Analysis and Design of Head-Tracked Compensation for Bilateral Ambisonics

Or Berebi, Zamir Ben-Hur, David Lou Alon, and Boaz Rafaely, *Senior Member, IEEE*

***Abstract*—Virtual and augmented reality technologies demand high-quality spatial sound recording and playback through headphones. However, achieving high-quality binaural reproduction requires a complex recording system and a large number of microphones. To address this issue, a recent study proposed Bilateral Ambisonics, which involves capturing the sound-field using two low-order microphone arrays located ear-distance apart. We present an analytical analysis of the limitation of a previously suggested head-tracking compensation solution to Bilateral Ambisonics. An alternative approach is proposed to overcome these limits in which the translation operation is band-limited. A subjective evaluation and a listening test are provided and complement the findings of the analytical analysis. Results indicate that in a static scenario, compensating for small lateral head-rotations up to $\pm 30^\circ$ with good accuracy is possible for microphone arrays of spherical harmonics (SH) order of 1 and for medium rotations of up to $\pm 60^\circ$ with SH order of 2. When a dynamic scenario is considered, Bilateral Ambisonics of order 2 were comparable to High-order Ambisonics, and Bilateral Ambisonics of order 1 provided performance comparable to third order Ambisonics with MagLS.**

***Index Terms*—Spatial sound, head-tracking.**

I. Introduction

BINAURAL reproduction, which involves the computation of headphone signals, is employed in spatial audio processing applications, including virtual and augmented reality [1], music playback, room auralization, and more [2], [3]. A common format for presenting the spatial audio signals in these applications is Ambisonics [4], which is typically derived from recordings with spherical microphone arrays [5]. In the process of binaural reproduction, Ambisonics signals are combined with a Head-Related Transfer Function (HRTF) [6] to compute the sound pressure at the listener's ears.

Manuscript received 17 October 2023; revised 3 December 2023; accepted 11 December 2023. Date of publication 22 December 2023; date of current version 4 January 2024. This work was supported by Reality Labs Research @ Meta. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Daniele Salvati. *(Corresponding author: Or Berebi.)*

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Human Subjects Research Committee of Ben-Gurion University under Application No. 1648-1.

Or Berebi and Boaz Rafaely are with the School of Electrical, Computer Engineering, Ben-Gurion University of the Negev, Beer-Sheva 84105, Israel (e-mail: berebio@post.bgu.ac.il; br@bgu.ac.il).

Zamir Ben-Hur and David Lou Alon are with Reality Labs Research, Meta, Menlo Park, CA 94025 USA (e-mail: zamirbh@meta.com; davidalon@meta.com).

Digital Object Identifier 10.1109/TASLP.2023.3345140

Although widely used, a significant limitation of the Ambisonics format arises from the fact that it may not lead to high-quality binaural signals when the number of microphones in the array is small, which is typically the case with commercially available spherical microphone arrays [7]. The limited quality, in the case of spherical harmonics (SH) encoded HRTFs, is mainly the result of HRTF SH order truncation errors [8], [9]. The latter may not apply to virtual loudspeaker decoding, but even with this approach low-order Ambisonics may still lead to limited spatial resolution in the reproduction [10]. One recent approach to overcome the limitations imposed by low-order Ambisonics binaural reproduction for headphones is by capturing the sound-field with two low-order spherical microphone arrays in a framework referred to as Bilateral Ambisonics [11], [12], [13]. This approach is based on the Binaural B-Format [14]. While the Binaural B-Format approach uses 1st order Ambisonics with a minimum-phase approximation of the HRTF, Bilateral Ambisonics is formulated to an arbitrary SH order and utilizes the ear-aligned HRTF representation, which has been demonstrated to reduce the effective SH order of the HRTF, preserving the HRTF phase information and the ITD [15]. In Bilateral Ambisonics reproduction, the two microphone arrays are positioned at the assumed locations of the listener's ears as, opposed to Ambisonics, where a single microphone array is positioned at the assumed listener's head center-position.

One drawback of Bilateral Ambisonics compared to Ambisonics is that compensating for the listener's head movements is not as straightforward as in the Ambisonics case [13], [16]. This is vital for compensating for the motion of the listener's head, and also in order to exploit dynamic cues, which are widely known to improve localization quality [17], [18]. Head-tracked rendering is inherently complicated for Bilateral Ambisonics, as the sound-field in Bilateral Ambisonics is defined around the ears, and rotation of the listener's head leads to a discrepancy between the microphone-array positions (which remain unchanged) and the listener's ear positions. Therefore, re-alignment at the locations of the ears and microphone-array is required to generate the binaural signals.

This paper presents a detailed analysis and design of a adaptive head-tracked compensation method for binaural reproduction utilizing Bilateral Ambisonics. The initial design was previously proposed by the same authors of this work in a conference paper [19], which included minimal evaluation and suggested that the proposed method significantly impairs the quality of binaural signals when low-order Bilateral Ambisonics signals are used, even with compensation for small head-rotation angles.

In this work, we enhance the original design by introducing a band-limited translation operation combined with a broadband rotation operation and show how this approach greatly improves the performance of the previously suggested method. Additionally, we provide a comprehensive analytical analysis that sheds light on the limitations of the method. Numerical and subjective listening tests support the findings of the research. The key contributions of this work include:

- Theoretical and numerical evaluation of performance for the approach presented in [19], leading to insight and formulations for the limits of performance, in particular the upper frequency for which this presented approach is expected to work well.
- The proposal and evaluation of an improved method for head-tracking compensation, that extends the limits of performance in terms of bandwidth under measures that are related to perception, i.e. magnitude error.
- Objective evaluation and a listening test evaluating the quality of the proposed method compared to the previous method, clearly demonstrating its perceptual benefits.

These contributions address the initial design's shortcomings and significantly improve the proposed head-tracking compensation method for binaural reproduction using Bilateral Ambisonics.

The paper is organized as follows. Section II reviews the mathematical formulation for binaural reproduction in the SH domain with standard Ambisonics and Bilateral Ambisonics, and includes the mathematical fundamentals of head-tracking compensation in the context of standard Ambisonics. Section III mathematically formulates the recently proposed method for head-tracking compensation with Bilateral Ambisonics. The limitations of this method are formulated in Section IV. The modified method is presented in Section V. Sections VI and VII present simulation studies evaluating the accuracy of the proposed method compared both to its original form, and to high-order Ambisonics reproduction and low-order MagLS Ambisonics reproduction [20]. A listening experiment complementing the findings of this research is presented in Section VIII. Section IX provides concluding remarks for the research.

II. Mathematical Background

This section presents the mathematical framework for binaural reproduction using standard Ambisonics and Bilateral Ambisonics, along with a discussion on head-tracking compensation with Ambisonics. It aims to establish the notation and provide the background equations for non-expert readers.

A. Ambisonics

Binaural signals, which represent the sound pressure observed at both ears, can be rendered by combining an Ambisonics signal [10] with a left and right SH domain representation of the HRTF [21]. This computation can be described as [22]

$$p^{L/R}(k) = \sum_{n=0}^{N} \sum_{m=-n}^{n} [\tilde{a}_{nm}(k)]^* h_{nm}^{L/R}(k) \tag{1}$$

where $p^{L/R}(k)$ are the left/right binaural signals with wavenumber $k = \frac{2\pi f}{c}$, f is the frequency, and c is the speed of sound. The Ambisonics signal, $a_{nm}(k)$ modified to $\tilde{a}_{nm}(k) = (-1)^m [a_{n(-m)}(k)]^*$, is obtained from sound pressure measurements using a spherical microphone array or simulations [10]. The left/right HRTF SH coefficients [23], $h_{nm}^{L/R}(k)$, can be measured or simulated independently. In practice, both $a_{nm}(k)$ and $h_{nm}(k)$ are available only up to finite orders, N_a and N_h, respectively. The maximum order N for the summation in (1) is determined as:

$$N = \min(N_a, N_h) \tag{2}$$

When $a_{nm}(k)$ is obtained from a spherical microphone array, N_a is typically smaller than N_h, leading to HRTF order truncation at N_a. This truncation affects spatial resolution and frequency content, impacting spatial sound perception [24].

B. Bilateral Ambisonics

To mitigate the effects of HRTF order truncation mentioned earlier, the concept of ear-aligned HRTF is introduced [15]. This approach reduces the HRTF's effective order and better preserves phase and magnitude information. Bilateral Ambisonics combines ear-aligned HRTF with two Ambisonics signals, one for each ear, resulting in the following binaural signal equation [11], [12]:

$$p^{L/R}(k) = \sum_{n=0}^{N} \sum_{m=-n}^{n} \left[\tilde{a}_{nm}^{L/R}(k)\right]^* h_{nm}^{aL/R}(k) \tag{3}$$

Here, $\tilde{a}_{nm}^{L/R}(k)$ represents the left/right ear Bilateral Ambisonics signals, and $h_{nm}^{aL/R}(k)$ are the left/right ear-aligned HRTF SH coefficients. Note that $h_{nm}^{aL/R}(k)$ can be directly computed from $h_{nm}^{L/R}(k)$ [15], however, $\tilde{a}_{nm}^{L/R}(k)$ are considered as known (measured or simulated) and not as estimated from the head-centered Ambisonics signal $a_{nm}(k)$.

C. Head-Tracking Compensation

With Ambisonics Head movement plays a crucial role in localizing stationary sound sources and creating a convincing binaural experience [17], [18]. Three-degrees-of-freedom (3DoF) head-tracking compensation involves adjusting the sound-field orientation to account for the listener's head movements [16], typically using sensors in VR headsets. This compensation corrects the difference between head-centered and world-centered coordinate systems. Ambisonics signals are initially based on the head's orientation, and as the listener's head rotates, sound sources rotate with it. This is corrected by counter-rotating the Ambisonics signal to maintain sound source orientation unaffected by head rotation.

The orientation change due to the head-rotation can be characterized using Euler angles, denoted as $((\alpha, \beta, \gamma))$ [25]. Here, the zyz sequence of rotation axes will be used. Ambisonics signals can be accurately rotated in the SH domain using a weighted sum of the Ambisonics signal and the Wigner-D function, denoted

as $D^n_{mm'}(\alpha, \beta, \gamma)$: [16], [26]

$$a^r_{nm}(k) = \sum_{m'=-n}^{n} a_{nm'}(k) D^n_{mm'}(\alpha, \beta, \gamma) \tag{4}$$

This compensation technique ensures that the sound sources remains spatially stationary with head movements.

