FORUM ACUSTICUM EURONOISE 2025

AMBISONICS ENCODER FOR WEARABLE ARRAY WITH IMPROVED BINAURAL REPRODUCTION

Yhonatan Gayer[1]   **Vladimir Tourbabin**[2]   **Zamir Ben-Hur**[2]
David Alon[2]   **Boaz Rafaely**[1,2]

[1] School of Electrical and Computer Engineering, Ben-Gurion University of the Negev, Beer-Sheva, Israel.
[2] Reality Labs Research at Meta, Redmond, WA, USA.

ABSTRACT

Ambisonics Signal Matching (ASM) is a recently proposed signal-independent approach to encoding Ambisonic signal from wearable microphone arrays, enabling efficient and standardized spatial sound reproduction. However, reproduction accuracy is currently limited due to the non-ideal layout of the microphones. This research introduces an enhanced ASM encoder that reformulates the loss function by integrating a Binaural Signal Matching (BSM) term into the optimization framework. The aim of this reformulation is to improve the accuracy of binaural reproduction when integrating the Ambisonic signal with Head-Related Transfer Functions (HRTFs), making the encoded Ambisonic signal better suited for binaural reproduction. This paper first presents the mathematical formulation developed to align the ASM and BSM objectives in a single loss function, followed by a simulation study with a simulated microphone array mounted on a rigid sphere representing a head-mounted wearable array. The analysis shows that improved binaural reproduction with the encoded Ambisonic signal can be achieved using this joint ASM-BSM optimization, thereby enabling higher-quality binaural playback for virtual and augmented reality applications based on Ambisonics.

Keywords: ambisonics, binaural reproduction, augmented reality

1. INTRODUCTION

Binaural reproduction is essential for creating immersive auditory experiences in virtual reality (VR) and augmented reality (AR), enabling spatial perception by enabling real-world acoustic cues [1]. Ambisonics [2] has emerged as a powerful method for spatial audio representation due to its ability to encode sound fields independent of the playback setup. When combined with individualized Head-Related Transfer Functions (HRTFs) [3], Ambisonics allows for accurate binaural reproduction over headphones, preserving directionality and depth cues.

Conventionally, Ambisonic signal are captured using spherical microphone arrays [4]. However, such capture rely on specialized arrays [5, 6], limiting its use in mobile or wearable VR/AR applications [7]. Therefore, due to limitations in Ambisonics encoding, alternative binaural rendering approaches such as Binaural Signal Matching (BSM) have been proposed [8]. These methods bypass the Ambisonics format and directly optimize the binaural signals to better match the target spatial cues. However, this approach is not as generic and versatile as Ambisonics, as it is both array and HRTF dependent.

To maintain the benefits of Ambisonics-based encoding while achieving performance that is close to BSM-based binaural reproduction, recent work has introduced the use of residual channels [9]. This approach demonstrated that encoding additional channels beyond standard Ambisonics can achieve BSM-level accuracy.

Building on this concept, this paper aims to develop a First-Order Ambisonics (FOA) encoding method specifically optimized for binaural reproduction, but through modification of the encoding loss function, rather than through the addition of channels.

2. BACKGROUND

This section provides the mathematical foundation for the signal model, presents the Ambisonics representation derived from this model, and outlines the formulations for encoding Ambisonics and rendering binaural signals using both spherical and arbitrary microphone arrays.

2.1 Signal Model

Consider an array consisting of M omnidirectional microphones, where each microphone is located at spherical coordinates (r_i, θ_i, ϕ_i) for $1 \leq i \leq M$. Where θ_i represents the elevation angle and ϕ_i represents the azimuth angle, and r_i is a radius in meters $\forall 1 \leq i \leq Q$. Additionally, assume the presence of Q plane waves, each arriving from a distinct direction (θ_q, ϕ_q), where $1 \leq q \leq Q$, forming the set Ω_Q. The array steering matrix, represented as $\mathbf{V}$, has dimensions $M \times Q$, where each element $[\mathbf{V}]_{i,q}$ corresponds to the frequency response of the i-th microphone to a plane wave incident from (θ_q, ϕ_q) at wave number k. The measured microphone signals are modeled as:

$$\mathbf{x}(k) = \mathbf{V}(k)\mathbf{s}(k) + \mathbf{n}(k) \tag{1}$$

where $\mathbf{x}(k) = [x_1(k), ..., x_M(k)]^T$ is the vector of microphone signals, and each x_i represents the measurement from the i-th microphone. The vector $\mathbf{s}(k) = [s_1(k), ..., s_Q(k)]^T$ contains the source signals, with each entry representing the amplitude of a plane wave at the origin. The noise vector $\mathbf{n}(k) = [n_1(k), ..., n_M(k)]^T$ consists of independent, identically distributed (i.i.d.) noise components, uncorrelated with $\mathbf{s}(k)$.

The binaural signal at the left or right ear can be computed using the spatial domain representation of the Head-Related Transfer Function (HRTF), denoted as $\mathbf{h}^{l,r}(k) = [h^{l,r}(\theta_1, \phi_1, k), ..., h^{l,r}(\theta_Q, \phi_Q, k)]^T$, and the source vector $\mathbf{s}$ [10]:

$$p^{l,r}(k) = \mathbf{h}^{l,r}(k)^T \mathbf{s}(k) \tag{2}$$

2.2 Ambisonics Representation

The Ambisonics representation of the signal $\mathbf{s}$ with respect to the Q plane waves can be written as [5]:

$$\mathbf{a_{nm}}(k) = \mathbf{Y}_{\mathbf{\Omega}_Q}^H \mathbf{s}(k) \tag{3}$$

where $\mathbf{Y}_{\mathbf{\Omega}_Q}$ is the Spherical Harmonics (SH) matrix of size $Q \times (N_a + 1)^2$. Each column vector $\mathbf{y_{nm}} = [Y_{nm}(\theta_1, \phi_1), ..., Y_{nm}(\theta_Q, \phi_Q)]^T$ represents SH functions of order n and degree m evaluated at (θ_q, ϕ_q). The

vector $\mathbf{a_{nm}}(k) = [a_{00}(k), ..., a_{N_a N_a}(k)]^T$ contains the Ambisonics coefficients up to order N_a and is of length $(N_a + 1)^2$.

Binaural reproduction based on Ambisonics encoding is given by [2]:

$$p^{l,r}(k) = \tilde{\mathbf{h}}_{\mathbf{nm}}^{l,r}(k)^T \mathbf{a_{nm}}(k) \tag{4}$$

where $\tilde{\mathbf{h}}_{\mathbf{nm}}^{l,r}$ is a modified SH-domain representation of the HRTF. This representation is derived from the HRTF vector $\mathbf{h}_{\mathbf{nm}}^{l,r}$ using the relation $\tilde{h}_{nm} = (-1)^m h_{n,-m}$. To ensure compatibility in Ambisonics reproduction, the orders of $\tilde{\mathbf{h}}_{\mathbf{nm}}^{l,r}$ and $\mathbf{a_{nm}}$ must be truncated such that $N_a = N_h$ [11].

For spherical arrays with near-uniform distributions, Ambisonics encoding can be performed accurately using Plane Wave Decomposition (PWD) [12], ensuring correct computation of Ambisonics coefficients for all $n \leq N_a$ and $-n \leq m \leq n$, subject to the condition:

$$(N_a + 1)^2 \leq M \tag{5}$$

This condition guarantees accurate Ambisonics encoding within the operational frequency range of the array [13].

