Binaural Reproduction Based on Bilateral Ambisonics and Ear-Aligned HRTFs

Zamir Ben-Hur, David Lou Alon, Ravish Mehra, and Boaz Rafaely, *Senior Member, IEEE*

***Abstract*—Reproduction of high quality spatial sound has gained considerable importance with the recent technology developments in the fields of virtual and augmented reality. Recently, the reproduction of binaural signals in the Spherical-Harmonics (SH) domain has been proposed. This is performed by using SH representations of the sound-field and the Head-Related Transfer Function (HRTF). These processes offer the flexibility to control the reproduced binaural signals, by manipulating the sound-field or the HRTFs using algorithms that operate directly in the SH domain. However, in most practical cases, the binaural reproduction is order-limited, which introduces truncation error that has a detrimental effect on the perception of the reproduced signals, mainly due to the truncation of the HRTF. A recent study showed that pre-processing of the HRTF by ear-alignment reduces its effective SH order, which may be beneficial for alleviating the above effect. In this paper, a method to incorporate the pre-processed ear-aligned HRTF into the binaural reproduction process is presented. The method uses Ambisonics representation of the sound-field formulated at the two ears, and is denoted here as Bilateral Ambisonics. The proposed method leads to a significant reduction in errors due to the limited-order reproduction, which yields a substantial improvement in perceived binaural reproduction quality even with SH as low as first order.**

***Index Terms*—Binaural reproduction, ambisonics, HRTF, head-related transfer function, spatial sound, spherical harmonics, spherical microphone array.**

I. Introduction

BINAURAL technology plays an important role in many applications, such as virtual and augmented reality [1], architectural acoustics [2] and hearing science [3]. Binaural reproduction of spatial sounds provides the listener with the sensation of being present in the 3D audio scene. Binaural signals can be obtained using microphones placed at the ears of a manikin, in which case the sound-field and the Head-Related Transfer Function (HRTF) are jointly captured, and the reproduced binaural signal is limited to the specific recording scenario [4]. Alternatively, more flexible reproduction can be achieved by synthesizing the binaural signals in post processing, which enables, for example, the use of personalized HRTFs and head-tracking, and is the approach followed in this paper. This requires the sound-field and the HRTF to be available separately. The HRTF, which is an individual function, can be measured acoustically or simulated numerically for each listener [5]. The sound-field can be captured using microphone array recordings or obtained from numerical simulations [6]–[8].

Recently, rendering of binaural signals in the spherical harmonics (SH) domain has been proposed [9]. This is performed by adding the products of the SH coefficients of the plane-wave (PW) density function (which encodes the directional information of the sound field, also referred to as Ambisonics [10]) with the SH representation of the free-field HRTFs. The advantage of using SH processing is the flexibility to manipulate the sound field, or the corresponding HRTFs, by directly employing algorithms that operate in the SH domain [11], [12].

The SH coefficients of the PW density function of a measured sound-field can be obtained from a spherical microphone array recording [13]. In practice, such arrays have a limited number of microphones, which limits the spatial bandwidth of the estimated PW decomposition [14]. Also, for practical applications where a simulated sound-field is used, a similar limitation may apply, due to memory usage or computational efficiency considerations [1], [15]. This, in turn, places a constraint on the maximum SH order of the employed HRTF, which is inherently a high-order function [16]. The truncation error, caused by the limited order of the HRTF, results in significant artifacts, both in space and in frequency. These artifacts have a detrimental effect on the perception of the reproduced binaural signals, for example, in timbre, localization, externalization, source width and stability of the virtual sound source [17]–[20].

Considerable efforts have been made in order to reduce the effect of the errors caused by the order limitation of the reproduced binaural signal. A recent study [18] suggested correction of the coloration caused by the truncation error by spectral equalization of the binaural signals. Hold *et al.* [21] presented an improvement to this correction by SH tapering, which also corrects some of the angle-dependent artifacts. These methods attempt to reduce the effect of the truncation error directly on the binaural signal, by means of post-processing. On the other hand, pre-processing of the HRTF has been shown to reduce its effective SH order [22], [23], which may provide some potential for reducing the truncation error. This order reduction is based on manipulating the phase component of the HRTF.

Manuscript received June 15, 2020; revised December 18, 2020; accepted January 25, 2021. Date of publication January 28, 2021; date of current version February 15, 2021. This work was supported by Facebook Reality Labs. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Stefan Bilbao. *(Corresponding author: Zamir Ben-Hur.)*

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

David Lou Alon and Ravish Mehra are with the Facebook Reality Labs, Facebook, 1 Hacker Way, Menlo Park, CA 94025 USA (e-mail: davidalon@fb.com; ravish.mehra@oculus.com).

This article has supplementary downloadable material available at https://doi.org/10.1109/TASLP.2021.3055038, provided by the authors.

Digital Object Identifier 10.1109/TASLP.2021.3055038

However, the use of such a pre-processed HRTF for binaural reproduction in the SH domain is not trivial due to the relation between the phases of the HRTF and the PW density function. In [24], Zaunschirm *et al.* presented a method that uses a pre-processed HRTF, obtained by means of frequency-dependent time-alignment, to reproduce binaural signals in the SH domain using constrained optimization. They suggested pre-processing of the HRTF by removing its linear-phase component at high frequencies. Schörkhuber *et al.* further developed this approach in [25], where they presented the Magnitude Least-Squares (MagLS) method that performs magnitude-only optimization at high frequencies. Although the linear-phase component at high frequencies may be less important for lateral localization [26], [27], its removal still introduces errors in the binaural signal, and may affect other perceptual attributes [28], [29]. In [30], Lübeck *et al.* showed that the MagLS method achieved similar perceptual improvement to the equalization methods for binaural reproduction with order 3 and above. In summary, despite of the significant advance in this field, current methods for binaural reproduction that are based on low-order HRTF and sound-field representations seem to degrade perception compared to a high-order reference. Therefore, high-quality binaural reproduction based on low-order Ambisonics remains an open problem.

In this paper, a method for the incorporation of a pre-processed HRTF with a reduced SH order into the binaural reproduction process is presented. The suggested method uses the ear-alignment technique that was recently developed for pre-processing of HRTFs [31], and computes the binaural signals directly at the location of the listener's ears, using Ambisonics representation of the sound-field at the two ears. The proposed method is based on the Binaural B-Format approach, suggested by Jot *et al.* [32] for 1st order Ambisonics using minimum-phase approximation of the HRTF. A similar approach was presented recently by Armstrong *et al.* [33] in the context of virtual loudspeaker reproduction. In the current contribution, the Binaural B-Format is formulated for an arbitrary SH order, enhancing its efficacy, and now incorporates the ear-aligned HRTF, which preserves the HRTF phase information. This method, denoted as Bilateral Ambisonics reproduction, significantly reduces the truncation error. An objective analysis shows that a binaural signal reproduced using Bilateral Ambisonics with a low SH order of $N = 4$ is comparable to a binaural signal that was rendered using Basic Ambisonics reproduction with an order as high as $N = 40$. Additionally, a perceptual test shows that Bilateral Ambisonics reproduction with an order as low as $N = 1$ may be sufficient for reproducing binaural signals that are highly similar to a high-order reference.

In summary, the contributions of this paper are as follows:

1) Mathematical formulation of the Bilateral Ambisonics reproduction by incorporating the ear-aligned HRTF. This generalizes the Binaural B-Format for an arbitrary SH order (Section III).
2) A comprehensive objective evaluation of the proposed Bilateral Ambisonics reproduction, providing insight into the effect of low order reproduction on errors in the binaural signals (Section IV).
3) Validation of the objective results by a listening test comparing the proposed method with the state-of-the-art Ambisonics reproduction method using MagLS (Section V).

The paper is arranged as follows: first, the mathematical formulation for reproducing Basic Ambisonics binaural signals is presented (Section II). Then, the formulation of Bilateral Ambisonics reproduction is developed (Section III). Objective and subjective evaluations of the suggested method are presented in Secs. IV and V, respectively. Section VI discusses the results presented in the paper and Section VII outlines the conclusions of the research.

II. Basic Ambisonics Reproduction

This section provides an overview of the currently used formulation for binaural reproduction using SH representation. Consider a sound-field composed of a continuum of PWs, described by a PW density function, $a(k, \Omega)$, where $\Omega \equiv (\theta, \phi) \in S^2$ is the spatial angle, represented by the elevation angle $\theta \in [0, \pi]$, which is measured downwards from the Cartesian z-axis, and the azimuth angle $\phi \in [0, 2\pi)$, which is measured counterclockwise from the Cartesian x-axis in the xy-plane. $k = 2\pi f/c$ is the wave number, f is the frequency, and c is the speed of sound. The binaural signal, which is the pressure observed at each of the listener's ears, can be calculated for the left ear by [7], [15]:

$$p^L(k) = \int_{\Omega \in S^2} a(k, \Omega) h^L(k, \Omega) \mathrm{d}\Omega, \tag{1}$$

where $h^L(k, \Omega)$ is the left ear HRTF [3]. The superscript L denotes the left ear (the formulation for the right ear can be similarly presented), $p^L(k)$ is the pressure at the ear and $\int_{\Omega \in S^2}(\cdot)\mathrm{d}\Omega \equiv \int_0^{2\pi}\int_0^{\pi}(\cdot)\sin(\theta)\mathrm{d}\theta\mathrm{d}\phi$.