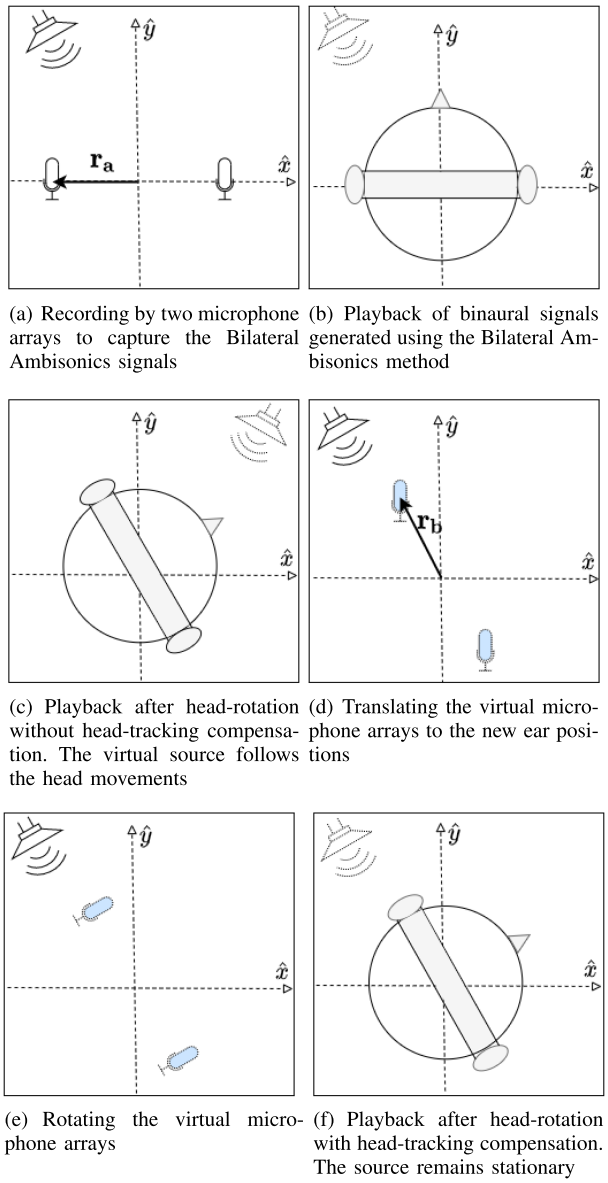
III. Head-Tracking Compensation

for bilateral Ambisonics Head-tracking compensation in the context of binaural reproduction with Bilateral Ambisonics is fundamentally different from head-tracking compensation with Ambisonics. The difference lies in the fact that the Bilateral Ambisonics signals are captured at the ear positions, and therefore head-rotation changes these ear positions. This section presents a method for compensating for this change.

A. An Overview of the Head-Tracking Compensation

Process Fig. 1 shows the entire process, from the original recording to compensation for head-rotation during playback. In the first stage of this process, as illustrated in Fig. 1(a), the sound-field is composed of a single source in free field, captured by two microphone arrays (represented by the microphone symbols). Then, Fig. 1(b) shows the listener wearing headphones, with his/her ears located at the position that represents the microphone array positions in the virtual environment. The headphones play the binaural signals generated using the Bilateral Ambisonics process described in Section II-B. The listener now perceives a virtual source from the direction of the real source (about 30° to the left), outlined in the figure with a dotted line.

Fig. 1(c) illustrates what happens next - the listener rotates his/her head, but the binaural signal remains the same, i.e. there is no compensation for head-rotation. The virtual source therefore changes its position instead of remaining stationary. However, the captured Bilateral Ambisonics signals of Fig. 1(a) can be manipulated to preserve the virtual source spatial stationarity. Fig. 1(d) and (e) illustrate this process in two stages. First, as illustrated in Fig. 1(d), the microphone arrays are translated to the new ear positions, denoted by the blue microphone symbols. In the second stage, both virtual microphone arrays are rotated to align with the new head orientation, as illustrated in Fig. 1(e). The translated and rotated virtual microphone arrays can now be used as Bilateral Ambisonics signals to generate new binaural signals to be played back via headphones, as illustrated in Fig. 1(f). With this head-tracking compensation applied, the virtual sound source is perceived to be stationary in space, despite head-rotation.



(a) Recording by two microphone arrays to capture the Bilateral Ambisonics signals

(b) Playback of binaural signals generated using the Bilateral Ambisonics method

(c) Playback after head-rotation without head-tracking compensation. The virtual source follows the head movements

(d) Translating the virtual microphone arrays to the new ear positions

(e) Rotating the virtual microphone arrays

(f) Playback after head-rotation with head-tracking compensation. The source remains stationary

Fig. 1. Head-rotation compensation for Bilateral Ambisonics recording played-back via headphones.

B. Translation of a Sound-Field

Having outlined the concept of head-tracking compensation for Bilateral Ambisonics, this process is now formulated mathematically. Note that the mathematical derivation is written only for the left ear Bilateral Ambisonics signal for notation simplicity. Sound-field representation as a infinite continuum of plane-waves is assumed here. The plane-wave amplitude density (PWAD) function $a(k, \Omega)$ represents the amplitude at the center of the coordinate system.

The listener head-center is defined at the coordinate system origin and the left ear position **before** head-rotation is denoted by the vector $\mathbf{r_a}$ (see Fig. 1(a)). The sound pressure at $\mathbf{r_a}$ can be computed by integrating over all plane-wave contributions:

$$p(k, \mathbf{r_a}) = \int_{\Omega \in \mathcal{S}^2} a(k, \Omega) e^{i\tilde{\mathbf{k}}(\Omega)\cdot \mathbf{r_a}} d\Omega \tag{5}$$

where $\tilde{\mathbf{k}}(\Omega)$ refers to the wave vector at wave-number k, and direction of arrival Ω [27]. $\Omega \equiv (\theta, \phi) \in \mathcal{S}^2$ denotes the spatial angle, composed of $\theta \in [0, \pi]$ (the elevation angle), which is defined as the angle measured downwards from the Cartesian z-axis, and $\phi \in [(0, 2\pi]$ (the azimuth angle), which is defined as

the angle measured counterclockwise from the x-axis in the xy plane. Additionally, $\int_{\Omega\in\mathcal{S}^2}(.)d\Omega \equiv \int_0^{2\pi}\int_0^{\pi}(.)sin(\theta)d\theta d\phi$.

In a similar manner, the pressure at the ear position **after** the head-rotation $\mathbf{r_b}$, as depicted in Fig. 1(d), can be written as

$$\begin{aligned} p(k,\mathbf{r_b}) &= \int_{\Omega\in\mathcal{S}^2} a(k,\Omega)e^{i\tilde{\mathbf{k}}(\Omega)\cdot\mathbf{r_b}}d\Omega \\ &= \int_{\Omega\in\mathcal{S}^2} a(k,\Omega)e^{i(\tilde{\mathbf{k}}(\Omega)\cdot\mathbf{r_b}-\tilde{\mathbf{k}}(\Omega)\cdot\mathbf{r_a}+\tilde{\mathbf{k}}(\Omega)\cdot\mathbf{r_a})}d\Omega \\ &= \int_{\Omega\in\mathcal{S}^2} [a(k,\Omega)e^{i\tilde{\mathbf{k}}(\Omega)\cdot\mathbf{r_a}}]e^{i\tilde{\mathbf{k}}(\Omega)\cdot(\mathbf{r_b}-\mathbf{r_a})}d\Omega \end{aligned} \tag{6}$$

From (5) and (6) we can formulate the translation operation as

$$a^{L\tau}(k,\Omega) = a^{L}(k,\Omega)e^{i\tilde{\mathbf{k}}(\Omega)\cdot(\mathbf{r_b}-\mathbf{r_a})} \tag{7}$$

and denote the translation function as

$$E^{L}(k,\Omega) = e^{i\tilde{\mathbf{k}}(\Omega)\cdot(\mathbf{r_b}-\mathbf{r_a})} \tag{8}$$

where $a^{L}(k,\Omega) = a(k,\Omega)e^{i\tilde{\mathbf{k}}(\Omega)\cdot\mathbf{r_a}}$. Note that $\mathbf{r_b}$ can be calculated from the initial ear position $\mathbf{r_a}$ and the rotation angles (α,β,γ) using Euler rotation matrices:

$$\mathbf{r_b} = \mathbf{R}_z(\alpha)\mathbf{R}_y(\beta)\mathbf{R}_z(\gamma)\mathbf{r_a} \tag{9}$$

where $\mathbf{R}_z$ and $\mathbf{R}_y$ are the 3×3 Euler rotation matrices [25]. The head-rotation angles (α,β,γ), are assumed to be available from a head-tracking device.

C. Rotation of the Translated Sound-Field

Similar to what has been discussed in Section II-C and illustrated in Fig. 1(e), the translated Bilateral Ambisonics signal also experiences a change in orientation. Changing the orientation of $a^{L\tau}(k,\Omega)$ can be performed in the SH domain, and applied to $a^{L\tau}_{nm}(k)$, the SH coefficients of $a^{L\tau}(k,\Omega)$:

$$\hat{a}_{nm}(k) = \sum_{m'=-n}^{n} a^{L\tau}_{nm'}(k)D^{n}_{mm'}(\alpha,\beta,\gamma) \tag{10}$$

Equation (10) results in $\hat{a}_{nm}(k)$, which represents the head-rotation-compensated Bilateral Ambisonics signals.

IV. Limits of Performance

This section presents the inherent limitations of the head-tracking compensation method derived in Section III. To compensate for head movements, $a(k,\Omega)$ undergoes translation with (7) and rotation with (10). The rotation operation on the Bilateral Ambisonics signals is accurate and does not introduce any error [26]. However, the translation operation does introduce error, which limits the accuracy of the proposed method. This error is formulated and analyzed next. Note that in the derivation that follows, the superscript $^{L/R}$ denoting the left/right ear is removed for brevity.

Let $a_\infty(k,\Omega)$ denote the full-order plane-wave amplitude density function. $a_\infty(k,\Omega)$ can be written as a superposition of an order N plane-wave amplitude density function, $a(k,\Omega)$, and a residual, $a_\epsilon(k,\Omega)$, representing orders $N+1$ and above:

$$\begin{aligned} a_\infty(k,\Omega) &= a(k,\Omega) + a_\epsilon(k,\Omega) \\ &= \sum_{n=0}^{N}\sum_{m=-n}^{n} a_{nm}(k)Y^m_n(\Omega) + \sum_{n=N+1}^{\infty}\sum_{m=-n}^{n} a^{\epsilon}_{nm}(k)Y^m_n(\Omega) \end{aligned} \tag{11}$$

with $Y^m_n(\Omega)$ denoting the SH functions [28].

Translating the full-order plane-wave amplitude density function $a_\infty(k,\Omega)$ has been shown to be exact [27], [29]. Therefore, translating $a_\infty(k,\Omega)$ with (7) should lead to an accurate representation of the sound-field at the new position. Substituting (11) into (7) and (8) leads to

$$\begin{aligned} a^{\tau}_\infty(k,\Omega) &= a(k,\Omega)E(k,\Omega) + a_\epsilon(k,\Omega)E(k,\Omega) \\ &= a^{\tau}(k,\Omega) + a^{\tau}_\epsilon(k,\Omega) \end{aligned} \tag{12}$$

with $a^{\tau}(k,\Omega)$ and $a^{\tau}_\epsilon(k,\Omega)$ denoting the translated plane-wave amplitude density functions. In the practical case, only $a^{\tau}(k,\Omega)$ is available, and so $a^{\tau}_\epsilon(k,\Omega)$ can be considered to be the translation error of $a^{\tau}(k,\Omega)$ with respect to $a^{\tau}_\infty(k,\Omega)$.