2.3 Ambisonics Encoding for Arbitrary Arrays

To maintain conciseness, the wave number k will be omitted in all sections of this paper from this point onward. To enable binaural reproduction from arbitrary microphone arrays, Ambisonics channels can be directly encoded from the array signals. Traditionally, Ambisonics encoding is performed using spherical microphone arrays or purposefully designed configurations [2, 5, 6]. This formulation be can be generalized to arbitrary array geometries, as proposed in [11, 14], which employs Tikhonov regularization [15], thus providing a linear mapping from the microphone signals to the Ambisonics representation:

$$\hat{a}_{nm} = \mathbf{c}_{nm}^H \mathbf{x}, \\ \forall 0 \leq n \leq N_a, \ -n \leq m \leq n \tag{6}$$

where $\hat{\mathbf{a}}_{\mathbf{nm}} = [\hat{a}_{00}, ..., \hat{a}_{N_a N_a}]^T$ denotes the estimated Ambisonics vector, of length $(N_a + 1)^2$, and $\mathbf{c}_{nm}$ is the encoding Ambisonics filter of the same length. This approach entails minimizing the following Normalized Mean Squared Error (NMSE) function to compute the optimal coefficients:

$$\varepsilon_{nm}^{\text{ASM}} = E\left[\|\hat{a}_{nm} - a_{nm}\|_2^2\right] \Big/ E\left[\|a_{nm}\|_2^2\right] \tag{7}$$

Minimization is achieved by substituting (1), (3), and (6) into (7). For a signal-independent formulation, we assume that the noise $\mathbf{n}$ is white with covariance $\mathbf{R}_n = \sigma_n^2 \mathbf{I}$ and uncorrelated with $\mathbf{s}$, and that the source covariance matrix satisfies $\mathbf{R_s} = \sigma_s^2 \mathbf{I}$, corresponding to a diffuse sound field of Q plane waves, the NMSE reduces to:

$$\varepsilon_{nm}^{\text{ASM}} = \frac{\sigma_s^2 \left\| \mathbf{V}^H \mathbf{c}_{nm} - \mathbf{y}_{nm} \right\|_2^2 + \sigma_n^2 \left\| \mathbf{c}_{nm} \right\|_2^2}{\sigma_s^2 \left\| \mathbf{y}_{nm} \right\|_2^2} \quad (8)$$

Solving (8) leads to the optimal filter coefficients:

$$\mathbf{c}_{nm}^{\text{ASM}} = [\mathbf{y}_{nm}]^H \mathbf{V}^H \left(\mathbf{V}\mathbf{V}^H + \frac{\sigma_n^2}{\sigma_s^2} \mathbf{I} \right)^{-1} \quad (9)$$

where $\mathbf{c}_{nm}^{\text{ASM}}$ represents the optimal filters that minimize (8) with respect to $[\mathbf{c}_{nm}]^H$. Equation (9) is valid provided that the inverse exists, which is ensured by the presence of the regularization term $\frac{\sigma_n^2}{\sigma_s^2}\mathbf{I}$. By substituting (9) into (6), an estimated Ambisonics representation is obtained via Ambisonics Signal Matching (ASM).

While this formulation is specific to each Ambisonics signal individually, it is extended here to reformulate the filter to encode all Ambisonics signals up to order $(N_a + 1)^2$:

$$[\mathbf{C_{nm}}^{\text{ASM}}]^H = [\mathbf{c}_{00}^{\text{ASM}}, \dots, \mathbf{c}_{N_a N_a}^{\text{ASM}}]^H$$
$$= [\mathbf{Y}_{\Omega_Q}]^H \mathbf{V}^H \left(\mathbf{V}\mathbf{V}^H + \frac{\sigma_n^2}{\sigma_s^2} \mathbf{I} \right)^{-1} \quad (10)$$

This reformulation provides a more convenient representation of the ASM method, to be used later in this paper.

2.4 Binaural Signal Matching (BSM) for Arbitrary Arrays

A direct approach for reproducing binaural signals from arbitrary microphone arrays is introduced in [10]. This method directly derives the binaural signals from the microphone recordings via BSM. The estimated binaural signal is given by:

$$\hat{p}_{\text{BSM}}^{l,r} = [\mathbf{c}_{\text{BSM}}^{l,r}]^H \mathbf{x} \quad (11)$$

where $\mathbf{c}_{\text{BSM}}^{l,r}$ is the BSM filter, optimized by minimizing the Mean Squared Error (MSE):

$$\varepsilon_{\text{BSM}} = E \left[\| \hat{p}_{\text{BSM}}^{l,r} - p^{l,r} \|_2^2 \right] \quad (12)$$