Alternatively, Rafaely and Avni [9] showed that, by applying Parseval's relation to (1), the binaural signal can be calculated in the SH domain [9] using the Basic Ambisonics reproduction:

$$p^L(k) = \sum_{n=0}^{\infty}\sum_{m=-n}^{n} [\tilde{a}_{nm}(k)]^* h^L_{nm}(k), \tag{2}$$

where $\tilde{a}_{nm}(k)$ is the spherical Fourier transform (SFT) of $[a(k, \Omega)]^*$, $[\cdot]^*$ denotes the complex conjugate, and $h^L_{nm}(k)$ are the SH coefficients of the left ear HRTF, which can be computed by applying the SFT to the HRTF, $h^L(k, \Omega)$. $\tilde{a}_{nm}(k)$, which are the Basic Ambisonics signals, can be calculated by capturing the sound-field using a spherical microphone array and applying PW decomposition in the SH domain [13], [34]. For spatially band-limited functions, $\tilde{a}_{nm}(k)$ and $h^L_{nm}(k)$ will only be available up to orders N_a and N_h, respectively, and, therefore, the infinite summation in Eq. (2) will be truncated to order $N = \min(N_a, N_h)$,

$$p^L(k) = \sum_{n=0}^{N}\sum_{m=-n}^{n} [\tilde{a}_{nm}(k)]^* h^L_{nm}(k). \tag{3}$$

In practice, when $\tilde{a}_{nm}(k)$ is derived from spherical microphone array recordings, its order will be limited by the number of microphones [35]. For example, by capturing a sound-field using the Eigenmike [36], which is a spherical microphone array with a radius of 4.2 cm and 32 microphones, the maximum order is about $N_a = 4$. Similar order limitation may also be introduced for a simulated sound-field in practical applications. However, the HRTF is inherently of high spatial order. Zhang *et al.* [16] showed that for physically accurate representation up to 20 kHz, an order of above $N_h = 40$ is required. Therefore, in this practical scenario, the HRTF will be truncated to order $N = 4$. This order truncation affects both the spatial and the spectral characteristics of the binaural signal, which has been shown to have a detrimental effect on the perceived spatial sound quality [17], [18].

III. Proposed Bilateral Ambisonics Reproduction

One approach for reducing truncation errors is to pre-process the HRTF in order to reduce its effective SH order. Several recent studies suggested methods to pre-process the HRTF in the context of SH interpolation of HRTFs. For example, by time-alignment [22], [24], using minimum-phase representation [37], using directional equalization [38], or by ear-alignment [31], [39]. All these methods are based on manipulating the linear-phase component of the HRTF, which makes a major contribution to the high-order nature of the HRTF [22]. While these pre-processing methods lead to efficient sampling of HRTFs, e.g. [31], incorporating the pre-processed HRTF in the computation of the binaural signal, as in Eq. (3), is not immediate due to the misalignment between the phase components of the HRTF and the sound-field. In other words, the HRTF and the Ambisonics signals in (3) must be represented in the same coordinate system and around the same origin. One way to align the two is to re-synthesize the HRTF phase before the computation of the binaural signal, which will increase its order back to the original high-order, and will cause similar truncation error to that in the original reproduction. Another way is to use the MagLS approach, which completely ignores the HRTF phase component at high frequencies [25]. An alternative way to align the two is to use the ear-aligned HRTFs together with Ambisonics signals which are also defined around each of the listener's ears, rather than around the center of the head. This approach was originally presented in 1998 by Jot *et al.* [32], and was denoted as Binaural B-Format. They suggested to use two B-Format recordings (e.g. by using two 4-channel sound-field microphones [40] at the ear locations) together with a minimum-phase approximation of the HRTF and an interaural time difference (ITD) estimation based on a spherical head model. The Binaural B-format can be extended such that high order Ambisonics signals are defined around the ear locations (the approach denoted here as Bilateral Ambisonics [41]). Methods to obtain such Bilateral Ambisonics signals and the practical application of this approach are discussed in Section VI.

Ear-alignment has been shown to be a robust method for reducing the effective SH order of the HRTF while preserving the HRTF phase information and the ITD [31]. The alignment is performed by translating the origin of the free-field component of the HRTF from the center of the head to the position of the ear. For the left ear, this is formulated by:

$$h_a^L(k,\Omega) = h^L(k,\Omega)e^{-ikr_a\cos\Theta_L}, \tag{4}$$

where $h_a^L(k,\Omega)$ is the ear-aligned HRTF, r_a is the radius of the head, Θ_L is the angle between the direction of the source, Ω, and the direction of the ear, Ω_L, and $\cos\Theta_L = \cos\theta\cos\theta_L + \cos(\phi-\phi_L)\sin\theta\sin\theta_L$. Note that the ear-aligned HRTF can be computed for any given HRTF (it is assumed that the HRTF is normalized by a free-field measurement at the origin).

Now, assuming that the PW density function is given at the position of the ear, denoted by $a^L(k,\Omega)$, then the binaural signal can be computed directly at the listener's ear, using the ear-aligned HRTF, by:

$$p^L(k) = \int_\Omega a^L(k,\Omega)h_a^L(k,\Omega)\mathrm{d}\Omega. \tag{5}$$

In the SH domain, truncated to order N (similar to the derivation of Eq. (3) from Eq. (1)), the Bilateral Ambisonics reproduction is defined as:

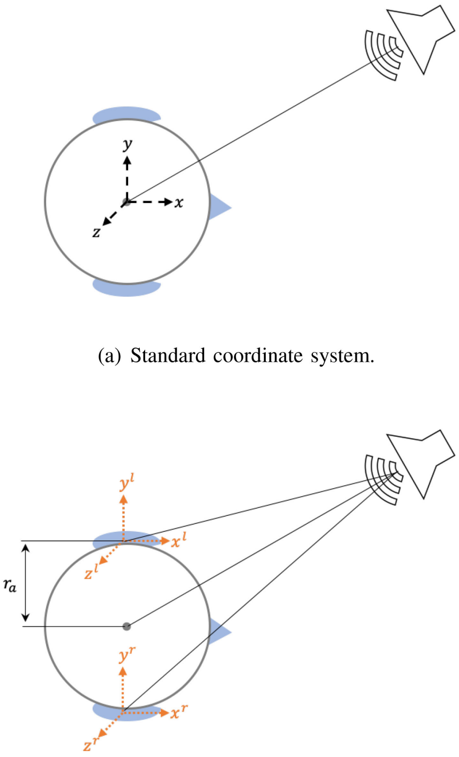
$$p^L(k) = \sum_{n=0}^{N}\sum_{m=-n}^{n}[\tilde{a}_{nm}^L(k)]^* h_{a\,nm}^L(k), \tag{6}$$

where $\tilde{a}_{nm}^L(k)$ and $h_{a\,nm}^L(k)$ are the SH coefficients of $[a^L(k,\Omega)]^*$ and $h_a^L(k,\Omega)$, respectively. A similar computation should also be performed for the right ear. It is important to note that $a^L(k,\Omega)$ is the PW density function of the sound-field as it is observed at the position of the left ear, in contrast to $a(k,\Omega)$, which is observed at the position of the center of the head. Fig. 1 illustrates the differences between the two coordinate systems. The standard coordinate system, denoted by black dashed axes with its origin at the center of the head, is used for the computations of the binaural signals in Eqs. (1) and (3) using the Basic Ambisonics signals, $\tilde{a}_{nm}(k)$, for both ears. The Bilateral coordinate systems, denoted by orange dotted axes with their origin at the positions of the ears, are used for the computation in Eq. (6) using the Bilateral Ambisonics signals, $\tilde{a}_{nm}^L(k)$ and $\tilde{a}_{nm}^R(k)$ for the left and right ears, respectively.

Theoretically, $a^L(k,\Omega)$ can be computed from $a(k,\Omega)$ by translation of the sound-field [35], which can be computed as $a^L(k,\Omega) = a(k,\Omega)e^{ikr_a\cos\Theta_L}$; however, in this case Eq. (5) is equivalent to Eq. (1), and, therefore, the binaural signals will be identical, i.e. the same truncation error as presented in Section II for the Basic Ambisonics reproduction is introduced. Alternatively, if a low-order PW density function is given directly at the position of the ear, the Bilateral Ambisonics-based signal, as in Eq. (6), potentially be more accurate than the Basic Ambisonics reproduction (as in Eq. (3)) because of the lower-order nature of the ear-aligned HRTF compared to an unprocessed HRTF.

IV. Objective Evaluation

This section presents an objective evaluation of the performance of the proposed Bilateral Ambisonics reproduction approach. The evaluation is performed by comparing the suggested method, as in Eq. (6), with the Basic Ambisonics reproduction method, as in Eq. (3).



(a) Standard coordinate system.

(b) Bilateral coordinate system.

Fig. 1. Illustration of the standard (a) and Bilateral (b) coordinate systems. The origin of the standard coordinate system is at the center of the head, while in the Bilateral coordinate system the origin is at the position of the ear.