An analysis of the highest and lowest SH orders of $a^{\tau}_\epsilon(k,\Omega)$ determines the range of orders which contain translation error. These extreme orders can be computed from the highest and lowest SH orders of $a_\epsilon(k,\Omega)$ and $E(k,\Omega)$, as $a^{\tau}_\epsilon(k,\Omega)$ is the product of these two functions [30]. Following (11),

$$\begin{aligned} \underline{N}_{a_\epsilon} &= N+1 \\ \overline{N}_{a_\epsilon} &= \infty \end{aligned} \tag{13}$$

are the low and high SH orders of $a_\epsilon(k,\Omega)$, respectively, and following (7) and (8),

$$\begin{aligned} \underline{N}_{E} &= 0 \\ \overline{N}_{E}(k) &\approx k|\mathbf{r_b}-\mathbf{r_a}| \end{aligned} \tag{14}$$

are expressions for the low SH order and high effective SH order of $E(k,\Omega)$. $\overline{N}_E(k)$ is referred to as an effective SH order due to the fact that $E(k,\Omega)$ of (8) can be written as a summation of SH and spherical Bessel functions $j_n(\cdot)$ [31]:

$$\begin{aligned} E(k,\Omega) &= e^{i\tilde{\mathbf{k}}\cdot(\mathbf{r_b}-\mathbf{r_a})} \\ &= \sum_{n=0}^{\infty}\sum_{m=-n}^{n} 4\pi i^n j_n(k|\mathbf{r_b}-\mathbf{r_a}|)[Y^m_n(\Omega_0)]^* Y^m_n(\Omega) \\ &\approx \sum_{n=0}^{\overline{N}_E(k)}\sum_{m=-n}^{n} 4\pi i^n j_n(k|\mathbf{r_b}-\mathbf{r_a}|)[Y^m_n(\Omega_0)]^* Y^m_n(\Omega) \end{aligned} \tag{15}$$

with Ω_0 denoting the direction of arrival of the plane-wave vector $\tilde{\mathbf{k}}$. The approximation in (15) can be relatively accurate if $\overline{N}_E(k)$ satisfies (14), because $j_n(k|\mathbf{r_b}-\mathbf{r_a}|)$ decays as a function of n for $n > k|\mathbf{r_b}-\mathbf{r_a}|$ [27].

Now, the highest effective SH order of $a^{\tau}_\epsilon(k,\Omega)$ can be calculated as the addition of the maximum orders of its two factors [30]:

$$\overline{N}_{a^{\tau}_\epsilon} = \overline{N}_{a_\epsilon} + \overline{N}_E = \infty \tag{16}$$

and the lowest order can be computed as the smallest difference in the orders of the two functions [30]. When $\overline{N}_E(k) > N$,

this difference becomes zero because $\underline{N}_{a_\epsilon} = N + 1$. However, when $\overline{N}_E(k) \leq N$, the difference is higher, leading to

$$\underline{N}_{a^\tau_\epsilon}(k) = \begin{cases} N + 1 - k|\mathbf{r_b} - \mathbf{r_a}| & , \text{ if } \overline{N}_E(k) \leq N \\ 0 & , \text{ otherwise} \end{cases} \tag{17}$$

Therefore, orders lower then $\underline{N}_{a^\tau_\epsilon}(k)$ can be considered accurate, since for these orders $a^\tau_\epsilon(k, \Omega)$ has negligible magnitude, and so

$$a^\tau_\infty(k, \Omega) \approx a^\tau(k, \Omega) \tag{18}$$

The analysis presented above leads to several important observations.

- The accuracy of $a^\tau(k, \Omega)$ decreases as k increases. This is because $\underline{N}_{a^\tau_\epsilon}(k)$ decreases as a function of k, leading to a higher translation error at the high frequencies.
- The accuracy of $a^\tau(k, \Omega)$ decreases as translation distance increases, due to the increase in $|\mathbf{r_b} - \mathbf{r_a}|$. This means that compensation for larger head-rotation is prone to higher error.
- $a^\tau(k, \Omega)$ can be considered as exact, i.e. (18) is only satisfied when $\underline{N}_{a^\tau_\epsilon}(k) > N$. Denote by f_1 the frequency at which this condition is satisfied. f_1 can be calculated by substituting (13) and (14) into (17) using the relation $f = \frac{kc}{2\pi}$ for $\overline{N}_E(k) = 0.5$:

$$f_1 = \frac{c}{4\pi|\mathbf{r_b} - \mathbf{r_a}|} \tag{19}$$

  with c denoting the speed of sound.
- Another useful parameter that can be deduced from (17) is the frequency f_2 above which **all** orders can be considered to be as inaccurate. f_2 can be calculated by substituting (13), (14) into (17) for the equality $\overline{N}_E(k) = N + 1$:

$$f_2 = \frac{c(N + 1)}{2\pi|\mathbf{r_b} - \mathbf{r_a}|} \tag{20}$$

V. Head-Tracking Compensation

With band-limited translation The inherent limitations of the proposed head-tracking compensation method have been presented in Section IV. According to (20) and (19), the translation can be considered exact for $f < f_1$, partially accurate for $f_1 \leq f \leq f_2$, and inaccurate for $f > f_2$. Inaccurate translation may impair the externalization and localization of the perceived binaural sound [32].

In this section a perceptually motivated approach is proposed to mitigate this problem. Previous research has provided evidence indicating that phase information may play a relatively less significant role in terms of spatial sound perception at high frequencies [20]. Since the translation of the Ambisonics signals involves phase manipulation, it may be perceptually beneficial to refrain altogether from altering the phase of the Ambisonics signals at frequencies for which translation is inaccurate and phase is less important for perception. This means compensating only via the rotation operation of (10). This approach is formulated next.

In Section III, the translation in (7) was applied for all values of f. It is now suggested to limit the bandwidth of (7) to frequencies lower then f_c. The bandwidth-limited translation is denoted as

$$a^\tau_{\text{BW}}(k, \Omega, f_c) = a(k, \Omega) E_{\text{BW}}(k, \Omega, f_c) \tag{21}$$

where

$$E_{\text{BW}}(k, \Omega, f_c) = \begin{cases} E(k, \Omega), & k \leq \frac{2\pi f_c}{c}\frac{1}{\sqrt{2}} \\ E(k, \Omega)(1 - \alpha) + \alpha, & \frac{2\pi f_c}{c}\frac{1}{\sqrt{2}} \leq k \leq \frac{2\pi f_c}{c}\sqrt{2} \\ 1, & k > \frac{2\pi f_c}{c}\sqrt{2} \end{cases} \tag{22}$$

and $E(k, \Omega)$ is taken from (7). To avoid audible artifacts due to a sharp transition at f_c, (22) includes a smooth transition interval of 1 octave at f_c, where α is the smoothing wight defined as

$$\alpha = 0.5 + \frac{\log f/f_c}{\log 2} \tag{23}$$

Note that (21) can directly replace (7), while the rest of the method remains the same. Additionally, f_c is considered to be a parameter and its value can be chosen depending on N and the head-rotation angle. For instance, f_c could be set to f_2, since for $f > f_2$ the translation is known to be inaccurate.

VI. Measures for Objective Evaluation

This section presents objective performance measures for the evaluation of the suggested head-tracking compensation method, i.e. Normalized Mean Square Error (NMSE), Normalized Magnitude Error, ITD, ILD, and SH Power Spectrum Error. Note that the measurements in this section are presented as a function of f instead of k. This change is made since the results presented in the following figures are plotted as a function of frequency.

A. NMSE

The NMSE between a reference binaural signal and an evaluated binaural signal is defined as

$$\epsilon_{\text{NMSE}}(f) = 10 \log_{10} \frac{|p_{\text{ref}}(f) - p(f)|^2}{|p_{\text{ref}}(f)|^2} \tag{24}$$

The NMSE is sensitive to both magnitude and phase errors; thus, it is considered to be a good measure of the physical accuracy of the binaural signals.

B. Normalized Magnitude Error

The Normalized Magnitude Error is defined as

$$\epsilon_{\text{mag}}(f) = 10 \log_{10} \frac{||p_{\text{ref}}(f)| - |p(f)||^2}{|p_{\text{ref}}(f)|^2} \tag{25}$$

where $p_{\text{ref}}(f)$ and $p(f)$ are the same as in (24). At high frequencies, accuracy in magnitude may be more important perceptually than accuracy in phase [33], which motivates the definition of this measure.

C. ITD

The ITD is an important spatial cue for sound localization [34]. In this work, the ITD as a function of direction has been computed by cross-correlating a 1.5 kHz low-pass-filtered version of the left and right binaural signals, and then computing the delay that corresponds to the strongest correlation [35]. The ITD is computed for the same signals as in (24). The ITD error averaged over all incident angles is defined as

$$\epsilon_{\text{ITD}}(\gamma) = \frac{1}{q}\sum_{i=1}^{q} |\text{ITD}_{\text{ref}}(\Omega_i) - \text{ITD}(\Omega_i, \gamma)| \tag{26}$$

with $\text{ITD}_{\text{ref}}(\Omega_i)$ denoting the ITD of the reference binaural signals, $p_{\text{ref}}^{L/R}(f)$, at incident angle Ω_i, the ITD of the evaluated signals after compensation with angle γ is denoted by $\text{ITD}(\Omega_i, \gamma)$.

D. ILD

Another important spatial cue is the ILD [34]. In this case, the ILD is calculated across 23 auditory filter bands as [21]

$$\text{ILD}(f_0, \Omega) = 10\log_{10}\frac{\int C(f, f_0)|p^L(f)|^2 df}{\int C(f, f_0)|p^R(f)|^2 df} \tag{27}$$

where $p^L(f), p^R(f)$ refers to the left and right ear binaural signal and $C(f, f_0)$ is a Gammatone filter with center frequency f_0, as implemented in the Auditory Toolbox. The integrals are evaluated from 1.5 kHz to 20 kHz and f_0 values are restricted accordingly. $\text{ILD}(f_0, \Omega)$ are averaged across the auditory filter bands to compute the ILD measure:

$$\text{ILD}(\Omega) = \frac{1}{23}\sum_{\text{f}_0} \text{ILD}(f_0, \Omega) \tag{28}$$

The ILD error averaged over all incident angles can then be defined by

$$\epsilon_{\text{ILD}}(\gamma) = \frac{1}{q}\sum_{i=1}^{q} |\text{ILD}_{\text{ref}}(\Omega_i) - \text{ILD}(\Omega_i, \gamma)| \tag{29}$$

with $\text{ILD}_{\text{ref}}(\Omega_i)$ denoting the ILD of the reference binaural signals, $p_{\text{ref}}^{L/R}(f)$, at incident angle Ω_i, the ILD of the evaluated signals after compensation with angle γ is denoted by $\text{ILD}(\Omega_i, \gamma)$. It is important to note that discrepancies or inaccuracies in the ILD and ITD of binaural signals have the potential to cause localization shifts, as the human auditory system relies on these cues, in combination with other factors, to determine the position of sound sources. The presence of conflicting cues may lead to ambiguity and confusion in the localization process, ultimately resulting in localization errors [36].

E. SH Power Spectrum Error

The SH Power Spectrum Error can be used to evaluate the accuracy of the Ambisonics signals for each order individually, enabling evaluation of the claims of Section IV, i.e. that the compensated Bilateral Ambisonics signals are expected to be accurate up to $\underline{N}_{a_\epsilon^t}$ of (17). The SH Power Spectrum Error is defined as

$$\epsilon_S(f, n) = \frac{\sum_{m=-n}^{n} |a_{nm}^{\text{ref}}(f) - a_{nm}(f)|^2}{\sum_{m=-n}^{n} |a_{nm}^{\text{ref}}(f)|^2} \tag{30}$$

where $a_{nm}^{\text{ref}}(f)$ is the reference Ambisonics signal and $a_{nm}(f)$ is the evaluated signal.

The SH Power Spectrum Error, $\epsilon_S(f, n)$, is used to numerically evaluate up to which order $a_{nm}(f)$ can be considered accurate. Here, 99% similarity is used. This means that the accumulated error over the orders up to the evaluated upper order should be lower than 1%. The parameter $N_b(f)$ describes the highest accurate order and is defined as

$$N_b(f) = \arg\min_N : \sum_{n=0}^{N} \epsilon_S(f, n) \leq 0.01 \tag{31}$$

VII. Simulation Study

In this section, the performance of the proposed head-tracking compensation approach for Bilateral Ambisonics is compared to head-tracked compensated Ambisonics, using the measures presented in Section VI.

