

Binaural Signal Matching with Wearable Arrays for Near-Field Sources and Directional Focus

Sapir Goldring¹, Zamir Ben Hur², David Lou Alon², Chad McKell², Sebastian Prepelitã², and Boaz Rafaely¹

¹*School of Electrical and Computer Engineering, Ben-Gurion University of the Negev*

²*Reality Labs Research, Meta, Redmond, WA, USA*

Abstract—This paper investigates the performance of Binaural Signal Matching (BSM) methods for near-field sound reproduction using a wearable glasses-mounted microphone array. BSM is a flexible, signal-independent approach for binaural rendering with arbitrary arrays, but its conventional formulation assumes far-field sources. In our previous work, we proposed a near-field extension of BSM (NF-BSM) that incorporates distance-dependent modeling and showed improved performance over far-field BSM using analytic data, though degradation persisted for sources very close to the array. In this study, we extend that analysis by using realistic simulated data of near-field Head-Related Transfer Functions (HRTFs) and Acoustic Transfer Functions (ATFs) of the array, accounting for listener head rotation and evaluating binaural cues such as interaural level and time differences (ILD and ITD). A key contribution is the introduction of a Field of View (FoV) weighting, designed to emphasize perceptually relevant directions and improve robustness under challenging conditions. Results from both simulation and a listening test confirm that NF-BSM outperforms traditional far-field BSM in near-field scenarios, and that the proposed NF-FoV-BSM method achieves the best perceptual and objective quality among all tested methods, particularly at close source distances and under head rotation. These findings highlight the limitations for far-field models in near-field sources and demonstrate that incorporating source distance and directional weighting can significantly improve binaural reproduction performance for wearable spatial audio systems.

Index Terms—Near-field, binaural reproduction, microphone arrays, higher-order ambisonics, binaural signal matching

I. INTRODUCTION

In recent years, applications such as Virtual Reality (VR) and teleconferencing have garnered significant interest due to their ability to provide immersive and interactive experiences [1]–[4]. Central to these applications is the binaural reproduction of sound, which aims to recreate an acoustic scene using headphones, as it would be perceived by a listener. This process requires information about both the sound field and the Head-Related Transfer Functions (HRTFs) [5].

One approach to binaural reproduction is to incorporate recorded acoustic scenes represented by Ambisonics signals, which can be computed from spherical microphone arrays [6]. Ambisonics has been widely studied for its ability to represent sound fields with high spatial accuracy [2], and can be directly convolved with HRTFs to compute binaural signals [7]. However, while Ambisonics can be readily encoded using spherical arrays, encoding it with arrays of arbitrary configuration remains challenging [2]. To address this limitation, several

methods have been proposed. One class of approaches is based on signal-independent processing. The beamforming-based binaural reproduction (BFBR) technique was introduced in [8], where a set of beamformers is applied to the microphone signals, followed by HRTF filtering. A theoretical framework for optimizing this method’s parameters was later presented in [9]. A recent followup to [9] is binaural signal matching (BSM) [5, 10]–[12], a method noted for its accuracy and flexibility with various array geometries, including a wearable array. BSM is based on estimating binaural signals by minimizing the mean-squared error (MSE) between the array steering vectors and the HRTFs, using a linear formulation. This method has shown promising results in reproducing acoustic scenes with high fidelity. Other approaches are signal-dependent. For example, the method described in [13] encodes microphone array signals into Ambisonic signals using spatial filtering that adapts to the signal content. This technique decomposes the sound field into direct and reverberant components, allowing higher spatial resolution and broader frequency coverage, and is particularly suited for irregular arrays. However, all these methods—signal independent and dependent approaches—share a significant limitation: they assume far-field plane wave sources and have not been specifically studied for near-field sources.

The far-field assumption is generally valid when the sound source is sufficiently distant from the array. Yet, in scenarios where the sound source is close to the array, such as in cases where a speaker is very close to a person with a wearable array, the own voice of the person wearing the array, or when playing a musical instrument, the far-field assumption may not hold. Instead, the sound behaves more like a point source with a radial wavefront, making the plane wave approximation less accurate and leading to errors in the reproduction process [14]. In such near-field scenarios, the development of algorithms that can accurately reproduce the binaural signal remains an open problem.

Previous work by the authors [15] presented the first investigation of BSM in near-field conditions using analytic data derived from rigid-sphere models. This study showed that while incorporating source distance information into the BSM framework (near-field BSM, or in short NF-BSM) substantially reduced binaural error compared to traditional far-field BSM (FF-BSM), performance for very close sources remained limited.

Building on that foundation, the present paper uses realistic

numerically simulated near-field HRTFs and steering data to conduct a more comprehensive evaluation of BSM in near-field scenarios. In addition to refining the NF-BSM framework by adopting a mixed error formulation for binaural reproduction, the study examines perceptual outcomes through a listening test and analyzes binaural cues (ITD and ILD) under head rotations. Furthermore, this paper introduces a Field of View (FoV) weighting approach designed to improve reproduction accuracy by prioritizing directions of interest.

The main contributions of this paper are as follows:

- A more realistic simulation study using realistic Finite Difference Time Domain (FDTD)-generated HRTFs and steering matrices tailored for wearable glasses-mounted arrays.
- An extended NF-BSM framework evaluated under head rotation and binaural cue metrics, with validation through a formal listening test.
- The introduction of a novel FoV weighting design for BSM in near-field conditions, demonstrating improved performance for relevant spatial regions.

II. MATHEMATICAL BACKGROUND

This section introduces the signal model used for microphone array processing, the binaural signal representation, and the framework for BSM, which form the foundation for understanding the proposed approach to near-field binaural reproduction.

A. Microphone Array Measurement Model

This subsection outlines the mathematical framework for modeling signals captured by a microphone array in the context of this study. A standard spherical coordinate system is adopted, represented by (r, θ, ϕ) , where r is the radial distance from the origin, θ is the elevation angle measured from the Cartesian z -axis downward toward the xy -plane, and ϕ is the azimuth angle measured from the positive x -axis in the direction of the positive y -axis. The wave number k is defined as $k = \frac{2\pi}{\lambda}$, where λ is the wavelength, and f represents the frequency. Let $\Omega = (\theta, \phi)$ denote the direction of arrival of a sound source in the sound [6].

The sound field is assumed to be composed of Q sound sources, each with source signals $s_q(k)$ assigned to the sources arriving from directions $\{\Omega_q\}_{q=1}^Q$. This sound field is captured by an M -element microphone array centered at the origin. The recorded signals can then be represented by the following narrowband model [10]:

$$\mathbf{x}(k) = \mathbf{V}(k, \Omega) \mathbf{s}(k) + \mathbf{n}(k), \quad (1)$$

where $\mathbf{x}(k) = [x_1(k), \dots, x_M(k)]^T$ is the M -microphone signal vector; $\mathbf{V}(k, \Omega) \in \mathbb{C}^{M \times Q}$ is the steering matrix whose q -th column is $\mathbf{v}_q(k, \Omega_q)$; $\mathbf{s}(k) = [s_1(k), \dots, s_Q(k)]^T$ is the source vector; $\mathbf{n}(k) = [n_1(k), \dots, n_M(k)]^T$ is additive noise; and $\Omega = \{\Omega_q\}_{q=1}^Q$. Superscript T denotes transpose.

The steering vectors can be derived analytically or numerically for specific array types [6], or obtained through direct measurement. The representation of these steering vectors for different source types, including plane waves and point sources, will be demonstrated in the simulation study.

B. Binaural Signal Model

Assuming that the listener's head is positioned at the origin, the binaural signal can be expressed as:

$$p^{l/r}(k) = [\mathbf{h}^{l/r}(k)]^T \mathbf{s}(k) \quad (2)$$

where $p^{l/r}(k)$ represents the binaural signal for the left (l) and right (r) ears, $[\mathbf{h}^{l/r}(k)]^T$ denotes the transposed HRTF vector for the left and right ears at wave number k , and $\mathbf{s}(k)$ is the source signal vector.

The HRTF vector $[\mathbf{h}^{l/r}(k)]^T$ can be defined as:

$$[\mathbf{h}^{l/r}(k)] = [h^{l/r}(k, \Omega_1) \quad h^{l/r}(k, \Omega_2) \quad \dots \quad h^{l/r}(k, \Omega_Q)]^T \quad (3)$$

where $h^{l/r}(k, \Omega_q)$ represents the HRTF from the acoustic source located at direction Ω_q to the left or right ear of the listener.

It is important to note that HRTFs vary with source distance (e.g., near-field vs far-field), which can influence the binaural signal measured at the listener's ears [16].

C. Binaural Signal Matching

In the BSM approach, the binaural signal is estimated by applying weights to the microphone signals and summing them in a manner similar to beamforming. The estimated binaural signal can be expressed as:

$$\hat{p}^{l/r}(k) = [\mathbf{c}^{l/r}(k)]^H \mathbf{x}(k) \quad (4)$$

where $\hat{p}^{l/r}(k)$ represents the estimated binaural signal for the left (l) and right (r) ears, $\mathbf{c}^{l/r}(k)$ is an $M \times 1$ complex vector of filter weights, $(\cdot)^H$ denotes the Hermitian (conjugate transpose) operator and $\mathbf{s}(k)$ is the source signal vector.

