

We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900

Open access books available

186,000

International authors and editors

200M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



Binaural Reproduction Based on Bilateral Ambisonics

*Zamir Ben-Hur, David Alon, Or Berebi, Ravish Mehra
and Boaz Rafaely*

Abstract

Binaural reproduction of high-quality spatial sound has gained considerable interest with the recent technology developments in virtual and augmented reality. The reproduction of binaural signals in the Spherical-Harmonics (SH) domain using Ambisonics is now a well-established methodology, with flexible binaural processing realized using SH representations of the sound-field and the Head-Related Transfer Function (HRTF). However, in most practical cases, the binaural reproduction is order-limited, which introduces truncation errors that have a detrimental effect on the perception of the reproduced signals, mainly due to the truncation of the HRTF. Recently, it has been shown that manipulating the HRTF phase component, by ear-alignment, significantly reduces its effective SH order while preserving its phase information, which may be beneficial for alleviating the above detrimental effect. Incorporating the ear-aligned HRTF into the binaural reproduction process has been suggested by using Bilateral Ambisonics, which is an Ambisonics representation of the sound-field formulated at the two ears. While this method imposes challenges on acquiring the sound-field, and specifically, on applying head-rotations, it leads to a significant reduction in errors caused by the limited-order reproduction, which yields a substantial improvement in the perceived binaural reproduction quality even with first order SH.

Keywords: binaural reproduction, HRTF, spherical-harmonics, 3D audio, spatial audio, ambisonics, head-tracking

1. Introduction

Recent developments in the field of virtual and augmented reality have increased the demand for high fidelity binaural reproduction technology [1]. Such technology aims to reproduce the spatial sound scene at the listener's ears through a pair of headphones, providing an immersive virtual sound experience [2]. The two main acoustic processes producing the binaural signals are the spatial sound-field result of the propagation from the sound source to the listener, and the interaction of this sound-field with the listener's body, which is described by the Head-Related Transfer Function (HRTF)¹ [3]. Binaural signals can be obtained directly using binaural microphones at the listener's ears [4]. In this way, the sound-field and the

¹ The term "HRTF" is used in this chapter to refer to the set of transfer functions for a set of source positions, unless stated otherwise.

HRTF are jointly captured and the reproduced binaural signals are limited to the recording scenario. More flexible reproduction, enabling, for example, the use of individual (personalized) HRTFs and head-tracking, can be obtained by rendering the binaural signals in post-processing. This requires the sound-field and the HRTF to be available separately. The HRTF could be obtained from an online database, or it could be measured acoustically or simulated numerically for an individual listener [5]. The sound field could also be simulated numerically, or captured using a microphone array [6–8].

In the past, the rendering of binaural signals using Ambisonics representation of the sound-field has been proposed [9–11]. The Ambisonics signals are the Spherical-Harmonics (SH) domain coefficients of the plane-wave amplitude density function, which encode the directional information of the sound-field. The binaural signals are computed by summing the products of the Ambisonics signals and the SH representation of the free-field HRTF. This offers the flexibility to manipulate either the sound field or the HRTF or both by employing algorithms that operate in the SH domain [12, 13].

The Ambisonics signals of a measured sound-field can be obtained from spherical microphone array recordings [14]. In practice, these arrays have a limited number of microphones, which limits the usable SH order [15]. A similar order limitation may also apply for a simulated sound-field due to computational efficiency considerations or memory usage [1, 16]. This order limitation places a constraint on the maximum SH order of the employed HRTF, which leads to truncation error [17]. Truncation error results in significant artifacts, both in frequency and in space, which have a detrimental effect on the perception of the reproduced binaural signals, for example, on the localization, source width, coloration and stability of the virtual sound source [18–21]. One way to overcome the limitations of low order Ambisonics is by a parametric representation of the sound field. For example, using DirAC [22], COMPASS [23], SPARTA [24] or HARPEX [25]. However, these approaches may introduce errors due to incomplete parameterization and thus do not provide ideal solution.

The HRTF truncation error can be reduced by pre-processing that lowers its effective SH order [26]. Evans et al. [27] suggested aligning the HRTF in the time domain prior to deriving its SH decomposition, and showed that this reduces the effective SH order significantly. They also showed that representing separately the magnitude and the unwrapped phase of the HRTF results in a lower SH order for both, compared to the complex-frequency representation. Romigh *et al.* [28] suggested using minimum-phase representation of the HRTF, together with logarithmic representation of the magnitude, and showed that a SH order as low as 4 is sufficient in order to achieve localization performance that is comparable with that of real sound sources in free-field. Brinkmann and Weinzierl [26] compared between these methods (among others), and concluded that the time-alignment method requires the lowest SH order in terms of SH energy distribution and Just Noticeable Difference (JND) in binaural models for source localization, coloration and correlation. Recently, a new method for efficient SH representation of HRTFs, which is based on ear-alignment, was presented [29]. This method proved to be more robust than the time-alignment method, while achieving a similar reduction in the effective SH order.

The order reduction of the HRTF using all the above methods is based on manipulating its phase component. However, the use of such a pre-processed HRTF for binaural reproduction using Ambisonics signals is not trivial due to the relation between the phases of the HRTF and the sound-field; hence, alternative solutions have also been explored. In [30], 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 [31], 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 [32, 33], its removal still introduces errors in the binaural signal, and may affect other perceptual attributes [34, 35]. In [36], Lübeck *et al.* showed that the MagLS method achieved similar perceptual improvement to previously suggested diffuse field equalization methods for binaural reproduction [19, 37]. In [38], Jot *et al.* presented the Binaural B-Format approach, which uses first order Ambisonics signals at the location of the listener's ears and a minimum-phase approximation of the HRTF to compute the binaural signals directly at the listener's ears. This approach was further studied in [39, 40], along with several other approaches also based on the linear decomposition of the HRTF over spatial functions. Recently, the Binaural B-Format was extended to an arbitrary SH order using Bilateral Ambisonics reproduction [41, 42], which uses the ear-aligned HRTF and preserves the HRTF phase information. This method significantly reduces the truncation error and was shown to outperform current state-of-the-art methods using MagLS with low SH order reproduction. However, using Bilateral Ambisonics imposes challenges on acquiring the sound-field, and, specifically, on applying head-rotations to the reproduced binaural signal.

This chapter presents a detailed description of the Bilateral Ambisonics method, from HRTF representation to reproduction, including a possible solution for head tracking. The performance of the method is evaluated and compared with current state-of-the-art methods.

2. Basic ambisonics reproduction

This section provides an overview of the currently used formulation for binaural reproduction using Ambisonics signals, denoted here as Basic Ambisonics. The binaural signal, which is the sound pressure observed at each of the listener's ears, can be calculated, in the general case of a sound-field composed of a continuum of plane-waves, by [7, 16]:

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

where $a(k, \Omega)$ is the plane-wave amplitude density function, $\Omega \equiv (\theta, \phi) \in S^2$ is the spatial angle in standard spherical coordinates, with elevation angle $\theta \in [0, \pi]$, which is measured downwards from the Cartesian z axis, and azimuth angle $\phi \in [0, 2\pi)$, which is measured counter-clockwise 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. $h^{L\backslash R}(k, \Omega)$ is the left ear, L , or right ear, R , HRTF, which is the acoustic transfer function from a far-field sound source to the listener's ear [3]. $p^{L\backslash R}(k)$ is the sound pressure at the ear and $\int_{\Omega \in S^2} (\cdot) d\Omega \equiv \int_0^{2\pi} \int_0^\pi (\cdot) \sin(\theta) d\theta d\phi$.

Alternatively, the binaural signal can be calculated in the SH domain, leading to the Basic Ambisonics reproduction formulation [10]:

$$p^{L\backslash R}(k) = \sum_{n=0}^{\infty} \sum_{m=-n}^n [\tilde{a}_{nm}(k)]^* h_{nm}^{L\backslash R}(k), \quad (2)$$

where $h_{nm}^{L\backslash R}(k)$ are the SH coefficients of the HRTF, which can be computed by applying the spherical Fourier transform (SFT) to the HRTF, $h^{L\backslash R}(k, \Omega)$. $\tilde{a}_{nm}(k)$ are the Basic Ambisonics signals, which are the SFT of $[a(k, \Omega)]^*$, where $[\cdot]^*$ denotes the complex conjugate. These Ambisonics signals can be calculated by capturing the sound-field using a spherical microphone array, and applying plane-wave decomposition in the SH domain [14, 43].