Under the assumptions of white noise $\mathbf{R}_n = \sigma_n^2 \mathbf{I}$ and an isotropic sound field where $\mathbf{R_s} = \sigma_s^2 \mathbf{I}$, the optimal BSM filter is obtained via Tikhonov regularization [15]:

$$[\mathbf{c}_{\text{BSM}}^{l,r}]^H = [\mathbf{h}^{l,r}]^T \mathbf{V}^H \left(\mathbf{V}\mathbf{V}^H + \frac{\sigma_n^2}{\sigma_s^2} \mathbf{I} \right)^{-1} \quad (13)$$

This formulation ensures robust binaural signal estimation for arbitrary array configurations.

3. PROPOSED METHOD

This paper introduces a unified approach for Ambisonics encoding by jointly minimizing ASM and BSM errors using a single loss function. Unlike existing methods that optimize these errors separately, the proposed method balances both objectives within a unified framework to compute an Ambisonic filter that encodes Ambisonics optimized for binaural reproduction.

3.1 ASM Design for Binaural Reproduction

Binaural reproduction using encoded Ambisonics can be achieved by substituting (10) into (4):

$$\hat{p}_{\text{ASM}}^{l,r} = [\tilde{\mathbf{h}}_{\mathbf{nm}}]^T [\mathbf{C_{nm}}^{\text{ASM}}]^H \mathbf{x} \quad (14)$$

where $\hat{p}_{ASM}^{l,r}$ represents the binaural signal reproduced from encoded Ambisonics via ASM. It is important to note that this formulation of the reproduced binaural signals does not minimize the binaural error. Instead, it represents the binaural signals reproduced using the encoded Ambisonics.

To derive the binaural error for the signal in (14), we can substitute $\hat{p}_{ASM}^{l,r}$ to replace $\hat{p}_{BSM}^{l,r}$ in (12). By making the same assumptions of a diffuse sound field and white, uncorrelated microphone noise as in Sec. 2.3, we obtain:

$$\varepsilon_{\mathbf{nm}}^{\text{BSM}} = \frac{\sigma_s^2 \left\| [\tilde{\mathbf{h}}_{\mathbf{nm}}^{l,r}]^T [\mathbf{C_{nm}}^{\text{ASM}}]^H \mathbf{V} - [\mathbf{h}^{l,r}]^T \right\|_2^2}{\sigma_s^2 \left\| \mathbf{h}^{l,r} \right\|_2^2}$$
$$+ \frac{\sigma_n^2 \left\| [\tilde{\mathbf{h}}_{\mathbf{nm}}^{l,r}]^T [\mathbf{C_{nm}}^{\text{ASM}}]^H \right\|_2^2}{\sigma_s^2 \left\| \mathbf{h}^{l,r} \right\|_2^2} \quad (15)$$

Although $\mathbf{C_{nm}}^{\text{ASM}}$, as defined above, minimizes the error in (8), it does not necessarily provide the coefficients that minimize (15). As a first step to formulating an Ambisonics encoder that minimizes the binaural error, the filter

11$^{\text{th}}$ Convention of the European Acoustics Association
Málaga, Spain • 23$^{\text{rd}}$ – 26$^{\text{th}}$ June 2025 •

4323

matrix $\mathbf{C}_{\mathbf{nm}}^{\text{ASM}}$ is restructured to the from of a long vector. Then, the product $[\tilde{\mathbf{h}}_{\mathbf{nm}}^{l,r}]^T[\mathbf{C}_{\mathbf{nm}}^{\text{ASM}}]^H$ is replaced by the following equivalent term:

$$[\tilde{\mathbf{h}}_{\mathbf{nm}}^{l,r}]^T[\mathbf{C}_{\mathbf{nm}}^{\text{ASM}}]^H = [\mathbf{c}_{\mathbf{nm,flat}}^{\text{ASM}}]^H \mathbf{H}_{\mathbf{nm}}^{l,r} \quad (16)$$