A. *Performance Measures*

Four objective performance measures are considered in this analysis, as detailed below: (i) Normalized Mean Square Error (NMSE), (ii) Energy Difference (ED), (iii) Interaural Time Difference (ITD), and (iv) Interaural Level Difference (ILD).

i) *NMSE:* The NMSE for the left ear is computed as:

$$\epsilon^L(f) = 10\log_{10}\frac{|p^L_{\text{ref}}(f) - p^L(f)|^2}{|p^L_{\text{ref}}(f)|^2}, \tag{7}$$

where p^L_{ref} is the reference high-order binaural signal computed as in Eq. (3) with $N = 41$, and p^L is the binaural signal computed according to the evaluated reproduction method by (3) or (6). The NMSE measures the overall error between the two signals, including both the magnitude and the phase components.

ii) *ED:* The ED for the left ear is computed in 39 auditory filter bands as:

$$\Delta G(p^L_{\text{ref}}, p^L, f_c) = 10\log_{10}\frac{\int C(f, f_c)\left|p^L_{\text{ref}}(f)\right|^2 \text{df}}{\int C(f, f_c)\left|p^L(f)\right|^2 \text{df}}, \tag{8}$$

where C is a Gammatone filter with center frequency f_c, as implemented in the Auditory Toolbox [42]. The integral is evaluated between 50 Hz and 20 kHz and f_c is restricted accordingly. This measure is similar to the calculation of the internal cochlea spectrum suggested by Salomons [43]. The broadband ED measure is defined by averaging the ED across the 39 auditory filters as:

$$\Delta G_{av} = \frac{1}{39}\sum_{f_c}|\Delta G|. \tag{9}$$

iii) *ITD:* The ITD is an important spatial cue for sound localization, which may be affected by the truncation error [31]. The ITDs, as a function of direction, were computed for a single PW sound field with an incident angle, Ω, that varies over 500 directions, distributed uniformly across the left horizontal half-plane ($\theta = 90^\circ$; $0^\circ \leq \phi \leq 180^\circ$). The threshold detection method was used, with a threshold of -30 dB applied to a 3 kHz low-pass filtered version of the signals. This method has been shown by Andreopoulou and Katz [44] to be the most perceptually relevant procedure for ITD estimation. The ITD error as a function of direction is computed as follows:

$$\epsilon_{\text{ITD}}(\Omega) = |\text{ITD}_{\text{ref}}(\Omega) - \text{ITD}(\Omega)|, \tag{10}$$

where $\text{ITD}(\Omega)$ is the ITD of the evaluated reproduction method, and ITD_{ref} is the ITD of the reference signal, $p^{L,R}_{\text{ref}}$.

iv) *ILD:* Another important cue in binaural rendering and sound localization is the ILD. The ILDs were computed for the same 500 binaural signals as for the ITD, each composed of a single PW with an incident angle Ω, in 39 auditory filter bands as [5]:

$$\text{ILD}(f_c, \Omega) = 10\log_{10}\frac{\int C(f, f_c)\left|p^L(f)\right|^2 \text{df}}{\int C(f, f_c)\left|p^R(f)\right|^2 \text{df}}. \tag{11}$$

This computation facilitates a perceptually motivated smoothing of the ILD across frequencies, which is required for appropriate comparison between ILDs. The ILD error as a function of both center frequency and direction is computed as follows:

$$\epsilon_{\text{ILD}}(f_c, \Omega) = |\text{ILD}_{\text{ref}}(f_c, \Omega) - \text{ILD}(f_c, \Omega)|, \tag{12}$$

where ILD_{ref} is the ILD for the reference signal, $p^{L,R}_{\text{ref}}$. The ILD and the ILD error were averaged across auditory filter bands by:

$$\text{ILD}_{av}(\Omega) = \frac{1}{39}\sum_{\text{f}_\text{c}}\text{ILD}(f_c, \Omega), \tag{13}$$

$$\epsilon_{\text{ILD}_{av}}(\Omega) = \frac{1}{39}\sum_{\text{f}_\text{c}}\epsilon_{\text{ILD}}(f_c, \Omega). \tag{14}$$

B. *HRTF of a Rigid Sphere and a Manikin*

The measures defined in the previous section were calculated for binaural signals composed of two different HRTFs, as detailed below: (i) rigid sphere approximation, (ii) simulated HRTF of KEMAR [45].

i) *Rigid Sphere:* As a first evaluation, the HRTF of an ideal rigid sphere is considered. This simplification makes it possible to analyze mathematically the binaural signal, and to evaluate it analytically [46].

The "left ear" HRTF of an ideal rigid sphere of radius r_a can be computed as [13]:

$$h^L(\Omega,k)=\sum_{n=0}^{\infty}\sum_{m=-n}^{n} b_n(kr_a)[Y_n^m(\Omega)]^* Y_n^m(\Omega_L), \tag{15}$$

where $Y_n^m(\Omega)$ is the complex SH function of order n and degree m, and b_n is given for a rigid sphere as follows:

$$b_n(kr_a)=4\pi i^n\left[j_n(kr_a)-\frac{j_n'(kr_a)}{h_n'(kr_a)}h_n(kr_a)\right], \tag{16}$$

where j_n and h_n are the spherical Bessel and Hankel functions, and j_n' and h_n' are their derivatives.

The rigid sphere HRTF was computed for a sphere with radius $r_a = 8$ cm, and for a position of the ear at $(\theta,\phi)=(90^\circ,90^\circ)$, for a total of $Q = 7396$ directions in accordance with an Equal-Angle sampling scheme [35], which can provide an HRTF up to a SH order of 42. The ear-aligned HRTF was computed as in Eq. (4), and its SH coefficients, $h^L_{a\,nm}(k)$, were computed using the SFT.

ii) *KEMAR:* Next, for a more realistic evaluation, an HRTF of KEMAR was used for the computation of the binaural signals. The HRTF was simulated using the Boundary Element Method (BEM) based on a 3D scan of the head of a KEMAR. A total of $Q = 4334$ directions were simulated in accordance with a Lebedev sampling scheme $(N = 56)$. For detailed information about this HRTF data set see Refs. [23], [47].

C. Binaural Signals for a Single PW Sound-Field

For an initial evaluation of the binaural signals, a simple sound-field composed of a single PW in free-field is considered. In this case, a^{PW}_{nm}, the Basic Ambisonics signals of this unit amplitude PW, arriving from direction Ω_0 relative to the center of the head, are given by [35]:

$$a^{\mathrm{PW}}_{nm}(k)=[Y_n^m(\Omega_0)]^*. \tag{17}$$

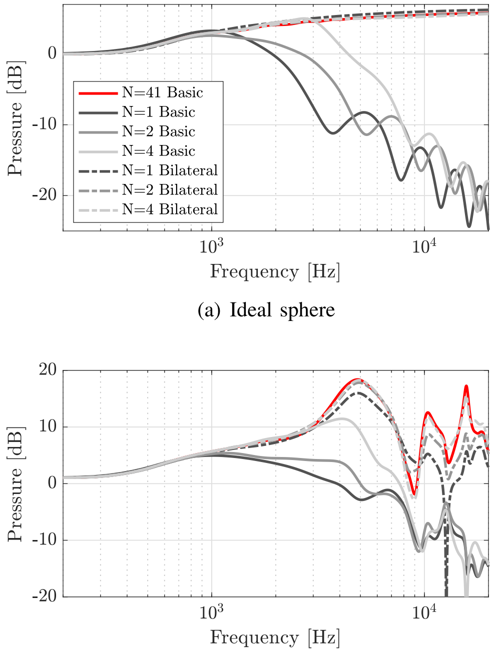
The SH coefficients of the same sound-field, but this time with the origin of the coordinate system aligned to the ear position, compose the Bilateral Ambisonics signals, and can be generally expressed by:

$$a^{L\,\mathrm{PW}}_{nm}(k)=A(\Omega_0,\Omega_L)[Y_n^m(\Omega_0)]^*, \tag{18}$$

where $A(\Omega_0,\Omega_L)$ is the complex amplitude of the PW. This amplitude is given by $A(\Omega_0,\Omega_L)=e^{ikr_a\cos\Theta_{L0}}$, where Θ_{L0} is the angle between Ω_L, the ear position, and Ω_0, the wave arrival direction.

The binaural signals for this PW sound-field can now be computed using the Basic Ambisonics reproduction of order N, as in Eq. (3), with a_{nm} and h_{nm} relative to the center of the head by:

$$p^L(k)=\sum_{n=0}^{N}\sum_{m=-n}^{n}[\tilde{a}^{\mathrm{PW}}_{nm}(k)]^* h^L_{nm}(k), \tag{19}$$



(a) Ideal sphere

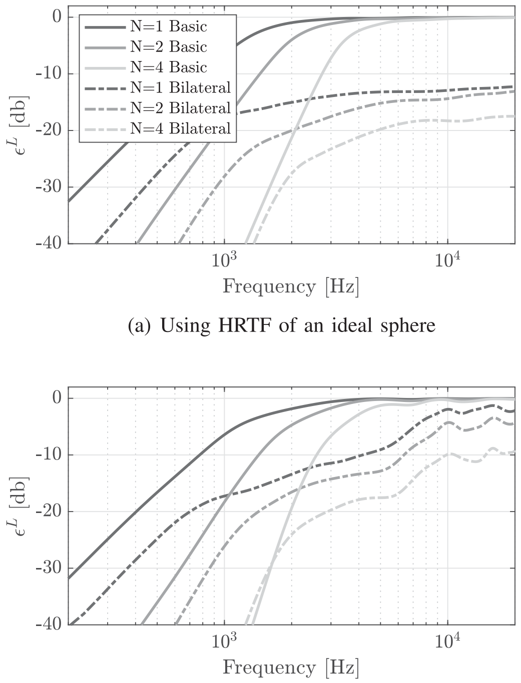
(b) KEMAR

Fig. 2. Magnitude of a left ear binaural signal of a single PW from direction $(\theta,\phi)=(90^\circ,40^\circ)$, with HRTF of an ideal rigid sphere (a), and KEMAR (b). Computed with Basic Ambisonics reproduction (solid lines) and with Bilateral Ambisonics reproduction (dashed lines), with $N = 1, 2, 4$, compared to a high-order reference with $N = 41$.

and using the Bilateral Ambisonics reproduction of order N, as in Eq. (6), with a^L_{nm} and $h_{a\,nm}$ relative to the position of the ear, by:

$$p^L(k)=\sum_{n=0}^{N}\sum_{m=-n}^{n}[\tilde{a}^{L\,\mathrm{PW}}_{nm}(k)]^* h^L_{a\,nm}(k). \tag{20}$$

