Binaural Ambisonics: Its optimization and applications for auralization

Makoto Otani*, Haruki Shigetani, Masataka Mitsuishi and Ryo Matsuda

Graduate School of Engineering, Kyoto University, Kyoto Daigaku-Katsura, Nishikyo-ku, Kyoto, 615–8540 Japan

Abstract: To better understand acoustic environment and the resulting auditory perception, it is essential to capture, analyze, and reproduce a sound field as a three-dimensional physical phenomenon because spatial aspects of auditory perception play important roles in various situations in our lives. Some approaches have been proposed to achieve the three-dimensional capture and reproduction of acoustic fields. Among them, Higher-Order Ambisonics (HOA) based on spherical harmonics expansion enables the capture and reproduction of a directivity pattern of incoming sound waves. On the basis of HOA, three-dimensional auditory space can be presented to a listener typically via a spherical loudspeaker array. In addition, binaural synthesis emulating the loudspeaker presentation enables HOA reproduction with a set of headphones or several loudspeakers by employing crosstalk cancellation. Thus, we are developing an HOA-based binaural reproduction/auralization system with head tracking. This system is aimed at realizing the reproduction and auralization of a sound field, including one excited by the listener’s own voice. In this paper, we review the topics related to the reproduction and auralization of the sound field and introduce the HOA-based binaural synthesis system we have developed, as well as our works on sweet-spot expansion in HOA decoding and self-voice reproduction/auralization.

Keywords: Sound field reproduction, Auralization, Higher-Order Ambisonics, Dynamic crosstalk cancellation, Self-voice auralization

PACS number: 43.60.Tj, 43.20.Bi [doi:10.1250/ast.41.142]

1. INTRODUCTION

To understand acoustic phenomena and the resulting auditory perception that we experience in our daily lives, we examine their spatial aspects, which play prominent roles. Therefore, it is essential to consider an acoustical phenomenon from the three-dimensional physical and perceptual viewpoints. Particularly in the field of the reproduction and auralization of sound fields, many researchers have been putting much effort into its theories and practices.

Technically, one of the ultimate goals of sound field reproduction/auralization is to control binaural signals at the listener’s two ears in the reproduced field so that they are identical to those observed in the primary field to be reproduced, that is a real acoustic field for sound field reproduction and a simulated acoustic field for auralization. To achieve this goal, some approaches have been proposed. One is based on boundary integral equations whose boundary surrounds a finite or half-infinite region in the sound field to be reproduced, for example, wave field synthesis (WFS) [1] or boundary surface control (BoSC) [2], by which the sound pressure and/or its gradient on the boundary is controlled to reproduce the interior sound field. Another is based on spherical harmonics presentation of the sound field, so-called Higher-Order Ambisonics (HOA) [3], which captures and reproduces the directivity of incident sound waves by employing spherical harmonics expansion. A third is binaural reproduction/synthesis, which reproduces the acoustic signals at the eardrum or entrance to the ear canal. In this review, we classify the former two approaches as field reproduction and the last one as binaural reproduction.

Field reproduction is intended to reproduce the sound field using a loudspeaker array equipped with a number of loudspeakers, whereas binaural reproduction is inherently suitable for headphone presentation as it requires only two output channels. Although the mathematical backgrounds established for field reproduction are quite useful and effective, their practical developments have been limited by the hardware resources available at the time of constructing the large-scale loudspeaker array. On the other hand, some hybrid approaches combining field reproduction and binaural reproduction have been proposed. Takane et al. proposed a theory for headphone presentation of field reproduction using an integral...
equation, and called it ADVISE (Acoustic Display based on the Virtual SpherE model) [4]. Furthermore, Noisternig et al. proposed virtual Ambisonics, which emulates HOA reproduction by binaural synthesis [5]. These hybrid approaches enable a system for headphone presentation of field reproduction, wherein physical sound wave propagation from secondary sound sources to the listener’s two ears in a reproduced field is replaced by binaural synthesis through signal processing using corresponding head-related transfer functions (HRTFs). Although such hybrid approaches suffer from the same problems as the binaural reproduction itself, such as interindividual differences in HRTFs, it is no longer restricted by the hardware resources for loudspeaker array implementation as far as the computation load required for signal processing does not exceed the available computational resources.