In practice, the infinite summation in Eq. (2) will be order limited:

$$p^{L\backslash R}(k) = \sum_{n=0}^N \sum_{m=-n}^n [\tilde{a}_{nm}(k)]^* h_{nm}^{L\backslash R}(k), \quad (3)$$

with $N = \min(N_a, N_h)$ [44], where N_a and N_h are the maximum available order of the Ambisonics signals and the HRTF, respectively. For example, when the Ambisonics signal is derived from spherical microphone array recordings, such as the Eigenmike [45], its order will be limited by the number of microphones; for the Eigenmike case with 32 microphones its order is around $N_a = 4$ [46]. A similar order limitation may also be introduced for a simulated sound-field in practical applications. On the other hand, Zhang *et al.* [17] showed that the HRTF is inherently of high spatial order. They concluded that for physically accurate representation up to 20 kHz, an order of above $N_h = 40$ is required. Therefore, in the practical scenario of $N_a = 4$, the HRTF will be severely truncated by the reproduction order, $N = 4$. This order truncation was shown to have a detrimental effect on the perceived spatial sound quality [18, 19], by affecting both the spectral and the spatial characteristics of the binaural signal.

3. Basic vs. ear-aligned HRTF representations

An efficient representation of the HRTF that reduces its effective SH order could provide a solution for reducing the effect of the truncation error on the reproduced binaural signal, caused by the limited order HRTF.

Recently, several pre-processing methods have been developed with the aim of reducing the effective SH order of the HRTF: for example, by time-alignment [27, 30], using directional equalization [47], using minimum-phase representation [28], or by ear-alignment [29, 48]. All these methods are based on manipulating the linear-phase component of the HRTF, which was shown to be the main contributor to the high-order nature of the HRTF [27].

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 Interaural Time Difference (ITD) [29], which are both important cues for sound source localization [5]. 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. This translation significantly reduces the effective SH order of the HRTF, as described next.

3.1 The effect of dual-centering on the basic SH representation of the HRTF

We denote the SH representation of the HRTF as the "basic representation". In this section, the effect of translating the origin of the free-field component of the HRTF on the basic representation is presented. This is performed by analyzing the simple case of a "free-field HRTF" as outlined in [29].

A pair of far-field HRTFs, h^L and h^R , is defined as a function of direction, Ω , and wave-number, k , by [3]:

$$h^{L\setminus R}(k, \Omega) = \frac{P^{L\setminus R}(k, \Omega)}{P_0(k, \Omega)}, \quad (4)$$

where P^L and P^R represent the sound pressure at the left and right ears, respectively, and P_0 represents the free-field sound pressure at the center of the head in the absence of the head.

Now, consider a single plane-wave in free-field arriving from direction Ω with unit amplitude and wave number k . The sound pressure at position (Ω_0, r) , can be written as [49]:

$$\begin{aligned} P_0(\Omega, k, \Omega_0, r) &= e^{ikr \cos \Theta} \\ &= \sum_{n=0}^{\infty} \sum_{m=-n}^n 4\pi i^n j_n(kr) [Y_n^m(\Omega)]^* Y_n^m(\Omega_0), \end{aligned} \quad (5)$$

where Θ is the angle between Ω and Ω_0 , $Y_n^m(\cdot)$ is the complex SH basis function of order n and degree m [50], and $j_n(\cdot)$ is the spherical Bessel function.

Defining the position of the ear to be at $(\Omega_{L\setminus R}, r_a)$, where r_a is the radius of the head, the free-field HRTF (an HRTF with the head absent) is defined by substituting Eq. (5) in Eq. (4):

$$h_0^{L\setminus R}(k, \Omega) = \frac{P_0(\Omega, k, \Omega_{L\setminus R}, r_a)}{P_0(\Omega, k, \Omega_0, 0)} = P_0(k, \Omega, \Omega_{L\setminus R}, r_a), \quad (6)$$

where $P_0(k, \Omega, \Omega_0, 0) = 1$ for all (Ω, k) . Thus, for a sound-field composed of plane-waves from directions $\Omega \in \mathbb{S}^2$ the free-field HRTF can be written as:

$$h_0^{L\setminus R}(k, \Omega) = \sum_{n=0}^{\infty} \sum_{m=-n}^n 4\pi i^n j_n(kr_a) [Y_n^m(\Omega)]^* Y_n^m(\Omega_{L\setminus R}). \quad (7)$$

From here, the SH coefficients of the free-field HRTF can be derived, as presented in [29]:

$$h_{nm0}^{L\setminus R}(k) = 4\pi i^n j_n(kr_a) [Y_n^m(\Omega_{L\setminus R})]^*. \quad (8)$$

This equation provides insight into the potential effect of the dual-centering measurement process of the HRTF. The free-field HRTF coefficients, as described in the equation, have energy at every order n , which means that the HRTF is of infinite SH order. Nevertheless, it can be considered to be approximately order limited by $N_h = kr_a$, where \cdot is the ceiling function, due to the behavior of the spherical Bessel function, which has a negligible magnitude for $kr > n$ [49, 51]. On the other hand, from Eq. (6) it is clear that if the position of the ear was defined as the origin of the coordinate system, with $r_a = 0$, the free-field HRTF would be constant with unity value, which is represented by a zero order SH. This demonstrates how a sound pressure measured at a distance r_a from the origin, when normalized by a sound pressure at the origin, can lead to an increase in the SH order by approximately $N = \lceil kr_a \rceil$. An example of this added order is illustrated in **Figure 1**, which demonstrates how the SH order increases up to 30 at high frequencies.

Note the similarity of the orders in **Figure 1** to the actual order of the HRTFs as presented in [17], which suggests that although the explanation presented in this section is theoretical, it gives an insight into the possible increase in SH order due to the dual-centering of the HRTF definition.

3.2 HRTF ear-alignment

To compensate for the effect described in the previous section, with the aim of reducing the effective SH order of the HRTF, ear-alignment of the HRTF is suggested.

The ear-aligned HRTF, h_a , is defined in a similar way to Eq. (4) as:

$$h_a^{L\backslash R}(k, \Omega) = \frac{P^{L\backslash R}(k, \Omega)}{P_0^{L\backslash R}(k, \Omega)}, \quad (9)$$

where $P_0^{L\backslash R}$ is the free-field pressure at the position of the left ear, L , or right ear, R , with the head absent. A measured HRTF can be aligned by translating the free-field pressure (denominator in (9)) from the center of the head to the position of the ear by:

$$h_a^{L\backslash R}(k, \Omega) = h^{L\backslash R}(k, \Omega) \cdot \frac{P_0(k, \Omega)}{P_0^{L\backslash R}(k, \Omega)}. \quad (10)$$

For a far-field HRTF, the free-field sound pressure can be computed using the plane-wave formulation as in Eq. (5), which leads to the ear-alignment formulation:

$$h_a^{L\backslash R}(k, \Omega) = h^{L\backslash R}(k, \Omega) e^{-ikr_a \cos \Theta_{L\backslash R}}, \quad (11)$$

where $\Theta_{L\backslash R}$ is the angle between the direction of the ear, $\Omega_{L\backslash R}$, and the direction of the sound source, Ω , and $\cos \Theta_{L\backslash R} = \cos \theta \cos \theta_{L\backslash R} + \cos(\phi - \phi_{L\backslash R}) \sin \theta \sin \theta_{L\backslash R}$. It is important to note that this ear-alignment process is invertible, which means that going from $h^{L\backslash R}$ to $h_a^{L\backslash R}$ and back can be performed without any loss of information.

Figure 2 presents an example of the SH spectrum of a KEMAR HRTF [26, 52], for the basic and ear-aligned HRTF representations. The SH spectrum, which is the energy of the SH coefficients at every order n , is computed as:

$$E_n(f) = \sum_{m=-n}^n |h_{nm}(f)|^2, \quad (12)$$

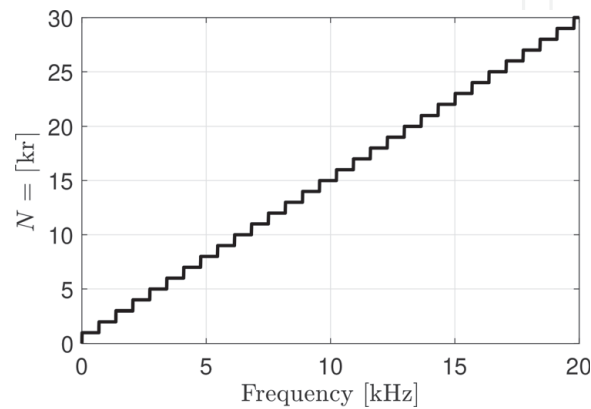


Figure 1.

Added SH order due to the dual-centering of the HRTF, $N = [kr_a]$, as a function of frequency ($[\cdot]$ is the ceiling function). Computed for the free-field HRTF with $r_a = 8$ cm and $c = 343$ m/s.