A. Setup

The signals in this section are generated using a computer simulation. In the simulation, the Ambisonics and Bilateral Ambisonics signals describe sound-fields composed of a single plane-wave in free field with unit amplitude. All of the binaural signals are rendered using the measured HRTF from the Cologne HRTF database for the Neumann KU100 dummy head [37].

B. Methodology

The following section provides the reader with a detailed description of the various binaural signals generated for the simulation study.

1) $p_{ref}(f)$ - High-Order Ambisonics: Consider a single plane-wave of unit amplitude in free-field arriving from direction Ω_0. The Ambisonics signal depicting this sound-field is [27]

$$a_{nm} = [Y_n^m(\Omega_0)]^* \tag{32}$$

Denote a_{nm} as the reference Ambisonics signal without head-rotation, calculated up to $\tilde{N} = 41$. Next, head-rotation described by Euler angles as in (4) is assumed. The head-tracked compensated reference Ambisonics signal is then calculated using (4). Finally, the left ear reference binaural signal, denoted by $p_{\text{ref}}(f)$, is rendered using (1) and the left HRTF SH coefficients $h_{nm}^L(f)$. This binaural signal will serve as a reference for comparison with other binaural signals.

2) $p_{loa}(f)$ - Low-Order Ambisonics: This signal is generated identically to the reference, except for a lower reproduction order, denoted by $N = 1 - 4$, replacing the high-order $\tilde{N}$. The low-order Ambisonics binaural signal is denoted as $p_{\text{loa}}(f)$. This signal will be used as a baseline for comparison of performance, representing an unprocessed low-order binaural signal.

3) $p_{mls}(f)$ - Low-Order Ambisonics With MagLS HRTF Preprocessing: The popular implementation of the Magnitude

Least Squares (MagLS) HRTF preprocessing suggested in [20] is used with an added smooth transition interval, as suggested in [38]. Additionally, as suggested in [10], a diffused-field covariance constraint equalization is applied to the HRTF. The MagLS HRTF SH coefficients are denoted as $h_{nm}^{\text{mls}L}(f)$. The transition interval used for all the MagLS signals is of 1 octave centered at $f_c = 2\,\text{kHz}$. Since the MagLS solution uses the Ambisonics format, its binaural signals are computed similarly to the binaural signals presented above. The MagLS binaural signal is denoted by $p_{\text{mls}}(f)$. The MagLS signal will be used as a benchmark for comparison to the head-tracked Bilateral signals.

4) $p_{blt}(f)$ - Bilateral Ambisonics: The left ear Bilateral Ambisonics representation of the single plane-wave sound-field can be written analytically (see [27]) using (32):

$$a_{nm}^{L} = e^{i\tilde{\mathbf{k}}(\Omega_0)\cdot\mathbf{r_a}} a_{nm} \tag{33}$$

where $\mathbf{r_a}$ refers to the listener's left ear position vector before the head-rotation, as seen in Fig. 1(a). In this simulation $|\mathbf{r_a}| = 8.75\,\text{cm}$ represents the head radius. a_{nm} is the Ambisonics signal from (32), and $\tilde{\mathbf{k}}(\Omega_0)$ is the wave vector at wave-number k and direction of arrival Ω_0.

a_{nm}^{L} is calculated up to order $N = 1, 3, 4, 15$, depending on the experiment. The head-tracked compensated Bilateral Ambisonics signals are calculated according to (3), (7), (10) from Sections II-B and II-C. The resulting binaural signal is denoted as $p_{\text{blt}}(f)$. This compensated binaural signal is the suggested method in Section II-C.

5) $p_{bl\text{-}Blt}(f)$ - Band-Limited Bilateral Ambisonics: Section V presents an alternative approach to head-tracking compensation of the Bilateral Ambisonics signals. Here, this signal is referred to as band-limited Bilateral Ambisonics. This signal is generated identically to the Bilateral Ambisonics signal, except for the translation from (7). Instead, the Bilateral Ambisonics signal is translated using (21), leading to the band-limited Bilateral binaural signal $p_{\text{bl-blt}}(f)$.

6) $p_{ideal}(f)$ - Ideal Bilateral Ambisonics: Section IV showed that translating finite-order bilateral signals leads to inaccuracies. To generate a reference signal that avoids the inaccuracies introduced by the translation operation, the Bilateral Ambisonics signals are re-calculated directly at the ear position after the head-rotation.

$$a_{nm}^{L_b} = e^{i\tilde{\mathbf{k}}(\Omega_0)\cdot\mathbf{r_b}} a_{nm} \tag{34}$$

where $\mathbf{r_b}$ of (9) is the left-ear vector **after** the head-rotation. Because the Bilateral Ambisonics signals are computed accurately at the translated ear position due to head-rotation, it is only necessary to rotate these signals using (10) and render the binaural signal with (3). The ideal Bilateral Ambisonics signal, denoted by $p_{\text{ideal}}(f)$, is used for comparing head-tracking compensation performance. The reader can refer to Table I for a summary of the signals involved in this section.

The objective measures defined in Section VI were calculated as follows:

- The NMSE, $\epsilon_{\text{NMSE}}(f)$, was calculated with (24) for all of the signals listed in Table I and Section VII-B. The evaluated signals were calculated for order $N = 4$ and three head-rotation angles $(\alpha = 0°, \beta = 0°, \gamma = 15°/45°/90°)$. The NMSE was averaged over 434 plane-waves directions taken form the Lebedev sampling scheme of order 17.
- The Normalized Magnitude Error, $\epsilon_{\text{mag}}(f)$, was calculated with (25) with the same reference and evaluated signals as in the NMSE, and also averaged identically.
- The average ITD error, $\epsilon_{\text{ITD}}(\gamma)$, was calculated with (26). $\text{ITD}_{\text{ref}}(\Omega_i)$ was evaluated using $p_{\text{ref}}(f)$ left and right ear binaural signals. $\text{ITD}(\Omega_i, \gamma)$ was evaluated for $p_{\text{blt}}(f)$, and $p_{\text{bl-blt}}(f)$ of order $N = 1, 2, 3, 4$ and $(\alpha = 0°, \beta = 0°, \gamma = 0° - 90°)$. Here, Ω_i varies over $q = 200$ directions, distributed uniformly on the left horizontal plane of the head $(\theta_i = 90°;\ 0 \le \phi_i \le 180°)$.
- The average ILD error was calculated with (27). $p_{\text{ref}}(f)$ was used to calculate $\text{ILD}_{\text{ref}}(\Omega_i)$. The evaluated ILD, $\text{ILD}(\Omega_i, \gamma)$, was calculated for $p_{\text{blt}}(f)$, and $p_{\text{bl-blt}}(f)$ of order $N = 1, 2, 3, 4$ with $(\alpha = 0°, \beta = 0°, \gamma = 0° - 90°)$. The plane-wave incident angle Ω_i was identical to in the ITD case.
- The upper accurate order $N_b(f)$, was calculated from (31) which is based on the SH Power Spectrum Error, $\epsilon_S(f, n)$ of (30). The reference signal was $a_{nm}^{\text{ref}}(f) = a_{nm}^{\text{ideal}}(f)$ from (34) and the evaluated signal $a_{nm}(f) = a_{nm}^{L_{\text{tr}}}(f)$ was the Bilateral Ambisonics signal of (33) translated by (7). Both the reference and the evaluated signals were of order $N = 15$ and the SH Power Spectrum Error was calculated for three rotation angles $(\alpha = 0°, \beta = 0°, \gamma = 15°, 45°, 90°)$. In the simulation, $\epsilon_S(f, n)$ was averaged over 434 different plane-wave directions. Additionally, the theoretical upper accurate order $\underline{N}_{a_\epsilon^\tau}(f)$ of (17) was also calculated and plotted to be compared with the numerical $N_b(f)$.

TABLE I
SIGNALS OF THE SIMULATIVE STUDY

Binaural signal	Signal name	Order	Sound-field
$p_{\text{ref}}(f)$	Reference	$\tilde{N} = 41$	$a_{nm}^{r}(f)$
$p_{\text{loa}}(f)$	Low-Order Ambisonics	$N = 1, \ldots, 4$	$a_{nm}^{r}(f)$
$p_{\text{mls}}(f)$	MagLS	$N = 1, \ldots, 4$	$a_{nm}^{r}(f)$
$p_{\text{blt}}(f)$	Bilateral Ambisonics	$N = 1, \ldots, 4$	$a_{nm}^{L_{\text{tr}}}(f)$
$p_{\text{bl-blt}}(f)$	Band-Limited Bilateral Ambisonics	$N = 1, \ldots, 4$	$a_{nm}^{\text{bl-blt}}(f)$
$p_{\text{ideal}}(f)$	Ideal Bilateral Ambisonics	$N = 1, \ldots, 4$	$a_{nm}^{\text{ideal}}(f)$

C. Results

1) Binaural Error Vs. SH: As shown in Section IV, the suggested head-tracking compensation method for Bilateral Ambisonics signals introduces error. Fig. 2 illustrates this error, by comparing the analytical measure $\underline{N}_{a_\epsilon^\tau}(f)$ (solid line) to the numerical measure $N_b(f)$ (dotted line), for three head-rotation values. The results of this simulation support the conclusions of Section IV. First, the error appears to be frequency dependent; specifically, it increases as the frequency increases, since more orders are observed to be inaccurate as the frequency increases. Second, the error appears to depend on the head-rotation angle, since the overall error increases as γ increases. Finally, both the analytical measure $\underline{N}_{a_\epsilon^\tau}(f)$ and the numerical measure $N_b(f)$

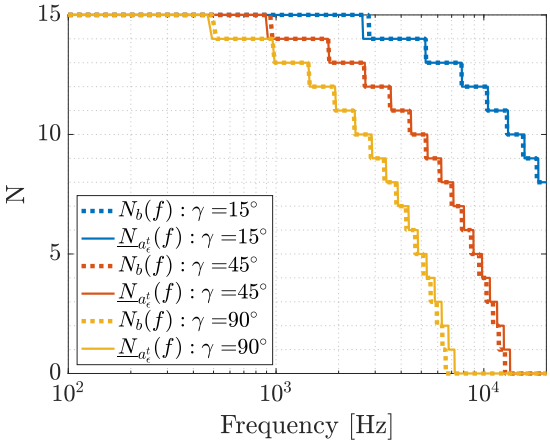
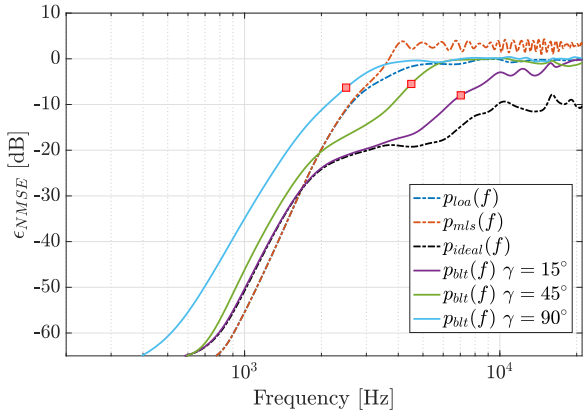
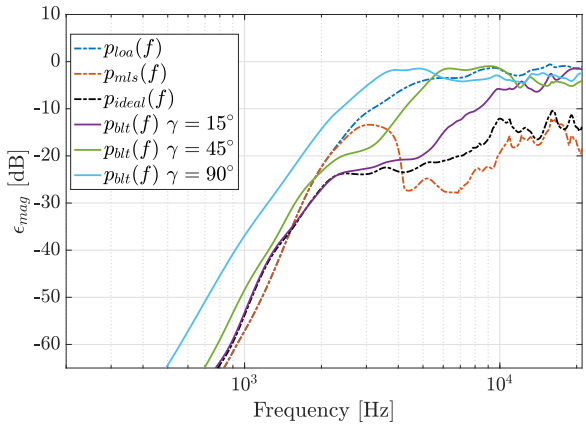


Fig. 2. Highest accurate order $N_b(f)$ and the theoretical highest accurate order $\underline{N}_{a_\epsilon^\tau}(f)$, for three head-rotations: 15°, 45°, and 90°.