This review focuses on binaural Ambisonics, or virtual Ambisonics, which is based on the theory of HOA for the capture and reproduction of the sound field combined with binaural synthesis that enables a headphone presentation of field reproduction based on HOA. In addition, we introduce our work on expanding the sweet spot, which is rather narrow and therefore a drawback to HOA reproduction, by introducing a least-norm solution and Max-r-E into HOA decoding. Next, we demonstrate the development of self-voice auralization using virtual Ambisonics, which enables a speaker to listen to the room acoustics response of his/her own voice.

2. BINAURAL AMBISONICS

2.1. Previous Works

Several methods for obtaining binaural signals to be reproduced at the listener’s ears have been studied that employ microphone array recordings and HRTF datasets, not limited to ones employing the HOA theory. These methods were comprehensively reviewed and introduced in [6]. These methods are categorized into the HRTF modeling approach and the microphone signal modeling approach, as mentioned in [6]. So-called binaural Ambisonics is categorized in the latter category, examples of which can be found in [5,7–13]. Audio output signals radiated from loudspeakers in a spherical array can be directly synthesized with HRTFs. Otherwise, it is also possible to synthesize the spatial information captured using the microphone array with HRTFs in the spherical harmonics domain by employing the spherical harmonics transformation of HRTF datasets, e.g., [11,14]. In the following subsections, we introduce the basic theory of HOA encoding and decoding and binaural synthesis when employing a spherical microphone array recording and direct synthesis of HRTF, as well as dynamic crosstalk cancellation that can be applied to realize binaural Ambisonics reproduction through a small number of loudspeakers instead of headphone presentation.

2.2. Higher-Order Ambisonics

2.2.1. Encoding from rigid spherical microphone array

The HOA theory mainly consists of encoding and decoding, which correspond to spherical harmonics expansion of recorded signals in the primary field and the determination of driving signals of secondary sources for the reproduced field, respectively. Here, for HOA encoding, let us assume that a rigid spherical microphone array is located in the primary field.

First, let us express the solution of the Helmholtz equation in a spherical coordinate system [15]. The sound pressure \( s_{\text{inc}} \) at observation point \( r = (r, \theta, \phi) \) driven by a unit amplitude plane wave, whose incident angle is \( k_{\text{inc}} = (\theta_{\text{inc}}, \varphi_{\text{inc}}) \), can be described as

\[
s_{\text{inc}}(r) = 4\pi \sum_{n=0}^{\infty} \sum_{m=-n}^{n} j_n(kr) Y_n^m(\theta, \phi) Y_n^m(\theta_{\text{inc}}, \varphi_{\text{inc}})^*,
\]

where \( k \) is the wave number; \( i \) is the imaginary unit; \( j_n \) is the spherical Bessel function; \( Y_n^m \) is the spherical harmonics function; and the superscript * denotes the complex conjugate. If we assume that a rigid sphere whose center is at the origin and radius is \( a \), the scattered field \( s_{\text{scatter}} \) is expressed as

\[
s_{\text{scatter}}(r) = \sum_{n=0}^{\infty} C_n h_n(kr) \sum_{m=-n}^{n} Y_n^m(\theta, \phi) Y_n^m(\theta_{\text{inc}}, \varphi_{\text{inc}})^*,
\]

where \( C_n \) is an unknown constant; \( h_n \) is the spherical Hankel function. The total sound pressure field is written as \( s_{\text{inc}} + s_{\text{scatter}} \) and, by applying the boundary condition

\[
\frac{\partial(s_{\text{inc}} + s_{\text{scatter}})}{\partial r} \bigg|_{r=a} = 0,
\]

the scattered field is derived as

\[
s_{\text{scatter}}(r) = -4\pi \sum_{n=0}^{\infty} \sum_{m=-n}^{n} j_n'(ka) h_n^0(ka) j_n(kr) Y_n^m(\theta, \phi) Y_n^m(\theta_{\text{inc}}, \varphi_{\text{inc}})^*.
\]

where \( j_n' \) and \( h_n^0 \) are the derivatives of the spherical Bessel and Hankel functions, respectively. Therefore, the total sound pressure field \( s \) is described as

\[
s(r) = 4\pi \sum_{n=0}^{\infty} \sum_{m=-n}^{n} \left( j_n(kr) - j_n'(ka) h_n^0(ka) \right) h_n^0(ka) \sum_{m=-n}^{n} Y_n^m(\theta, \phi) Y_n^m(\theta_{\text{inc}}, \varphi_{\text{inc}})^*.
\]