where

$$\mathbf{c}_{\mathbf{nm,flat}}^{\text{ASM}} = \left[\left[\mathbf{c}_{00}^{\text{ASM}}\right]^T, \ldots, \left[\mathbf{c}_{N_a N_a}^{\text{ASM}}\right]^T\right]^T \quad (17)$$

is a vectorized form of the matrix $\mathbf{C}_{\mathbf{nm}}^{\text{ASM}}$, changing its dimensions from $M \times (N_a + 1)^2$ to $M(N_a + 1)^2 \times 1$, and

$$\mathbf{H}_{\mathbf{nm}}^{l,r} = \left[\tilde{h}_{00}^{l,r} \, \mathbf{I}, \ldots, \tilde{h}_{N_a N_a}^{l,r} \, \mathbf{I}\right]^T \quad (18)$$

is a Matrix of size $M(N_a + 1)^2 \times M$, where $\mathbf{I}$ is the unit matrix of size $M \times M$. Substituting (17) and (18) in (16) leads to equalities in these equations. Now, (15) can be rewritten using (16) and by further replacing $\mathbf{c}_{\mathbf{nm,flat}}^{\text{ASM}}$ with $\mathbf{c}_{\mathbf{nm,flat}}$, to denote that it is a free parameter to be optimized:

$$\varepsilon_{\mathbf{nm}}^{\text{BSM}} = \frac{\sigma_s^2 \left\|[\mathbf{c}_{\mathbf{nm,flat}}]^H \mathbf{H}_{\mathbf{nm}}^{l,r} \mathbf{V} - [\mathbf{h}^{l,r}]^T\right\|_2^2}{\sigma_s^2 \left\|[\mathbf{h}^{l,r}]^T\right\|_2^2}$$
$$+ \frac{\sigma_n^2 \left\|[\mathbf{c}_{\mathbf{nm,flat}}]^H \mathbf{H}_{\mathbf{nm}}^{l,r}\right\|_2^2}{\sigma_s^2 \left\|[\mathbf{h}^{l,r}]^T\right\|_2^2} \quad (19)$$

Now, $\mathbf{c}_{\mathbf{nm,flat}}^{\text{BSM}}$, can be computed as a solution to the minimization of (19), which can be achieved using Tikhonov regularization [15]:

$$[\mathbf{c}_{\mathbf{nm,flat}}^{\text{BSM}}]^H = \arg\min_{\mathbf{c}_{\mathbf{nm,flat}}} \varepsilon_{\mathbf{nm}}^{\text{BSM}}$$
$$= [\mathbf{h}^{l,r}]^T [\mathbf{H}_{\mathbf{nm}}^{l,r} \mathbf{V}]^H \times$$
$$\left(\mathbf{H}_{\mathbf{nm}}^{l,r} \mathbf{V}[\mathbf{H}_{\mathbf{nm}}^{l,r} \mathbf{V}]^H + \frac{\sigma_n^2}{\sigma_s^2} \mathbf{H}_{\mathbf{nm}}^{l,r}[\mathbf{H}_{\mathbf{nm}}^{l,r}]^H\right)^{-1} \quad (20)$$

where, similar to (17), and (10), $[\mathbf{c}_{\mathbf{nm,flat}}^{\text{BSM}}]$ can be rearenged to the same form as in (15):

$$\mathbf{C}_{\mathbf{nm}}^{\text{BSM}} = \left[\mathbf{c}_{00}^{\text{BSM}}, \ldots, \mathbf{c}_{N_a N_a}^{\text{BSM}}\right] \quad (21)$$

3.2 Joint ASM-BSM design

The authors of this paper demonstrated in [9] that when the encoded Ambisonics order is small, i.e. $N_a \ll N_h$, the minimization of (8) typically leads to high binaural error.

