{ITD}} = \frac{1}{Q}\sum_{\Omega}|\text{ITD}_{\text{ref}}(\Omega) - \text{ITD}_{\text{recon}}(\Omega)|, \tag{13}$$

where $\text{ITD}_{\text{ref}}(\Omega)$ is the reference ITD at direction Ω and $\text{ITD}_{\text{recon}}(\Omega)$ is the ITD computed from the binaural signals generated by one of the evaluated methods.

Finally, the mean absolute ILD error across evaluated directions is:

$$\epsilon_{\text{ILD}} = \frac{1}{Q}\sum_{\Omega}|\text{ILD}_{\text{ref}}(\Omega) - \text{ILD}_{\text{recon}}(\Omega)|, \tag{14}$$

where $\text{ILD}_{\text{ref}}(\Omega)$ is the reference ILD at direction Ω, and $\text{ILD}_{\text{recon}}(\Omega)$ is the ILD computed from the binaural signals generated by one of the evaluated methods.

TABLE I
OBJECTIVE PERFORMANCE MEASURES

Methods	ϵ_{NMSE} (dB)	ϵ_{NME} (dB)	ϵ_{ITD} (ms)	ϵ_{ILD} (dB)
Basic	-0.73	-3.91	0.223	3.66
MLS	2.04	-10.4	0.337	1.26
Proposed	1.87	-10.6	0.303	0.93

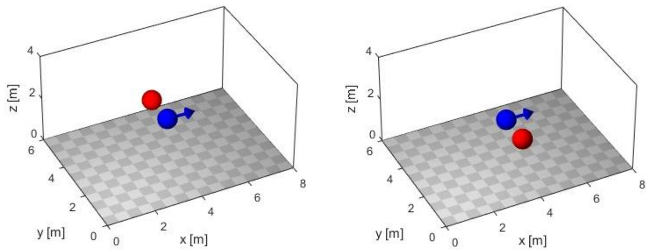

Fig. 3. Room layout and source/listener position used for simulation via McRoomSIM.

2) Results: As summarized in Table I, the proposed DH method achieves the lowest binaural reproduction error across all objective metrics when compared with both the baseline LS method and the state-of-the-art MLS method. Fig. 1 further illustrates the spatial distribution of ϵ_{NMSE} and ϵ_{NME}. Crucially, the proposed DH method effectively suppresses errors in contralateral regions (e.g., $225° < \phi < 315°$ for the left ear), validating the efficacy of its directional hybrid optimization approach. Fig. 2 presents the accuracy of spatial cues. The proposed DH method demonstrates improved reconstruction of both ITDs and ILDs, particularly at lateral positions ($45° < \phi < 135°$), which are critical for sound localization.

B. Subject Listening Experiments

1) Experimental Setup: To perceptually validate the objective results, a controlled listening experiment was conducted using the MUSHRA methodology [19]. A virtual shoebox room (dimensions $8 \times 6 \times 4$ m) was simulated using the multichannel room acoustics simulation toolbox (McRoomSIM) [20]. All room boundaries were assigned a uniform absorption coefficient of 0.28, resulting in a mean reverberation time of $T_{60} = 0.45s$. A simulated spherical microphone array was positioned at the listener location $(x, y, z) = (4, 3, 1.7)$ m. Two discrete omnidirectional sound sources were tested at $(4, 1.6, 1.7)$ m and $(4, 4.4, 1.7)$ m, which are a distance of 1.4 m and the immediate left and right of the array, as shown in Fig. 3. Castanets were selected as the audio source signal due to their rich high-frequency content and prominent transients. The same pre-processed KU-100 HRTFs evaluated in the objective measures (Section IV-A) were employed for binaural rendering of the simulated sound fields. The reference signal used the high-order ($N = 41$) LS optimization.

We recruited six normal-hearing subjects, including 2 females and 4 males, who are aged 24–26 years, with prior experience to evaluate spatial audio content. Participants used their own computers and headphones. Appropriate headphone equalization (compensation) was applied to ensure a flat frequency response [21]. Participants were instructed to perform the test in as quiet an environment as possible. Participants rated the perceived degradation of the test signals relative to the hidden reference signal using a continuous scale from 0 to 5, where a score of 5 means that the difference from the reference is imperceptible, and 0 means a very large perceived difference from the reference. The experiment comprised four separate tests that two tests (one for each source position) for spherical harmonics order $N = 1$ and $N = 4$, respectively. Each test presented four signals for rating, including the hidden reference and the three test conditions.