Here, by assuming that the observation point is on the rigid sphere and by applying

\[
j_n(x) - j_n'(x) h_n^0(x) = \frac{-i}{h_n^0(x)x^2},
\]

143
derived from the Wronskian relationship, we obtain the sound pressures on the rigid sphere surface as

\[ s(r) = 4\pi \sum_{n=0}^{\infty} \sum_{m=-n}^{n} B_n^m Y_n^m(\theta, \varphi), \]

where \( B_n^m \) is the spherical harmonics coefficient, which is unknown.

Now, if we assume \( Q \) observation points at \( r_q = (\alpha_q, \theta_q, \varphi_q) \) \( (q = 1, \ldots, Q) \) on the rigid sphere and let \( W_n = 4\pi r_0^{n-1}/h_n(ka)^2 \), the above equation is rewritten as

\[ s(r_q) = \sum_{n=0}^{N} W_n \sum_{m=-n}^{n} B_n^m Y_n^m(\theta_q, \varphi_q), \quad (1) \]

where the order of spherical harmonics expansion is truncated to \( N \). The radial function \( W_n \) is defined in accordance with the type of spherical array, such as open array or dual array [16]. Let \( s = [s(r_1), \ldots, s(r_Q)] \) be the vector consisting of the observed sound pressure at each observation point for wave number \( k \). Then, Eq. (1) is expressed in matrix form as

\[ s = Y \cdot \text{diag}[W_n] \cdot B, \]

where \( \text{diag}[\cdot] \) is the diagonal matrix; \( Y \) and \( B \) are matrix forms of the spherical harmonics function \( Y_n^m(\theta, \varphi) \) and its coefficient \( B_n^m \). Then, the spherical harmonics coefficient, or expansion coefficient, \( B \) is derivable as

\[ B = \text{diag} \left( \frac{1}{W_n} \right) \cdot Y^+ \cdot s, \]

where \( Y^+ \) is the Moore–Penrose pseudo inverse matrix of \( Y \).

2.2.2. Decoding to spherical loudspeaker array

In HOA decoding, the driving signal \( G \) of secondary sound sources is derived from \( B \) as

\[ G = D \cdot B, \quad \text{and} \quad G = [G_1, \ldots, G_L]^T. \]

\[ D = \begin{cases} \begin{bmatrix} Y_L^1 \\ Y_L^{H}(Y_L^1)^{H}^{-1} \end{bmatrix} & L = (N + 1)^2, \\ \begin{bmatrix} Y_L^0(\theta_l, \varphi_l) & \cdots & Y_L^0(\theta_L, \varphi_L) \\ \vdots & \ddots & \vdots \\ Y_L^N(\theta_L, \varphi_L) & \cdots & Y_L^N(\theta_L, \varphi_L) \end{bmatrix} & L > (N + 1)^2. \end{cases} \quad (3) \]

where \( D \) is the decoding matrix; \( Y_L \) is the matrix consisting of spherical harmonics functions for the sound sources, which radiate plane waves, located in the polar directions \( \theta_l = (\theta_l, \varphi_l) \) \( (l = 1, \ldots, L) \); the superscript \( H \) denotes the conjugate transpose. The HOA decoding of an \( N \)-th order Ambisonics signal requires no less than \((N + 1)^2\) sound sources. As written in Eq. (3), if \( L = (N + 1)^2 \), the decoding matrix \( D \) is derived as the inverse of \( Y_L \), whereas, if \( L > (N + 1)^2 \), \( D \) is obtainable as a least-norm solution.

In contrast to the above-mentioned conventional HOA decoding, Max-rE decoding was proposed to improve sound localization at higher frequencies by optimizing an energy vector in the reproduced field. In Max-rE decoding, the driving signal \( G \) is obtained as

\[ G = D \cdot \text{diag}[a_n] \cdot B, \quad (4) \]

where \( a_n = P_n \left( \cos \left( \frac{137.9^\circ}{N + 1.51} \right) \right) \); \( P_n(x) \) is the Legendre polynomial for the expansion order \( n \) [17,18].