(a) Normalized mean square error



(b) Normalized magnitude error

Fig. 3. Left ear binaural NMSE (a) and magnitude (b) error from (24) and (25), computed for $N = 4$, three head-rotations 15°, 45°, 90° and four methods: $p_{\text{loa}}(f)$-low-order Ambisonics, $p_{\text{mls}}(f)$-MagLS, $p_{\text{ideal}}(f)$−Ideal Bilateral Ambisonics, and p_{blt}-Bilateral Ambisonics. The Reference signal is an $N = 41$ binaural signal, rendered using (1). The red markers in (a) indicate f_2 from (20), computed to be 7 kHz, 4.5 kHz, 2.5 kHz for $\gamma = 15°, 45°, 90°$, respectively.

seem to concur. These measures could be useful in finding up to which frequency and order the suggested compensation method can be considered accurate. For example, in Fig. 2 with $\gamma = 45°$ for $f \geq 14$kHz every order in the compensated signal is inaccurate. Similar behavior was also observed for other reproduction orders and head-rotation angles.

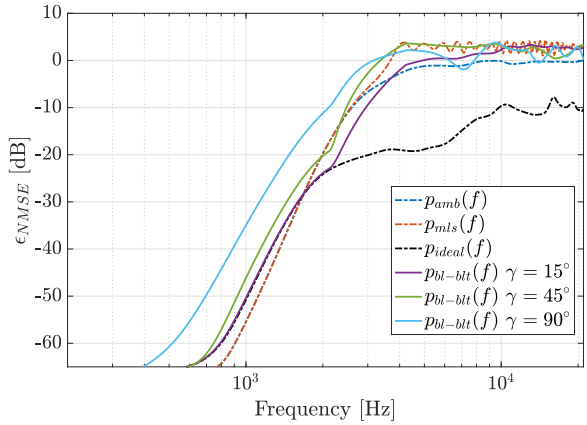
2) Binaural Error Vs. Frequency: Fig. 3 presents the NMSE curves of the evaluated left ear compensated binaural signals for three head-rotation angles. Fig. 3(a) depicts the NMSE of $p_{\text{loa}}(f)$, $p_{\text{mls}}(f)$, $p_{\text{ideal}}(f)$, $p_{\text{blt}}(f)$. An error of less than -10 dB is achieved for $p_{\text{blt}}(f)$ up to about 6.3 kHz, 3.7 kHz, 2 kHz for $\gamma = 15°, 45°, 90°$, respectively. Because the error typically increases monotonically with frequency, the value of -10 dB was selected as a cut-off point to identify the transition between frequency regions of low and high error.

Several observations can be make when considering the NMSE curves in Fig. 3(a). First, the NMSE increases significantly for large rotation angles compared to $p_{\text{ideal}}(f)$. Second, f_2 in (20), marked by red squares in Fig. 3(a), was computed to be 7 kHz, 4.5 kHz, 2.5 kHz for $\gamma = 15°, 45°, 90°$ and $N = 4$, which coincides here with the -5 dB to -10 dB error values. Note that as $p_{\text{loa}}(f)$, $p_{\text{mls}}(f)$, $p_{\text{ideal}}(f)$ are not expected to change as a function of head-rotation angle, they are presented without head-rotation. Finally, The LOA and MagLS approaches show an NMSE error of less than -10 dB up to about 2.6 kHz. The MagLS NMSE is high above the cutoff frequency, as expected, since it is optimized to lower the magnitude error and ignore the phase information.

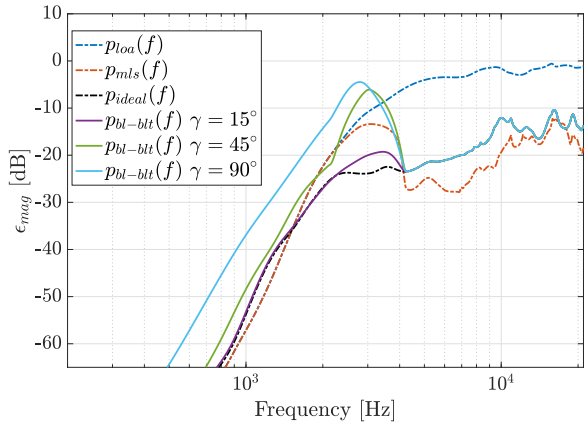
Fig. 3(b) presents the magnitude error of the same signals. An error of less than -10 dB is observed up to 8.2 kHz, 4.2 kHz, 2.4 kHz for $p_{\text{blt}}(f)$ under the three rotations. The magnitude error increases significantly for large rotation angles, compared to $p_{\text{ideal}}(f)$, which presents an error lower than -10 dB for the entire range. $p_{\text{loa}}(f)$ shows low accuracy with an error of less than -10 dB only up to 3.1 kHz. The MagLS signal presents the best accuracy over most of the frequency range.

3) Binaural Error for Band-Limited Bilateral Ambisonics: Fig. 4 presents the binaural NMSE error 4(a) and binaural magnitude error 4(b) of $p_{\text{bl-blt}}(f)$ for the same three rotation angles presented above, and for $p_{\text{loa}}(f)$, $p_{\text{mls}}(f)$, and $p_{\text{ideal}}(f)$ as in Fig. 3. The NMSE of $p_{\text{bl-blt}}(f)$ increases compared to $p_{\text{blt}}(f)$ in Fig. 3(a). However it remains lower than -10 dB up to 2 kHz for all rotation angles. In Fig. 4(b), the $p_{\text{bl-blt}}(f)$ error curves merge with $p_{\text{ideal}}(f)$ above 4.2 kHz, leading to a significant reduction in magnitude error over this frequency range compared to the magnitude error of $p_{\text{blt}}(f)$ in Fig. 3(b). These results may suggest perceptual benefits of $p_{\text{bl-blt}}(f)$ compared to $p_{\text{blt}}(f)$. The ITD and ILD analysis presented next aims to further investigate this behavior.

4) ILD With Band-Limited Bilateral Ambisonics: Performance analysis of $p_{\text{blt}}(f)$ compared to $p_{\text{bl-blt}}(f)$ in terms of ILD is presented next. Fig. 5(a) presents the ILD curves of $p_{\text{blt}}(f), p_{\text{bl-blt}}(f)$ of $N = 4$ with $\gamma = 0°, 90°$. Additionally the ILD of $p_{\text{ref}}(f)$ is also depicted. Fig. 5(a) shows that the ILD of, $p_{\text{blt}}(f)$ is degraded for $\gamma = 90°$, while the ILD of $p_{\text{bl-blt}}(f)$ remains fairly accurate. For example, in Fig. 5(a) $p_{\text{blt}}(f)$ with $\gamma = 90°$ at incident angle 50° has a ILD of 10 dB while the reference curve has a ILD of 15 dB. This 5 dB discrepancy could

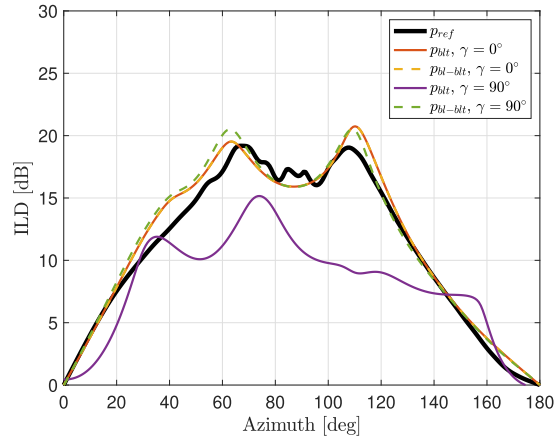


(a) Normalized mean square error

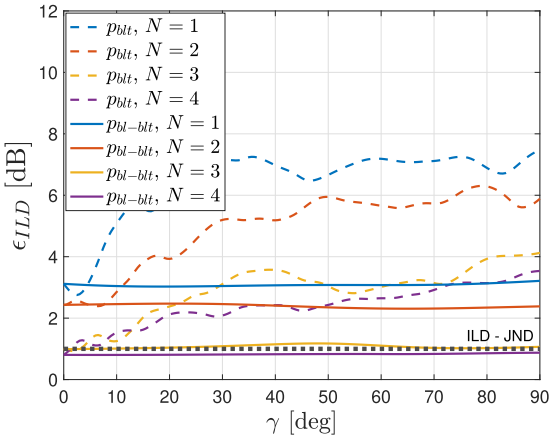


(b) Normalized magnitude error

Fig. 4. Left ear binaural NMSE (a) and magnitude (b) error from (24) and (25), computed for $N = 4$, three head-rotations $15°$, $45°$, $90°$ and four methods: $p_{\text{loa}}(f)$-low-order Ambisonics, $p_{\text{mls}}(f)$-MagLS, $p_{\text{ideal}}(f)$−Ideal Bilateral Ambisonics, and $p_{\text{bl-blt}}$-Band-Limited Bilateral Ambisonics. The Reference signal is an $N = 41$ binaural signal, rendered using (1).