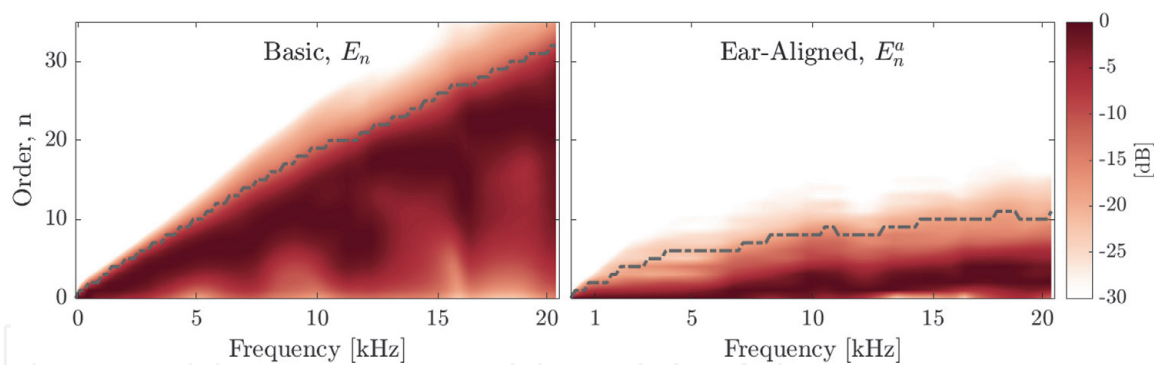


Figure 2. Normalized SH spectra, E_n , of basic and ear-aligned KEMAR HRTF representations, computed according to Eq. (12). The dashed gray line represents the order at which 99% of the energy is contained.

and normalized by the maximum value for each frequency. The figure shows how the energy of the high-order SH coefficients of the ear-aligned HRTF is significantly reduced compared to the basic HRTF. This validates the finding from Section 3.1, in which the high orders of the basic HRTF actually originate from the translation from the origin. In particular, the order at which 99% of the energy is contained is reduced to be below order 10 for all frequencies.

It is interesting to note that the SH order reduction of the ear-aligned HRTF can explain the reduced order of the time-alignment method. This is discussed in detail in [26, 27]. The ear-alignment can be interpreted as “virtually” removing the inherent delay in an HRTF caused by normalizing the pressure at the ear by the pressure at the origin. This is evident from Eq. (11), where the phase in the exponent represents a delay from the origin to the ear due to a source at Ω . The main difference between the time-alignment and ear-alignment methods is as follows. Performing time-alignment requires numerical estimation of the time delays; this may be challenging and its accuracy may depend on the HRTF direction and on the quality of the measurements [53, 54]. In contrast, ear-alignment can be performed parametrically with the parameters r_a and $\Omega_{L \setminus R}$. Moreover, using the ear-alignment with fixed parameters makes it data-independent, which improves its robustness to measurement noise (as discussed comprehensively in [29]).

4. Binaural reproduction based on bilateral ambisonics and ear-aligned HRTFs

While the ear-alignment method leads to efficient SH representation of the HRTF, incorporating the pre-processed ear-aligned HRTF in a binaural reproduction process is not trivial. The computation of the binaural signal (Eq. (3)) requires the HRTF and the Ambisonics signals to be represented in the same coordinate system and around the same origin. One way to align them 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 Basic Ambisonics reproduction. Another way is to use the MagLS approach, which completely ignores the HRTF phase component at high frequencies [31]. Alternatively, the Binaural B-Format approach, presented by Jot *et al.* [38], can be used. In this approach, two B-Format recordings at the ear locations are used, together with a minimum-phase approximation of the HRTF and an ITD estimation based on a spherical head model. The Binaural B-format can be extended by using the ear-aligned HRTF together with high-order Ambisonics signals that are defined

around the ear locations. This approach is denoted as Bilateral Ambisonics reproduction [41, 42].

Assuming that the plane-wave amplitude density function, denoted by $a^{L\backslash R}(k, \Omega)$, is given at the position of the ear, then the binaural signal can be computed directly at the listener's ears, using the ear-aligned HRTF, similarly to in Eq. (1):

$$p^{L\backslash R}(k) = \int_{\Omega} a^{L\backslash R}(k, \Omega) h_a^{L\backslash R}(k, \Omega) \Omega. \quad (13)$$

From here, the Bilateral Ambisonics reproduction of order N can be formulated as:

$$p^{L\backslash R}(k) = \sum_{n=0}^N \sum_{m=-n}^n \left[\tilde{a}_{nm}^{L\backslash R}(k) \right]^* h_{anm}^{L\backslash R}(k), \quad (14)$$

where $\tilde{a}_{nm}^{L\backslash R}(k)$ and $h_{anm}^{L\backslash R}(k)$ are the SH coefficients of $[a^{L\backslash R}(k, \Omega)]^*$ and $h_a^{L\backslash R}(k, \Omega)$, respectively. It is important to note that, in contrast to $a(k, \Omega)$, which is the plane-wave amplitude density function of the sound-field as observed at the position of the center of the head, $a^{L\backslash R}(k, \Omega)$ is observed at the position of the ears. **Figure 3** demonstrates 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 red dotted axes with their origin at the positions of the ears, are used for the computation in Eq. (14) using the Bilateral Ambisonics signals, $\tilde{a}_{nm}^{L\backslash R}(k)$, which are different for each ear. **Figure 4** demonstrates the signal-flow of the Basic and Bilateral Ambisonics.

Theoretically, the plane-wave amplitude density function at the position of the ear can be computed from the center function by translation of the sound-field [46], which is computed as $a^{L\backslash R}(k, \Omega) = a(k, \Omega) e^{ikr_a \cos \Theta_{L\backslash R}}$; however, this will lead to equivalence between Eq. (13) and Eq. (1), which means that the binaural signals

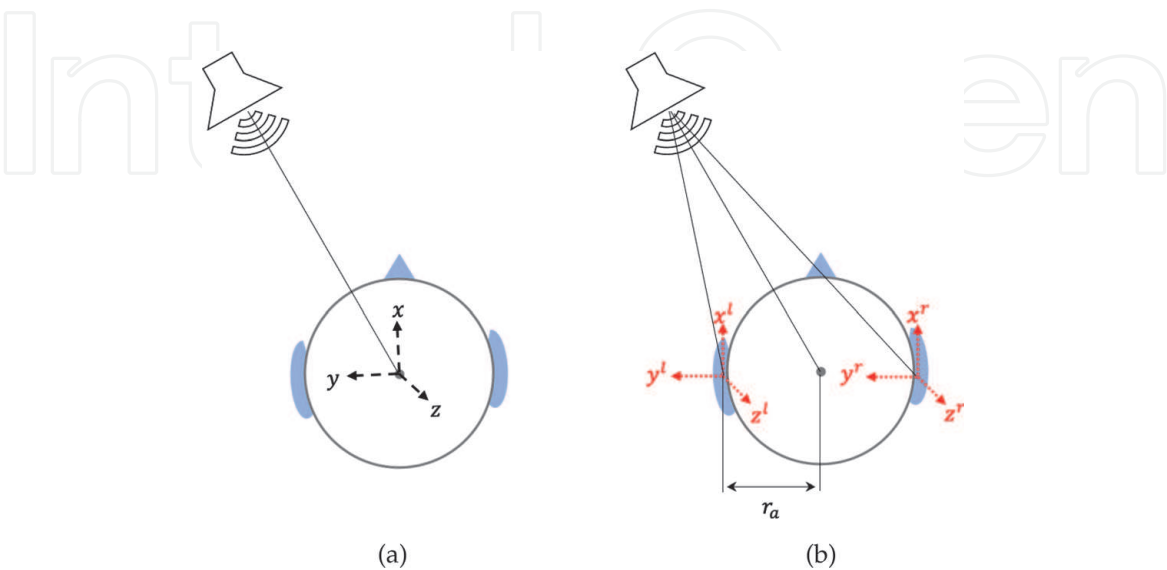


Figure 3. Diagram of the standard (a) and Bilateral (b) coordinate systems. The origin of the standard coordinate system is at the center of the head, while in the bilateral coordinate system the origin is at the position of the ear.