2.3. Binaural Synthesis

The HOA reproduction system can be emulated by binaural synthesis under the assumption that a listener’s head is located at the center of the loudspeaker array in the reproduction field. Binaural signals to be reproduced at the listener’s two ears are generated by convolving driving signals \( D \) with the head-related impulse responses (HRIRs) for each loudspeaker position and then by summing all the resulting signals.

The HRIR between each loudspeaker and the listener varies in accordance with the listener’s head rotation. Therefore, usually, to allow the listener’s head rotation and realize head-tracking binaural synthesis, a set of HRIRs with dense spatial resolution is required. On the other hand, as in [5], if we assume a virtual loudspeaker array that rotates in sync with the listener’s head, a head-tracking binaural synthesis is realizalbe with a fixed set of HRIRs by generating decoding matrix \( D \) and driving signal \( B \) in real time. This approach drastically reduces the required number of HRIRs and avoids auditory artifacts due to HRIR switching [19].

2.4. Development

In this subsection, we introduce the HOA-based binaural synthesis system that we developed for the reproduction and auralization of a sound field.

Figure 1 illustrates the overview of the binaural Ambisonics system we developed. As mentioned above, the sound pressure signals are recorded using a spherical microphone array in real or simulated sound fields and
then are encoded to expansion coefficients \( B \). Next, the decoding matrix \( D \) and the driving signals \( G \) are computed in real time in accordance with the listener’s head rotation, which is captured using a head-tracking device. Finally, the driving signals \( G \) are convolved with the corresponding HRIRs to obtain the binaural signals that are to be presented through a set of headphones.

To capture the sound field, a spherical microphone array can be employed, such as the ViReal™ Mic (YAMAHA) [20], which is a rigid sphere with 64 microphones installed on its surface. Audio signal processing, including real-time decoding and HRIR convolution, is conducted in Max (Cycling ’74). The listener’s head rotation is captured by Colibri (Trivisio) or any head-tracking device. The HRIRs for binaural synthesis are measured or numerically simulated in advance for given locations of secondary sources or virtual loudspeakers. In our current work, HRIRs are numerically simulated by the boundary element simulation using three-dimensional models of real and dummy heads such as KEMAR (GRAS) or 4128C (Brüel & Kjær) [21]. In the boundary element simulation, we computed a set of HRIRs for plane waves, which can be directly applied to the general HOA theory utilizing plane wave decomposition of the sound field. Note that HRTFs are typically measured for point sources but in the current work, we employ HRTFs for plane waves. This is because, the decoding equations, i.e., Eqs. (2) and (3), are independent of the frequency for plane wave decomposition but dependent on frequency in decoding to spherical waves, which necessitates another FIR filtering process and therefore a considerable computational cost. Furthermore, HRTFs can be computed for plane waves when using numerical methods such as described in [21].

In addition, it has been reported that HRTFs for spherical waves are no longer dependent on the source distance when it is larger than 1 m and are almost identical to those for plane waves [22].

2.5. Loudspeaker Presentation through Dynamic Crosstalk Cancellation

Binaural signals can also be presented via a small number of loudspeakers by utilizing crosstalk cancellation (CTC) [23], which cancels unwanted acoustic transmission from the loudspeakers to the listener’s two ears. The conventional CTC requires a listener to remain still because its inverse filters depend on acoustic transfer functions between the loudspeakers and the listener’s two ears. However, dynamic CTC has been developed so that the listener can move freely, by employing the head-tracking device and real-time inverse filter computation [24–26]. Such a dynamic CTC system can be applied to HOA-based binaural synthesis, enabling its presentation with a small number of loudspeakers, as shown in Fig. 2. The CTC filter, in addition to the decoding matrix, are generated in real time in accordance with the head rotation detected by a head-tracking device. Because the listener no longer requires a set of headphones in this case, it would be preferable to use a noncontact head-tracking device.