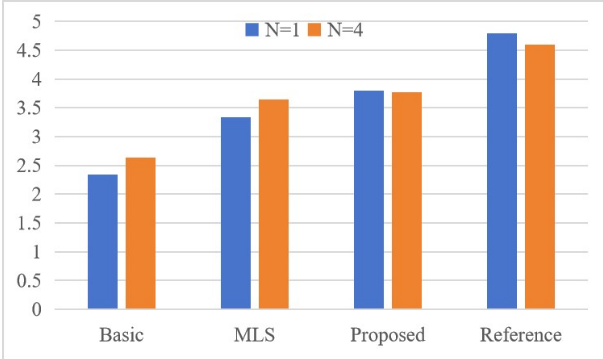

Fig. 4. The histogram of listening test scores.

2) Results: The results of the MUSHRA listening experiment for all conditions and subjects are shown in Fig. 4. One test result (out of four total tests) from a single subject was excluded from the analysis because this subject rated the hidden reference signal below 4 (specifically lower than 4). Importantly, the hidden reference scores for the included data were all above 4.5, confirming that participants were consistently able to identify the reference signal. Each column in Fig. 4 shows the results for a different pre-processing method at a specific spherical harmonics order. The results confirm the superiority of the proposed DH method. Both the MLS and DH methods scored over one point higher on average than the baseline LS method. Moreover, the DH method scored approximately 0.5 points higher than the MLS method at order $N = 1$ and approximately 0.15 points higher at order $N = 4$. The reduction of the performance gap between DH and MLS is due to that, as the spherical harmonics order increases, the relative improvement offered by DH becomes smaller. This is likely attributable to the fact that higher spherical harmonics orders inherently capture more spatial phase information, reducing the impact of the specific optimization strategy.

Both objective and subjective results demonstrate that the proposed DH method effectively mitigates rendering errors in contralateral directions while preserving essential spatial cues. This enables high-quality binaural reproduction even only using low-order spherical harmonics.

V. Conclusions

This work tackles persistent challenges in using low-order ambisonic signals for binaural reproduction, specifically reducing contralateral reconstruction errors and spatial cue inaccuracies. The proposed DH method employs directional spatial partitioning optimization, where contralateral regions utilize the LS optimization, ipsilateral regions apply the magnitude-only MLS processing, and transition zones implement the linear blending to ensure continuity.

Objective results validate the effectiveness of DH , showing much lower errors in all evaluated metrics. Visualizations demonstrate the suppression of contralateral errors and the improved ITD/ILD accuracy at lateral source positions. Subjective evaluations further reveal consistent superiority, where the obtained DH scores exceed MLS by 0.5 MUSHRA points at order $N = 1$ and 0.15 points at $N = 4$ across source locations. The diminished improvement at higher orders correlates with increased phase-information encoding in spherical harmonics.

The proposed method provides an alternative solution for low-order spatial audio rendering when only using low-order ambisonic signals, balancing physical accuracy with perceptual quality. Future work will investigate dynamic spatial partitioning thresholds and individualized HRTF adaptation strategies.

References

[1] F. Zotter and M. Frank, *Ambisonics: A Practical 3D Audio Theory for Recording, Studio Production, Sound Reinforcement, and Virtual Reality*. 2019.

[2] M. A. Gerzon, "Ambisonics in multichannel broadcasting and video," *j.audio eng.soc*, 1985.

[3] B. Rafaely, *Fundamentals of spherical array processing*. Fundamentals of Spherical Array Processing, 2015.

[4] Z. Ben-Hur, F. Brinkmann, J. Sheaffer, S. Weinzierl, and B. Rafaely, "Spectral equalization in binaural signals represented by order-truncated spherical harmonics," *Journal of the Acoustical Society of America*, vol. 141, no. 6, p. 4087, 2017.