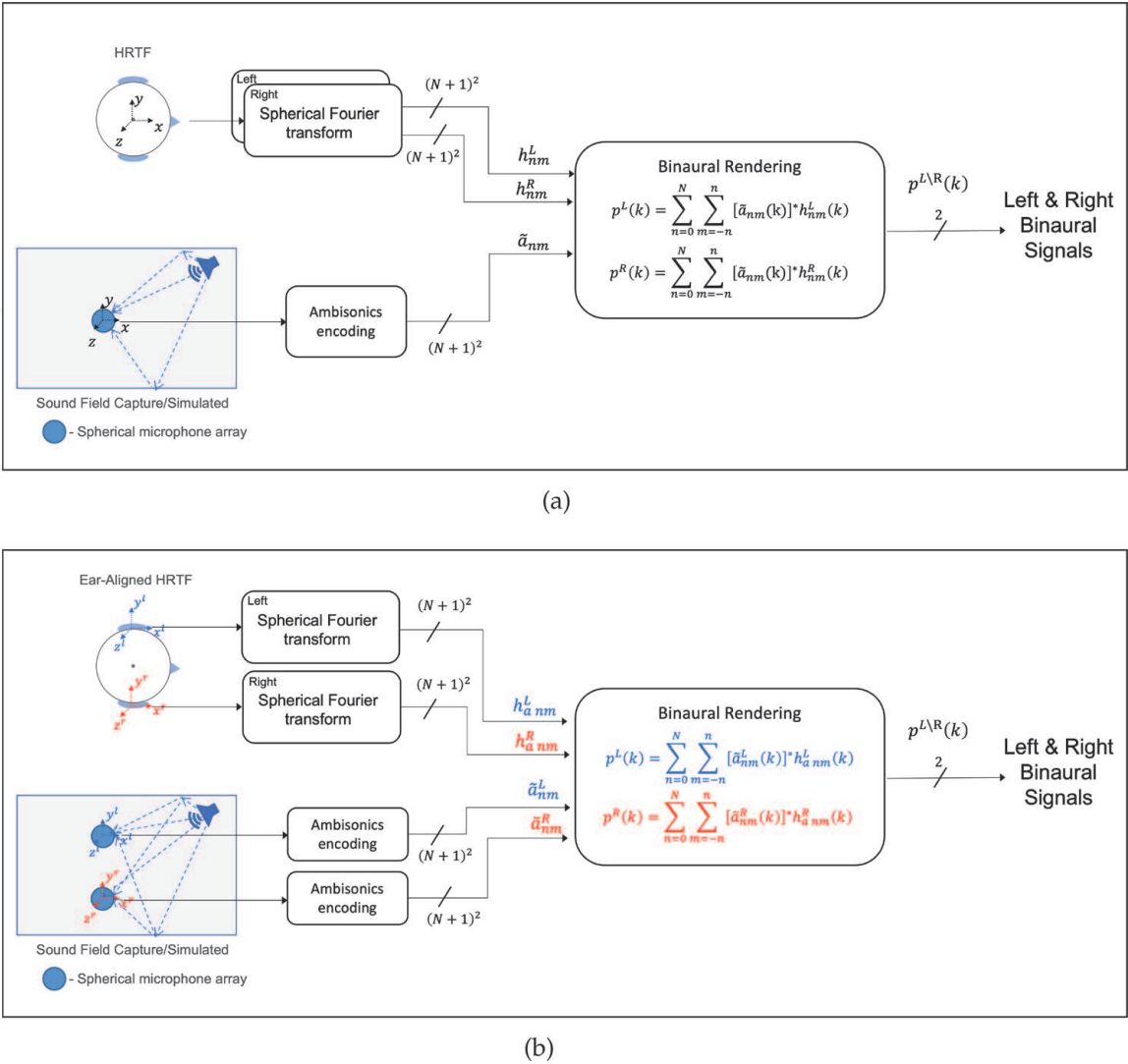


Figure 4.
Binaural reproduction signal-flow of the Basic (a) and Bilateral (b) Ambisonics.

will be identical. Thus, the same truncation error as in the Basic Ambisonics reproduction is introduced. Alternatively, if a low-order plane-wave amplitude density function is given directly at the position of the ear, the Bilateral Ambisonics-based signals (from Eq. (14)) may potentially be more accurate than the Basic Ambisonics reproduction (from Eq. (3)) due to the lower-order nature of the ear-aligned HRTF compared to the unprocessed basic HRTF.

Figure 5 demonstrates the improved accuracy of the Bilateral Ambisonics reproduction. The figure shows the magnitude response of the binaural signals for a single plane-wave of unit amplitude arriving from direction $(\theta, \phi) = (90^\circ, 20^\circ)$, using a KEMAR HRTF, with $N = 1, 4$, and a high-order reference of $N = 40$. For the low-order signals computed using Basic Ambisonics reproduction, a high-frequency roll-off above the sphere cut-off frequency, $kr_a = N$ [14], is clearly observed. This is discussed further in [19]. Additionally, amplitude distortion is also observed at these high frequencies. The Bilateral Ambisonics-based signals seem significantly more accurate in terms of both frequency roll-off and distortion, where reproduction of order $N = 4$ seems to preserve the signal magnitude up to almost 20 kHz, including the important spectral cues (peaks and notches). Further evaluation of the performance of the Bilateral Ambisonics reproduction is presented in Section 6.

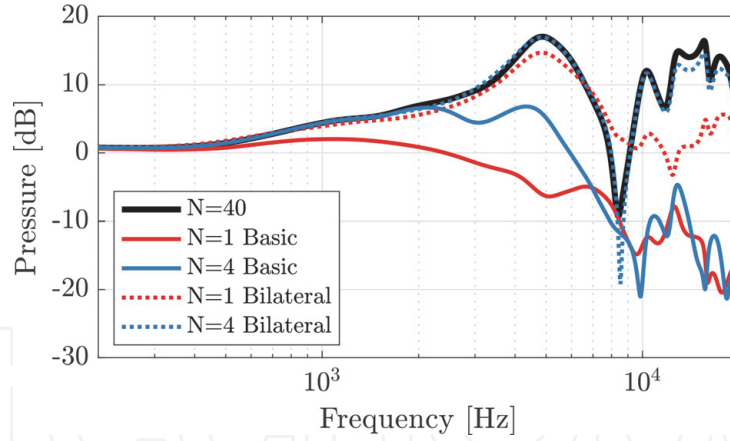


Figure 5. Magnitude of a left ear binaural signal of a single plane-wave from direction $(\theta, \phi) = (90^\circ, 20^\circ)$, with HRTF of KEMAR, computed with Basic Ambisonics reproduction (solid lines) and with Bilateral Ambisonics reproduction (dashed lines), with $N = 1, 4$, compared to a high-order reference with $N = 40$.

5. Head-tracking in bilateral ambisonics reproduction

While Bilateral Ambisonics leads to a more efficient representation of the spatial audio signal and more accurate binaural reproduction, such a procedure will result in a static binaural reproduction. In contrast to the Basic Ambisonics reproduction, where head-rotations can be incorporated in post-processing by a simple rotation of the Ambisonics signals using Wigner-D functions [55], performing this operation in Bilateral Ambisonics is not straightforward. A method to incorporate head-rotations in Bilateral Ambisonics reproduction is presented in this section.

Consider the specific case where a binaural signal is played via headphones to a listener, representing a spatial acoustic scene composed of a single sound source. According to the Bilateral Ambisonics format, the scene is represented by two Ambisonics signals with their origin at the listener's expected ear positions, as seen in **Figure 6a**. Note that the microphone symbols in **Figure 6** represent the left and right Ambisonics signals origin. Upon playback of the acoustic scene, the listener is expected to perceive a virtual source from the direction of the real source (in this example about 30° to the left). Next, the listener rotates his/her head while

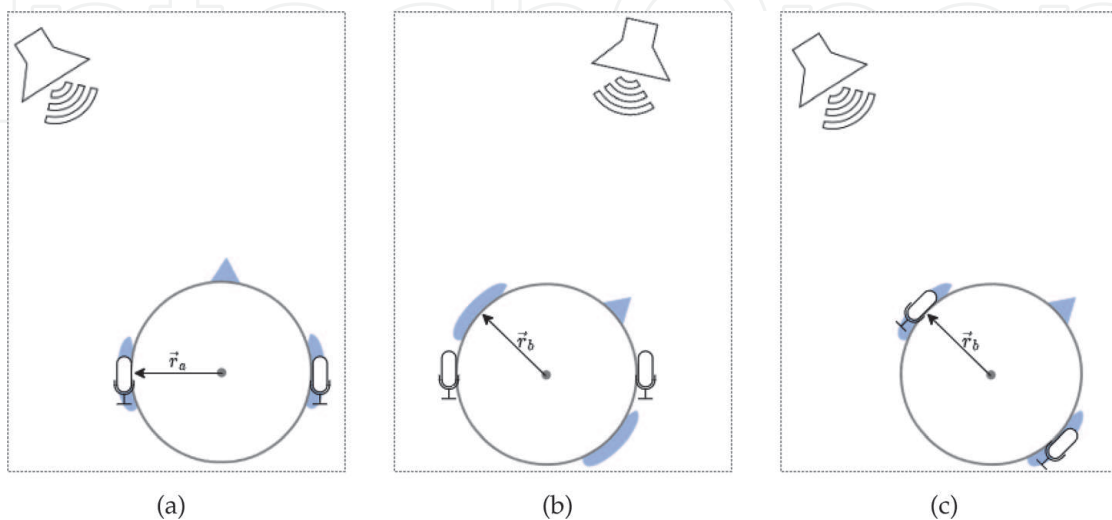


Figure 6. Demonstration of the head-tracking method in bilateral coordinate systems, (a) before the head-rotation, (b) after head-rotation and without head-tracking and (c) after head-rotation and with head-tracking. \vec{r}_a and \vec{r}_b are the head-rotation ear vectors before and after head rotation, respectively.