To address this limitation, this paper proposes minimizing the joint ASM-BSM error, defined as:

$$\varepsilon_{\mathbf{nm}}^{\text{joint}} = \alpha \sum_{n=0}^{N_a} \sum_{m=-n}^{n} \varepsilon_{nm}^{\text{ASM}} + (1 - \alpha)\varepsilon_{\mathbf{nm}}^{\text{BSM}} \quad (22)$$

where $\varepsilon_{\mathbf{nm}}^{\text{joint}}$ represents the objective error function, and α acts as a balancing factor. Both $\varepsilon_{nm}^{\text{ASM}}$ as in (8) and $\varepsilon_{\mathbf{nm}}^{\text{BSM}}$ as in (19) depend on the Ambisoniocs encoder coefficients, as the variables to be optimized. The solution to the error in (22) is a linear combination of the solutions to (15) and (8), expressed as:

$$\mathbf{C}_{\mathbf{nm}}^{\text{joint}} = \alpha \, \mathbf{C}_{\mathbf{nm}}^{\text{ASM}} + (1 - \alpha) \, \mathbf{C}_{\mathbf{nm}}^{\text{BSM}} \quad (23)$$

where $\mathbf{C}_{\mathbf{nm}}^{\text{joint}}$ represents the joint ASM-BSM filter.

4. EXPERIMENT

In this experiment, the joint ASM-BSM model is evaluated by measuring the ASM and BSM errors. The model is tested with different α values, allowing us to analyze the trade-off between the ASM and BSM solutions, highlighting the benefits of their combination.

4.1 Setup

Steering functions and HRTF were simulated in Matlab based on a spherical head model. A rigid sphere of radius $10\,\text{cm}$ was simulated, and the ear and microphones were positioned on its surface as detailed below. Ambisonics of order $N_a = 1$ was chosen for encoding, with a reference order of $N_a = 20$. Far-field sources were simulated at 240 directions distributed nearly-uniformly in the entire directional space.

The microphone array consists of $M = 5$ microphones arranged along a semi-circle on the rigid sphere. The location of each microphone is defined in spherical coordinates (θ, ϕ). The microphone positions are given by: $\{(90°, -70°), (72°, -35°), (108°, 0°), (72°, 35°), (90°, 70°)\}$ as illustrated in Fig. 1. These positions aim to represent an array mounted on glasses, such as the one used for the EasyCom database [16].

4.2 Methodology

Joint ASM-BSM filters were as follows:

- Joint - ASM: Joint design with $\alpha = 1$, optimizing only for Ambisonics.

- Joint - BSM: Joint design with $\alpha = 0$, optimizing only for BSM.

- Joint - ASM-BSM: Joint design with $\alpha = 0.5$, balancing between ASM and BSM.

- Std BSM: The standard BSM approach as described in Sec. 2.4.

4.3 Performance Measures

To avoid confusion between the errors used for filter design and those used for performance evaluation, we define the relevant error measures separately in this subsection. The per-channel error for the ASM, similar to (8) is defined as:

$$\xi_{nm}^{\text{ASM}} = \frac{\left\| \mathbf{V}^H \mathbf{c}_{nm}^{\text{joint}} - \mathbf{y}_{nm} \right\|_2^2}{\|\mathbf{y}_{nm}\|_2^2} \qquad (24)$$

This represents the reconstruction error for the n, m Ambisonics channel using filter $\mathbf{c}_{nm}^{\text{joint}}$ which is one of the columns of filter $\mathbf{c}_{nm}^{\text{joint}}$ corresponding to the n, m channel. The error for the binaural reproduction when jointly designing filters, similar to (19) is given by:

$$\xi_{\mathbf{nm}}^{\text{BSM}} = \frac{\left\| [\tilde{\mathbf{h}}_{\mathbf{nm}}^{l,r}]^T [\mathbf{C}_{\mathbf{nm}}^{\text{joint}}]^H \mathbf{V} - [\mathbf{h}^{l,r}]^T \right\|_2^2}{\|\mathbf{h}^{l,r}\|_2^2} \qquad (25)$$

where $\mathbf{C}_{\mathbf{nm}}^{\text{joint}}$ is computed as in Eq. (23). Serving as a reference for binaural reproduction, the performance of the Standard BSM is evaluated using:

$$\xi^{\text{BSM}} = \frac{\left\| [\mathbf{c}_{\text{BSM}}^{l,r}]^H \mathbf{V} - [\mathbf{h}^{l,r}]^T \right\|_2^2}{\|\mathbf{h}^{l,r}\|_2^2} \qquad (26)$$

where $\mathbf{c}_{\text{BSM}}^{l,r}$ represents the standard BSM filter as in (13).