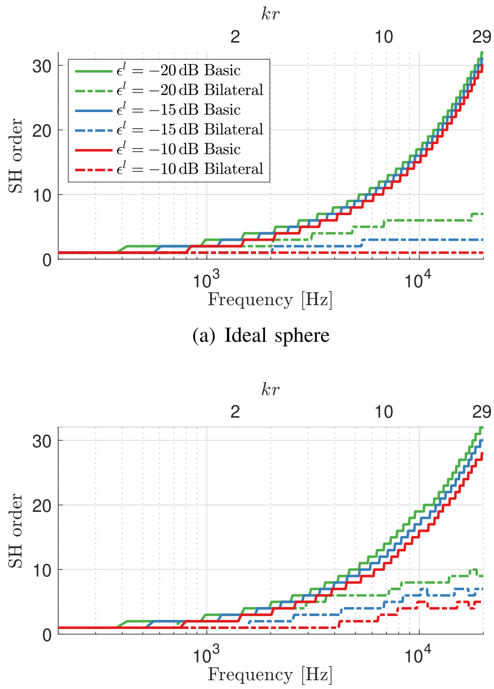
Fig. 2 shows the magnitude response of the binaural signals, using both rigid sphere and KEMAR HRTFs, computed with Basic Ambisonics reproduction (Eq. (19)) and with Bilateral Ambisonics reproduction (Eq. (20)), with $N = 1, 2, 4$, compared to the high-order reference of $N = 41$ (computed using Eq. (19)), for a PW arrival direction $(\theta,\phi)=(90^\circ,40^\circ)$. Note that for this simple case of a single-PW sound-field, the resulting signal is the truncated version of the HRTF from this direction. High frequency roll-off, as described in [18], is clearly observed for the low-order signals computed using the Basic Ambisonics reproduction. Furthermore, amplitude distortion is also observed at these high frequencies, above the sphere cut-off frequency, $kr_a = N$ [13]. The Bilateral Ambisonics-based signals seem to correct for both frequency roll-off and distortion. For the rigid sphere, the signal magnitude is very similar to the reference signal, even with an order of $N = 1$. For the KEMAR HRTF, using Bilateral Ambisonics reproduction of order $N = 4$ seems to preserve the signal magnitude up to almost 20 kHz. Note that the Bilateral Ambisonics-based signal also seems to preserve the important spectral cues (peaks and notches) at the high frequencies. It is important to note that while the results are shown only for the left ear, similar results were also observed



(a) Using HRTF of an ideal sphere

(b) Using HRTF of KEMAR

Fig. 3. NMSE of binaural signals computed for sound-fields composed of a single PW, averaged over 770 PW directions (distributed according to a Lebedev grid), with HRTF of an ideal rigid sphere (a), and KEMAR (b). The NMSE is computed using Eq. (7), with Basic Ambisonics reproduction (solid lines) and with Bilateral Ambisonics reproduction (dashed lines), with $N = 1, 2, 4$, and a high-order reference with $N = 41$.



(a) Ideal sphere

(b) KEMAR

Fig. 4. The minimum SH order to achieve NMSEs lower than -20 dB, -15 dB, -10 dB. The minimal order is computed for Basic Ambisonics reproduction (solid lines) and with Bilateral Ambisonics reproduction (dashed lines), and presented as a function of frequency and kr. The computed NMSEs were averaged across a range of single PW sound fields as in Fig. 3.

for the right ear. This also applies in the remainder of the paper.

D. NMSE Analysis

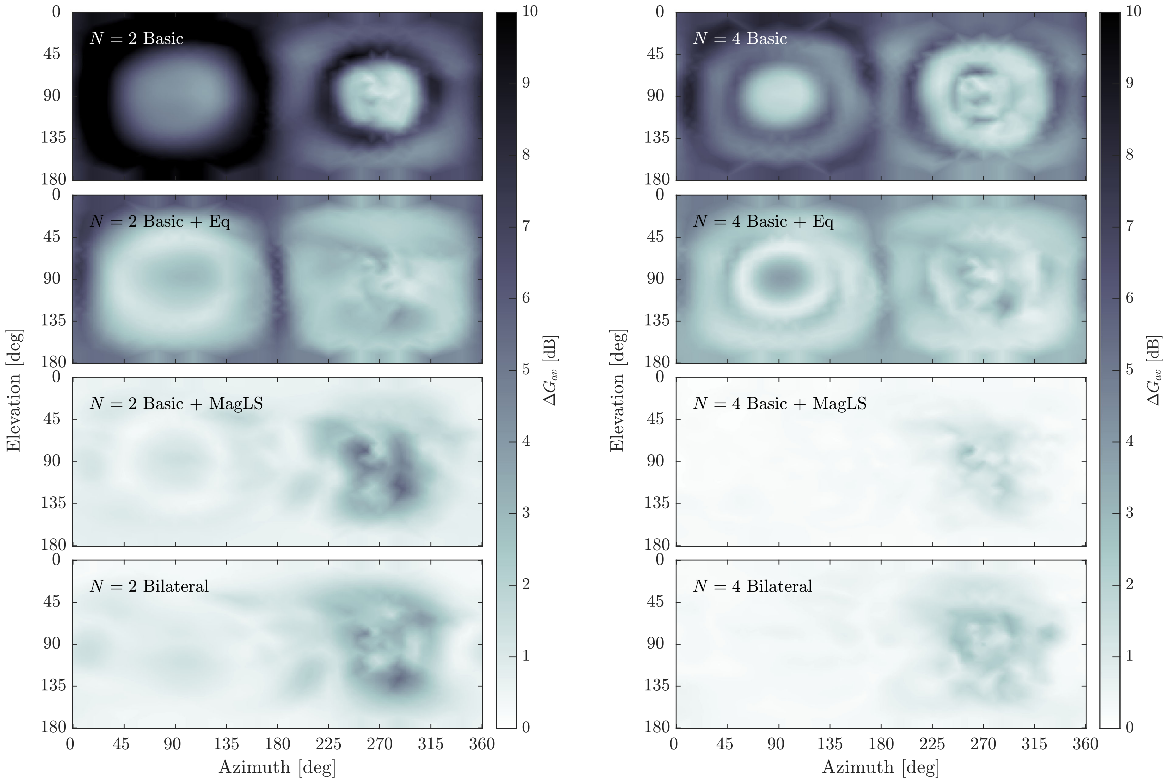
The evaluation presented in the previous section for a single PW is repeated here, but this time the NMSE as defined in Eq. (7) is computed and averaged over a range of PWs. 770 single PWs with incidence angles distributed nearly-uniformly over the sphere, using the Lebedev sampling scheme of order 23, were used for the averaging. The NMSE for both evaluated reproduction methods, as in Eqs. (19) and (20), was computed using the high-order reference, $p^L_{\text{ref}}(f)$, with $N = 41$.

Fig. 3 presents the average NMSE, with Basic and Bilateral Ambisonics reproductions. The figure shows that the error increases at frequencies above the cut-off frequency, where for the signal with the Basic Ambisonics reproduction the error approaches 0 dB. The Bilateral Ambisonics-based signals produce a substantially smaller error. With an HRTF of an ideal sphere the error is below -10 dB up to 20 kHz, even with order $N = 1$. Using the KEMAR HRTF, the error is less than -10 dB up to around 4 kHz for $N = 1$, and up to 15 kHz for $N = 4$.

Fig. 4 shows the minimum SH order required to guarantee NMSE as in Eq. (7) of less than a value of -20 dB, -15 dB and -10 dB, as a function of frequency. The figure demonstrates the significant reduction in the required SH order for achieving a target error level when using Bilateral Ambisonics reproduction. For the rigid sphere the required order is reduced from about $N = 30$ in the Basic Ambisonics reproduction, down to $N = 3$ in the Bilateral Ambisonics reproduction, for an error of -15 dB. For the KEMAR HRTF, the orders similarly decrease, from about $N = 30$ in Basic Ambisonics reproduction, down to $N = 7$, for an error of -15 dB over the entire audible frequency bandwidth.

E. ED Analysis

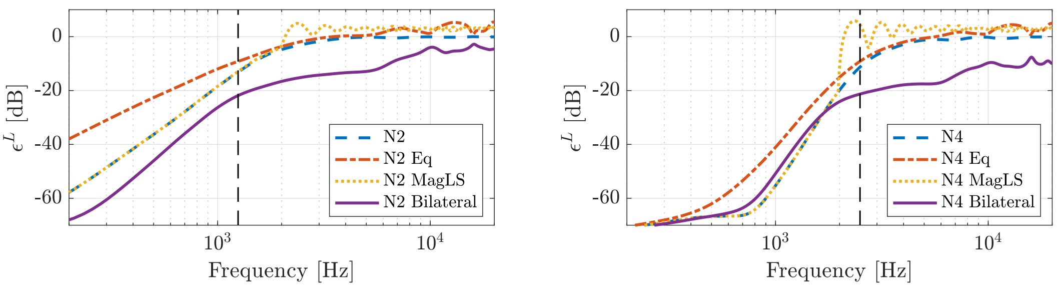
The broadband ED, as defined in Eq. (9), was computed for both evaluated reproduction methods with SH order $N = 4$ relative to the reference signal ($N = 41$), for a sound-field similar to that presented in IV-D, separately for each one of the 770 single-PW sound-fields. In addition, the ED for the Basic Ambisonics reproduction with equalization and tapering, as suggested in [21], was also computed, as it was reported to reduce reproduction errors for low-order binaural signals with similar perceptual performance as other suggested methods [30]. The results of the Basic Ambisonics reproduction using the MagLS approach are also presented. The MagLS HRTF was computed as described in [48] with a cut-off frequency of 2 kHz, as suggested in [25]. Fig. 5 shows the broadband ED as a function of the PW incident angle, for $N = 2$ and $N = 4$. The figure shows that the EDs across PW incident angles are significantly reduced with the Bilateral Ambisonics reproduction compared to the Basic Ambisonics reproduction method; similar outcomes



(a) $N = 2$

(b) $N = 4$

Fig. 5. Broadband ED, as in Eq. (9), as a function of PW arrival direction, for order truncated ((a) $N = 2$, (b) $N = 4$) left-ear binaural signals, relative to a high-order ($N = 41$) reference, calculated according to Eq. (19) using Basic Ambisonics reproduction (top), Basic Ambisonics reproduction with equalization [21] (second row), Basic Ambisonics reproduction with MagLS [25] (third row), and according to Eq. (20) using Bilateral Ambisonics reproduction (bottom), with HRTF of KEMAR. Color represents difference in dB.