listening; this action will result in the virtual source changing its position in space, remaining at about 30° to the left, as illustrated in **Figure 6b**. One way to compensate for the head rotation is to acquire new Bilateral Ambisonics signals located at the listener ears' new locations, and also to rotate them according to the angle of rotation of the listener's head, as illustrated in **Figure 6c**. This, of course, may not be a practical option since acquiring new Bilateral Ambisonics signals requires re-synthesizing the sound-field, in the case of a simulation, or adjusting the position of the physical microphone arrays, in the case of sound field capture. The former may be computationally expensive, while the latter is practically infeasible since recording is typically performed independently from the listener's head orientation. Note that a multi-microphone binaural recording method could be employed, similar to the Motion-Tracked Binaural recording method [56], though this solution will be complex in terms of the recording resources. Hence, developing methods that compensate for the listener head movements using head-tracking is of great importance for Bilateral Ambisonics recording and reproduction. **Figure 6** shows that as a result of the rotation in this case, the ears (which are the Ambisonics reference point) change their orientation while also translating in space. Proper compensation for head-rotation needs to take both movements into account.

Now, consider the general case, where an arbitrary sound-field is represented by a plane-wave amplitude density function, denoted by $a(k, \Omega)$, given at the position of the ear. Note that $a(k, \Omega)$ represents the same function as $a^{L/R}(k, \Omega)$ from Eq. (13), but the superscript L/R is left out for notation simplicity since the operation is similar for both ears. Assuming that the ear position with respect to the head center is known before and after the rotation, denoted as \vec{r}_a and \vec{r}_b , respectively, head-tracking can be performed by translation of the plane-wave amplitude density function $a(k, \Omega)$, accordingly. This translation can be performed by a phase-shifting operation, as follows [46, 57]:

$$a^t(k, \Omega) = a(k, \Omega)e^{-i\vec{k} \cdot (\vec{r}_b - \vec{r}_a)} = a(k, \Omega)e^{-ikr_a(\cos \Theta_a - \cos \Theta_b)}, \quad (15)$$

where $a^t(k, \Omega)$ is the translated plane-wave amplitude density function, which represents the plane-wave amplitude density around the ear of the listener after head-rotation (but with the pre-rotation orientation). $r_a = |\vec{r}_a|$ is the head radius, \vec{k} is the wave vector, Θ_a is the angle between the sound source direction, Ω , and the pre-rotation ear position, \vec{r}_a , and Θ_b is the angle between Ω and \vec{r}_b . **Figure 7**

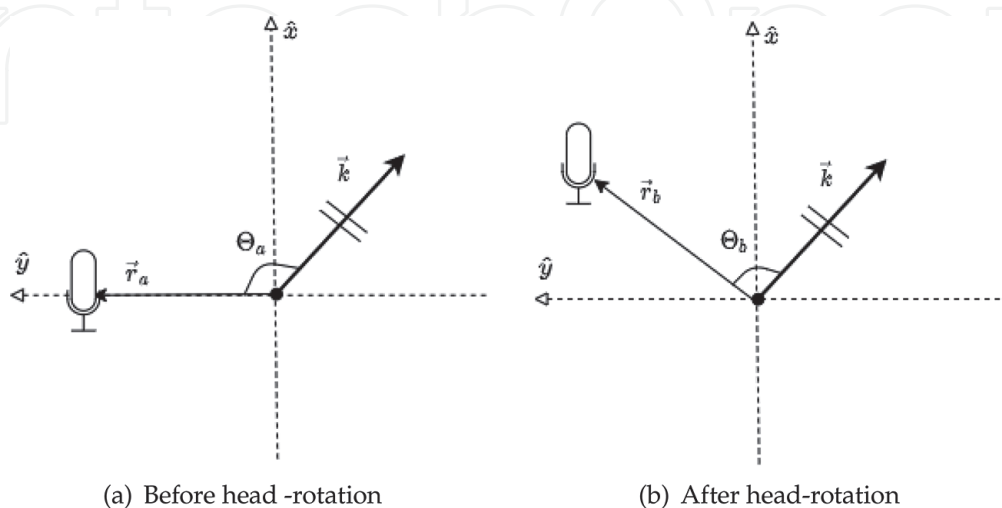


Figure 7.

Schematic illustration for left ear microphone array translation due to head rotation: \vec{r}_a is the left ear vector with respect to the head center, \vec{r}_b is the left ear vector after a clock-wise rotation, \vec{k} is the wave vector of the plane-wave, where Θ_a and Θ_b represent the angle between the ear vector and the wave vector.

demonstrates this translation for a simple case where the sound-field is comprised of a single plane-wave and the microphone symbols represent the measurement position of $a(k, \Omega)$ made by microphone arrays.

Next, the orientation of the translated plane-wave amplitude density function is corrected by applying rotation. This is formulated in the SH domain by:

$$a_{nm}^r(k) = \sum_{m'=-n}^n a_{nm'}^t(k) D_{mm'}^n(\alpha, \beta, \gamma), \quad (16)$$

where $a_{nm}^t(k)$ are the SH coefficients of $a^t(k, \Omega)$, $D_{mm'}^n$ denotes the Wigner D functions, and a_{nm}^r are the rotated Bilateral Ambisonics signals. (α, β, γ) are the Euler angles [58] of the head-rotation, which are assumed to be known, for example from a head-tracker. Note that this procedure needs to be applied to both left and right ears.

In practice, the Bilateral Ambisonics signals will be order limited due to the constraints mentioned in Section 2. The finite order representation, in turn, leads to limitations in the accuracy of the suggested method. These limitations will be presented and demonstrated in numerical simulations in Section 6.

6. Performance analysis

This section presents an objective evaluation of the performance of the proposed Bilateral Ambisonics reproduction approach, and compares it to that of the Basic Ambisonics+MagLS reproduction method.

A binaural signal for a sound-field composed of a single plane-wave of unit amplitude, as presented in **Figure 5**, is computed, and the Normalized Mean Square Error (NMSE) for the left ear is evaluated as:

$$\epsilon^L(f) = 10 \log_{10} \frac{|p_{\text{ref}}^L(f) - p^L(f)|^2}{|p_{\text{ref}}^L(f)|^2}, \quad (17)$$

where p_{ref}^L is the reference high-order binaural signal computed using Eq. (3) with $N = 40$, and p^L is the binaural signal computed using Eq. (3) or (14). The NMSE, although positive and real, is sensitive to both the magnitude and the phase

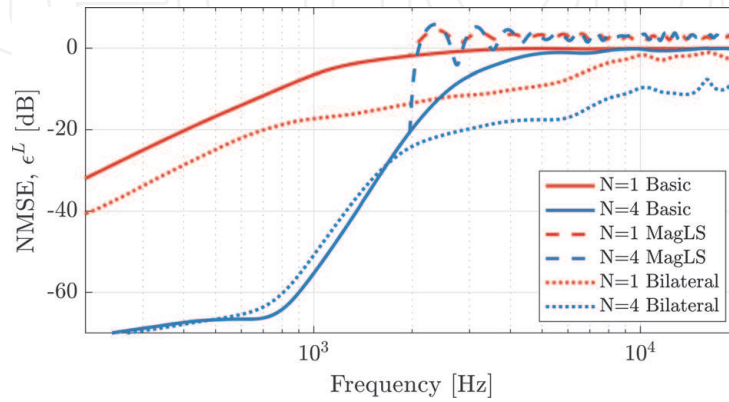


Figure 8.

NMSE of binaural signals computed for sound-fields composed of a single plane-wave, averaged over 434 plane-wave directions (distributed according to a Lebedev grid), with HRTF of KEMAR. The NMSE is computed using Eq. (17), with Basic Ambisonics reproduction (solid lines), with Basic Ambisonics reproduction with MagLS [31] (dashed lines), and with Bilateral Ambisonics reproduction (dotted lines), with $N = 1, 4$, and a high-order reference with $N = 40$.

errors in the binaural signal. The NMSE is averaged over a range of 434 plane-waves with incidence angles distributed nearly-uniformly over the sphere, using the Lebedev sampling scheme of order 17 [59].