The filter coefficients $\mathbf{c}^{l/r}(k)$ are calculated as the solution to the minimization of the mean squared error $\epsilon^{l/r}$ defined by:

$$\mathbf{c}^{l/r} = \arg \min_{\mathbf{c}} \epsilon^{l/r} \quad (5)$$

where the error $\epsilon^{l/r}$ is given by:

$$\epsilon^{l/r} = \mathbb{E} \left[|p^{l/r}(k) - \hat{p}^{l/r}(k)|^2 \right] \quad (6)$$

with $\mathbb{E}[\cdot]$ representing the expectation operator. The solution to this minimization problem, assuming the sound sources and the noise are i.i.d (independent and identically distributed), is given by:

$$\mathbf{c}^{l/r} = \left(\mathbf{V}\mathbf{V}^H + \frac{\sigma_n^2}{\sigma_s^2} \mathbf{I}_M \right)^{-1} \mathbf{V}[\mathbf{h}^{l/r}]^* \quad (7)$$

Here, σ_n^2 and σ_s^2 are the noise and signal power respectively, $\mathbf{I}_M$ is the $M \times M$ identity matrix and $[\mathbf{h}^{l/r}]^*$ denotes the complex conjugate of the HRTF vector.

The BSM reproduction accuracy is measured by the normalized error, defined as:

$$\epsilon_{LS}^{l/r}(k) = \frac{\mathbb{E} \left[|p^{l/r}(k) - \hat{p}^{l/r}(k)|^2 \right]}{\mathbb{E} \left[|p^{l/r}(k)|^2 \right]} \quad (8)$$

This yields the following expression for the normalized error:

$$\epsilon_{\text{LS}}^{l/r}(k) = \frac{\sigma_s^2 \left\| \mathbf{V}^T \left(\mathbf{c}_{\text{LS}}^{l/r} \right)^* - \mathbf{h}^{l/r} \right\|_2^2 + \sigma_n^2 \left\| \left(\mathbf{c}_{\text{LS}}^{l/r} \right)^* \right\|_2^2}{\sigma_s^2 \left\| \mathbf{h}^{l/r} \right\|_2^2} \quad (9)$$

Here, $\| \cdot \|_2$ denotes the L_2 -norm. For frequencies above approximately 1500 Hz, it has been shown that the perceptual interaural level difference (ILD) becomes more dominant than the interaural time difference (ITD) [10]. Consequently, at high frequencies, the binaural signal matching criterion based on the MSE of the complex signal is replaced by an alternative criterion that matches the magnitude of the binaural signals instead.

This leads to the definition of the magnitude-based normalized error:

$$\epsilon_{\text{MagLS}}^{l/r}(k) = \frac{\left| \left\| \mathbf{V}^T \left(\mathbf{c}_{\text{MagLS}}^{l/r} \right)^* \right\|_2 - \left\| \mathbf{h}^{l/r} \right\|_2 \right|^2}{\left\| \mathbf{h}^{l/r} \right\|_2^2} \quad (10)$$

Here, $\mathbf{c}_{\text{MagLS}}^{l/r}$ denotes the filter weights optimized to minimize this magnitude-based error.

Another error is defined that is composed of a smooth transition between the complex-valued error (important at low frequencies) and the magnitude-based error (important at high frequencies), denoted as mixed error criterion:

$$\epsilon_{\text{mix}}^{l/r}(k) = \alpha(k) \cdot \epsilon_{\text{LS}}^{l/r}(k) + [1 - \alpha(k)] \cdot \epsilon_{\text{MagLS}}^{l/r}(k) \quad (11)$$

where $\epsilon_{\text{LS}}^{l/r}(k)$ is defined as in Eq. (9), using the optimal weights $\mathbf{c}_{\text{LS}}^{l/r}$, and $\epsilon_{\text{MagLS}}^{l/r}(k)$ is computed using the corresponding optimal weights $\mathbf{c}_{\text{MagLS}}^{l/r}$.

The weighting function $\alpha(k)$ is defined as:

$$\alpha(k) = \begin{cases} 1, & f_k < 800 \text{ Hz}, \\ \frac{1500 - f_k}{700}, & 800 \text{ Hz} \leq f_k \leq 1500 \text{ Hz}, \\ 0, & f_k > 1500 \text{ Hz}. \end{cases} \quad (12)$$

where f_k is the frequency corresponding to index k . This definition ensures that the error measure transitions smoothly from complex-valued matching at low frequencies to magnitude-based matching at high frequencies.

III. NEAR-FIELD BSM: MATHEMATICAL FRAMEWORK AND MODELING

This chapter presents the mathematical framework for applying the BSM algorithm to near-field sources [15]. The BSM formulation for near-field sources is introduced, showing the modifications required for steering functions and HRTFs to account for near-field effects. Two evaluation approaches for BSM are then described: one using filter coefficients designed assuming far-field sources but applied to near-field sources and the other using filters designed assuming near-field sources.

A. BSM Formulation for Near-Field Sources

Assume that the sound field is composed of Q point sources. In this case, the signal model specified in Eq. (1) can be written as:

$$\mathbf{x}(k) = \mathbf{V}_{\text{nf}}(k, \Omega, r_s) \mathbf{s}(k) + \mathbf{n}(k), \quad (13)$$

where $\mathbf{V}_{\text{nf}}(k, \Omega, r_s)$ is the point source steering matrix, with r_s denoting the source distance. Each $\mathbf{v}_q(k, \Omega_q, r_q)$ column in this matrix represents the steering vector of a point source located at (Ω_q, r_q) relative to the M -element microphone array. The point source steering matrix differs from the plane wave steering matrix due to the radial wavefront of sound propagation in the near field, as opposed to a planar wavefront in the far field.

The true binaural signal, specified in Eq. ((2)), becomes:

$$p^{l/r}(k) = [\mathbf{h}_{\text{nf}}^{l/r}(k, r_s)]^T \mathbf{s}(k), \quad (14)$$

where $\mathbf{h}_{\text{nf}}^{l/r}(k, r_s)$ denotes the near-field HRTF vector for the left/right ear. It differs from its far-field counterpart because the wavefront curvature is distance-dependent and interacts with the head-torso-pinnae anatomy. In addition, distance-related mechanisms modify the cues: (i) proximity effects increase head shadowing and overall level (via the $1/r$ decay), yielding larger interaural level differences (ILDs) in the near field [17]; and (ii) geometric parallax changes monaural spectral cues because, as a source approaches the head along a radial path, the source-to-ear angle varies with distance (a condition that generally does not hold in the far field) [17, 18]. In practice, real emitters are not ideal point sources, so HRTFs can exhibit source dependence; in our simulations we assume ideal point sources to isolate distance-related effects.

The estimated binaural signal is calculated as detailed in Eq. (4), using BSM filter coefficients $\mathbf{c}^{l/r}(k)$ that are computed based on two modeling assumptions: far-field and near-field propagation. In both cases, the filter coefficients are frequency-dependent and derived using different optimization strategies depending on the frequency range, as governed by the mixed error criterion defined in Eq. (11).

Far-Field BSM Filters: For the far-field configuration, the BSM filter coefficients are computed assuming a plane-wave propagation model. The steering matrix $\mathbf{V}_{\text{ff}}$ corresponds to far-field sources (at a distance of 1.5 m), and the HRTF vector $[\mathbf{h}_{\text{ff}}^{l/r}]$ represents the corresponding far-field HRTFs. Two filters are computed: The LS filter $\mathbf{c}_{\text{LS,ff}}^{l/r}$ is obtained using Eq. (7), by substituting $\mathbf{V} = \mathbf{V}_{\text{ff}}$ and $\mathbf{h} = [\mathbf{h}_{\text{ff}}^{l/r}]$, and used in the complex-valued error criterion (Eq. (9)).

The filter $\mathbf{c}_{\text{MagLS,ff}}^{l/r}(k)$ is computed by minimizing the magnitude-based error defined in Eq. (10). After computing these two filters, they are used in the calculation of the mixed error $\epsilon_{\text{mix}}^{l/r}$ detailed in Eq.(11), with the weighting function $\alpha(k)$ from Eq. (12).

Near-Field BSM Filters: For the near-field configuration, the BSM filters incorporate a point-source propagation model. The steering matrix $\mathbf{V}_{\text{nf}}$ models spherical wavefronts from sources at finite distances (< 1.5 m), and the HRTF vector $[\mathbf{h}_{\text{nf}}^{l/r}]$ corresponds to the near-field HRTFs.

Two filters are computed: the LS filter $\mathbf{c}_{\text{LS},\text{nf}}^{l/r}$, obtained using Eq. (7) with $\mathbf{V} = \mathbf{V}_{\text{nf}}$ and $\mathbf{h} = [\mathbf{h}_{\text{nf}}^{l/r}]$, and the magnitude-based filter $\mathbf{c}_{\text{MagLS},\text{nf}}^{l/r}$, minimizing the criterion in Eq. (10). As in the far-field case, the mixed error $\epsilon_{\text{mix}}^{l/r}$ is then formed by linearly combining the two errors using $\alpha(k)$ (Eq. (11)).

IV. NEAR-FIELD BSM DESIGN WITH LIMITED FIELD OF VIEW

In many practical scenarios, sources are not uniformly distributed across the full spatial domain but are instead concentrated within a limited angular region—for example, in front of the listener during conversation or interaction with a device. To model such directional priors, we introduce a design framework that incorporates a defined *Field of View* (FoV) into the BSM formulation.

The FoV in this work is selected as the region within some aperture in both azimuth and elevation with respect to the frontal direction of the head. This region is intended to model the listener's primary visual region in realistic scenarios, such as face-to-face communication or interactive environments.