3. OPTIMIZATION FOR AURALIZATION

3.1. Sweet Spot in HOA Reproduction

In practice, HOA reproduction suffers from its small sweet-spot size. Namely, the region in which the reproduction error is below a certain criterion is small. The radius of a spherical sweet spot, formed around the reproduction point, is proportional to the expansion order
and inverse of frequency [27]. Therefore, at higher frequencies, the listener’s two ears would no longer be within the sweet spot, depending on the expansion order. This leads to inaccurate auditory spatial perception, especially at higher frequencies or when employing low-order expansion, and is one of the major drawbacks to HOA reproduction. Because such auditory artifacts originate from HOA theory, which is the basis of binaural Ambisonics, they are also observed in reproduction using binaural Ambisonics.

Researchers have been putting effort into evaluating the perceptual effects of such artifacts of binaural Ambisonics owing to low order or a small narrow sweet spot (e.g., [11,28]). Some researchers have proposed the optimization of binaural Ambisonics, such as spherical harmonics domain tapering and coloration compensation [29], interaural level difference optimization [30], diffuse field equalization [31], spectral equalization [32], timbre correction [12], adjustment of the sweet spot by time alignment [33], and other approaches [34,35].

In HOA reproduction, it is important to position the listener’s ears inside the sweet spot for physically and perceptually accurate reproduction. Particularly for binaural Ambisonics, in which the binaural signals to be reproduced at both ears can be controlled separately, the sweet spot can be optimized for each ear by employing some modifications in the decoding process, e.g., the time alignment [33]. However, the following subsection introduces our investigation on how the conventional approaches in HOA decoding, i.e., Max-rE decoding and/or least-norm solution, contribute to the expansion of the sweet spot in HOA reproduction and binaural Ambisonics.

3.2. Sweet-Spot Expansion

It has been reported that the sweet spot can be broadened by employing Max-rE in HOA decoding [3,36], where the energy vector in the reproduction field is optimized, as formulated in Eq. (4). In addition, it is indicated that the use of a larger number of secondary sources than the minimum requirement for deriving a decoding matrix, namely, employing the least-norm solution for the case of $L > (N+1)^2$ in Eq. (3), also contributes to a broader sweet spot [37].

Figure 3 shows the results of a numerical comparison. The primary field is excited by a single plane wave at 1 kHz. The reproduced fields are generated with and without Max-rE and the least-norm solution. The secondary sources are distributed on a sphere as uniformly as possible on the basis of [38]. The figure illustrates that the conventional case, $L = (N+1)^2$, yields a spherical sweet spot whose radius roughly corresponds to the one derived in the literature [27] as $N = kx$, where $x$ is the radius. In contrast, the sweet spot is expanded when Max-rE [$L = (N+1)^2$ w/ Max-rE] or the least-norm solution [$L = (N+2)^2$] is employed. Furthermore, note that the distortions outside the sweet spot are significantly improved in these cases compared with the conventional case.

Finally, in the case of employing both Max-rE and the least-norm solution, the sweet spot is further expanded along the incident direction.

In addition to the reproducibility of the sound field by each of the decoding approaches with and without Max-rE and/or the least-norm solution, the reproducibility of binaural signals has also been examined by a boundary element simulation of binaural signals in HOA reprodu-
The computer model of the dummy head used for the numerical simulation is shown in Fig. 4. Figure 5 illustrates the magnitude spectra of binaural signals observed at the left ear of the dummy head for \( N = 10 \) and \( L = (N + 1)^2 \) and \( (N + 2)^2 \) in the primary and reproduced fields. The abscissa and ordinate represent the frequency in kHz and azimuthal angle in degrees, respectively. The color represents the magnitude in dB in accordance with the color bar. Figure 5 also reveals that both the Max-rE and the least-norm solution lead to sweet-spot expansion.

4. SELF-VOICE AURALIZATION

4.1. Background

Hearing sounds that are generated by the listener and are reflected by the surrounding environment as feedback is also important in spatial auditory perception. For example, it is known that a singer may adjust his/her way of singing depending on the auditory feedback from room acoustics. However, in the acoustic design of architectural spaces for music performances, often only the audience’s perception is taken into account. Thus, auralization of the voices of the speaker or singer would lead to more appropriate acoustic designs for them.