Figure 8 shows this averaged NMSE. For the MagLS approach, a cutoff frequency of 2 kHz was used, as indicated by the increased error above this frequency, where the phase is completely inaccurate. The figure demonstrates the improvement in the accuracy of the Bilateral Ambisonics reproduction, compared to the Basic Ambisonics reproduction methods, where at high frequencies, up to about 5 kHz for $N = 1$ and 15 kHz for $N = 4$, the errors are lower by 10–20 dB. Two important spatial cues for sound source localization are the Interaural Time Difference (ITD) and Interaural Level Difference (ILD). Both were shown to be affected by the truncation error due to low-order reproduction [29]. **Figure 9** shows the ITDs, ILDs and their corresponding errors relative to a high-order reference ($N = 40$). The ITDs and ILDs were computed for binaural signals of a single plane-wave sound-field with incident angles across the left horizontal half-plane ($\theta = 90^\circ$; $0^\circ \leq \phi \leq 180^\circ$) with 1° resolution, and with a KEMAR HRTF. The ITDs were esti-

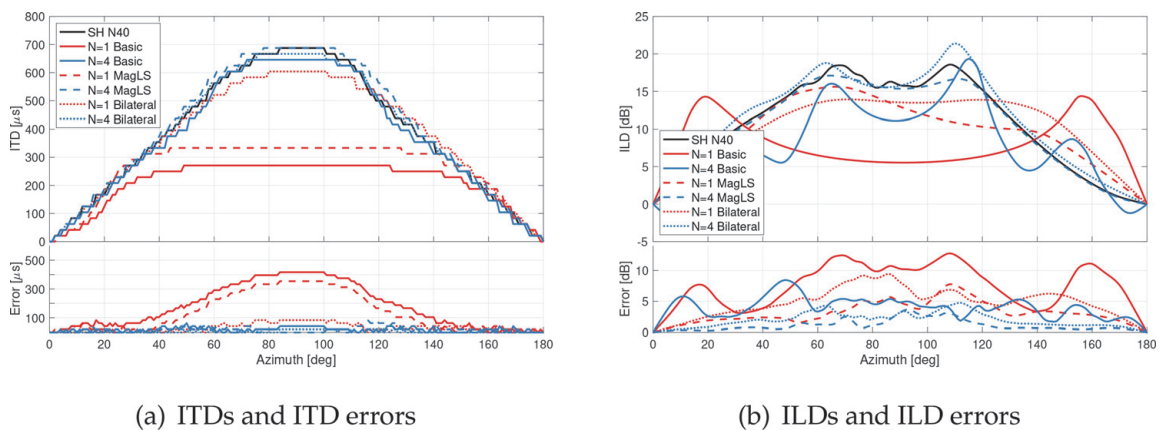


Figure 9.

ITDs, ILDs and their corresponding errors as a function of azimuth angle for binaural signals computed for sound-fields composed of a single plane-wave from 180 directions on the left horizontal plane (the right side is symmetrical), with HRTF of KEMAR. The signals were computed using Basic Ambisonics reproduction with and without MagLS, and Bilateral Ambisonics reproduction.

mated using the onset detection method, applied to a 2 kHz low-pass filtered version of the signals [54]. The ILDs were calculated and averaged across 18 auditory filter bands as [5]:

$$\text{ILD}(f_c, \Omega) = 10 \log_{10} \frac{\int C(f, f_c) |p^L(f)|^2 df}{\int C(f, f_c) |p^R(f)|^2 df}, \quad (18)$$

$$\text{ILD}_{av}(\Omega) = \frac{1}{18} \sum_{f_c} \text{ILD}(f_c, \Omega), \quad (19)$$

where C is a Gammatone filter with center frequency f_c , as implemented in the Auditory Toolbox [60]. The integral is evaluated between 1.5 kHz and 16 kHz and f_c is restricted accordingly. This computation facilitates a perceptually motivated smoothing of the ILD across frequencies, which is required for appropriate comparison between ILDs.

Comparison of the ITD errors with the Just Notable Differences (JND) values reported by Andreopoulou and Katz in [54] ($40 \mu\text{s}$ for the frontal directions and about $100 \mu\text{s}$ for the lateral directions) reveals the main advantage of the Bilateral

Ambisonics approach, where the phase information is preserved and the ITD errors are below the JND even at $N = 1$.

Figure 9b shows that both the MagLS and the Bilateral approaches achieve significant improvement in the ILD accuracy compared to the Basic Ambisonics reproduction. While with the Basic Ambisonics reproduction the ILD errors are above the JND (~ 1 dB [61, 62]) even with $N = 4$, with the MagLS and the Bilateral Ambisonics reproduction the errors for $N = 4$ are below the JND for most angles. Relatively high errors can be seen at the lateral angles compared to the front and back directions. This can be explained by the fact that the ILD at the front and back directions is close to zero, where the errors are expected to be small due to the symmetry of the HRTF model. Nevertheless, both the MagLS and the Bilateral Ambisonics reproduction led to substantially lower ILD errors compared to the Basic Ambisonics reproduction.

As discussed in Section 5, a limitation of the Bilateral Ambisonics method compared to Basic Ambisonics is found in terms of the incorporation of head-tracking in post-processing. In Section 5, a method to overcome this limitation was suggested. To evaluate the performance of this method, a simulation study was conducted. The simulation results aim to evaluate the NMSE introduced by the head rotation and its dependence on the Bilateral Ambisonics signal order and the head rotation angle. In the simulation, a head was positioned in free-field, facing the \hat{x} direction with the ears positioned on the xy plane. A sound-field was generated, consisting of a single plane-wave with unit amplitude arriving from directions that are taken from the Lebedev sampling scheme, using the same sampling scheme mentioned earlier. The Bilateral Ambisonics signal, $a_{nm}^L(k)$, is calculated with respect to the left ear position \vec{r}_a up to order N . Note that the superscript L denoting the left ear will be removed for brevity from now on. The signal is then transformed to $a(k, \Omega)$ with the discrete inverse spherical Fourier transform (DISFT) [46]. Next, the head is rotated by γ degrees clockwise in the horizontal plane, as shown in **Figure 6b**, resulting in a new rotated left ear position \vec{r}_b . The translated plane-wave amplitude density function, $a^t(k, \Omega)$, is computed using Eq. (15). Next, Eq. (16) is used to calculate $a_{nm}^r(k)$ from $a_{nm}^t(k)$, the discrete spherical Fourier transform (DSFT) [46] of $a^t(k, \Omega)$. The signal $a_{nm}^r(k)$ represents the head-rotated left ear Bilateral Ambisonics signal; note that the right ear signal can be calculated in a similar manner. Finally, the left ear binaural signal with head-tracked Bilateral Ambisonics, $p(f)$, is calculated using Eq. (14) with $a_{nm}^r(k)$ and a KEMAR HRTF. The reference binaural signal, $p_{\text{ref}}(f)$, is calculated using Eq. (14) with an accurately generated Bilateral Ambisonics signal $a_{nm}^{\text{ref}}(k)$ of order N at the head-rotated position. The NMSE is calculated using Eq. (17), and averaged over the 434 sampling scheme directions.

Figure 10a shows the NMSE between $p_{\text{ref}}(f)$ and $p(f)$ for a head rotation of $\gamma = 30^\circ$ and different reproduction orders, $N = 1, 4, 10$. The figure demonstrates the improvement in the NMSE as the order increases. Additionally, the figure demonstrates how the error increases with frequency. For $N = 1, 4, 10$ an error of less than -10 dB is achieved up to about 1, 5 kHz and 11 kHz, respectively. This result indicates that, for example, with order $N = 4$ and a rotation angle of 30° the suggested rotation method will experience a noticeable loss in accuracy above 5 kHz, compared to the reference. To evaluate the performance of the suggested method for different head rotation angles, the order was kept at $N = 4$ and various values of head rotation angle, γ , were used. **Figure 10b** illustrates how the performance deteriorates as the rotation angle increases. For $\gamma = 30^\circ, 60^\circ, 90^\circ$ an error of less than -10 dB is achieved up to about 5 kHz, 3 kHz and 2 kHz, respectively.

We now compare between binaural reproduction performance with head-tracked Bilateral Ambisonics, head-tracked MagLS and with head-tracked Basic Ambisonics.