To incorporate this directional prior, we modify both the steering matrix and the HRTF vector used in the BSM formulation. Specifically, we apply a spatial weighting function $w(\theta, \phi)$ to each direction of arrival (θ, ϕ) , where:

$$w(\theta, \phi) = \begin{cases} 1, & \text{if } (\theta, \phi) \in \text{FoV} \\ \beta, & \text{otherwise} \end{cases}$$

Here, $0 \leq \beta < 1$ is a small positive constant introduced to avoid completely ignoring the contribution of out-of-FoV directions, which would otherwise lead to instability or excessively large errors in these directions. The value of β can be chosen through a grid search as the optimal tradeoff between minimizing errors within the FoV and avoiding excessive errors outside the FoV.

The modified steering matrix and HRTF vector are then defined as:

$$\tilde{\mathbf{V}} = \mathbf{W}\mathbf{V} \quad (15)$$

$$\tilde{\mathbf{h}}^{l/r} = \mathbf{W}\mathbf{h}^{l/r} \quad (16)$$

where $\mathbf{W} = \text{diag}(w(\theta_1, \phi_1), \dots, w(\theta_Q, \phi_Q))$ is a diagonal matrix of weights applied per DOA, and Q is the number of sampled directions.

These weighted versions are then used in the BSM filter computation exactly as before, whether in the far-field or near-field model. The goal of this design is to determine whether restricting attention to a perceptually relevant region of space enhances the performance differences between the two models.

V. PERFORMANCE MEASURES

In this section, the performance measures used to evaluate the accuracy of BSM reproduction are defined. The metrics include the normalized BSM error, ILD error, ITD error, and the effective spherical harmonics (SH) order. Each measure captures a different aspect of the reproduction quality.

A. Binaural Error: Near-Field BSM

As an initial study of the reproduction accuracy for near-field BSM, the normalized binaural signal matching MSE is defined in Eq. (9). Substituting the near-field model, the normalized MSE can be written as:

$$\epsilon_{\text{complex}}^{l/r}(k) = \frac{\sigma_s^2 \left\| \mathbf{V}_{\text{nf}}^\top \left(\mathbf{c}_{\text{BSM}}^{l/r} \right)^* - \mathbf{h}_{\text{nf}}^{l/r} \right\|_2^2 + \sigma_n^2 \left\| \left(\mathbf{c}_{\text{BSM}}^{l/r} \right)^* \right\|_2^2}{\sigma_s^2 \left\| \mathbf{h}_{\text{nf}}^{l/r} \right\|_2^2} \quad (17)$$

where $\mathbf{V}_{\text{nf}}$ is the near-field steering matrix, and $\mathbf{h}_{\text{nf}}^{l/r}$ are the near-field HRTF vectors.

Following the same rationale described earlier, the near-field mixed error is defined by substituting the far-field quantities $\mathbf{V}$ and $\mathbf{h}^{l/r}$ with their near-field counterparts $\mathbf{V}_{\text{nf}}$ and $\mathbf{h}_{\text{nf}}^{l/r}$ in the expressions for both $\epsilon_{\text{complex}}^{l/r}(k)$ and $\epsilon_{\text{MagLS}}^{l/r}(k)$ defined in Eq.(10).

B. Interaural Level Difference (ILD) Error

The ILD error quantifies the deviation in interaural level differences between the reproduced and reference binaural signals. The ILD is estimated by processing the binaural signals through Equivalent Rectangular Bandwidth (ERB) filter banks, as outlined in [19]. Specifically, the ILD in decibels for each filter band is calculated as:

$$ILD_f(f_c, \Omega) = 10 \log_{10} \frac{\sum_{f=0}^{f_c^{\max}} |G(f, f_c) p^l(f)|^2}{\sum_{f=0}^{f_c^{\max}} |G(f, f_c) p^r(f)|^2}, \quad (18)$$

where $G(f, f_c)$ represents the ERB filter with central frequency f_c at frequency f , and $f_c^{\max}$ denotes the maximum frequency of this ERB filter. An average over 32 filter bands spanning the range 1.5–20 kHz, generated using the Auditory Toolbox [20], is used to compute the ILD as:

$$ILD(\Omega) = \frac{1}{32} \sum_{f_c} ILD_f(f_c, \Omega) \quad (19)$$

The ILD error is then calculated as the absolute difference between the ILD of the reproduced and reference signals:

$$\epsilon_{\text{ILD}}(\Omega) = \frac{1}{32} \sum_{f_c} |ILD_{\text{rep}}(f_c, \Omega) - ILD_{\text{ref}}(f_c, \Omega)| \quad (20)$$

where $ILD_{\text{ref}}(f_c, \Omega)$ represents the ILD of the reference binaural signal, and $ILD_{\text{rep}}(f_c, \Omega)$ is the ILD estimated from the reproduced binaural signal.

C. Interaural Time Difference (ITD) Error

The ITD error measures deviations in the interaural time differences between the reproduced and reference binaural signals. ITD is estimated using the group delay of the binaural signals [21]. The group delay in seconds for the left and right ears is calculated as:

$$\tau_{l/r}^{(s)}(f) = \frac{1}{f_s} \operatorname{Re} \left[\frac{\operatorname{DFT}\{n p^{l/r}(n)\}}{\operatorname{DFT}\{p^{l/r}(n)\}} \right] \quad (21)$$

where $p^{l/r}(n)$ is the binaural signal for the left (l) or right (r) ear (either true or estimated), $\operatorname{DFT}\{\cdot\}$ denotes the discrete Fourier transform, and f_s is the sampling rate (Hz). The ITD is then obtained by subtracting the group delay of the right ear from that of the left ear:

$$\operatorname{ITD}(f) = \tau_l(f) - \tau_r(f). \quad (22)$$

To evaluate the ITD error, the ITD is averaged over frequencies up to 1.5 kHz, and the error is computed as the absolute difference between the reference and estimated ITD:

$$\epsilon_{\operatorname{ITD}}(r) = |\operatorname{ITD}_{\operatorname{ref}}(r) - \operatorname{ITD}_{\operatorname{rep}}(r)|. \quad (23)$$

D. Null Space Projection Measure

The null space projection measure, introduced in [22], quantifies the degree to which a vector resides within the null space of a matrix, or, equivalently, its deviation from the column space of the matrix. In the context of this work, this measure is applied to determine whether the HRTF vector, $\mathbf{h}_{\operatorname{nf}}^{l/r}$, can be spanned by the column space of the steering matrix $\mathbf{V}^\top$. The null space projection measure is formally defined as:

$$\xi_{\operatorname{null}} = 10 \log_{10} \left(\frac{\|\mathbf{V}_0^\top \mathbf{h}_{\operatorname{nf}}^{l/r}\|_2^2}{\|\mathbf{h}_{\operatorname{nf}}^{l/r}\|_2^2} \right), \quad (24)$$

where the null space of matrix $\mathbf{V}^\top$ denoted by $\mathbf{V}_0^\top$ is obtained from the singular value decomposition (SVD) of $\mathbf{V}$. Specifically, $\mathbf{V}_0^\top$ comprises the eigenvectors corresponding to singular values below a threshold of -20 dB relative to the largest singular vector, representing an approximation of the null space of $\mathbf{V}$.

With this definition, lower values of the measure $\xi_{\operatorname{null}}$ reflect a stronger alignment of $\mathbf{h}$ with the column space, thereby facilitating a more accurate reconstruction of the binaural signals.

VI. SIMULATION STUDY

This section presents a simulation study aimed at objectively quantifying the performance of the BSM algorithm for near-field sources. Unlike the analytical modeling approach used in our earlier work [15], the present study relies on numerically simulated near-field data generated with the HATS manikin model, for both HRTFs and array steering functions. Therefore, this evaluation is more realistic compared to the sphere-based models.

The structure of this chapter is as follows: Section VI-A presents the simulation setup, Section VI-B describes the methods of the simulation, Section VI-C reports the main findings, and Section VI-D discusses the results of the Field-of-View (FoV) analysis.

A. Setup

To generate realistic near-field steering matrices and HRTFs, finite-difference time-domain (FDTD) simulations were performed, as detailed in [23]. The validation results in the cited work indicate that the simulated HRTFs agree with measured data within 1–2 dB up to 17 kHz, making them a reliable source for the present analysis. Our numerical simulation results are expected to extrapolate to HRTFs measured on humans because the boundary conditions of human skin are very similar to the rigid boundary conditions applied to the HATS surface in our simulations [24]. An ideal point source was employed, providing high omnidirectionality and strong time coherence - features critical for accurate near-field HRTF characterization. Six grids were used in the convergence study and asymptotic prediction, with the grid spacing ΔX varying from 0.6 mm to 0.96 mm.

The simulations also included a wearable array configuration, in which five microphones were mounted on a glasses frame positioned on the HATS manikin head [25]. This allowed the extraction of both HRTFs and acoustic transfer functions (ATFs) of the array required for the BSM formulation. The geometry, source and receiver configurations, and solver parameters are outlined next.

A Brüel & Kjær Type 5128 Head-and-Torso Simulator (HATS) [25] was used as both the measurement and simulation geometry. The HATS mesh used in the simulations was an accurate (error ≤ 1 mm) 3D scan of the HATS. Figure 1 shows the glasses array mounted on the HATS head.

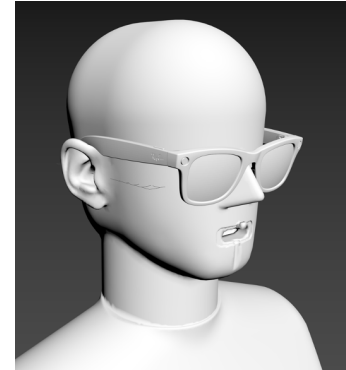


Fig. 1: Glasses microphone array mounted on the HATS manikin.