4.4 Results

Fig. 2 shows the performance of the joint ASM-BSM design. For $\alpha = 1$ (upper), the Ambisonics representation is optimized, exhibiting minimal errors at low frequencies which rises only above 1 kHz. However, the binaural error is considerably high for this method. In contrast, for $\alpha = 0$ (middle), the binaural matching is near optimal with errors comparable to those of the standard BSM method while the Ambisonics reconstruction incurs larger errors. For the intermediate case of $\alpha = 0.5$ (lower), the joint ASM-BSM design strikes a balance: Ambisonics errors are reduced relative to the the case of $\alpha = 0$, but are

higher than in the $\alpha = 1$ case. On the other hand, the binaural errors show only a slight increase compared to the $\alpha = 0$ case. Overall, these results highlight the intrinsic tradeoff between high-fidelity Ambisonics reconstruction and effective binaural signal matching.

5. CONCLUSION AND FUTURE WORK

A framework for designing an Ambisonics encoder for a wearable array was presented. Matching for Ambisonics in this case typically leads to a high binaural error, significantly higher than what can usually be achieved by BSM. The joint ASM-BSM design enables the optimization of ASM filters that are also tailored for BSM. These filters achieve a much lower BSM error but at the cost of increased Ambisonics error. The proposed formulation has a closed-form solution and provides a trade-off between ASM and BSM. Currently, only a single listener head orientation relative to the array is considered, and so future research should investigate the effects of head rotation and head tracking. Additionally, future work could explore this approach with more realistic HRTFs and conduct listening tests to assess the advantages of the joint ASM-BSM design. It would also be valuable to understand the implications of errors in the ambisonics signals on applications beyond binaural reproduction.

6. REFERENCES

[1] B. Rafaely, V. Tourbabin, E. Habets, Z. Ben-Hur, H. Lee, H. Gamper, L. Arbel, L. Birnie, T. Abhayapala, and P. Samarasinghe, "Spatial audio signal processing for binaural reproduction of recorded acoustic scenes - review and challenges," vol. 6, 10 2022.

[2] F. Zotter and M. Frank, "Ambisonics: A Practical 3D Audio Theory for Recording, Studio Production, Sound Reinforcement, and Virtual Reality," 01 2019.

[3] B. Masiero, M. Pollow, and J. Fels, "Design of a fast broadband individual head-related transfer function measurement system," vol. 97, pp. 136–136, 2011.

[4] B. Rafaely, "Plane-wave decomposition of the sound field on a sphere by spherical convolution," vol. 116, pp. 2149–2157, 1 2004.

[5] B. Rafaely, "Fundamentals of spherical array processing," vol. 8, Springer, 2015.

[6] J. Meyer and G. W. Elko, "Analysis of the high-frequency extension for spherical eigenbeamforming

microphone arrays.," vol. 127, pp. 1979–1979, Acoustical Society of America, 2010.

[7] B. Rafaely, "Analysis and design of spherical microphone arrays," vol. 13, pp. 135–143, 2005.

[8] L. Madmoni, J. Donley, V. Tourbabin, and B. Rafaely, "Binaural Reproduction From Microphone Array Signals Incorporating Head-Tracking," in *2021 Immersive and 3D Audio: from Architecture to Automotive (I3DA)*, pp. 1–5, 2021.

[9] Y. Gayer, V. Tourbabin, Z. Ben-Hur, J. Donley, and B. Rafaely, "Ambisonics encoding for arbitrary microphone arrays incorporating residual channels for binaural reproduction," in *2024 IEEE International Conference on Acoustics, Speech, and Signal Processing Workshops (ICASSPW)*, pp. 244–248, 2024.

[10] L. Madmoni, Z. Ben-Hur, J. Donley, V. Tourbabin, and B. Rafaely, "Design and analysis of binaural signal matching with arbitrary microphone arrays," *arXiv preprint arXiv:2408.03581*, 2024.