Yadav et al. proposed a method for the auralization of the listener’s own voice using oral-binaural room impulse responses (OBRIRs) that represent acoustic transfer functions between the mouth and both ears of the speaker/singer followed by real-time binaural synthesis and headphone presentation [40]. Because the OBRIR is obtainable through measurement using a real or dummy head in the actual sound field, a number of OBRIR measurements with various facing directions is necessary to realize dynamic binaural presentation allowing listener’s head rotation.

4.2. Self-Voice Auralization Using HOA-based Binaural Synthesis

Aiming at the development of a self-voice auralization system that allows easier dynamic binaural presentation, we have attempted to introduce binaural Ambisonics to self-voice auralization. The employment of HOA-based sound-field capture using a spherical microphone array enables us to realize dynamic binaural synthesis for self-voice auralization more easily. Namely, instead of OBRIR measurement using a real/dummy head, room impulse responses (RIRs) from a loudspeaker, which is located close to the microphone array and thereby corresponds to the mouth of the speaker/singer, to the microphones are measured using a spherical microphone array. Then, the RIRs can be encoded and decoded on the basis of HOA theory and then converted to the binaural signals that will be reproduced at the ears of the listener (speaker/singer), through dynamic binaural synthesis.

Figure 6(a) illustrates an overview of the self-voice auralization system based on OBRIRs. In the primary field, an OBRIR \( h(t) \) consists of direct components from the
mouth to both ears, including air- and body-conducted paths, \( h_d(t) \), and indirect components that are feedback from the room, \( h_i(t) \). Assuming that \( x(t) \) denotes the speech signal, binaural signal \( y(t) \) observed at the speaker’s ear is expressed as

\[
y(t) = h_d(t) * x(t) \quad \text{and} \quad y(t) = h_i(t) * x(t)
\]

where \( * \) denotes the convolution. If we assume that the reproduction is performed in an anechoic environment and the set of headphones is acoustically transparent, the direct sound \( h_d(t) * x(t) \) is also naturally reproduced. Therefore, only the indirect sound \( h_i(t) * s(t) \) is to be reproduced artificially. The direct component \( h_d(t) \) can be easily removed from the OBRIR \( h(t) \).

The one drawback to the OBRIR-based self-voice auralization is that it requires many OBRIR measurements with various facing directions of the speaker to allow the speaker to rotate their head. On the other hand, with a spherical microphone array, the RIR measurement needs to be performed only once, because the binaural Ambisonics deals with the speaker’s head rotation in the reproduction phase. Figure 6(b) depicts an overview of the self-voice auralization system based on binaural Ambisonics. The RIRs from a loudspeaker located near the spherical microphone array to each microphone in the array are measured. Then, indirect components are extracted from the RIRs, resulting in \( s(t) \), which corresponds to \( s \) in Fig. 1. Then, \( s \) is encoded to the spherical harmonics coefficients \( B \), which are stored for the reproduction phase. In the reproduction, \( B \) is decoded with decoding matrix \( C^+ \) generated in real time in accordance with the speaker’s head rotation detected by a head tracker, and then the driving signals \( p \) for a virtual loudspeaker array are obtained. The driving signals are subsequently converted to indirect components of OBRIRs that reflect the direction in which the speaker is facing. These are called HOA-synthesized OBRIRs in the figure. Finally, the speech signal, \( x(t) \), recorded by a microphone located near the speaker’s mouth, is convolved with them in real time, and then the feedback from the room is presented to the speaker him/herself through a set of headphones.

One disadvantage of using binaural Ambisonics for self-voice auralization is that the loudspeaker emulating the speaker’s mouth in the primary field is located at a fixed position whereas the actual position of the speaker’s mouth varies with the speaker’s head rotation. Furthermore, the directional radiation pattern of the loudspeaker must be different from that of the actual speaker. In addition, the
audio signal processing results in nonnegligible system latency that leads to excessively delayed reflections, which may affect the spatial perception of room acoustics. These issues should be addressed in future work.

5. CONCLUSIONS

We first reviewed the background of the reproduction and auralization of sound fields focusing on the theory of HOA and binaural synthesis. Then, we introduced our development of a binaural Ambisonics system aimed at the realization of the reproduction and auralization of sound field using the HOA theory while enabling binaural presentation through a set of headphones and several loudspeakers with dynamic crosstalk cancellation. In addition, we introduced our works on expanding the sweet spot through HOA decoding to overcome the narrow reproduction area, which is the major disadvantage of HOA. Finally, we also introduced self-voice reproduction/auralization using the binaural Ambisonics approach.