(a) $N = 2$

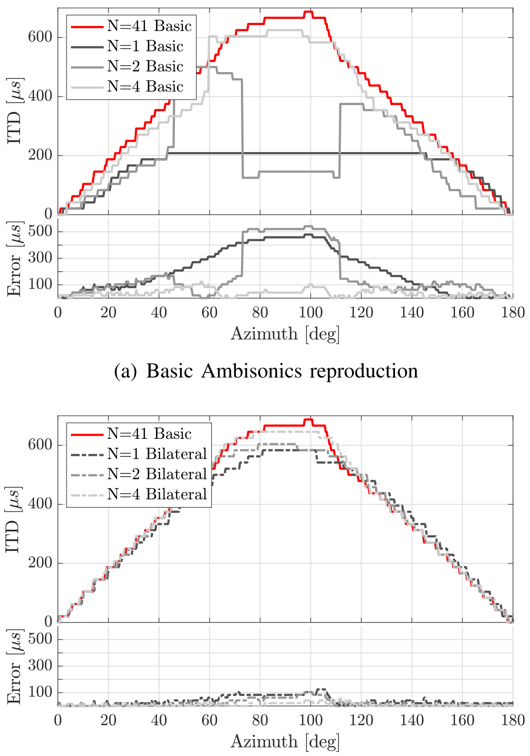
(b) $N = 4$

Fig. 6. NMSE of binaural signals computed for sound-fields composed of a single PW, averaged over 770 PW directions (distributed according to a Lebedev grid), with HRTF of KEMAR. The NMSE is computed using Eq. (7), with Basic Ambisonics reproduction, Basic Ambisonics reproduction with equalization [21], Basic Ambisonics reproduction with MagLS [25], and with Bilateral Ambisonics reproduction (dashed lines), with (a) $N = 2$ and (b) $N = 4$, and a high-order reference with $N = 41$. The dashed vertical line represents the cut-off frequency at $kr = N$.

are also found for the MagLS approach. With both the Bilateral Ambisonics reproduction and the MagLS, the differences at the ipsilateral incident angles ($0^\circ \leq \phi \leq 180^\circ$) are nearly 0 dB for $N = 4$, compared to 5 dB for the equalized signal, and less than 1.5 dB for $N = 2$, compared to 7 dB for the equalized signal. Some larger differences are observed at the contralateral incident angles ($180^\circ \leq \phi \leq 360^\circ$), with up to 3 dB for $N = 4$ and up to 5 dB for $N = 2$.

Fig. 6 shows the NMSE for the same binaural signals, computed using Eq. (7) averaged across all directions. The figure demonstrates the improvement in the accuracy of the Bilateral Ambisonics reproduction, compared to the Basic Ambisonics reproduction methods, where at high frequencies, above the cut-off frequency at $kr = N$ and up to about 6 kHz for $N = 2$ and 15 kHz for $N = 4$, the errors are lower by 10-20 dB.

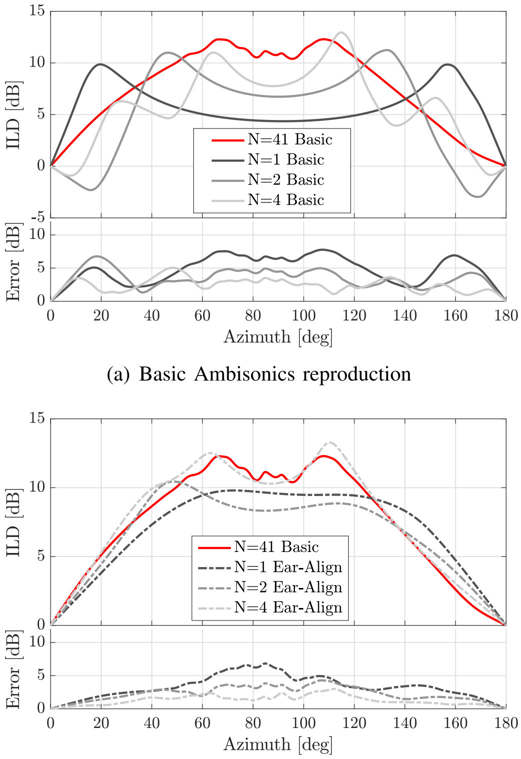
These ED and NMSE results suggest that the main advantage of the Bilateral Ambisonics reproduction compared to the state-of-the-art Ambisonics reproduction with MagLS is its ability to preserve the binaural signal phase information with Ambisonics of low SH order.



(a) Basic Ambisonics reproduction

(b) Bilateral Ambisonics reproduction

Fig. 7. ITDs and ITD errors as a function of azimuth angle, calculated as described in Section IV-A, for binaural signals computed for sound-fields composed of a single PW from 500 directions on the left horizontal plane (the right side is symmetrical), with HRTF of KEMAR. The signals were computed using (a) the Basic Ambisonics reproduction, as in Eq. (19), and (b) the Bilateral Ambisonics reproduction, as in Eq. (20).



(a) Basic Ambisonics reproduction

(b) Bilateral Ambisonics reproduction

Fig. 8. ILDs and ILD errors as a function of azimuth angle, calculated as described in Section IV-A using Eq. (13), for binaural signals computed for sound-fields composed of a single PW from 500 directions on the left horizontal plane (the right side is symmetrical), with HRTF of KEMAR. The signals were computed using (a) the Basic Ambisonics reproduction, as in Eq. (19), and (b) the Bilateral Ambisonics reproduction, as in Eq. (20).

F. *ITD and ILD Analysis*

The ITD and ITD error, as defined in IV-A-(iii) and Eq. (10), were computed for both evaluated reproduction methods with low-orders ($N = 1, 2, 4$), and relative to the high-order reference ($N = 41$), for a single PW with an incident angle that varies over 500 directions, distributed uniformly across the left horizontal plane ($\theta = 90^\circ$; $0^\circ \leq \phi \leq 180^\circ$). The HRTF of KEMAR was employed in this case (and not the ideal sphere), to better model human hearing. Fig. 7 shows the ITDs and the ITD error, as a function of azimuth angle, for both the Basic and the Bilateral Ambisonics-based signals. It may be interesting to compare these errors to just notable differences (JND) values. As reported by Andreopoulou and Katz [44], the JND for the ITD ranges between about 40 μ s (for the frontal direction), and about 100 μ s (for the lateral directions). The figure shows that with Basic Ambisonics reproduction, ITD error is below the JND only with $N = 4$. Surprisingly, with Bilateral Ambisonics reproduction, ITD error below JND is achieved already at $N = 1$. This clearly shows the superiority of the Bilateral Ambisonics reproduction with respect to this important cue.

The ILD and ILD error, as defined in Eqs. (13) and (14), were computed in a similar manner, i.e. for both evaluated reproduction methods with low-orders and relative to the high-order reference, and for the same 500 PWs and HRTFs. Fig. 8 shows the ILD and ILD error. The figure shows that with Basic Ambisonics reproduction, ILD errors are above the JND, which is shown to be around 1 dB [49], [50], even with $N = 4$. With Bilateral Ambisonics reproduction, the errors for $N = 4$ are below the JND for most angles. The relatively high errors at the lateral angles compared to the frontal and backward angles are expected because the ILD in the front and back is close to zero, and, because the HRTF model is symmetric, the errors are expected to be small at these angles. Nevertheless, the Bilateral Ambisonics reproduction led to substantially lower ILD errors compared to the Basic Ambisonics reproduction.

V. Subjective Evaluation

The objective results presented in Section IV, suggest that using the proposed Bilateral Ambisonics reproduction method to reproduce binaural signals in the SH domain with low SH orders achieves better performance compared to the Basic Ambisonics reproduction. With the aim of validating these results perceptually, a listening test was conducted.

A. *Methodology*

Room impulse responses in a rectangular room of dimensions $8 \times 6 \times 4$ m were simulated using the image method [51]. Room boundaries were characterized with a reflection coefficient of 0.8, leading to a reverberation time of $T_{60} = 0.53$ s and a critical distance $r_{cd} = 1.07$ m. A simulated spherical microphone array was positioned at $(x, y, z) = (4, 3, 1.7)$ m in the room, and an omni-directional source at $(5.2, 3.7, 1.7)$ m, which is a distance of 1.4 m and an angle of 30° to the left of the array, relative to the HRTF coordinate system. For the Bilateral Ambisonics reproduction, two simulated arrays were placed at the position of the ears, as described in Fig. 1, with $r_a = 8.75$ cm. The PW density functions, $a_{nm}(k)$ and $a_{nm}^{L/R}(k)$, were computed directly in the SH domain using the data on image sources provided by the image method and the source signal. These were computed at the desired SH order, such that no spatial aliasing is introduced in the simulation.