Five microphones were integrated into the glasses frame at predefined locations. Their Cartesian coordinates are listed in Table I. No special microphone porting was done: the pressure was acquired on the surface of the glasses.

TABLE I: Microphone coordinates on the glasses frame, rounded to the nearest mm, relative to the head center.

Label	x	y	z
Nose	101	-17	-5
Left mid-temple	31	77	21
Right mid-temple	31	-77	21
Left logo	86	73	29
Right logo	86	-73	29

Finally, the point sources were positioned on a spherical grid around the HATS head at 2702 directions on a lebedev grid (azimuth–elevation pairs) and at six radial distances

$$r_s = \{0.15, 0.20, 0.45, 0.70, 1.00, 1.50\} \text{ m}$$

from the head center. In contrast to the dataset used in our previous research [15], where the far-field distance was 3.2 m, the far-field HRTF here is defined at 1.5 m due to simulation domain constraints. The ear canals of the HATS mesh were manually blocked such that blocked-meatus simulated HRTFs were used in this work. The Signal-to-Noise Ratio (SNR) was set to 20 dB by appropriately choosing σ_s^2 and σ_n^2 .

B. Methods

In this study, performance was evaluated using two formulations of the BSM filter coefficients ($c_{\text{BSM}}^{l/r}$), as detailed in the previous section: one using far-field BSM filters and the other using near-field BSM filters.

To simulate head rotations of the listener during playback and analyze their influence on the BSM performance, we adopted a method similar to the one proposed in [10]. Specifically, the head rotation was modeled by modifying the azimuth angles of the HRTFs, assuming a rotation of the head in the horizontal plane. For a given rotation angle $\Delta\phi$, the rotated HRTF vector $\mathbf{h}_{\text{rot}}^{l/r}(k, r_s)$ is defined as:

$$\mathbf{h}_{\text{rot}}^{l/r}(k, r_s) = [h^{l/r}(k, \theta_1, \phi_1 + \Delta\phi), h^{l/r}(k, \theta_2, \phi_2 + \Delta\phi), \dots, h^{l/r}(k, \theta_Q, \phi_Q + \Delta\phi)]^T \quad (25)$$

Since only azimuthal rotations were considered, the elevation angles θ_q were kept constant, and only the azimuth components ϕ_q of the source directions were updated.

C. Results: Analysis of Binaural Measures for Far-Field and Near-Field BSM

Performance was evaluated based on the measures detailed in Sec. IV, and is presented in this section.

1) *Binaural Error*: Figure 2 presents the normalized Mixed MSE of the left ear, as defined in Eq. (11), for both far-field and near-field BSM filters, computed as detailed in sec. III. Since the results are identical for both ears in this symmetric scenario, only left ear is shown. The near-field results are indicated by dashed lines, while the far-field results are shown using solid lines. The analysis covers frequencies from 75 Hz to 10 kHz for several source distances. The distance $r_s = 1.5$ m corresponds to the far-field distance in this simulation.

Across most of the frequency range and source distances, the far-field BSM—which assumes a fixed source distance of 1.5 m—yields slightly higher normalized MSE values than the near-field BSM. This suggests that under ideal listening conditions (i.e., without head rotation), incorporating distance-dependent information via the near-field model provides a marginal improvement in reproduction accuracy. For very close sources ($r_s = 0.15$ m and $r_s = 0.2$ m), approximately 5–10 cm from the surface of the head, the error increases

significantly for both filters, highlighting the limitations of BSM in extreme near-field conditions.

Figure 3 shows the normalized Mixed MSE under azimuthal head rotation of 40° . The asymmetry introduced by rotation is evident in the differences between the left and right ears. Despite this, the performance gap between the near-field and far-field BSM remains relatively small under head rotation, suggesting that the added value of incorporating source-distance information is limited in dynamic scenarios as well.

Overall, the results indicate that while the near-field BSM consistently outperforms its far-field counterpart, the improvement is modest under most conditions. Both filters exhibit degraded performance at very short distances, underscoring the challenge of achieving accurate binaural reproduction for extremely close sources. This aspect will be investigated in more detail in Section VI-C4.

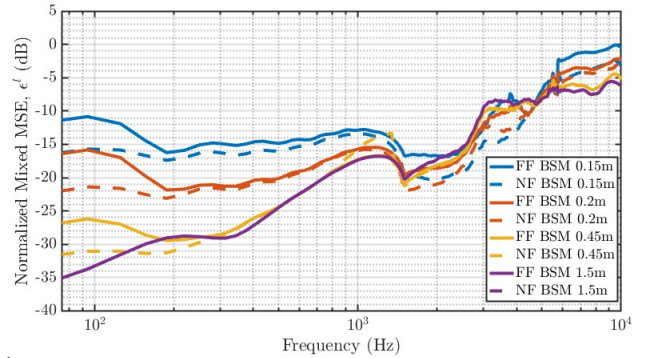


Fig. 2: Normalized Mixed MSE of the left ear for far-field (solid lines) and -field (dashed lines) BSM filters as a function of frequency for source distances ranging from 0.15 to 1.5 meters, without head rotations.

2) *ILD error*: The ILD values and ILD errors were calculated according to Eqs. (18)–(20). The results for both far-field and near-field BSM are shown in Fig. 4 for source distances of 0.15 m and 0.45 m. Small differences are observed between the two methods, suggesting that both far-field and near-field BSM achieve comparable performance in preserving ILD. This consistency indicates that incorporating either far-field or near-field BSM leads to similar accuracy in ILD. Note that both errors are significantly above the just noticeable difference (JND) threshold of 1 dB [26, 27], suggesting that the errors are clearly audible.

3) *ITD error*: The ITD was calculated following the method described in Eqs. (21) and (22). Fig. 5 shows the frequency-averaged ITD of the reference and BSM signals. The corresponding frequency-averaged ITD error, defined in Eq. (23), is also displayed as a function of azimuth and computed using both far-field and near-field BSM filters for the two tested source distances (0.15 m and 0.45 m).

The ITD errors are relatively small for both far-field and near-field BSM, remaining below the JND threshold of $20 \mu\text{s}$ for frontal directions and $100 \mu\text{s}$ for lateral directions [28, 29].

4) *Null Space projection*: As observed in the previous section, source distances of only a few centimeters from the

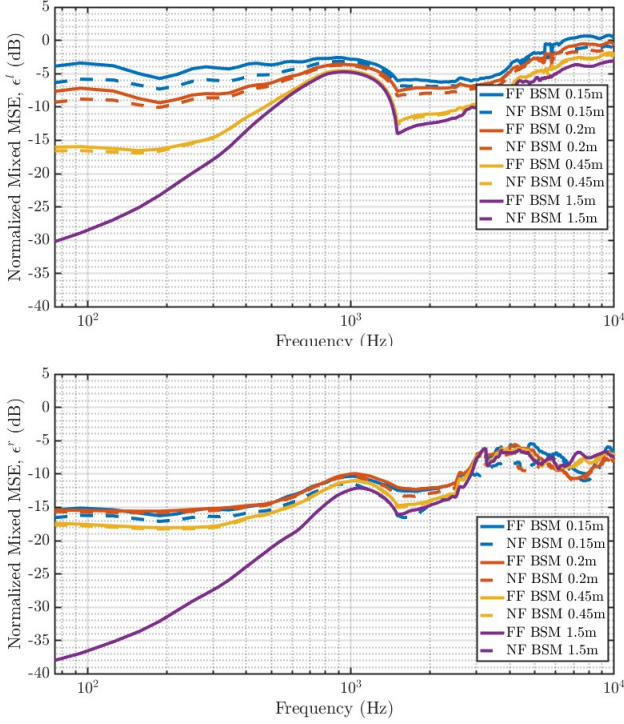


Fig. 3: Same as figure 2 but under head rotation of 40° for the left ear (top) and right ear (bottom).

microphone array result in significant binaural errors for both far-field and near-field BSM. Even when the source distance is known and near-field BSM is applied, substantial performance degradation persists. This indicates that incorporating source-distance information alone is insufficient to fully resolve the inaccuracies in extreme near-field conditions.

As described in Eq. (7), BSM aims to find a combination of microphone steering vectors that reconstructs the HRTF. When the HRTF lies within the steering matrix range, the BSM filters can achieve an accurate reconstruction, and a small error.

This condition can be quantified using the *Null Space Projection Measure* defined in Eq. (24). This measure evaluates how much of the HRTF vector lies within the *null space* of the steering matrix—essentially, the portion of the HRTF that cannot be represented using the given microphone array.

Fig. 6 presents the null space projection values of the near-field HRTF onto the near-field steering matrix across different source distances, along with the projection of the far-field HRTF onto the far-field steering matrix. The results reveal that for sources located at very short distances, the null space projection is significantly larger, indicating that a considerable portion of the HRTF is not representable by the steering matrix. The underlying reason for this dependency on source distance is not yet fully understood. Further investigation into this phenomenon is left for future research.

D. Results: BSM with limited FoV

In Chapter IV, we introduced the FoV design approach, which incorporates a directional prior based on the assumption that sources of interest are located within a frontal region.