(a) ILD curves

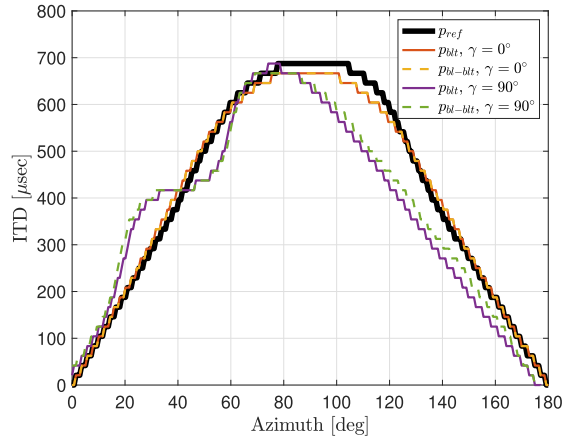


(b) ILD error

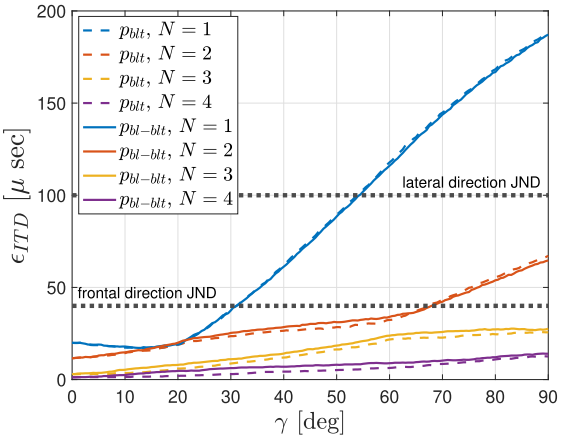
Fig. 5. (a) ILD curves of $p_{\text{ref}}(f), p_{\text{blt}}(f), p_{\text{bl-blt}}(f)$ calculated by (28) for $N = 4$ and $\gamma = 0°, 90°$. (b) Depicts the ILD error of (29), averaged with respect to the incident angle Ω_i, calculated for $p_{\text{blt}}(f), p_{\text{bl-blt}}(f)$ with $N = 1, 2, 3, 4$ and $\gamma = 0° - 90°$. The dotted black line represents the JND value.

cause a significant localization shift of about $\pm 30°$[36]. On the other hand, $p_{\text{bl-blt}}(f)$ only has a 1 dB difference, suggesting that no localization shifts may occur. To further demonstrate this phenomena, Fig. 5(b) depicts the error, $\epsilon_{\text{ILD}}(\gamma)$ of (29), of $p_{\text{blt}}(f)$ and $p_{\text{bl-blt}}(f)$ for $\gamma \in [0, 90]°$ with $N = 1, 2, 3, 4$. Here, $p_{\text{bl-blt}}(f)$ seems unaffected by γ, as opposed to $p_{\text{blt}}(f)$ whose ILD error depends on the compensation angle. Maintaining the initial ILD error is especially preferable in the $N = 3, 4$ case where the average error is close the JND value (1 dB) [39], as seen Fig. 5(b) and in [12]. This behavior can be explained by the fact that for frequencies above $\sqrt{2} f_c \approx 2.8$kHz from (22), for which ILD is important, $p_{\text{bl-blt}}(f)$ avoids the inaccurate operation of translation. This behavior is also supported by Figs. 4(b) and 3(b). The results in this section demonstrated the advantage of the band-limited Bilateral Ambisonics method in terms of ILD preservation.

5) ITD With Band-Limited Bilateral Ambisonics: Fig. 6(a) presents ITD curves of $p_{\text{bl-blt}}(f)$ and $p_{\text{blt}}(f)$, calculated for $N = 2$ and $\gamma = 0°, 90°$. The figure clearly shows the degradation in ITD at 90 deg. The ITD of $p_{\text{bl-blt}}(f)$ and $p_{\text{blt}}(f)$ are similar because ITD is computed up to 1.5 kHz, while for $f \leq \frac{1}{\sqrt{2}} f_c \approx 1.4$kHz, $p_{\text{blt}}(f)$ and $p_{\text{bl-blt}}(f)$ are identical. Next, the ITD error is evaluated as a function of the rotation angle γ, and presented in Fig. 6(b), detailing $\epsilon_{\text{ITD}}(\gamma)$ for $p_{\text{blt}}(f)$ and $p_{\text{bl-blt}}(f)$ and for $\gamma \in [0, 90]°$ and $N = 1, 2, 3, 4$. The figure shows that for orders $N = 1, 2$ the error exhibits a sharp rise at $\gamma = 20°, 60°$, respectively, which is not apparent for $N = 3, 4$. Fig. 6(b) shows that the average ITD error for orders $N = 3$ and $N = 4$ remains below the frontal direction JND of 40μ sec for all compensation angles. Notice that the curves represents the average error and although the average is below the JND, for some incident angles the error could be above the JND value. In Fig. 5(b), the average ILD error of $p_{\text{bl-blt}}(f)$ for these orders is also below the JND value, suggesting that perceptual performance may remain intact even when head rotation angles are significant. However, it should be noted that for orders $N = 1$ and $N = 2$, the average ITD error exceeds the frontal JND value at $\gamma = 30°$ and $\gamma = 70°$, respectively [35]. Additionally, the average ITD error for $N = 1$ crosses the lateral JND of 100μ sec at $\gamma = 55°$. These results offer an indication of the



(a) ITD curves



(b) ITD error

Fig. 6. (a) ITD curves of $p_{\text{ref}}(f), p_{\text{blt}}(f), p_{\text{bl-blt}}(f)$ for $N = 2$ and $\gamma = 0°, 90°$. (b) Depicts the ITD error of (26) calculated for $p_{\text{blt}}(f), p_{\text{bl-blt}}(f)$ with $N = 1, 2, 3, 4$ and $\gamma = 0° - 90°$, averaged with respect to the incident angle. The dotted two black line represents the value for the lateral and frontal JND.

limited potential range of perceptual accuracy for compensation at these particular orders. The degradation in ITD for $N = 1, 2$ could be explained by considering (20) for $f_2 = 1.5$kHz, leading to $|r_b - r_a|$ that is equivalent to $\gamma = 29°, 61°$ for $N = 1, 2$, respectively. As presented in Section IV, the translation at angles higher than γ can be considered highly inaccurate, therefore negatively affecting the ITD.

In summary, the simulation study presented above showed how the performance of the head-tracking compensation methods proposed in Section III is dependent on the head-rotation angle, reproduction order, and frequency. The performance was compared to Ambisonics and MagLS reproduction in terms of NMSE and Normalized Magnitude Error. $p_{\text{bl-blt}}(f)$, derived in Section V, was observed to have a lower Normalized Magnitude Error at high frequencies compared to $p_{\text{blt}}(f)$, and a similar error at low frequencies. Additionally, $p_{\text{bl-blt}}(f)$, NMSE and Normalized Magnitude Error were shown to be non-dependent on γ in the $f \geq \sqrt{2} f_c$ frequency range. This lead to improved ILD performance while exhibiting similar ITD performance as $p_{\text{blt}}(f)$.

VIII. Listening Tests

This section introduces two listening tests designed to subjectively assess the performance of the proposed head-tracking compensation technique. The first test focuses on evaluating compensation accuracy for various rotation angles, by isolating compensation dependency on specific rotation angles in a static setting. The second test introduces head-tracking capabilities, and seeks to investigate performance differences between Bilateral and MagLS Ambisonics, across different SH orders in a dynamic setting. differences between Bilateral and MagLS Ambisonics.

A. Test I - Static Compensation

1) Experimental Setup: The signals were generated in a simulated acoustic environment composed of a rectangular room of dimensions $8 \times 6 \times 4$m. The image method [40] was employed for the acoustic simulation, and implemented in MATLAB (2020b). The room had a reverberation time of $T_{60} = 0.51$sec with reflection density of 30000 reflections per second, a critical distance of $r_{\text{cd}} = 1.4$m, and a maximum reflection order of 44. A single omni-directional point source was positioned at $(x, y, z) = (4.19, 1.44, 1.7)$m. Two types of audio source stimuli where used: English male speaker and castanets. The male speaker represents typical content for binaural reproduction applications, while the castanets were chosen due to their high-frequency content and strong transients. The Ambisonics signals were calculated using a spherical microphone array with its center located at (3.5,3,1.7)m, which is at a distance of 1.7 m from the source at $-66°$ (azimuth angle) relative to the coordinate system depicted in Fig. 1(b). The Bilateral Ambisonics signals were calculated from two spherical microphone arrays located at (3.5,3.0857,1.7)m, and (3.5,2.9125,1.7)m. These locations correspond to the left/right ear position of a listener orientated in the positive x-axis direction with a head radius of $r = 8.75$cm, as illustrated in Fig. 1(b). To simulate different listener head orientations, Bilateral Ambisonics signals were also calculated from two spherical microphone arrays positioned and oriented to match four head-rotation angles $\gamma = 15°, 30°, 60°, 90°$. Measurements of the Neumann KU100 HRTF [37] where used to render the signals in this section. All signals were convolved with matching headphone compensation filters, taken from the Cologne database [37], which were measured on the Neumann KU100 dummy head, and the loudness of all signals was equalized to the same level.

2) Methodology: The evaluation focused on signals $p_{\text{bl-blt}}$ - Bilateral Ambisonics of orders $N = 1, 2$, calculated using (3) and listed in Table I. Compensation for head rotation followed the procedure detailed in Section IV, covering rotation angles $\gamma = 15°, 30°, 60°, 90°$ using (10) and (21), with a cutoff frequency of $f_c = 2$ kHz. The objective was to create signals that would be perceived by listeners as originating from the initial source location, as described in Section VIII-A1. The listening tests followed the MUltiple Stimuli with Hidden Reference and Anchor (MUSHRA) protocol [41]. Signals were divided into four MUSHRA screens, each containing all rotation angles $\gamma = 0°, 15°, 30°, 60°, 90°$, for orders $N = 1$ or $N = 2$ and

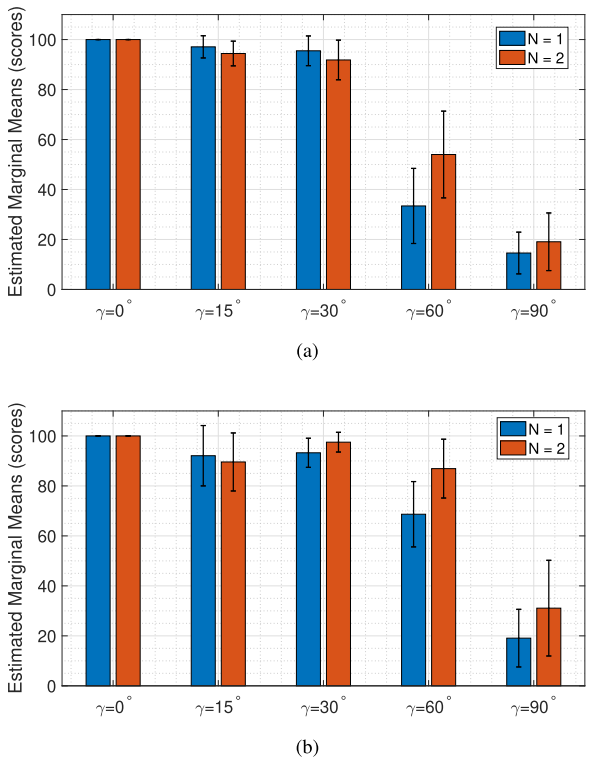


Fig. 7. Test I results: Estimated marginal means and 95% confidence interval of the compensated $p_{\text{bl-blt}}$ signals for the γ and N variables, (a) for castanets, and (b) for speech stimulus.

source signals (speech or castanets). The hidden reference signal was $p_{\text{bl-blt}}$ with $\gamma = 0°$, leading to five signals per screen.

The test involved nine male and three female participants without known hearing impairments, all familiar with spatial listening tests. Participants conducted the test in a quiet environment and were instructed to keep their heads still during listening due to the absence of head tracking during playback. A training phase and signal familiarization phase preceded the actual test. The training phase acquainted participants with the test equipment and grading scale. In the signal familiarization phase, participants became accustomed to all test signals. Participants were required to rate the relative differences between reference and test signals on a scale of 0 to 100, where 100 indicates no audible difference from the reference. Lower scores indicate more significant differences in comparison to the reference. Participants assessed signals based on overall quality, encompassing spatial artifacts, time-varying artifacts, and spectral artifacts.

3) Results: The scores reported by the participants where analysed using a Repeated Measures ANalysis of VAriance (RM-ANOVA) [42]. The within-subject variables were defined as follows: Order ($N = 2, N = 1$), Angle difference ($\gamma = 0°, 15°, 30°, 60°, 90°$) and stimuli (speaker, castanets). The estimated marginal means and 95% confidence interval for the interactions are presented in Fig. 7. The analysis uncovered statistically significant main effects for γ, $F(4, 44) = 137.44, p < .001, \eta_p^2 = .93$, the stimuli, $F(1, 11) = 13.31, p = .004, \eta_p^2 = .55$, and partially statistically significant main effect for N, $F(1, 11) = 4.74, p = .052, \eta_p^2 = .3$. Additionally, the interaction of γ and N was also statistically significant, $F(4, 44) = 10.508, p < .001, \eta_p^2 = .49$.