The HRTF set used in this experiment is the measured HRTF from the Cologne HRTF database for the Neumann KU100 dummy head [52]. The ear-aligned HRTF was computed as in Eq. (4) with the parameters $r_a = 8.75$ cm and the positions of the ears at $\theta = 90°$, $\phi = 90°, 270°$ for the left and right ears, respectively.

The binaural signals were computed using the Basic Ambisonics reproduction in the SH domain, as described in Section II, Eq. (3), with MagLS as outlined in Eqs. (4.57-4.59) in [48] using a cutoff frequency of 2 kHz, as suggested in [25], and using the suggested Bilateral Ambisonics reproduction, as described in Section III, Eq. (6). All signals were convolved with matching headphone compensation filters, taken from the Cologne database [52], which were measured on the Neumann KU100 dummy head.

Three SH orders were selected for perceptual testing based on the results of the objective analysis and on informal listening, $N = 1$, 2 and 4, for both the MagLS and Bilateral Ambisonics reproductions. These represent low and mid-range order Ambisonics. A high-order reference, rendered using the Basic Ambisonics reproduction with $N = 41$, was used. Two types of audio source signals were used: castanets and English male speech. The castanets were chosen because of their high frequency content and strong transients. The speech was included as typical real life content for binaural reproduction applications. The loudness of all signals was equalized to the same level [53].

13 normal hearing subjects (2 females, 11 males, aged 27–50, mean age 34.8), with experience in critical listening to spatial audio content, participated in a multiple stimuli with hidden reference and anchor (MUSHRA) listening experiment [54]. The experiment comprised four separate tests, two for each audio source signal, with seven test signals in each test (2 reproduction methods × 3 orders + a hidden reference).[1] An explicit anchor was not included to avoid the consequences of using an inappropriate anchor that may widen the ranking range, thus unnecessarily compressing the differences between test signals. The experiment was developed in Matlab [55] and was shared with the participants using a shared folder. Each participant performed the experiment using their own computer and headphones with an appropriate headphone compensation filter from the Cologne database [52]. The participants were instructed to run the experiment in as quiet an environment as possible.

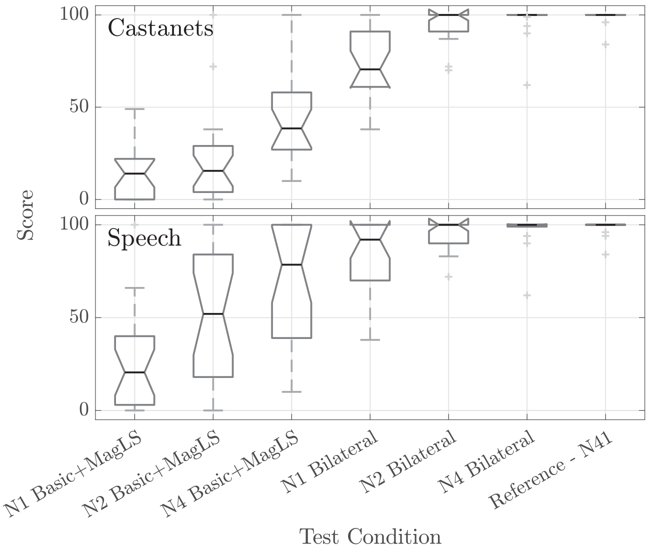


Fig. 9. Box plot of listening test scores. Each row represents a different audio source signal. each box represents a different test condition. The box bounds the interquartile range (IQR) divided by the median, and Tukey-style whiskers extend to a maximum of 1.5×IQR beyond the box. The notches refer to a 95% confidence interval [57].

At the beginning of the experiment, a training session was conducted. The training comprised two parts: user interface training and signals familiarization. The purpose of the training session was to assist the participants in achieving the following: (i) user interface training: to learn how to use the test equipment and the grading scale; and (ii) signals familiarization: to become familiar with all the test signals and their quality level ranges. The participants were asked to rate the relative differences between the reference and the test signals on a scale from 0 to 100, where a score of 100 means that the signal is not differentiable from the reference, and values downwards towards 0 mean larger differences with respect to the reference. As both spatial and spectral artifacts are expected due to low order reproduction [17]–[19], the possible differences were described to the participants as either spatial artifacts, time varying artifacts or spectral artifacts, with the aim of evaluating the overall differences between the test signals. The participants were instructed not to move their head while listening, as no head tracking was available during playback.

B. Results

Listening test results for all conditions and 11 subjects are shown in Fig. 9 by means of standard box plots. 2 subjects were excluded from the results because in at least one of the tests they rated the hidden reference with a score lower than 80. Each row shows the results of a different audio source signal. Results indicate a clear perceptual improvement for the proposed Bilateral Ambisonics reproduction method compared to the Basic+MagLS Ambisonics reproduction. A three-factorial

[1] The signals are available for listening in the supplementary material, with the addition of other audio source signals

repeated measure ANOVA with the factors “SH order” ($N = 1$, 2, 4), “Reproduction” (Basic+MagLS, Bilateral) and “Audio source signal” (Castanets, Speech), paired with a Tukey-Kramer post hoc test at a confidence level of 95%, was used to determine the statistical significance of the results [54], [56]. The main effects for all factors are statistically significant ($F_{(2204)} = 74.1$, $F_{(1204)} = 162.66$ and $F_{(1204)} = 126.91$ for “SH order,” “Reproduction” and “Audio source signal,” respectively, with $p_{\text{val}} < 0.001$ for all of them). The interaction effect between the “SH order” and “Reproduction” is also statistically significant ($F_{(2204)} = 16.06$, $p_{\text{val}} < 0.001$), which means that the effect of SH order on the perceived difference under different reproduction methods is significantly different.

Following the main effect of the “Reproduction” factor, pairwise comparisons between the test signals for this factor reveal that with both audio source signals 1st order Bilateral Ambisonics reproduction achieved statistically significantly higher scores compared to 1st and 2nd order Basic+MagLS Ambisonics reproduction ($p_{\text{val}} < 0.001$). In addition, the median scores for the 1st order Bilateral Ambisonics reproduction are comparable to, or even higher than, the 4th order Basic+MagLS Ambisonics reproduction (70.5 compared to 38.5, $p_{\text{val}} < 0.001$, for Castanets and 92 compared to 78.5, $p_{\text{val}} = 0.21$, for speech).

Pairwise comparisons between the test signals and the reference reveal that with Bilateral Ambisonics reproduction the ratings of all tested signals (except for Castanets with $N = 1$) were not significantly different from the reference ($p_{\text{val}} > 0.31$). On the other hand, using the Basic+MagLS Ambisonics reproduction, orders $N = 1$ and 2 are significantly different from the reference in all audio source signals, and with Castanets, even order $N = 4$ is significantly different from the reference. This further demonstrate the perceptual benefits of Bilateral Ambisonics compared to Basic+MagLS Ambisonics.

VI. Discussion

Both objective and subjective results presented in this paper suggest that Bilateral Ambisonics reproduction can produce high quality virtual sounds using relatively low-order reproduction. This section will summarize the results presented in the paper and discuss their implications for practical use of the proposed method.

A. Results Summary

Section IV provided a comprehensive objective analysis that demonstrates the advantages of the proposed Bilateral Ambisonics reproduction method, in terms of reproduction errors and binaural cues preservation, over Basic binaural reproduction in the SH domain. The perceptual evaluation reported in Section V verified the results of the objective analysis and revealed that using Bilateral Ambisonics reproduction with SH orders as low as $N = 1$ has significant perceptual benefits over the Basic Ambisonics reproduction. The configuration with $N = 1$ can be seen as being equivalent to the Binaural B-Format configuration [32]. Results showed that the 1st order Bilateral Ambisonics reproduction achieved similar, or even higher, scores than those for the 4th order Basic+MagLS Ambisonics reproduction. Moreover, 2nd order Bilateral Ambisonics reproduction achieved similar scores to those of the high-order reference for all tested audio source signals.

The perceptual results obtained for the Basic+MagLS Ambisonics reproduction are consistent with those found in previous literature. In [18], the Basic Ambisonics reproduction with a spectral equalization filter was evaluated, and it was shown that for a speech signal in a simulated environment, participants in the experiment claimed perceptual indifference in the case of signals of order 6 and above, compared to a high-order reference. A recent study by Ahrens and Andersson [20] showed that for binaural reproduction from spherical microphone array recordings, order 8 or higher is required for reproducing virtual sound that is deemed perceptually indifferent compared to a ground truth dummy head measurement. Additionally, they reported that the room has no effect on the ratings, which suggests that the perceptual results presented in Section V can be generalized to other acoustic scenarios. However, they also concluded that the virtual sound source location has a significant effect on the perception at low orders, which is in agreement with the results presented in Secs. IV-E and IV-F, where the errors are shown to be direction dependent. Perception evaluation of the proposed method with different sound source locations is suggested for future work.

Recently, Zaunschirm *el al.* [24] proposed a method for low-order binaural reproduction in the SH domain using frequency-dependent time-aligned HRTFs followed by a diffuse-field optimization. As presented in [31], time-alignment of the HRTF has a similar influence to that of ear-alignment in terms of SH order reduction. However, in their time-aligned reproduction, the linear-phase component of the HRTF is completely removed at high frequencies, while in the proposed Bilateral Ambisonics reproduction method, presented in the current paper, the complex phase component of the HRTF is preserved over the entire frequency bandwidth. Moreover, the perceptual results presented in [24] led to the conclusion that, using their reproduction method, a SH order of $N = 5$ is required in order to achieve similar performance to that of a high-order reference.