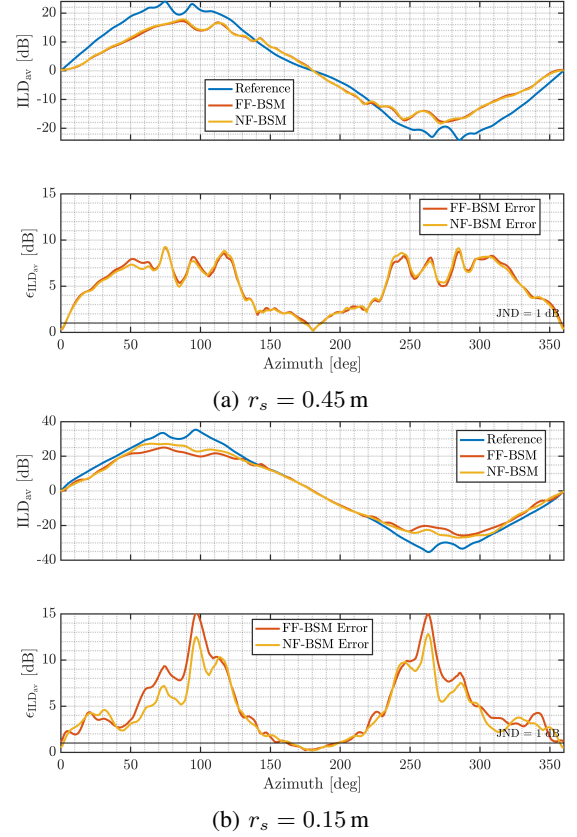


Fig. 4: ILD values and ILD errors for far-field and near-field BSM filters at source distances of 0.15 m and 0.45 m.

This design introduces a tunable parameter β that controls the trade-off between accurate reproduction inside and outside the FoV. In the following results, we present the performance for $\beta = 0.2$, which was found to offer the best trade-off between reducing errors within the FoV while avoiding excessive degradation outside it. A FoV of $\pm 45^\circ$ was selected in both azimuth and elevation around the front looking direction.

1) *Binaural Error with FoV weighting*: Fig. 7 presents the normalized Mixed MSE for the left ear, comparing the performance of far-field and near-field BSM filters under FoV weighting. The results are shown separately for DOAs inside the FoV and outside the FoV, in two subfigures.

In the FoV region (top figure), the differences between the near-field and far-field BSM are significantly more pronounced compared to the unweighted (spatially uniform) case shown in Fig. 2. The near-field BSM yields notably lower binaural errors, especially below approximately 2 kHz, while the far-field BSM performance degrades sharply in this frequency range. This result highlights the importance of incorporating source-distance information when source directions are focused within a limited region.

In contrast, the results for sources outside the FoV (bottom figure) remain largely similar to those in the unweighted case. The performance gap between the far-field and near-field BSM is relatively small, and both filters show comparable trends to those in Fig. 2. This consistency suggests that the directional prior primarily enhances performance within the attended region but does not significantly distort the reproduction quality

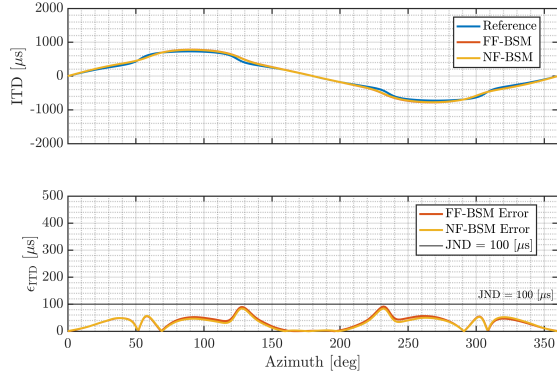
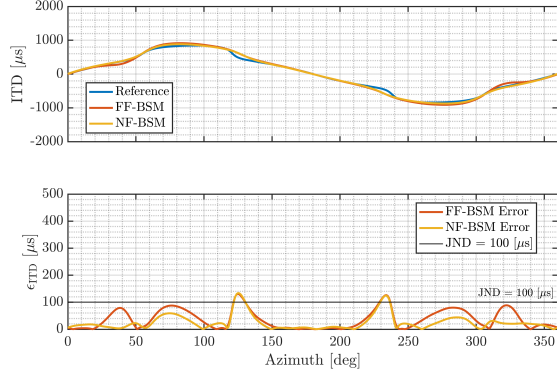
(a) $r_s = 0.45$ m(b) $r_s = 0.15$ m

Fig. 5: ITD values and ITD errors for far-field and near-field BSM filters at source distances of 0.15 m and 0.45 m.

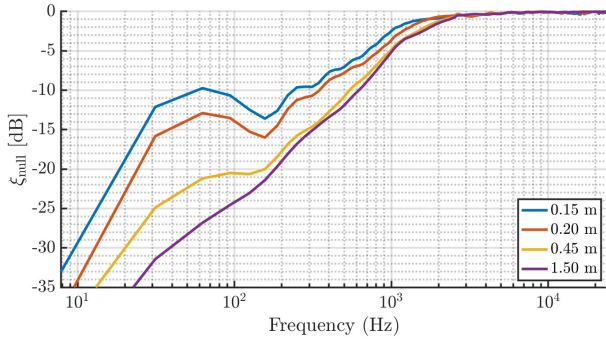
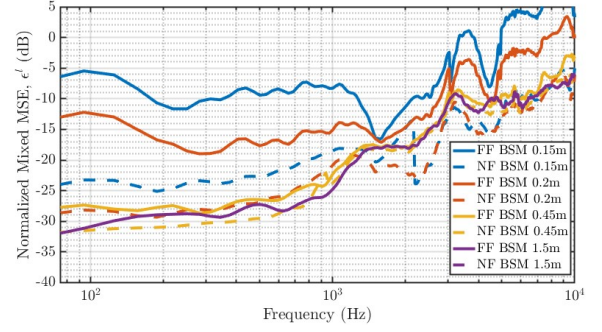
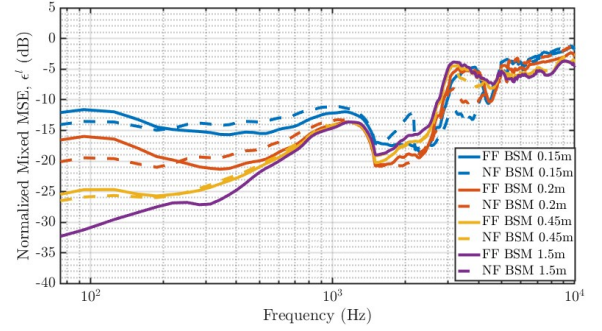


Fig. 6: Null space projection values, as in Eq. (24), of the near-field HRTF onto the near-field steering matrix, evaluated across different source distances, where the distance of 1.5 m corresponds to the far-field case.

outside it.

Fig. 8 shows the Mixed error results under a head rotation of 40° . Since the head is no longer front-facing, the left-right symmetry is broken, and results are presented for both ears. In the FoV region (top figures), the difference between the far-field and near-field BSM filters becomes more pronounced when head rotation is introduced, particularly for the ear oriented away from the microphone array. The near-field BSM continues to yield superior performance, especially in the low-frequency range, while the far-field filter shows notable

(a) Directions within FoV ($\pm 45^\circ$ azimuth and elevation)

(b) Directions outside FoV

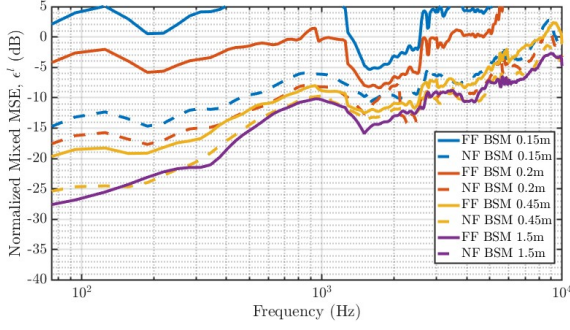
Fig. 7: Normalized Mixed MSE of the left ear for far-field and near-field BSM filters under Field of View weighting, (a) inside the FoV, and (b) outside the FoV.

degradation.

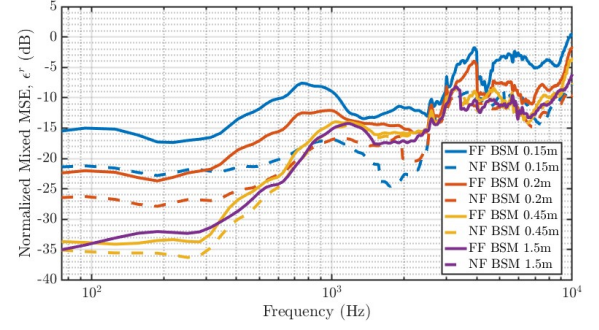
Outside the FoV (bottom figures), the results remain largely consistent with those observed in the spatially uniform case under 40° head rotation (see Fig. 3), suggesting that the directional weighting has minimal impact on performance for off-FoV sources.

2) *ILD Error with FoV weighting*: Fig. 9 shows the ILD values and ILD errors for near-field and far-field BSM filters at source distances of 0.15 m and 0.45 m without head rotation with FoV weighting. The dashed vertical lines in the figure indicate the boundaries of the FoV region. The near-field BSM consistently achieves better results, with lower ILD errors overall. Inside the FoV, the error is significantly smaller, and similarly, reduced errors are observed at the back of the head, possibly due to the symmetry between the front and back regions in this scenario. In contrast, outside the FoV—particularly at the side directions—the ILD errors are higher with both filters. The differences between near-field and far-field BSM are more pronounced compared to the spatially uniform case, shown in Fig. 4 especially for close source distances.