[5] T. Lübeck, J. M. Arend, C. Prschmann, H. Helmholz, and J. Ahrens, "Perceptual evaluation of mitigation approaches of impairments due to spatial undersampling in binaural rendering of spherical microphone array data: Dry acoustic environments," in *International Conference on Digital Audio Effects 2020*, 2020.

[6] C. Hold, H. Gamper, V. Pulkki, N. Raghuvanshi, and I. Tashev, "Improving binaural ambisonics decoding by spherical harmonics domain tapering and coloration compensation," in *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2019.

[7] F. Brinkmann and S. Weinzierl, "Comparison of head-related transfer functions pre-processing techniques for spherical harmonics decomposition," in *2018 AES International Conference on Audio for Virtual and Augmented Reality: Science, Technology, Design, and Implementation: Redmond, Washington, USA, 20-22 August 2018*, 2018.

[8] Z. Markus, S. Christian, and H. Robert, "Binaural rendering of ambisonic signals by head-related impulse response time alignment and a diffuseness constraint," *The Journal of the Acoustical Society of America*, vol. 143, no. 6, pp. 3616–3627, 2018.

[9] C. Schrkhuber, M. Zaunschirm, and R. Hldrich, "Binaural rendering of ambisonic signals via magnitude least squares," in *Fortschritter der Akustik (DAGA)*, 2018.

[10] O. Berebi, F. Brinkmann, S. Weinzierl, and B. Rafaely, "Ambisonics binaural rendering via masked magnitude least squares," *arXiv preprint arXiv:2501.18224*, 2025.

[11] M. Noisternig, A. Sontacchi, T. Musil, and R. Holdrich, "A 3d ambisonic based binaural sound reproduction system," in *Audio Engineering Society Conference: 24th International Conference: Multichannel Audio, The New Reality*, Audio Engineering Society, 2003.

[12] B. Bernschütz, A. V. Giner, C. Pörschmann, and J. Arend, "Binaural reproduction of plane waves with reduced modal order," *Acta Acustica united with Acustica*, vol. 100, no. 5, pp. 972–983, 2014.

[13] Z. Ben-Hur, D. L. Alon, R. Mehra, and B. Rafaely, "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*, vol. 27, no. 12, pp. 2249–2262, 2019.

[14] W. M. Hartmann, B. Rakerd, Z. D. Crawford, and P. X. Zhang, "Transaural experiments and a revised duplex theory for the localization of low-frequency tones," *The Journal of the Acoustical Society of America*, vol. 139, no. 2, pp. 968–985, 2016.

[15] E. A. Macpherson and A. T. Sabin, "Binaural weighting of monaural spectral cues for sound localization," *The Journal of the Acoustical Society of America*, vol. 121, no. 6, pp. 3677–3688, 2007.

[16] O. Berebi, Z. Ben-Hur, D. L. Alon, and B. Rafaely, "Analysis and design of head-tracked compensation for bilateral ambisonics," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 32, pp. 959–972, 2023.

[17] F. L. Wightman and D. J. Kistler, "The dominant role of low-frequency interaural time differences in sound localization," *The Journal of the Acoustical Society of America*, vol. 91, no. 3, pp. 1648–1661, 1992.

[18] B. Bernschütz, "A spherical far field hrir/hrtf compilation of the neumann ku-100," in *Proceedings of the 40th Italian (AIA) annual conference on acoustics and the 39th German annual conference on acoustics (DAGA) conference on acoustics*, German Acoustical Society (DEGA) Berlin, vol. 29, 2013.

[19] ITU-R, *Methods for the subjective assessment of small impairments in audio systems including multichan-*

nel sound systems, Recommendation, Rev. 1, Geneva, Switzerland, 1997.

[20] A. Wabnitz, N. Epain, C. Jin, and A. Van Schaik, "Room acoustics simulation for multichannel microphone arrays," in *Proceedings of the International Symposium on Room Acoustics*, Citeseer, 2010, pp. 1–6.

[21] X. Li, Z. Cai, C. Zheng, and X. Li, "Equalization of loudspeaker response using balanced model truncation," *The Journal of the Acoustical Society of America*, vol. 137, no. 4, EL241–EL247, 2015.