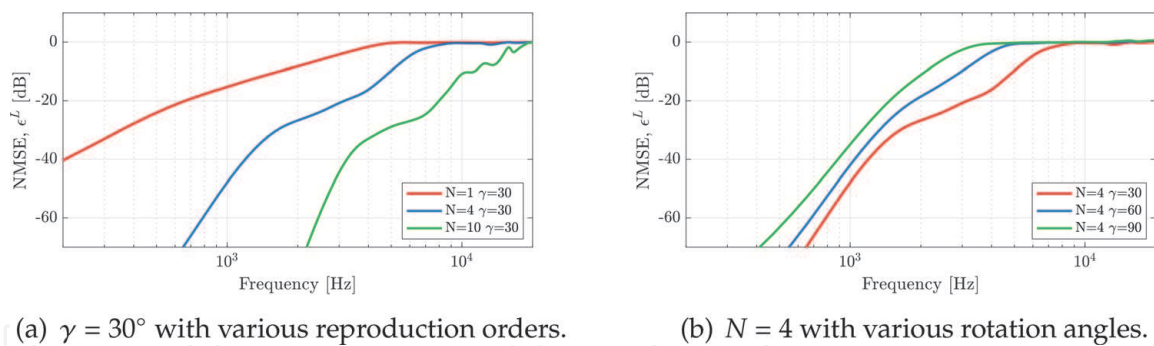


Figure 10.
NMSE of binaural signals computed using Bilateral Ambisonics reproduction with head rotation as in Eq. (16), for various orders (a) and rotation angles (b).

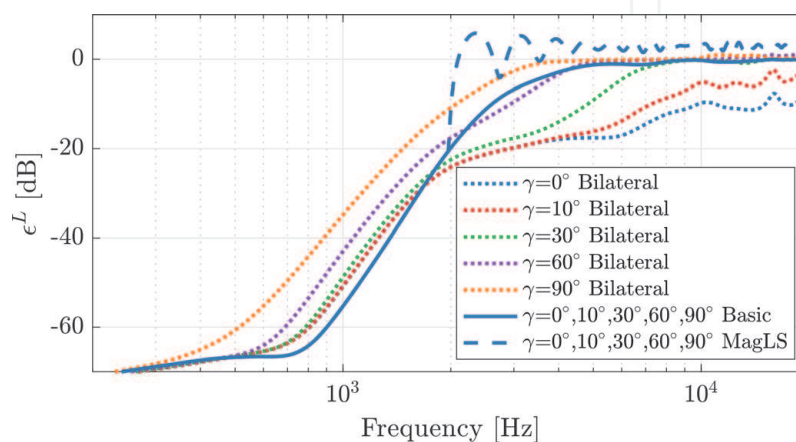


Figure 11.
NMSEs of binaural signals computed using rotated Basic, MagLS and Bilateral Ambisonics signals with order $N = 4$ relative to a high-order Basic Ambisonics reproduction with $N = 40$, with various rotation angles, γ . The NMSE is averaged over 434 plane-wave directions.

In the simulation (which is identical to the previously described simulation), the NMSE is measured for head-tracked binaural signals computed using Basic, MagLS and Bilateral Ambisonics reproductions with order $N = 4$, and compared with a high-order reference computed using Basic Ambisonics reproduction with order $N = 40$. The head-tracked Bilateral Ambisonics signals are calculated with the suggested method, using Eqs. (15) and (16), and both the head-tracked Basic Ambisonics and MagLS signals are calculated in the SH domain using Eq. (16). Note that for head-tracking with Basic Ambisonics and MagLS, the error is independent of the rotation angle, γ . **Figure 11** presents the results for different head-rotation angles, γ . As expected, the rotation procedure compromises the accuracy of the binaural signal with Bilateral Ambisonics at high frequencies. For $\gamma = 10^\circ$, the Bilateral Ambisonics reproduction retains its advantage in accuracy compared to the Basic Ambisonics reproduction up to around 20 kHz. However, for a head-rotation of $\gamma = 30^\circ$, the Bilateral Ambisonics reproduction retains its advantage only up to about 7 kHz. For a head-rotation of $\gamma = 60^\circ$, the two reproduction schemes are equally accurate. For a head-rotation of $\gamma = 90^\circ$, the Bilateral Ambisonics reproduction results in an error of less than -10 dB up to about 2 kHz, compared to 2.5 kHz for Basic Ambisonics. Similar behavior was also observed for other reproduction orders. These results indicate that, in this case, the suggested rotation method is mainly beneficial for head rotations up to 60° . Note that 60° means that the listener can turn his/her head 60° both to the left and to the right. The inaccuracies depicted in **Figures 10** and **11** relating to the reproduction order N and head-rotation angle γ , can be explained by errors due to the translation operation described in Eq. (15) [46, 57].

Further evaluation of head-tracking compensation is the subject of ongoing research. The study could include evaluation of ITD/ILD reconstruction, Lateral Error, Polar error in median plane, Coloration error [26] and subjective listening tests.

7. Conclusions

This chapter presented a detailed description of the Bilateral Ambisonics reproduction method. The method incorporates a pre-processed ear-aligned HRTF, which provides an efficient representation of the HRTF in the SH domain, with bilateral representation of the Ambisonics signals. The method was shown to improve the accuracy of low-order binaural reproduction in comparison to Basic Ambisonics reproduction in terms of reduced errors in the binaural signals, as well as more accurate ITD and ILD. The two main limitations of this method are the requirement for two Ambisonics signals at the positions of the ears, and the difficulty of incorporating head-tracking. The latter has been addressed in this chapter by presenting a method to incorporate head-tracking in post-processing. Ways should be sought to mitigate the requirement for two different Ambisonics signals, for example by transforming a Basic Ambisonics signal into a Bilateral Ambisonics signal.

Conflict of interest

The authors declare no conflict of interest.

Author details


Zamir Ben-Hur^{1*}, David Alon¹, Or Berebi², Ravish Mehra¹ and Boaz Rafaely²

¹ Facebook Reality Labs Research, Facebook, USA

² School of Electrical and Computer Engineering, Ben-Gurion University of the Negev, Israel

*Address all correspondence to: zamirbh@fb.com

IntechOpen

© 2021 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. 

References

- [1] Begault DR. 3-D sound for virtual reality and multimedia. NASA, Ames Research Center, Moffett Field, California. 2000:132–136.
- [2] Vorländer M. Auralization: fundamentals of acoustics, modeling, simulation, algorithms and acoustic virtual reality. Springer Science & Business Media; 2007.
- [3] Blauert J. Spatial hearing: the psychophysics of human sound localization. MIT press; 1997.
- [4] Møller H. Fundamentals of binaural technology. Applied acoustics. 1992;36(3):171–218.
- [5] Xie B. Head-related transfer function and virtual auditory display. J. Ross Publishing; 2013.
- [6] Brandstein M, Ward D. Microphone arrays: signal processing techniques and applications. Springer Science & Business Media; 2013.
- [7] Duraiswami R, Zotkin DN, Li Z, Grassi E, Gumerov NA, Davis LS. High order spatial audio capture and binaural head-tracked playback over headphones with HRTF cues. 119th Convention of Audio Engineering Society. 2005.
- [8] Sheaffer J, Van Walstijn M, Rafaely B, Kowalczyk K. Binaural reproduction of finite difference simulations using spherical array processing. IEEE/ACM Transactions on Audio, Speech, and Language Processing. 2015;23(12):2125–2135.
- [9] Gerzon MA. Ambisonics in multichannel broadcasting and video. Journal of the Audio Engineering Society. 1985;33(11):859–871.
- [10] Rafaely B, Avni A. Interaural cross correlation in a sound field represented by spherical harmonics. The Journal of the Acoustical Society of America. 2010;127(2):823–828.
- [11] Zotter F, Frank M. Ambisonics: A practical 3D audio theory for recording, studio production, sound reinforcement, and virtual reality. Springer Nature; 2019.
- [12] Jeffet M, Shabtai NR, Rafaely B. Theory and perceptual evaluation of the binaural reproduction and beamforming tradeoff in the generalized spherical array beamformer. IEEE/ACM Transactions on Audio, Speech, and Language Processing. 2016;24(4):708–718.
- [13] Alon DL, Rafaely B. Beamforming with optimal aliasing cancelation in spherical microphone arrays. IEEE/ACM Transactions on Audio, Speech, and Language Processing. 2015;24(1):196–210.
- [14] Rafaely B. Plane-wave decomposition of the sound field on a sphere by spherical convolution. The Journal of the Acoustical Society of America. 2004;116(4):2149–2157.
- [15] Rafaely B. Analysis and design of spherical microphone arrays. Speech and Audio Processing, IEEE Transactions on. 2005;13(1):135–143.
- [16] Noisternig M, Sontacchi A, Musil T, Holdrich R. A 3D ambisonic based binaural sound reproduction system. Journal of the Audio Engineering Society. 2003 June.
- [17] Zhang W, Abhayapala TD, Kennedy RA, Duraiswami R. Insights into head-related transfer function: Spatial dimensionality and continuous representation. The Journal of the Acoustical Society of America. 2010;127(4):2347–2357.
- [18] Avni A, Ahrens J, Geier M, Spors S, Wierstorf H, Rafaely B. Spatial

perception of sound fields recorded by spherical microphone arrays with varying spatial resolution. *The Journal of the Acoustical Society of America*. 2013;133(5):2711–2721.