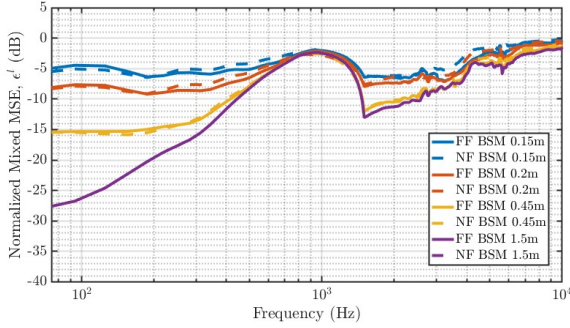
Fig. 10 presents the ILD and ILD error results under a 40° head rotation. As seen, there is a clear distinction between the inside-FoV and outside-FoV regions. Within the FoV, the differences between near-field and far-field BSM become even larger, particularly for the very close source



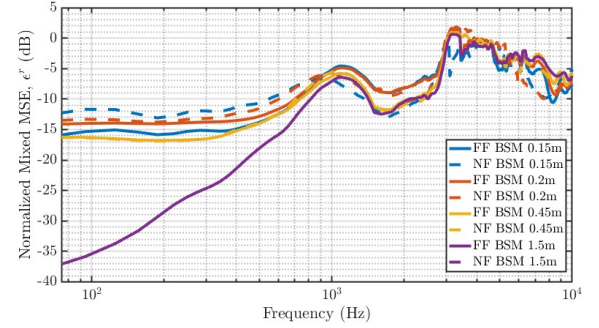
(a) Directions within FoV – Left Ear



(b) Directions within FoV – Right Ear



(c) Directions outside FoV – Left Ear



(d) Directions outside FoV - Right Ear

Fig. 8: Normalized Mixed MSE for far-field and near-field BSM filters under Field of View weighting and a 40° head rotation, (a) inside the FoV – left ear, (b) inside the FoV – right ear, (c) outside the FoV – left ear, and (d) outside the FoV – right ear.

distance. Additionally, the errors are generally larger compared to the case without head rotation, highlighting the increased challenges introduced by head movement.

3) *ITD Error with FoV weighting*: Fig. 11 shows the ITD values and ITD errors for near-field and far-field BSM filters under FoV weighting without head rotation. The dashed vertical lines in the figure indicate the boundaries of the FoV region. The errors achieved with the near-field BSM are generally quite small—usually below the JND line of $100 \mu\text{s}$ inside the FoV, and slightly above it outside the FoV. The differences between near-field and far-field BSM are typically not very dominant in this configuration.

However, Fig. 12 shows the corresponding results under a head rotation of 40° . In this case, the ITD errors are noticeably larger than in the static head condition across all scenarios. In this case, the near-field BSM yields noticeably smaller errors, particularly in the directions inside the FoV, emphasizing its robustness and accuracy even under head movement.

VII. LISTENING EXPERIMENT

This section presents the listening experiment conducted to evaluate the perceptual performance of the proposed BSM methods. The chapter is structured as follows: first, the experimental setup is described, followed by the listening test methodology, and finally, the results are reported.

A. Listening Test Setup

The listening test setup¹ was largely based on the acoustic simulation framework described in Section VI-A, using the same source modeling and simulated HRTFs. In this subsection, we highlight only the elements specific to the listening experiment, including the source-positioning, head orientations, and the signal rendering methods used for the test.

The simulated acoustic environment was modeled as a rectangular room with dimensions $[10 \times 6 \times 3]$ m, using the image source method [30], implemented in MATLAB 2024b. The room was configured with a reverberation coefficient of $R = 0.7$ leading to a reverberation time of ~ 0.3 s.

For the listening test, a single omnidirectional point source was placed at coordinates $(x, y, z) = (5, 3.7, 1.7)$ m, playing a speech signal. The microphone array signals $\mathbf{x}(k)$ were computed using the ATFs described in Section VI-A.

Two source-listener configurations were evaluated, for which the center of the microphone array was positioned at source distances of $r_s = 0.15$ m and $r_s = 0.45$ m, located at an azimuth of 30° to the left of the source and elevation 0° . Two listener head orientations were tested: one with no head rotation, such that the listener head is aligned with the head of the recording scene, and one with a 40° rotation of the listener head to the right. For the rotated case, the source azimuth was adjusted to -10° to maintain the same relative source direction of 30° to the right after rotation. The HRTFs

¹The listening study was conducted at Ben-Gurion University of the Negev.

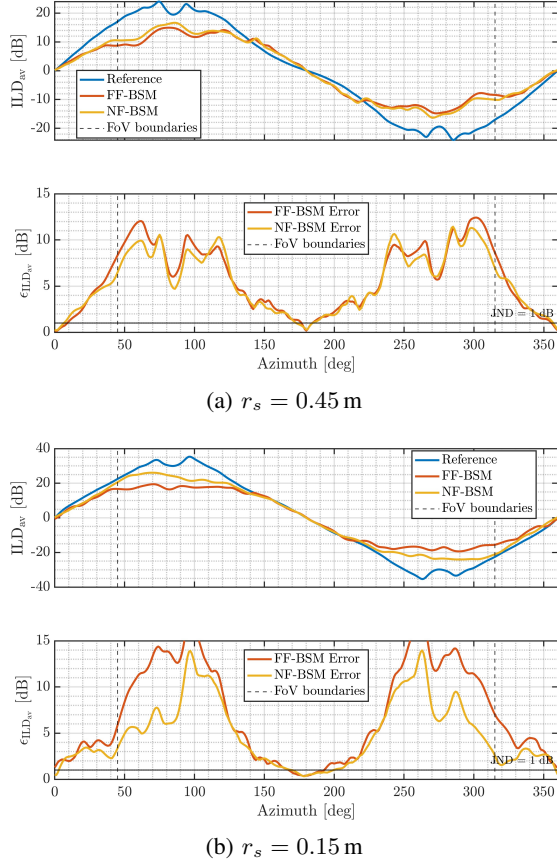


Fig. 9: Average ILD values and ILD errors for far-field and near-field BSM filters under Field of View weighting, shown for source distances of 0.15 m and 0.45 m. The dashed vertical lines in the figure indicate the boundaries of the FoV region ($\pm 45^\circ$ from the center, i.e., at 45° and 315°).

used for binaural rendering were also obtained from the same simulation framework.

BSM filter coefficients were computed using four methods detailed in Secs. III and IV : (1) Far-Field BSM (FF-BSM), (2) Near-Field BSM (NF-BSM), (3) FF-BSM with FoV weighting (FF-FoV-BSM), and (4) NF-BSM with FoV weighting (NF-FoV-BSM). All filters used the MagLS criterion for frequencies above 1.5 kHz, assuming a high SNR of +20 dB as detailed in sec. VI-A.

As a reference, an additional binaural signal was rendered using high-order Ambisonics (order 14) with simulated HRTFs under the same room conditions. Participants listened to all stimuli over AKG K702 open-back reference headphones, with headphone equalization applied.

B. Methodology

The listening tests followed the Multiple Stimuli with Hidden Reference and Anchor (MUSHRA) protocol [31]. Participants were asked to rate the similarity of each stimulus to the reference, with respect to *Overall Quality*. The experiment consisted of four MUSHRA screens, corresponding to two source distances ($r_s = 0.15$ m and $r_s = 0.45$ m) and two head orientations (with and without 40° head rotation). Each screen

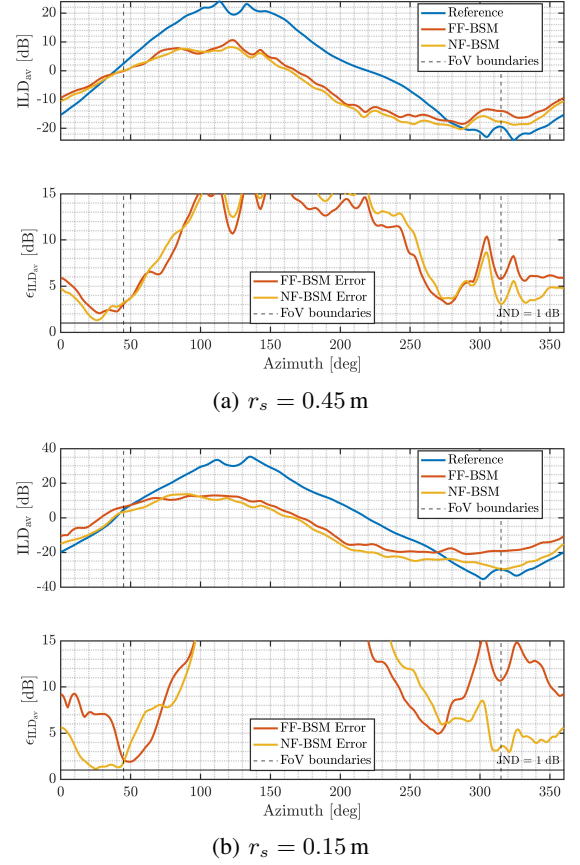


Fig. 10: Same as figure 9 but under head rotation of 40° .

contained five signals: a hidden reference, and four BSM-based binaural renderings (FF-BSM, NF-BSM, FF-FoV-BSM, and NF-FoV-BSM). While there was not a formal anchor signal in this experiment, the FF-BSM method represents a low quality baseline, as it is not specifically designed for near field sources and does not apply any directional weighting.

14 participants (12 male and 2 female), aged 24 to 37 years, took part in the study. All participants reported normal hearing and were familiar with spatial audio or binaural listening tests.

The experiment was conducted in a quiet environment and consisted of two phases prior to the main evaluation: a training phase and a familiarization phase. In the training phase, participants were introduced to the test interface, headphones, and rating scale. In the familiarization phase, they were allowed to freely explore and listen to all stimuli before starting the formal test. Participants were encouraged to use the full rating scale and to rely on their perceptual impression of spatial realism, coloration, and overall fidelity. No time limit was imposed, and participants could replay signals as needed before submitting their scores.