ACKNOWLEDGMENT

This work is partly supported by grants-in-aid from JSPS, Japan (No. 16H02857 and 17KT0137).

REFERENCES

[1] A. J. Berkhourt, “A holographic approach to acoustic control,” J. Audio Eng. Soc., 36, 977–995 (1988).
[2] S. Ise, “A principle of sound field control based on the Kirchhoff-Helmholtz integral equation and the theory of inverse systems,” Acustica, 85, 78–87 (1999).
[3] J. Daniel, R. Nicol and S. Moreau, “Further investigations of High Order Ambisonics and wavefield synthesis for holographic sound imaging,” Proc. 114th Audio Eng. Soc. Conv., 5788 (2003).
[4] S. Takane, Y. Suzuki, T. Miyajima and T. Sone, “A new theory for high definition virtual acoustic display named ADVISE,” Acoust. Sci. & Tech., 24, 276–283 (2003).
[5] M. Noisternig, A. Sontacchi, T. Musil and R. Holdrich, “A 3D Ambisonic based binaural sound reproduction system,” Proc. Audio Eng. Soc. Int. Conf., 1 (2003).
[6] C. D. Salvador, S. Sakamoto, J. Treviño and Y. Suzuki, “Design theory for binaural synthesis: Combining microphone array recording and head-related transfer function databases,” Acoust. Sci. & Tech., 38, 51–62 (2017).
[7] R. Duraiswami, D. N. Zotkin, Z. Li, E. Grassi, N. A. Gumerov and L. S. Davis, “High order spatial audio capture and its binaural head-tracked playback over headphones with HRTF cues,” Proc. 119th Audio Eng. Soc. Conv. (2005).
[8] A. Avni, J. Ahrens, M. Geier, S. Spors, H. Wierstorf and B. Rafaeley, “Spatial perception of sound fields recorded by spherical microphone arrays with varying spatial resolution,” J. Acoust. Soc. Am., 133, 2711–2721 (2013).
[9] C. D. Salvador, S. Sakamoto, J. Treviño and Y. Suzuki, “Numerical evaluation of binaural synthesis from rigid spherical microphone array recording,” Proc. Audio Eng. Soc. Int. Conf. Headphone Tech (2016).
[10] C. D. Salvador, S. Sakamoto, J. Treviño and Y. Suzuki, “Spatial accuracy of binaural synthesis from rigid spherical microphone array recordings,” Acoust. Sci. & Tech., 38, 23–30 (2017).
[11] B. Bernschatz, A. V. Giner, C. Porschmann and J. Arend, “Binaural reproduction of plane waves with reduced modal order,” Acta Acust. united Ac., 85, 250–257 (2014).
[12] J. Sheaffer, S. Villeval and B. Rafaeley, “Rendering binaural room impulse responses from spherical microphone array recordings using timbre correction,” Proc. EAA J. Symp. Auralization and Ambisonics (2014).
[13] J. Sheaffer, M. van Walstijn, B. Rafaeley and K. Kowalczyk, “Binaural reproduction of finite difference simulation using spherical array processing,” IEEE/ACM Trans. Audio Speech Lang. Process., 23, 2125–2135 (2015).
[14] E. Rasumow, M. Blau, M. Hansen, S. van de Par, S. Dolco, V. Mellert and D. Puschel, “Smoothing individual head-related transfer functions in the frequency and spatial domain,” J. Acoust. Soc. Am., 135, 2012–2025 (2014).
[15] E. G. Williams, Fourier Acoustics: Sound Radiation and Nearfield Acoustic Holography (Academic Press, London, 1999).
[16] L. Balmages and B. Rafaeley, “Open-sphere designs for spherical microphone arrays,” IEEE Trans. Audio Speech Lang. Process., 15, 727–732 (2007).
[17] J. Daniel, J. B. Rault and J. D. Polack, “Ambisonics encoding of other audio formats for multiple listening conditions,” Proc. 105th Audio Eng. Soc. Conv., 4795 (1998).
[18] F. Zotter and M. Frank, “All-round Ambisonics panning and decoding,” J. Audio Eng. Soc., 60, 807–820 (2012).
[19] M. Otani and T. Hirahara, “Auditory artifacts due to switching head-related transfer functions of a dynamic virtual auditory display,” IEICE Trans. Fundam., E91-A, 1320–1328 (2008).
[20] S. Kaneko, T. Suenaga, H. Akiyama, Y. Miyake, S. Tominaga, F. Shirakihara and H. Okumura, “Development of a 64-channel spherical microphone array and a 122-channel loudspeaker array system for 3D sound field capturing and reproduction technology research,” Proc. 144th Audio Eng. Soc. Conv., 10021 (2018).
[21] M. Otani and S. Ise, “Fast calculation system specialized for head-related transfer function based on boundary element method,” J. Acoust. Soc. Am., 119, 2589–2598 (2006).
[22] M. Otani, T. Hirahara and S. Ise, “Numerical study on source distance dependency of head-related transfer functions,” J. Acoust. Soc. Am., 125, 3253–3261 (2009).
[23] M. R. Schroeder, “Digital simulation of sound transmission in reverberant spaces,” J. Acoust. Soc. Am., 47, 424–431 (1970).
[24] W. G. Gardner, “3-D audio using loudspeakers,” Massachusetts Institute of Technology, Ph.D. Thesis (1997).
[25] T. Lentz, “Dynamic crosstalk cancellation for binaural synthesis in virtual reality environment,” J. Audio Eng. Soc., 54, 283–294 (2006).
[26] H. Kurabayashi, M. Otani, K. Itoh, M. Hashimoto and M. Kayama, “Sound image localization using dynamic transaural reproduction with non-contact head tracking,” IEICE Trans. Fundam., E97-A, 1849–1858 (2014).
[27] D. Ward and T. Abhayapala, “Reproduction of a plane-wave sound field using an array of loudspeakers,” IEEE Trans. Speech Audio Process., 9, 697–707 (2001).
[28] H. Lee, M. Frank and F. Zotter, “Spatial and timbral fidelties of binaural Ambisonics decoders for main microphone array recordings,” Proc. Audio Eng. Soc. Int. Conf., 75 (2019).
[29] C. Hold, H. Gamper, V. Pulikki, N. Raghuvarshi and I. Tashev, “Improving binaural Ambisonics decoding by spherical harmonics domain tapering and coloration compensation,” Proc. ICASSP 2019, pp. 261–265 (2019).
[30] T. McKenzie, D. T. Murphy and G. Kearney, “Interaural level difference optimization of binaural Ambisonic rendering,”
[31] T. McKenzie, D. T. Murphy and G. Kearney, “Diffuse-field equalisation of binaural Ambisonic rendering,” Appl. Sci., 8, 1226 (2019).