A recent study by Lübeck *et al.* [30] presented a comprehensive comparison between different reproduction methods. They evaluated the perceived differences between a dummy-head recording, as the reference, and binaural signals reproduced from spherical microphone arrays of different orders. As part of thier subjective evaluation they compared the Basic Ambisonics reproduction with equalization and tapering (same as used in Section IV) and with the MagLS approach (same as used in Secs. IV and V), which is an advancement of the frequency-dependent time-alignment method [24] discussed above. They concluded that both methods perform similarly in improving the raw Basic Ambisonics reproduction. However, small differences compared to the reference are still observed up to order $N = 7$, with both methods. These results are in agreement with the results presented in the perceptual evaluation in Section V. Furthermore, a preliminary experiment, where the Basic Ambisonics reproduction with equalization and tapering was evaluated in comparison to the proposed Bilateral Ambisonics reproduction, showed similar results as with the MagLS approach.[2]

It is important to note here that the results presented in Section V for the MagLS approach, where the low order signals are perceived as significantly different than the reference (especially for the Castanets source signal), can be explained, to some extent, by the specific implementation of the approach. As described in [25], the implementation of the MagLS approach requires an optimization process for the estimation above the chosen cutoff frequency. The implementation performed in Section V followed the steps described in [48], which eliminates the need for optimization by an iterative process that smooths the HRTF phase above the cutoff frequency. Although this results in a smoothed phase, the process still introduces errors around the cutoff frequency (as can be seen in Fig. 6). These errors cause audible artifacts, that were described by some of the participants as "high-frequency beeping or hissing." These artifacts are more noticeable in the Castanets signal due to its high energy around the chosen cutoff frequency of 2 kHz, and might be also emphasized by the transients in the stimulus. Choosing a different cutoff frequency tailored to a specific signal, reduce some of the audible artifacts, but may also introduce new artifacts. In this work the MagLS as published in [25] and [48] was employed, while the design of signal-tailored MagLS is beyond the scope of this paper and is left for future research.

B. Application of the Proposed Method

The proposed method could be relevant for a wide range of binaural reproduction applications. While the SH coefficients of the ear-aligned HRTF can be readily computed from a typical HRTF [31], the SH coefficients of the sound-field at the position of the ears is typically not available. In the case where the sound-field is generated using numerical simulations, the Bilateral Ambisonics signals can be directly simulated at the two ear positions. This will require twice the number of SH coefficients compared to Basic Ambisonics reproduction. However, the perceptual results presented in this paper show that Bilateral Ambisonics reproduction using only the 1st order is comparable to, or even better than, Basic Ambisonics reproduction using the 4th order (in the sense of lower difference compared to a high order reference). Therefore, the number of required coefficients for Bilateral Ambisonics reproduction of the same (or better) quality compared to Basic Ambisonics could be smaller by about 70%, i.e. 8 coefficients representing the 1st order at the two ears, compared to 25 coefficients representing the Basic Ambisonics 4th order.

In the case of binaural reproduction based on sound-field measurements, the Bilateral Ambisonics signals can be derived from two microphone array measurements at the positions of the ears. For 1st order recordings, as suggested by Jot *et al.* [32], one can use two 4-channel sound-field microphones [40], and for 3 rd or 4th order recordings, one can use the 32-channel Eigenmike [36]. Although requiring two microphone arrays may complicate the recording procedure, it may lead to a more efficient representation of the spatial audio recording and to a more accurate binaural reproduction. It is important to note here that using such a procedure will result in a static binaural reproduction, i.e. incorporating head-tracking will require further processing, e.g. by using post-processing algorithms such as in [58]. Another possibility is to estimate the Bilateral Ambisonics signals from standard recordings at the center of the head, e.g. by sound-field translation [59]–[62]. The incorporation of the sound-field estimation and head-tracking processing algorithms in Bilateral Ambisonics reproduction is suggested for future work.

VII. Conclusion

The current paper presented a method for binaural reproduction in the SH domain based on Bilateral Ambisonics and ear-aligned HRTFs, which may be seen as an extension to the Binaural B-Format method. The proposed method has been shown to overcome the limitation imposed by low-order reproduction, where the main source of errors is the truncation of the HRTF. Using an ear-aligned HRTF together with a sound-field measured or simulated at the position of the ears, one can reproduce a low-order binaural signal with relatively small reproduction errors. Results of objective and subjective evaluations indicate that Bilateral Ambisonics reproduction with an order as low as $N = 1$ achieves similar results to the state-of-the-art 4th order Basic Ambisonics reproduction with MagLS, and that 2nd order Bilateral Ambisonics reproduction can reproduce virtual sounds that are highly similar to a high-order reproduction.

Acknowledgment

The authors would like to thank Tim Lübeck for his help in the implementation of the MagLS.

References

[1] D. R. Begault and L. J. Trejo, "3-D Sound for Virtual Reality and Multimedia," NASA, Ames Research Center, Mountain View, CA, USA, pp. 132–136, 2000.

[2] M. Vorländer, *Auralization: Fundamentals of Acoustics, Modelling, Simulation, Algorithms and Acoustic Virtual Reality*, Berlin, Germany: Springer Science & Business Media, 2007.

[3] J. Blauert, *Spatial Hearing: The Psychophysics of Human Sound Localization*, Cambridge, MA, USA: MIT Press, 1997.

[4] H. Møoller, "Fundamentals of binaural technology," *Appl. Acoust.*, vol. 36, no. 3, pp. 171–218, 1992.

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

[6] M. Brandstein and D. Ward, *Microphone Arrays: Signal Processing Techniques and Applications*. Berlin, Germany: Springer Science & Business Media, 2013.

[7] R. Duraiswami, D. N. Zotkin, Z. Li, E. Grassi, N. A. Gumerov, and L. S. Davis, "High order spatial audio capture and binaural head-tracked playback over headphones with HRTF cues," in *Proc. Audio Eng. Soc. Conf.: 119 AES Convention*, 2005, pp. 1–16.

[8] J. Sheaffer, M. Van Walstijn, B. Rafaely, and K. Kowalczyk, "Binaural reproduction of finite difference simulations using spherical array processing," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 23, no. 12, pp. 2125–2135, Dec. 2015.

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

[10] M. A. Gerzon, "Ambisonics in multichannel broadcasting and video," *J. Audio Eng. Soc.*, vol. 33, no. 11, pp. 859–871, 1985.

[2] The signals are available for listening in the supplementary material at http://www.ee.bgu.ac.il/~acl

[11] M. Jeffet, N. R. Shabtai, and B. Rafaely, "Theory and perceptual evaluation of the binaural reproduction and beamforming tradeoff in the generalized spherical array beamformer," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 24, no. 4, pp. 708–718, Apr. 2016.

[12] D. L. Alon and B. Rafaely, "Beamforming with optimal aliasing cancellation in spherical microphone arrays," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 24, no. 1, pp. 196–210, Jan. 2016.

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

[14] B. Rafaely, "Analysis and design of spherical microphone arrays, speech and audio processing," *IEEE Trans. Speech, Audio, Process.*, vol. 13, no. 1, pp. 135–143, Jan. 2005.

[15] M. Noisternig, A. Sontacchi, T. Musil, and R. H'oldrich, "A 3D ambisonic based binaural sound reproduction system," in *Proc. Audio Eng. Soc. Conf.: 24th Int. Conf.: Multichannel Audio, New Reality. Audio Eng. Soc.*, 2003, pp. 1–5.

[16] W. Zhang, T. D. Abhayapala, R. A. Kennedy, and R. Duraiswami, "Insights into head-related transfer function: Spatial dimensionality and continuous representation," *J. Acoust. Soc. Am.*, vol. 127, no. 4, pp. 2347–2357, 2010.

[17] A. Avni, J. Ahrens, M. Geier, S. Spors, H. Wierstorf, and B. Rafaely, "Spatial perception of sound fields recorded by spherical microphone arrays with varying spatial resolution," *J. Acoust. Soc. Am.*, vol. 133, no. 5, pp. 2711–2721, 2013.

[18] 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. Am.*, vol. 141, no. 6, pp. 4087–4096, 2017.

[19] 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, Mar. 2019.

[20] J. Ahrens and C. Andersson, "Perceptual evaluation of headphone auralization of rooms captured with spherical microphone arrays with respect to spaciousness and timbre," *J. Acoust. Soc. Am.*, vol. 145, no. 4, pp. 2783–2794, 2019.

[21] C. Hold, H. Gamper, V. Pulkki, N. Raghuvanshi, and I. J. Tashev, "Improving binaural ambisonics decoding by spherical harmonics domain tapering and coloration compensation," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2019, pp. 261–265.

[22] M. J. Evans, J. A. Angus, and A. I. Tew, "Analyzing head-related transfer function measurements using surface spherical harmonics," *J. Acoust. Soc. Am.*, vol. 104, no. 4, pp. 2400–2411, 1998.

[23] F. Brinkmann and S. Weinzierl, "Comparison of head-related transfer functions pre-processing techniques for spherical harmonics decomposition," in *Proc. Audio Eng. Soc. AES Int. Conf. Audio Virtual Augmented Reality*. 2018, pp. 1–10.

[24] M. Zaunschirm, C. Schörkhuber, and R. Höldrich, "Binaural rendering of ambisonic signals by head-related impulse response time alignment and a diffuseness constraint," *J. Acoust. Soc. Am.*, vol. 143, no. 6, pp. 3616–3627, 2018.

[25] C. Schörkhuber, M. Zaunschirm, and R. Höldrich, "Binaural rendering of ambisonic signals via magnitude least squares," in *Proc. DAGA*, vol. 44, 2018, pp. 339–342.