C. Results

The participants' MUSHRA scores were analyzed using a repeated-measures Analysis of Variance (RM-ANOVA) [32] with three within-subject factors: *Method* (Reference, NF-FOV-BSM, NF-BSM, FF-FOV-BSM, FF-BSM), *Source Distance* (0.15 m, 0.45 m), and *Listener Head Orientation* (0° , 40°). The distributions of ratings for each distance/orientation

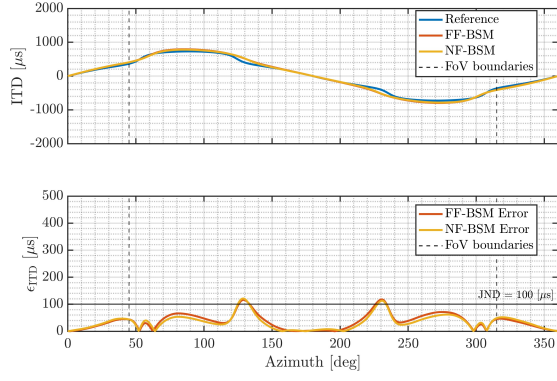
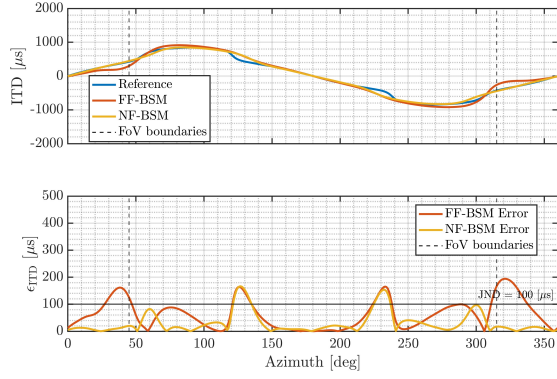
(a) $r_s = 0.45$ m(b) $r_s = 0.15$ m

Fig. 11: Average ITD values and ITD errors for far-field and near-field BSM filters under Field of View weighting without head rotation, shown for source distances of 0.15 m and 0.45 m. The dashed vertical lines in the figure indicate the boundaries of the FoV region ($\pm 45^\circ$ from the center, i.e., at 45° and 315°).

combination are illustrated in Fig. 13, which presents four boxplots representing the four combinations of source distance and head orientation.

The RM-ANOVA revealed statistically significant main effects of all three factors:

- **Method:** $F(4, 52) = 72.23$, $p < .001$, $\eta^2 = .967$;
- **Source Distance:** $F(1, 13) = 30.09$, $p < .001$, $\eta^2 = .698$;
- **Head Orientation:** $F(1, 13) = 29.41$, $p < .001$, $\eta^2 = .693$.

These main effects indicate that listener ratings varied significantly across methods, differed between source distances, and decreased when the head was rotated by 40° .

Significant interaction effects were also found:

- **Method $\times$ Distance:** $F(4, 52) = 89.45$, $p < .001$, $\eta^2 = .973$;
- **Method $\times$ Orientation:** $F(4, 52) = 21.16$, $p < .001$, $\eta^2 = .894$;
- **Distance $\times$ Orientation $\times$ Method:** $F(4, 52) = 5.28$, $p = .015$, $\eta^2 = .679$.

These interactions demonstrate that the relative ranking of methods depended on both source distance and head orientation. Pairwise comparisons with Bonferroni correction

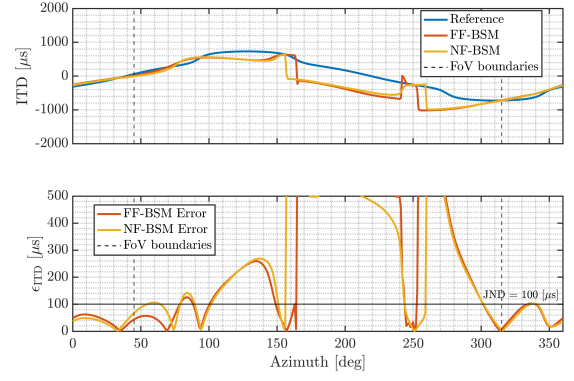
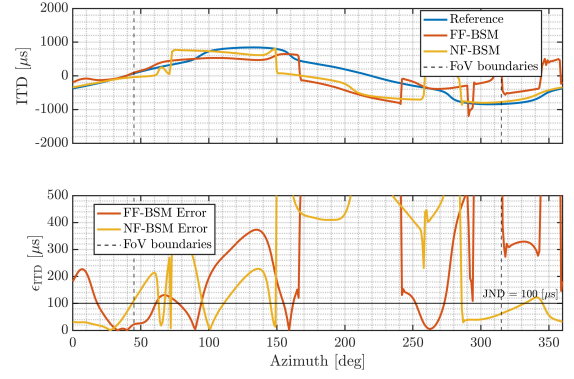
(a) $r_s = 0.45$ m(b) $r_s = 0.15$ m

Fig. 12: Same as figure 11 but under head rotation of 40° for the left ear (top) and right ear (bottom).

revealed clear differences between methods. Reference scored significantly higher than all BSM methods, including NF-FOV-BSM ($p < .001$, mean difference = 17.4 points), NF-BSM ($p < .001$, mean difference = 55.5 points), FF-FOV-BSM ($p < .001$, mean difference = 52.3 points), and FF-BSM ($p < .001$, mean difference = 67.8 points). Among the BSM methods, NF-FOV-BSM achieved the highest ratings. It scored significantly higher than FF-BSM ($p < .001$, mean difference = 50.4 points), FF-FOV-BSM ($p < .001$, mean difference = 34.9 points), and NF-BSM ($p < .001$, mean difference = 38.1 points). FF-BSM (the traditional BSM method) consistently received the lowest ratings. The differences between FF-BSM and NF-FOV-BSM ($p < .001$) and Reference ($p < .001$) were very large, while the differences with NF-BSM ($p = .024$) and FF-FOV-BSM ($p = .010$) were smaller but still statistically significant. These results align with the objective findings presented in the simulation study in Section VI-C, which demonstrated improvements of NF-BSM over FF-BSM, particularly for near-field sources at close distances. Notably, for the closest source distance (0.15 m), the perceptual gap between NF-BSM and FF-BSM was more pronounced than anticipated from the simulations in Section VI-C, suggesting that listener perception accentuated the benefits of near-field modeling at very close ranges. In particular, the inclusion of the field-of-view weighting (NF-FOV-BSM) further enhanced performance, as demonstrated in Sec. VI-D, resulting in ratings that were closest to the Reference signal. Consistent with the simulations, NF-FOV-BSM was shown to be the most robust

method for very close source distances (0.15 m) and across head orientations.

VIII. CONCLUSIONS

This work investigated the performance of BSM methods for near-field sources using realistic simulated data and perceptual evaluation. Building upon our previous study on near-field BSM, this paper extended the analysis to include listener head rotations, evaluation of binaural cues (ILD and ITD), and a newly introduced FoV weighting strategy.

Results from both the simulation study and the listening experiment consistently showed that NF-BSM outperforms traditional FF-BSM, particularly for sources close to the array, where distance-dependent effects are significant. While the inclusion of source-distance information in NF-BSM reduced binaural reproduction errors, performance degradation remains evident for very close sources, especially under head rotation, highlighting the inherent limitations of wearable arrays in such extreme conditions.

A key contribution of this work is the introduction of FoV weighting for near-field scenarios. This strategy improved robustness and perceptual quality by prioritizing directions of interest. When combined with NF-BSM, the resulting NF-FoV-BSM method substantially outperformed other BSM variants and provided reproduction quality closest to the reference among the evaluated methods.

These findings underscore the limitations of BSM methods that assume far-field plane waves when applied to near-field sources. At the same time, they demonstrate that by incorporating source-distance information and directional weighting, these limitations can largely be overcome, enabling accurate and robust binaural reproduction even for challenging near-field conditions.

Future work may involve incorporating deep neural networks to enhance binaural reproduction in near-field scenarios, potentially without explicit estimation of source distance or direction. Extending the analysis beyond point sources to more realistic models, such as vibrating pistons that better approximate loudspeakers or human talkers, could provide deeper insight into the physical limitations of current methods. Further investigation into the underlying causes of degradation in near-field reproduction—particularly under extreme proximity or head rotation—may also guide the development of improved algorithms and array designs.