[19] Ben-Hur Z, Brinkmann F, Sheaffer J, Weinzierl S, Rafaely B. Spectral equalization in binaural signals represented by order-truncated spherical harmonics. *The Journal of the Acoustical Society of America*. 2017;141(6):4087–4096.

[20] Ben-Hur Z, Alon DL, Rafaely B, Mehra R. Loudness stability of binaural sound with spherical harmonic representation of sparse head-related transfer functions. *EURASIP Journal on Audio, Speech, and Music Processing*. 2019 Mar;2019(1):5.

[21] Ahrens J, Andersson C. Perceptual evaluation of headphone auralization of rooms captured with spherical microphone arrays with respect to spaciousness and timbre. *The Journal of the Acoustical Society of America*. 2019; 145(4):2783–2794.

[22] Politis A, McCormack L, Pulkki V. Enhancement of ambisonic binaural reproduction using directional audio coding with optimal adaptive mixing. In: 2017 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA). IEEE; 2017. p. 379–383.

[23] Politis A, Tervo S, Pulkki V. COMPASS: Coding and Multidirectional Parameterization of Ambisonic Sound Scenes. In: 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP); 2018. p. 6802–6806.

[24] McCormack L, Politis A. SPARTA & COMPASS: Real-time implementations of linear and parametric spatial audio reproduction and processing methods. *Journal of the Audio Engineering Society*. 2019 March.

[25] Barrett N, Berge S. A new method for B-format to binaural transcoding. *Journal of the Audio Engineering Society*. 2010 October.

[26] Brinkmann F, Weinzierl S. Comparison of head-related transfer functions pre-processing techniques for spherical harmonics decomposition. *Journal of the Audio Engineering Society*. 2018 August.

[27] Evans MJ, Angus JA, Tew AI. Analyzing head-related transfer function measurements using surface spherical harmonics. *The Journal of the Acoustical Society of America*. 1998;104(4):2400–2411.

[28] Romigh GD, Brungart DS, Stern RM, Simpson BD. Efficient real spherical harmonic representation of head-related transfer functions. *IEEE Journal of Selected Topics in Signal Processing*. 2015;9(5):921–930.

[29] Ben-Hur Z, Alon DL, Mehra R, Rafaely B. Efficient Representation and Sparse Sampling of Head-Related Transfer Functions Using Phase-Correction Based on Ear Alignment. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*. 2019; 27(12):2249–2262.

[30] Zaunschirm M, Schörkhuber C, Höldrich R. Binaural rendering of Ambisonic signals by head-related impulse response time alignment and a diffuseness constraint. *The Journal of the Acoustical Society of America*. 2018; 143(6):3616–3627.

[31] Schörkhuber C, Zaunschirm M, Höldrich R. Binaural rendering of ambisonic signals via magnitude least squares. In: *Proceedings of the DAGA*. vol. 44; 2018. p. 339–342.

[32] Wightman FL, Kistler DJ. The dominant role of low-frequency interaural time differences in sound localization. *The Journal of the*

Acoustical Society of America. 1992;91 (3):1648–1661.

[33] Macpherson EA, Middlebrooks JC. Listener weighting of cues for lateral angle: the duplex theory of sound localization revisited. *The Journal of the Acoustical Society of America*. 2002;111 (5):2219–2236.

[34] Minnaar P, Christensen F, Moller H, Olesen SK, Plogsties J. Audibility of all-pass components in binaural synthesis. *Journal of the Audio Engineering Society*. 1999 may.

[35] Benichoux V, Rébillat M, Brette R. On the variation of interaural time differences with frequency. *The Journal of the Acoustical Society of America*. 2016;139(4):1810–1821.

[36] Lübeck T, Helmholtz H, Arend JM, Pörschmann C, Ahrens J. Perceptual Evaluation of Mitigation Approaches of Impairments due to Spatial Undersampling in Binaural Rendering of Spherical Microphone Array Data. *Journal of the Audio Engineering Society*. 2020;68(6):428–440.

[37] Hold C, Gamper H, Pulkki V, Raghuvanshi N, Tashev IJ. Improving binaural ambisonics decoding by spherical harmonics domain tapering and coloration compensation. In: *ICASSP 2019–2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE; 2019. p. 261–265.

[38] Jôt JM, Wardle S, Larcher V. Approaches to binaural synthesis. *Journal of the Audio Engineering Society*. 1998 september.

[39] Jot JM, Larcher V, Pernaux JM. A comparative study of 3-d audio encoding and rendering techniques. *Journal of the Audio Engineering Society*. 1999 march.

[40] Larcher V, Warusfel O, Jot JM, Guyard J. Study and comparison of

efficient methods for 3-d audio spatialization based on linear decomposition of hrtf data. *Journal of the Audio Engineering Society*. 2000 february.

[41] Ben-Hur Z, Alon D, Mehra R, Rafaely B. Binaural reproduction using bilateral ambisonics. *Journal of the Audio Engineering Society*. 2020 august.

[42] Ben-Hur Z, Alon DL, Mehra R, Rafaely B. Binaural Reproduction Based on Bilateral Ambisonics and Ear-Aligned HRTFs. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*. 2021;29:901–913.

[43] Park M, Rafaely B. Sound-field analysis by plane-wave decomposition using spherical microphone array. *The Journal of the Acoustical Society of America*. 2005;118(5):3094–3103.

[44] Ben-Hur Z, Sheaffer J, Rafaely B. Joint sampling theory and subjective investigation of plane-wave and spherical harmonics formulations for binaural reproduction. *Applied Acoustics*. 2018;134:138–144.

[45] mh acoustics. em32 Eigenmike microphone array release notes; 2009. 25 Summit Ave Summit NJ 07901, <http://www.mhacoustics.com/products#eigenmike1>.

[46] Rafaely B. *Fundamentals of Spherical Array Processing*. vol. 8. Springer; 2015.

[47] Pörschmann C, Arend JM, Brinkmann F. Directional Equalization of Sparse Head-Related Transfer Function Sets for Spatial Upsampling. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*. 2019; 27(6):1060–1071.

[48] Ben-Hur Z, Alon DL, Mehra R, Rafaely B. Sparse Representation of HRTFs by Ear Alignment. In: *2019 IEEE Workshop on Applications of Signal*

Processing to Audio and Acoustics (WASPAA). IEEE; 2019. p. 70–74.

[49] Williams EG. Fourier acoustics: sound radiation and nearfield acoustical holography. Academic press; 1999.

[50] Arfken GB, Weber HJ, Harris FE. Mathematical Methods for Physicists: A Comprehensive Guide. Elsevier; 2012. Available from: <https://books.google.com/books?id=qLFoZ-PoGIC>.

[51] Ward DB, Abhayapala TD. Reproduction of a plane-wave sound field using an array of loudspeakers. IEEE Transactions on speech and audio processing. 2001;9(6):697–707.

[52] Dinakaran M, Brinkmann F, Harder S, Pelzer R, Grosche P, Paulsen RR, et al. Perceptually motivated analysis of numerically simulated head-related transfer functions generated by various 3D surface scanning systems. In: 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE; 2018. p. 551–555.

[53] Katz BF, Noisternig M. A comparative study of interaural time delay estimation methods. The Journal of the Acoustical Society of America. 2014;135(6):3530–3540.

[54] Andreopoulou A, Katz BF. Identification of perceptually relevant methods of inter-aural time difference estimation. The Journal of the Acoustical Society of America. 2017;142(2):588–598.

[55] Rafaely B, Kleider M. Spherical microphone array beam steering using Wigner-D weighting. IEEE Signal Processing Letters. 2008;15:417–420.

[56] algazi vr, duda ro, thompson dm. motion-tracked binaural sound. Journal of the audio engineering society. 2004 november;52(11):1142–1156.

[57] Berebi O, Ben-Hur Z, Alon D, Rafaely B. Enabling Head-Tracking for Binaural Sound Reproduction Based on Bilateral Ambisonics. International Conference on Immersive and 3D Audio (I3DA). 2021.

[58] Arfken GB, Weber HJ. Mathematical methods for physicists. American Association of Physics Teachers; 1999.

[59] Lecomte P, Gauthier PA, Langrenne C, Garcia A, Berry A. On the use of a Lebedev grid for Ambisonics. Journal of the Audio Engineering Society. 2015 October.

[60] Slaney M. Auditory toolbox. Interval Research Corporation, Tech Rep. 1998;10:1998.

[61] Mills AW. Lateralization of high-frequency tones. The Journal of the Acoustical Society of America. 1960;32(1):132–134.

[62] Yost WA, Dye Jr. RH. Discrimination of interaural differences of level as a function of frequency. The Journal of the Acoustical Society of America. 1988;83(5):1846–1851.