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

[27] E. A. Macpherson and J. C. Middlebrooks, "Listener weighting of cues for lateral angle: The duplex theory of sound localization revisited," *J. Acoust. Soc. Am.*, vol. 111, no. 5, pp. 2219–2236, 2002.

[28] P. Minnaar, F. Christensen, H. Moller, S. K. Olesen, and J. Plogsties, "Audibility of all-pass components in binaural synthesis," in *Audio Eng. Soc. Conv.,* pp. 4911–4926, May 1999.

[29] V. Benichoux, M. Rébillat, and R. Brette, "On the variation of interaural time differences with frequency," *J. Acoust. Soc. Am.*, vol. 139, no. 4, pp. 1810–1821, 2016.

[30] T. Lübeck, H. Helmholz, J. M. Arend, C. Pörschmann, and J. Ahrens, "Perceptual evaluation of mitigation approaches of impairments due to spatial undersampling in binaural rendering of spherical microphone array data," *J. Audio Eng. Soc.*, vol. 68, no. 6, pp. 428–440, 2020.

[31] 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. Audio, Speech, Lang. Process.*, vol. 27, no. 12, pp. 2249–2262, Dec. 2019.

[32] J.-M. Jôt, S. Wardle, and V. Larcher, "Approaches to binaural synthesis," *Audio Eng. Soc. Conv.,* pp. 4861–4874, 1998.

[33] C. Armstrong, D. Murphy, and G. Kearney, "A bi-radial approach to ambisonics," in *Proc. Audio Eng. Soc. Conf.*: *AES Int. Conf. Audio Virtual Augmented Reality*, 2018, pp. 1–10.

[34] M. Park and B. Rafaely, "Sound-field analysis by plane-wave decomposition using spherical microphone array," *J. Acoust. Soc. Am.*, vol. 118, no. 5, pp. 3094–3103, 2005.

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

[36] mh acoustics, "em32 Eigenmike Microphone Array Release Notes," Feb. 2009, 25 Summit Ave Summit NJ 07901. [Online]. Available: http://www.mhacoustics.com/products#eigenmike1.

[37] G. D. Romigh, D. S. Brungart, R. M. Stern, and B. D. Simpson, "Efficient real spherical harmonic representation of head-related transfer functions," *IEEE J. Sel. Top. Signal Process.*, vol. 9, no. 5, pp. 921–930, Aug. 2015.

[38] C. Pörschmann, J. M. Arend, and F. Brinkmann, "Directional equalization of sparse head-related transfer function sets for spatial upsampling," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 27, no. 6, pp. 1060–1071, Jun. 2019.

[39] Z. Ben-Hur, D. L. Alon, R. Mehra, and B. Rafaely, "Sparse representation of hrtfs by ear alignment," in *Proc. IEEE Workshop Appl. Signal Process. Audio Acoust.*, 2019, pp. 70–74.

[40] K. Farrar, "Soundfield microphone," *Wireless World*, vol. 85, no. 1526, pp. 48–50, 1979.

[41] Z. Ben-Hur, D. Alon, R. Mehra, and B. Rafaely, "Binaural reproduction using bilateral ambisonics," in *Proc. Audio Eng. Soc. AES Int. Conf. Audio Virtual Augmented Reality*, 2020.

[42] M. Slaney, "Auditory toolbox," Interval Res. Corporation, Tech. Rep. #1998-010, vol. 10, 1998.

[43] A. M. Salomons, Coloration and Binaural Decoloration of Sound Due to Reflections TU Delft, Delft, The Netherlands Delft Univ. Technol., 1995.

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

[45] M. Burkhard and R. Sachs, "Anthropometric Manikin for acoustic research," *J. Acoust. Soc. Am.*, vol. 58, no. 1, pp. 214–222, 1975.

[46] R. O. Duda and W. L. Martens, "Range dependence of the response of a spherical head model," *J. Acoust. Soc. Am.*, vol. 104, no. 5, pp. 3048–3058, 1998.

[47] M. Dinakaran *et al.*, "Perceptually motivated analysis of numerically simulated head-related transfer functions generated by various 3D surface scanning systems," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2018, pp. 551–555.

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

[49] A. W. Mills, "Lateralization of high-frequency tones," *J. Acoust. Soc. Am.*, vol. 32, no. 1, pp. 132–134, 1960.

[50] W. A. Yost and R. H. Dye Jr., "Discrimination of interaural differences of level as a function of frequency," *J. Acoust. Soc. Am.*, vol. 83, no. 5, pp. 1846–1851, 1988.

[51] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *J. Acoust. Soc. Am.*, vol. 65, no. 4, pp. 943–950, 1979.

[52] B. Bernschütz, "A spherical far field HRIR/HRTF compilation of the neumann KU 100," in *Proc. 40th Italian Annu. Conf. Acoust. 39th German Annu. Conf. Acoust.*, 2013, pp. 592–595.

[53] ITU-R-BS.1770-4, *Algorithms to measure audio programme loudness and true-peak audio level.* (Electronic Publication, Geneva, 2015).

[54] ITU-R-BS, *1534-3: Method for the subjective assessment of intermediate quality level of audio systems*. (Electronic Publication, Geneva, 2015).

[55] MATLAB. version 9.7 (R2019b). Natick, Massachusetts: The MathWorks Inc.; 2019.

[56] J. Breebaart, "Evaluation of statistical inference tests applied to subjective audio quality data with small sample size," *IEEE/ACM Trans. Audio, Speech Lang. Process.*, vol. 23, no. 5, pp. 887–897, May 2015.

[57] J. W. Tukey, *Exploratory Data Analysis*, vol. 2, Boston, MA, USA: Addison-Wesley, Reading, Massachusetts, USA, 1977, pp. 39–42.

[58] S. Nagel and P. Jax, "Dynamic binaural cue adaptation," in *Proc. 16th Int. Workshop Acoust. Signal Enhancement*, 2018, pp. 96–100.

[59] R. Duraiswami, Z. Li, D. N. Zotkin, E. Grassi, and N. A. Gumerov, "Plane-wave decomposition analysis for spherical microphone arrays," in *Proc. IEEE Workshop Appl. Signal Process. Audio Acoust.*, 2005, pp. 150–153.

[60] T. Pihlajamaki and V. Pulkki, “Synthesis of complex sound scenes with transformation of recorded spatial sound in virtual reality,” *J. Audio Eng. Soc.*, vol. 63, no. 7/8, pp. 542–551, 2015.

[61] L. Birnie, T. Abhayapala, P. Samarasinghe, and V. Tourbabin, “Sound field translation methods for binaural reproduction,” in *IEEE Workshop Appl. Signal Process. Audio Acoust.*, 2019, pp. 140–144.

[62] M. Kentgens, A. Behler, and P. Jax, “Translation of a higher order ambisonics sound scene based on parametric decomposition,” in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2020, pp. 151–155.

Zamir Ben-Hur received the B.Sc. degree (summa cum laude) and M.Sc. degrees in electrical and computer engineering, in 2015 and 2017, respectively, from the Ben Gurion University of the Negev, Beer-Sheva, Israel, where he is currently working toward the Ph.D. degree in electrical and computer engineering. His current research focuses on audio signal processing for binaural reproduction with improved spatial perception. He was the recipient of the Ben Gurion University High-Tech Fellowship.

David Lou Alon received the B.Sc., M.Sc., and Ph.D. degrees in electrical engineering from Ben Gurion University, Beer-Sheva, Israel, in 2009, 2013, and 2017, respectively. He is currently a Research Scientist with Facebook Reality Labs, investigating efficient representations of head-related transfer functions and headphone equalization for spatial audio application.

Ravish Mehra received the Ph.D. degree in computer science from the University of North Carolina, Chapel Hill, NC, USA, in the field of acoustics and spatial audio. He is the Research Science Lead for the Audio Team, Facebook Reality Labs. His team is responsible for the research and advanced development of new audio techniques to push the state-of-the-art for audio in VR and AR. In his doctoral work, he worked on novel physically based simulation techniques for simulating complex acoustic phenomena arising out of propagation of sound waves in large environments. His research interests include audio, acoustics, signal processing, and virtual and augmented reality. His work in acoustics and spatial audio has generated considerable interest in the audio community, and his sound propagation and spatial sound system has been integrated into virtual reality systems, Oculus Head Mounted Display, with demonstrated benefits.

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. With the ISVR, he was appointed as a Lecturer in 1997 and a Senior Lecturer in 2001, working on active control of sound and acoustic signal processing. In 2002, he was a Visiting Scientist for six months with Sensory Communication Group, Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA, USA, investigating speech enhancement for hearing aids. In 2003, he joined the Department of Electrical and Computer Engineering, Ben Gurion University, as a Senior Lecturer, and was appointed as an Associate Professor in 2010 and a Professor in 2013. He is currently heading the acoustics laboratory, investigating methods for audio signal processing and spatial audio. From 2010 to “2014, he was an Associate Editor for the IEEE TRANSACTIONS ON AUDIO, SPEECH AND LANGUAGE PROCESSING, from 2013 to ”2018, as a Member of the IEEE Audio and Acoustic Signal Processing Technical Committee, from 2015 to 2019, an Associate Editor for the IEEE SIGNAL PROCESSING LETTERS , from 2016 to 2019, for the IET *Signal Processing*, and currently for the Acta Acustica united with Acustica. From 2013 to 2016, he was the Chair of the Israeli Acoustical Association, and is currently the Chair of the Technical Committee on Audio Signal Processing in the European Acoustical Association. He was the recipient of the British Councils Clore Foundation Scholarship.