[32] 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., 141, 4087–4096 (2017).

[33] 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., 143, 3616–3627 (2018).

[34] F. Brinkmann and S. Weinzierl, “Comparison of head-related transfer functions pre-processing techniques for spherical harmonics decomposition,” Proc. Audio Eng. Soc. Int. Conf. Audio for Virtual and Augmented Reality (2018).

[35] C. Schörkhuber, M. Zaunschirm and R. Höldrich, “Binaural rendering of Ambisonic signals via magnitude least squares,” Proc. DAGA (2018).

[36] D. Murillo, F. Fazi and M. Shin, “Evaluation of Ambisonics decoding methods with experimental measurements,” Proc. 114th Audio Eng. Soc. Conv., 5788 (2003).

[37] M. Otani and H. Shigetani, “Reproduction accuracy of Higher-Order Ambisonics with Max-rE and/or least norm solution in decoding,” Acoust. Sci. & Tech., 40, 23–28 (2019).

[38] H. Vogel, “A better way to construct the sunflower head,” Math. Biosci., 44(3–4), 179–189 (1979).

[39] H. Shigetani and M. Otani, “Accuracy of binaural signals in Higher-Order Ambisonics reproduction with different decoding approaches,” Acoust. Sci. & Tech., 40, 144–147 (2019).

[40] M. Yadav, D. Cabrera and W. Martens, “A system for simulating room acoustical environments for one’s own voice,” Appl. Acoust., 73, 409–414 (2012).