[11] A. Bastine, L. Birnie, T. D. Abhayapala, P. Samarasinghe, and V. Tourbabin, "Ambisonics Capture using Microphones on Head-worn Device of Arbitrary Geometry," in *2022 30th European Signal Processing Conference (EUSIPCO)*, pp. 309–313, 2022.

[12] I. Balmages and B. Rafaely, "Room acoustics measurements by microphone arrays," in *2004 23rd IEEE Convention of Electrical and Electronics Engineers in Israel*, pp. 420–423, 2004.

[13] B. Rafaely, "Analysis and design of spherical microphone arrays," *IEEE Transactions on speech and audio processing*, vol. 13, no. 1, pp. 135–143, 2004.

[14] L. McCormack, A. Politis, R. Gonzalez, T. Lokki, and V. Pulkki, "Parametric Ambisonic Encoding of Arbitrary Microphone Arrays," vol. 30, pp. 2062–2075, 2022.

[15] G. H. Golub, P. C. Hansen, and D. P. O'Leary, "Tikhonov regularization and total least squares," vol. 21, pp. 185–194, SIAM, 1999.

[16] J. Donley, V. Tourbabin, J.-S. Lee, M. Broyles, H. Jiang, J. Shen, M. Pantic, V. K. Ithapu, and R. Mehra, "Easycom: An augmented reality dataset to support algorithms for easy communication in noisy environments," *arXiv preprint arXiv:2107.04174*, 2021.
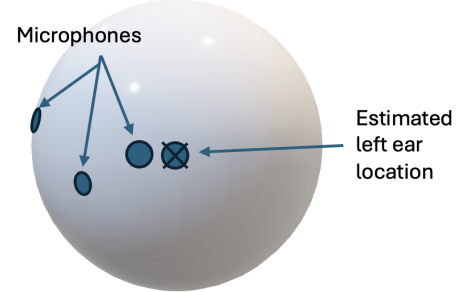
Figure 1: An illustration of the array from a horizontal view. The microphone positions on a rigid sphere, given in spherical coordinates (θ, ϕ), are $\{(90°, -70°), (72°, -35°), (108°, 0°), (72°, 35°), (90°, 70°)\}$, and the array is symmetric. The ears are located at $(90°, \pm 90°)$.
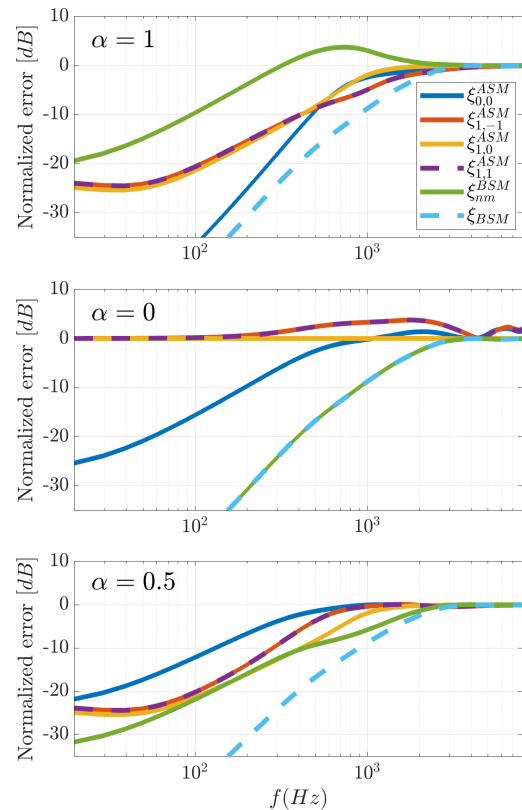

Figure 2: Normalized errors for the case of $\alpha = 1$ (top), showing ξ_{nm}^{ASM} for $(n, m) = (0, 0)$ to $(1, 1)$, ξ_{nm}^{BSM} and ξ^{BSM}. The same errors are shown for $\alpha = 0$ (middle) and $\alpha = 0.5$ (bottom).