REFERENCES

- [1] L. Madmoni and B. Rafaely, "Direction of arrival estimation for reverberant speech based on enhanced decomposition of the direct sound," *IEEE Journal of Selected Topics in Signal Processing*, vol. 13, no. 1, pp. 131–142, 2018.
- [2] 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," *Acta Acustica*, vol. 6, p. 47, 2022.
- [3] G. Richard, P. Smaragdis, S. Gannot, P. A. Naylor, S. Makino, W. Kellermann, and A. Sugiyama, "Audio signal processing in the 21st century: The important outcomes of the past 25 years," *IEEE Signal Processing Magazine*, vol. 40, no. 5, pp. 12–26, 2023.
- [4] Z. Ben-Hur, F. Brinkmann, J. Sheaffer, S. Weinzierl, and B. Rafaely, "Spectral equalization in binaural signals represented by order-truncated spherical harmonics," *The Journal of the Acoustical Society of America*, vol. 141, no. 6, pp. 4087–4096, 2017.
- [5] L. Madmoni, Z. Ben-Hur, J. Donley, V. Tourbabin, and B. Rafaely, "Design and analysis of binaural signal matching with arbitrary microphone arrays and listener head rotations," *EURASIP Journal on Audio, Speech, and Music Processing*, vol. 2025, no. 1, p. 11, 2025.
- [6] B. Rafaely, *Fundamentals of spherical array processing*. Springer, 2015, vol. 8.
- [7] M. A. Gerzon, "Ambisonics in multichannel broadcasting and video," *Journal of the Audio Engineering Society*, vol. 33, no. 11, pp. 859–871, 1985.
- [8] W. Song, W. Ellermeier, and J. Hald, "Using beamforming and binaural synthesis for the psychoacoustical evaluation of target sources in noise," *The Journal of the Acoustical Society of America*, vol. 123, no. 2, pp. 910–924, 2008.
- [9] I. Ifergan and B. Rafaely, "On the selection of the number of beamformers in beamforming-based binaural reproduction," *EURASIP Journal on Audio, Speech, and Music Processing*, vol. 2022, no. 1, p. 6, 2022.
- [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] T. Deppisch, H. Helmholtz, and J. Ahrens, "End-to-end magnitude least squares binaural rendering of spherical microphone array signals," in *2021 Immersive and 3D Audio: from Architecture to Automotive (I3DA)*. IEEE, 2021, pp. 1–7.
- [12] J. Ahrens, H. Helmholtz, D. L. Alon, and S. V. Amengual Garí, "Spherical harmonic decomposition of a sound field based on observations along the equator of a rigid spherical scatterer," *The Journal of the Acoustical Society of America*, vol. 150, no. 2, pp. 805–815, 2021.
- [13] L. McCormack, A. Politis, R. Gonzalez, T. Lokki, and V. Pulkki, "Parametric ambisonic encoding of arbitrary microphone arrays," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 30, pp. 2062–2075, 2022.
- [14] E. Fisher and B. Rafaely, "Near-field spherical microphone array processing with radial filtering," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, no. 2, pp. 256–265, 2010.
- [15] S. Goldring, Z. B. Hur, D. L. Alon, and B. Rafaely, "Binaural signal matching with wearable arrays for near-field sources," 2025. [Online]. Available: <https://arxiv.org/abs/2507.15517>
- [16] A. Kan, C. Jin, and A. van Schaik, "A psychophysical evaluation of near-field head-related transfer functions synthesized using a distance variation function," *The Journal of the Acoustical Society of America*, vol. 125, no. 4, pp. 2233–2242, 2009.
- [17] D. S. Brungart and W. M. Rabinowitz, "Auditory localization of nearby sources. head-related transfer functions," *The Journal of the Acoustical Society of America*, vol. 106, no. 3, pp. 1465–1479, 1999.
- [18] D. S. Brungart, "Near-field virtual audio displays," *Presence*, vol. 11, no. 1, pp. 93–106, 2002.
- [19] B. Xie, *Head-related transfer function and virtual auditory display*. J. Ross Publishing, 2013.
- [20] M. Slaney, "Auditory toolbox," *Interval Research Corporation, Tech. Rep*, vol. 10, no. 1998, p. 1194, 1998.
- [21] B. F. Katz and M. Noisternig, "A comparative study of interaural time delay estimation methods," *The Journal of the Acoustical Society of America*, vol. 135, no. 6, pp. 3530–3540, 2014.
- [22] Y. Gayer, V. Tourbabin, Z. Ben-Hur, J. Donley, and B. Rafaely, "Ambisonics encoding for arbitrary microphone arrays incorporating residual channels for binaural reproduction," *arXiv preprint arXiv:2402.17362*, 2024.
- [23] J. Gómez Bolaños, M. Geronazzo, R. Mehra, L. Savioja *et al.*, "Pinna-related transfer functions and lossless wave equation using finite-difference methods: Validation with measurements," *The Journal of the Acoustical Society of America*, vol. 147, no. 5, pp. 3631–3645, 2020.
- [24] S. Hajarolasvadi, M. Khaleghimeybodi, P. Razavi, M. Smirnov, and S. T. Prepeljčič, "Effect of sound-induced vibrations of the pinna on head-related transfer functions: Experimental and numerical investigations," *The Journal of the Acoustical Society of America*, vol. 155, no. 4, pp. 2875–2890, 2024.
- [25] C. Oreinos and J. M. Buchholz, "Measurement of a full 3d set of hrtfs for in-ear and hearing aid microphones on a head and torso simulator (hats)," *Acta Acustica united with Acustica*, vol. 99, no. 5, pp. 836–844, 2013.
- [26] J. Blauert, *Spatial hearing: the psychophysics of human sound localization*. MIT press, 1997.
- [27] W. A. Yost and R. H. Dye Jr, "Discrimination of interaural differences of level as a function of frequency," *The Journal of the Acoustical Society of America*, vol. 83, no. 5, pp. 1846–1851, 1988.

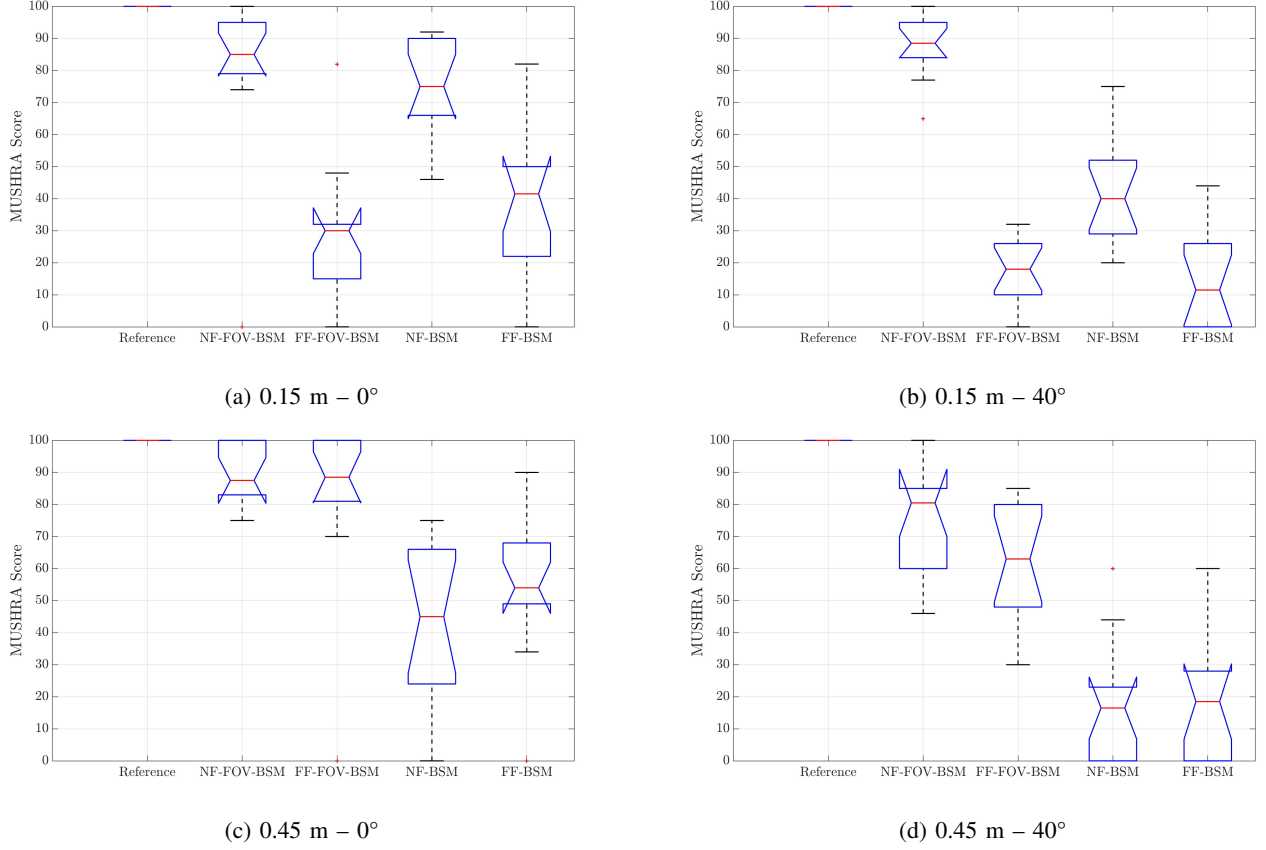


Fig. 13: Boxplots of MUSHRA scores for each distance and head orientation combination. Each subplot shows ratings for Reference (High Order Ambisonics), NF-FOV-BSM, NF-BSM, FF-FOV-BSM, and FF-BSM. The boxes represent the interquartile range (IQR) with a line indicating the median. Tukey-style whiskers extend up to $1.5 \times \text{IQR}$ beyond the box, and notches indicate the 95% confidence interval of the median.

- [28] J. E. Mossop and J. F. Culling, “Lateralization of large interaural delays,” *The Journal of the Acoustical Society of America*, vol. 104, no. 3, pp. 1574–1579, 1998.
- [29] A. Andreopoulou and B. F. Katz, “Identification of perceptually relevant methods of inter-aural time difference estimation,” *The Journal of the Acoustical Society of America*, vol. 142, no. 2, pp. 588–598, 2017.
- [30] J. B. Allen and D. A. Berkley, “Image method for efficiently simulating small-room acoustics,” *The Journal of the Acoustical Society of America*, vol. 65, no. 4, pp. 943–950, 1979.
- [31] B. Series, “Method for the subjective assessment of intermediate quality level of audio systems,” *International Telecommunication Union Radiocommunication Assembly*, vol. 2, 2014.
- [32] H. Keselman, J. Algina, and R. K. Kowalchuk, “The analysis of repeated measures designs: a review,” *British Journal of Mathematical and Statistical Psychology*, vol. 54, no. 1, pp. 1–20, 2001.