To further understand the dependence on γ, a post-hoc test with Bonferroni correction was performed. The pairwise comparisons between the estimated marginal means of the reference ($\gamma = 0°$) and the rest of the angles resulted in a mean difference of $6.71, 5.48, 39.25, 76.08$ points with $p = .38, .07, .001, .001$ for $\gamma = 15°, 30°, 60°, 90°$, respectively. These results seem to concur with the conclusions of Section VII, suggesting decreasing accuracy with increasing head-rotation angles. However, for small angles ($15°, 30°$), the performance drop seems to be less severe compared to the rest of the angles.

Analyzing the statistically significant interaction of γ and N could help to identify how N affects the overall quality with respect to γ. The pairwise comparisons between $N = 1, 2$ for a fixed $\gamma = 60°$ revealed a statistically significant mean difference of 19.41 points with $p < 0.001$. All other pairs demonstrated non-statistically significant differences. The $N = 2, \gamma = 60°$ signal mean was 70.46 and for $N = 1, \gamma = 60°$ was 51.04, which positions the former at "Good" and the latter at "Fair" in the verbal grade scale attached to the test. This result agrees with the conclusions outlined in Sections V and VII, specifically with (20), which can provide an estimation for the accuracy of the compensation. Plugging the simulation parameters into (20) results in $f_2 = 1247$Hz for $N = 1, \gamma = 60°$, which is lower than the cutoff frequency, 2 kHz, used in (22), compared to $f_2 = 1871$Hz for $N = 2, \gamma = 60°$, which is closer to the cutoff, thus leading to better compensation. It may be also important to note that while the rating is related to overall quality, most participants informally reported on differences relating to spatial perception between signals, which does suggest that the difference shown in the results of the listening test are, to some extent, due to spatial attributes.

B. Test II - Dynamic Compensation

1) Experimental Setup: The signals where generated in a simulated anechoic environment, with a single source positioned at $-30°$ azimuth and $0°$ elevation. Binaural signals were computed and stored for a dense grid of 2702 diverse head orientations. These signals were played back according to the head-tracking device to support compensation for head rotation, enabling 3DoF binaural audio experience. The Meta Quest 2 VR headset was used to track participants' head orientations. Finally, the AKG K702 headphones were utilized for playback, including headphone compensation filters. The visual VR environment included the test menus, a chair and a table, and an image of a loudspeaker located $-30°$ to the listener initial orientation.

2) Methodology: The listening tests involved three binaural reproduction methods, leading to the following test signals:

- p_{ref} - order $N = 41$ Ambisonics without HRTF preprocessing, calculated using (1), and rotated using (4). This signal is used as the hidden reference.
- p_{mls} - Ambisonics of orders $N = 1, 2, 3$, calculated using (1) and rotated using (4). For these signals, a MagLS preprocessed HRTF was used, with a cutoff frequency, octave

TABLE II
HEAD MOTION STATISTICS SHOWING MEAN, STANDARD DEVIATION AND FULL RANGE, AVERAGED ACROSS ALL SUBJECTS, FOR LISTENING TEST II

Orientation	Mean	STD	Range
Elevation [degrees]	-2.3	16.2	95.8
Azimuth [degrees]	6.5	44.3	215.3
Elevation Angular Velocity [degrees/sec]	0.0	0.8	18.0
Azimuth Angular Velocity [degrees/sec]	0.0	0.5	10.0

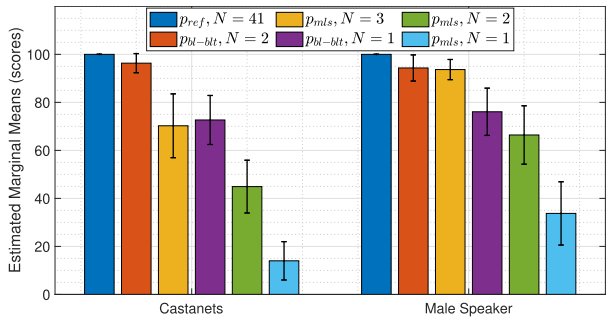


Fig. 8. Test II results: Estimated marginal means and 95% confidence interval for methods p_{ref}, p_{mls}, $p_{\text{bl-blt}}$ including head tracking and two types of stimulus.

smoothing interval and diffuse-field covariance constraint equalization, as described in Section VII-B.

- $p_{\text{bl-blt}}$ - Bilateral Ambisonics of orders $N = 1, 2$, calculated using (3). Compensation for head-rotation was applied as detailed in Section IV using (10) and (21).

The listening tests followed the MUSHRA protocol with two screens for speech (three second long sentence) and castanet signals (seven seconds long), each incorporating all evaluated methods. Twelve participants from test I also participated in test II, undergoing training and signal familiarization. During the test, they rated the overall quality differences between reference and test signals on a 0-100 scale. In each screen, the chosen signal was played in a loop, and participants were allowed to listen to each signal as many times as they wished before assigning scores. Participants conducted the test in a quiet environment and were encouraged to move their heads while listening to evaluate signals across a wide range of orientations. The participants performed the experiment while seated on a swivel chair, providing easy access to a broad range of azimuth angles. The average time it took to finish the entire test was 19 minutes.

3) Results: Head motion data gathered during the listening tests is analyzed in Table II, offering details on participants head movements. The mean, standard deviation (STD), and range of motion, for elevation, azimuth, elevation angular velocity, and azimuth angular velocity readings from the Quest II head tracker were calculated across subjects to derive the results in Table II. Notably, mean values for all variables are near zero, aligning with the expectation of no specific directional bias. Noteworthy are the relatively high STD values for azimuth and elevation at 44.3 and 16.2 degrees, respectively, accompanied by ranges of 215.3 and 95.8 degrees. Conversely, angular velocities exhibit low STD values of 0.5 and 0.8. These findings could suggest that participants engaged in diverse head movements across a wide range of angles while maintaining stability during most of the listening test. The high values of the range of the angular velocities of both elevation and azimuth could indicate that participants tried to examine fast head movements.

Two within-subject variables where analyzed: (a) the stimuli (speaker, castanets), and (b) the method (p_{ref} of $N = 41$, $p_{\text{bl-blt}}$ of $N = 1, 2$, p_{mls} of $N = 1, 2, 3$). The estimated marginal means and 95% confidence interval of the results are presented in Fig. 8. The RM-ANOVA analysis uncovered statistically significant main effect for the stimuli $F(1, 11) = 22, p < .001, \eta_p^2 = .7$, method $F(5, 55) = 105, p < .001, \eta_p^2 = .9$, and the interactions $F(5, 55) = 6, p < .001, \eta_p^2 = .4$. A post-hoc test with Bonferroni correction was conducted for pairwise comparisons between estimated marginal means regarding stimuli and method variables.

Regarding the stimuli variable, a statistically significant mean difference of 11 points ($p < .001$) suggests performance differences concerning stimuli, albeit small. Pairwise comparisons between estimated marginal means of p_{ref} and $p_{\text{bl-blt}}$ $N = 1$, p_{mls} $N = 3, 2, 1$ were statistically significant, with mean differences of $26, 18, 44, 76$ points ($p < .001$). The points difference between p_{ref} and $p_{\text{bl-blt}}$ $N = 2$ was not statistically significant, with a small mean difference of 5 points ($p = .21$). Comparing $p_{\text{bl-blt}}$ of $N = 2$ to p_{mls} of $N = 3, 2, 1$ yielded statistically significant points differences of $13, 40, 71$ ($p = .21, p < .001, p < .001$) in favor of $p_{\text{bl-blt}}$ $N = 2$. Comparatively, $p_{\text{bl-blt}}$ of $N = 1$ against p_{mls} of $N = 3, 2, 1$ resulted in points differences of $-7, 19, 50$ ($p = .12, p = .07, p < .001$).

From this analysis, it can be concluded that binaural signals reproduced from lower order Bilateral Ambisonics, compensated using the proposed method, exhibited similar performance to the high-order Ambisonics reference for $N = 2$, and performance distinguishable from the reference for $N = 1$. These findings are in partial agreement with previous research [11], [12], [43], suggesting that even at $N = 1$, Bilateral Ambisonics can be comparable to high-order Ambisonics. The degradation in Bilateral Ambisonics performance for $N = 1$ is aligned with the results of Section VIII-A and the simulative study in Section VII. Bilateral Ambisonics $N = 2$ did not display performance degradation, hinting that in dynamic scenarios, this order might suffice for reproduction that is indistinguishable from a high-order reference, however, further investigation is warranted to confirm this result. Furthermore, Bilateral $N = 1, 2$ outperformed MagLS $N = 1, 2$ in this scenario, with Bilateral $N = 2$ also outperforming MagLS $N = 3$.

It is important to note that the subjective study results presented should be regarded as preliminary. This is because, given the comprehensive scope of this work, both Test I and II solicited overall quality ratings without delving into specific attributes like coloration, localization, and source stability. We recommend exploring these specific attributes in future studies to better understand the perceptual artifacts associated with the proposed method.

IX. Conclusion

This paper presented a method for the head-tracking compensation of the Bilateral Ambisonics format. The method is based on rotations in the SH domain and translation in the space domain of the Bilateral Ambisonics signals. An alternative frequency-dependent head-tracking compensation method is also proposed based on a limitation study presented in the paper. Results of objective and subjective evaluations indicate that a compensation error is present and depends on the signal order and compensation angle. An objective limitation analysis and static listening tests showed that for $N = 1, 2$, the perceived error should be low up to $\pm 30^\circ, \pm 60^\circ$ degrees of head-rotation, respectfully. Notably, when evaluated with active head-tracking, Bilateral Ambisonics with $N = 2$ showed good performance comparable to high-order Ambisonics, and $N = 1$ performed better than MagLS with $N = 2$. The proposed head-tracking compensation method could be beneficial in scenarios where high-quality binaural sound is needed, and two spherical microphone arrays are available. In particular, results suggest that the proposed head-tracking compensation approach could be suited for VR applications where head rotation motion resemble the one examined in the test, which could be characterized by large head rotation angles with stationary listening. This could be applicable in scenarios like VR teleconferencing or media consumption in VR, where users focus on specific points of interest without extensive head movements. Furthermore, Bilateral Ambisonics with $N = 2$ emerges as a versatile option for a broader range of VR applications, benefiting from more relaxed constraints on head rotations with performance in the objective and listening tests nearly comparable to that of HOA.

References

[1] D. R. Begault and L. J. Trejo, "3-D sound for virtual reality and multimedia," NASA, Ames Research Center, Moffett Field, CA, USA, Tech. Rep. NASA/TM-2000-209606, pp. 132–136, 2000.

[2] M. Vorländer, D. Schröder, S. Pelzer, and F. Wefers, "Virtual reality for architectural acoustics," *J. Building Perform. Simul.*, vol. 8, no. 1, pp. 15–25, 2015.

[3] E. Bates and F. Boland, "Spatial music, virtual reality, and 360 media," in *Proc. Audio Eng. Soc. Conf.: Conf. Audio Virtual Augmented Reality*, 2016. [Online]. Available: https://www.aes.org/e-lib/browse.cfm?elib=18496

[4] M. A. Gerzon, "Periphony: With-height sound reproduction," *J. Audio Eng. Soc.*, vol. 21, no. 1, pp. 2–10, 1973.

[5] S. Moreau, J. Daniel, and S. Bertet, "3D sound field recording with higher order ambisonics–Objective measurements and validation of a 4th order spherical microphone," in *Proc. 120th Conv. AES*, 2006, pp. 20–23.

[6] J. S. Bamford, "An analysis of ambisonic sound systems of first and second order," Ph.D. dissertation, University of Waterloo, Waterloo, ON, Canada, 1995.

[7] E. Bates, S. Dooney, M. Gorzel, H. O'Dwyer, L. Ferguson, and F. M. Boland, "Comparing ambisonic microphones—Part 2," in *Proc. Audio Eng. Soc. Conf.: 2016 AES Int. Conf. Sound Field Control.*, 2017. [Online]. Available: https://www.aes.org/e-lib/browse.cfm?elib=18607

[8] Z. Ben-Hur, D. L. Alon, B. Rafaely, and R. Mehra, "Loudness stability of binaural sound with spherical harmonic representation of sparse head-related transfer functions," *EURASIP J. Audio, Speech, Music Process.*, vol. 2019, no. 1, pp. 1–14, 2019.

[9] D. L. Alon, Z. Ben-Hur, B. Rafaely, and R. Mehra, "Sparse head-related transfer function representation with spatial aliasing cancellation," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2018, pp. 6792–6796.

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

[11] Z. Ben-Hur, D. Alon, R. Mehra, and B. Rafaely, "Binaural reproduction using Bilateral Ambisonics," in *Proc. Audio Eng. Soc. Conf.: AES Int. Conf. Audio Virtual Augmented Reality*, 2020. [Online]. Available: https://www.aes.org/e-lib/browse.cfm?elib=20871

[12] Z. Ben-Hur, D. L. Alon, R. Mehra, and B. Rafaely, "Binaural reproduction based on bilateral ambisonics and ear-aligned HRTFs," *IEEE/ACM Trans. Audio, Speech, Lang. Process*, vol. 29, pp. 901–913, 2021.

[13] Z. Ben-Hur, D. Alon, O. Berebi, R. Mehra, and B. Rafaely, "Binaural reproduction based on bilateral ambisonics," in *Advances in Fundamental and Applied Research on Spatial Audio*. London, U.K.: IntechOpen, 2021, pp. 155–174.

[14] J.-M. Jôt, S. Wardle, and V. Larcher, "Approaches to binaural synthesis," in *Proc. Audio Eng. Soc. Conv. 105th AES*, 1998. [Online]. Available: https://www.aes.org/e-lib/browse.cfm?elib=8319

[15] 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 Trans.*, vol. 27, no. 12, pp. 2249–2262, Dec. 2019.

[16] B. Rafaely and M. Kleider, "Spherical microphone array beam steering using Wigner-D weighting," *IEEE Signal Process. Lett.*, vol. 15, pp. 417–420, 2008.

[17] S. Perrett and W. Noble, "The effect of head rotations on vertical plane sound localization," *J. Acoust. Soc. Amer.*, vol. 102, no. 4, pp. 2325–2332, 1997.

[18] D. R. Begault, E. M. Wenzel, and M. R. Anderson, "Direct comparison of the impact of head tracking, reverberation, and individualized head-related transfer functions on the spatial perception of a virtual speech source," *J. Audio Eng. Soc.*, vol. 49, no. 10, pp. 904–916, 2001.

[19] O. Berebi, B. Rafaely, Z. Ben-Hur, and D. L. Alon, "Enabling head-tracking for binaural sound reproduction based on Bilateral Ambisonics," in *Proc. IEEE Immersive 3D Audio: Architecture Automot.*, 2021, pp. 1–7.

[20] C. Schörkhuber, M. Zaunschirm, and R. Höldrich, "Binaural rendering of ambisonic signals via magnitude least squares," in *Proc. DAGA*, 2018, pp. 339–342.

[21] B. Xie, *Head-Related Transfer Function and Virtual Auditory Display*. Fort Lauderdale, FL, USA: J. Ross Publishing, 2013.

[22] B. Rafaely and A. Avni, "Interaural cross correlation in a sound field represented by spherical harmonics," *J. Acoust. Soc. Amer.*, vol. 127, no. 2, pp. 823–828, 2010.

[23] B. Rafaely, "Plane-wave decomposition of the sound field on a sphere by spherical convolution," *J. Acoust. Soc. Amer.*, vol. 116, no. 4, pp. 2149–2157, 2004.

[24] Z. Ben-Hur, F. Brinkmann, J. Sheaffer, S. Weinzierl, and B. Rafaely, "Spectral equalization in binaural signals represented by order-truncated spherical harmonics," *J. Acoust. Soc. Amer.*, vol. 141, no. 6, pp. 4087–4096, 2017.

[25] G. B. Arfken, H. J. Weber, and D. Spector, "Mathematical methods for physicists," *Amer. J. Phys.*, vol. 67, pp. 165–169, 1999.

[26] P. J. Kostelec and D. N. Rockmore, "FFTs on the rotation group," *J. Fourier Anal. Appl.*, vol. 14, no. 2, pp. 145–179, 2008.

[27] B. Rafaely, *Fundamentals of Spherical Array Processing*. vol. 8. Berlin, Germany: Springer, 2015.

[28] E. G. Williams, *Fourier Acoustics: Sound Radiation and Nearfield Acoustical Holography*. Cambridge, MA, USA: Academic Press, 1999.

[29] V. Rokhlin, "Diagonal forms of translation operators for the Helmholtz equation in three dimensions," *Appl. Comput. Harmon. Anal.*, vol. 1, no. 1, pp. 82–93, 1993.

[30] J. R. Driscoll and D. M. Healy, "Computing fourier transforms and convolutions on the 2-sphere," *Adv. Appl. Math.*, vol. 15, no. 2, pp. 202–250, 1994.

[31] J. D. Jackson, *Classical Electrodynamics*. vol. 67, American Journal Of Physics, 1999, pp. 841–842.

[32] V. Best, R. Baumgartner, M. Lavandier, P. Majdak, and N. Kopčo, "Sound externalization: A review of recent research," *Trends Hear.*, vol. 24, 2020, Art. no. 2331216520948390.

[33] F. L. Wightman and D. J. Kistler, "The dominant role of low-frequency interaural time differences in sound localization," *J. Acoust. Soc. Amer.*, vol. 91, no. 3, pp. 1648–1661, 1992.

[34] L. H. Loiselle, M. F. Dorman, W. A. Yost, S. J. Cook, and R. H. Gifford, "Using ILD or ITD cues for sound source localization and speech understanding in a complex listening environment by listeners with bilateral and with hearing-preservation cochlear implants," *J. Speech, Lang., Hear. Res.*, vol. 59, no. 4, pp. 810–818, 2016.

[35] A. Andreopoulou and B. F. Katz, "Identification of perceptually relevant methods of inter-aural time difference estimation," *J. Acoust. Soc. Amer.*, vol. 142, no. 2, pp. 588–598, 2017.

[36] B. F. Katz and R. Nicol, "Binaural spatial reproduction," in *Sensory Evaluation of Sound*. Boca Raton, FL, USA: CRC Press, 2018, pp. 349–388.
[37] B. Bernschütz, "A spherical far field HRIR/HRTF compilation of the neumann KU 100," in *Proc. 40th Italian (AIA) Annu. Conf. Acoust. 39th German Annu. Conf. Acoust. (DAGA) Conf. Acoust.*, 2013, pp. 29–32.
[38] I. Engel, D. Goodman, and L. Picinali, "Improving binaural rendering with bilateral ambisonics and MagLS," in *Proc. Annu. German Conf. Acoust.*, vol. 99, no. 1, 2021, pp. 10–13.
[39] W. A. Yost and R. H. Dye Jr., "Discrimination of interaural differences of level as a function of frequency," *J. Acoust. Soc. Amer.*, vol. 83, no. 5, pp. 1846–1851, 1988.
[40] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *J. Acoust. Soc. Amer.*, vol. 65, no. 4, pp. 943–950, 1979.
[41] *Method for the Subjective Assessment of Intermediate Quality Level of Audio Systems*, Recommendation ITU-R Standard BS.1534-3, International Telecommunication Union Radiocommunication Assembly, Geneva, Switzerland, 2014.
[42] H. Keselman, J. Algina, and R. K. Kowalchuk, "The analysis of repeated measures designs: A review," *Brit. J. Math. Stat. Psychol.*, vol. 54, no. 1, pp. 1–20, 2001.
[43] I. Engel, D. F. Goodman, and L. Picinali, "Assessing HRTF preprocessing methods for Ambisonics rendering through perceptual models," *Acta Acustica*, vol. 6, 2022, Art. no. 4.

Or Berebi received the B.Sc. and M.Sc. degrees in electrical and computer engineering in 2019 and 2022, respectively, from the Ben-Gurion University of the Negev, Beer-Sheva, Israel, where he is currently working toward the Ph.D. degree. His research focuses on improving spatial audio perception for low order binaural reproduction.

Zamir Ben-Hur received the B.Sc. (*summa cum laude*), M.Sc. and Ph.D. degrees in electrical and computer engineering from Ben-Gurion University of the Negev, Beer-Sheva, Israel, in 2015, 2017, and 2020, respectively. He is currently a research Scientist with Meta Reality Labs Research, Meta, CA, USA, working on spatial audio technologies. His research interests include spatial audio signal processing for binaural reproduction with improved spatial perception.

David Lou Alon received the Ph.D. degree in electrical engineering from Ben Gurion University, Beer-Sheva, Israel, in 2017, in the field of spherical microphone array processing. He is currently a Research Scientist with Meta Reality Labs Research, Menlo Park, CA, USA, investigating spatial audio technologies. His research interests include head-related transfer functions, spatial audio capture, binaural reproduction, and headphone equalization for VR and AR application.

Boaz Rafaely (Senior Member, IEEE) received the B.Sc. degree (*cum laude*) in electrical engineering from Ben-Gurion University, Beer-Sheva, Israel, in 1986, the M.Sc. degree in biomedical engineering from Tel-Aviv University, Tel Aviv, Israel, in 1994, and the Ph.D. degree from the Institute of Sound and Vibration Research (ISVR), Southampton University, Southampton, U.K., in 1997. At the ISVR, he was appointed Lecturer in 1997 and Senior Lecturer in 2001, working on active control of sound and acoustic signal processing. In 2002, he spent six months as a Visiting Scientist with the Sensory Communication Group, Research Laboratory of Electronics, Massachusetts Institute of Technology (MIT), Cambridge, MA, USA, investigating speech enhancement for hearing aids. He then joined the Department of Electrical and Computer Engineering, Ben-Gurion University, Beersheba, Israel, as a Senior Lecturer in 2003, and appointed an Associate Professor in 2010, and Professor in 2013. He is currently heading the acoustics laboratory, investigating methods for audio signal processing and spatial audio. During 2010–2014, he was an Associate Editor for IEEE TRANSACTIONS ON AUDIO, SPEECH AND LANGUAGE PROCESSING, and during 2013–2018, a Member of the IEEE Audio and Acoustic Signal Processing Technical Committee. He was also an Associate Editor for IEEE SIGNAL PROCESSING LETTERSduring 2015–2019, *IET Signal Processing* during 2016–2019, and currently of *Acta Acustica*. During 2013–2016, he was the Chair of the Israeli Acoustical Association, and is currently chairing the Technical Committee on Audio Signal Processing in the European Acoustical Association. Dr. Rafaely was awarded the British Council's Clore Foundation Scholarship.