Audio Engineering Society

Convention Paper 9421

Presented at the 139th Convention
2015 October 29–November 1 New York, USA

Comparison of Techniques for Binaural Navigation of Higher-Order Ambisonic Soundfields

Joseph G. Tylka and Edgar Y. Choueiri

3D Audio and Applied Acoustics Laboratory, Princeton University, Princeton, NJ, 08544, USA

Correspondence should be addressed to Joseph G. Tylka (josephgt@princeton.edu)

ABSTRACT

Soundfields that have been decomposed into spherical harmonics (i.e., encoded into higher-order ambisonics – HOA) can be rendered binaurally for off-center listening positions, but doing so requires additional processing to translate the listener and necessarily leads to increased reproduction errors as the listener navigates further away from the original expansion center. Three techniques for performing this navigation (simulating HOA playback and listener movement within a virtual loudspeaker array, computing and translating along plane-waves, and re-expanding the soundfield about the listener) are compared through numerical simulations of simple incident soundfields and evaluated in terms of both overall soundfield reconstruction accuracy and predicted localization. Results show that soundfield re-expansion achieves arbitrarily low reconstruction errors (relative to the original expansion) in the vicinity of the listener, whereas errors generated by virtual-HOA and plane-wave techniques necessarily impose additional restrictions on the navigable range. Results also suggest that soundfield re-expansion is the only technique capable of accurately generating high-frequency localization cues for off-center listening positions, although the frequencies and translation distances over which this is possible are strictly limited by the original expansion order.

1. INTRODUCTION

Binaural navigation of three-dimensional (3D) higher-order ambisonic soundfields (i.e., soundfields that have been decomposed into spherical harmonics through modal beamforming) enables a listener to virtually explore an acoustic space and experience a spatially-accurate perception of the soundfield. Applications of binaural navigation may be found in virtual-reality reproductions of real-world spaces. For example, to reproduce an orchestral performance in virtual reality, binaural navigation of an acoustic recording of the performance is likely to yield superior spatial and tonal fidelity com-

pared to that produced through acoustic simulation of the performance. Binaural navigation of acoustic recordings may also be preferable when reproducing real-world spaces for which computer modeling of complex wave-phenomena and room characteristics may be too computationally intensive for real-time playback and interaction.

Traditional binaural recordings (such as those made with binaural dummy heads) can provide a listener with an accurate spatial perception of a 3D soundfield, but are inherently limited in two ways: 1) the perspective experienced by the listener during playback is restricted to the vantage point of the recording individual in the original soundfield and 2) the 3D localization cues embedded in the recording are only ideally suited for playback to the recording individual, as the recording individual's unique morphology (i.e., that individual's head-related transfer function) has already filtered the incoming sound waves in a highly idiosyncratic and direction-dependent manner.

The latter issue may be resolved by post-processing a higher-order ambisonics (HOA) recording of a soundfield (made, for example, using a spherical microphone array [1]) with an individual's particular head-related transfer functions (HRTFs), in a process known as *binaural decoding* of HOA. However, existing binaural decoding techniques have primarily been developed to place the listener at the position of the recording array [1, 2, 3, 4], so the former issue remains largely unaddressed.

One possible explanation for this is that a finite-order spherical-harmonic expansion of a soundfield yields only an approximation to that soundfield, the accuracy of which decreases with increasing frequency and distance from the expansion center [5], so the prospect of navigating such a soundfield is inherently limited. However, synthetic soundfields can be generated to arbitrarily high orders, and microphone array technology is rapidly advancing, such that it may soon be practical to capture very high-order expansions of real soundfields.

Another, more fundamental limitation of the spherical-harmonic description of soundfields is that the region over which the expansion is valid is limited by the nearest sound source to the expansion center (see Section 2.2), so near-field sources pose a particularly limiting problem to navigation. However, many real soundfields are free of extremely near-field sources, and alternative solutions for rendering synthetic near-field sources exist [6]. Therefore, despite the challenges facing navigational techniques, the problem of binaural navigation remains compelling.

1.1. Background and Previous Work

Higher-order ambisonics (HOA) provides a multichannel framework for representing measured soundfields, in which each signal (hereafter referred to as an "ambisonic signal") represents a different term of that soundfield's spherical-harmonic expansion. Generally, binaural decoding of HOA aims to generate the appropriate binaural signals for a listener at a point (with any orientation) in a recorded (or synthesized) soundfield, such that the resulting perception of the soundfield is identical to that which would have occurred in the real soundfield. At the origin, this task may be accomplished by:

1. decoding the ambisonic signals to a fixed set of virtual loudspeakers and filtering each loudspeaker's signal by the appropriate HRTF [3, 4],

2. transforming the ambisonic signals into a plane-wave expansion of the soundfield and filtering each plane-wave term by the appropriate HRTF [1, 6],

3. decomposing the (first-order only) ambisonic signals into plane-wave components using non-linear, parametric techniques (e.g., HARPEX) and filtering each component by the appropriate HRTF [2], or

4. using a spherical-harmonic decomposition of the listener's HRTFs to filter each ambisonic signal directly [7, 8].

Rotation of the listener (or, equivalently, of the soundfield) in HOA is straightforward and has been well-established in the literature [9, 10], so we will not discuss it in detail. Briefly, it involves the application of rotation matrices to change coordinates and frequency-independent mixing of the ambisonic signals. However, techniques to translate the listener to any other point in the soundfield require additional processing and are not as well-established.

One technique to allow navigation throughout the soundfield is to decode the ambisonic signals to a given loudspeaker array and simulate playback and navigation within that array. For real HOA playback, Frank *et al.* showed that so-called "max-r_E" decoding schemes

yield more accurate localization at off-center positions [11], and Satongar *et al.* showed that off-center localization improves with increasing HOA expansion order [12]. Also, to account for the finite distances of loudspeakers from the listening position, Daniel developed a near-field-compensated decoding scheme which treats the loudspeakers as point-sources and consequently reconstructs the soundfield more accurately at off-center locations [13]. In Section 3.1, we describe a method which employs these optimized decoding techniques and simulates navigation relative to virtual point-source loudspeakers.

An alternative navigational technique is to compute a plane-wave expansion of the soundfield and translate along each plane-wave term. Menzies and Al-Akaidi first derived the mathematical operations required for this technique, although they did so while developing a technique to more accurately render synthetic near-field sources binaurally by way of a plane-wave expansion and translation [6]. Schultz and Spors later formulated the plane-wave translation technique for the purpose of binaural navigation and examined the time- and frequency-domain consequences of the translation operation [14]. The localization properties of this technique were explored by Winter *et al.*, who showed that the range over which accurate localization is possible increases with HOA expansion order and that increasing the number of plane-wave-expansion terms beyond a certain threshold (also set by the expansion order) does not improve localization [15]. This technique is reviewed in Section 3.2.

The final navigational technique we consider is to translate the HOA expansion center by re-expanding the soundfield about the desired point. Gumerov and Duraiswami derived recurrence relations which enable fast computation of such re-expansions [9] and Zotter extended those derivations to real-valued spherical harmonics [10]. Menzies and Al-Akaidi in particular described how this technique can be used to allow a listener to virtually navigate a higher-order ambisonic soundfield [16], although a detailed analysis was not performed. This soundfield re-expansion technique is described in Section 3.3.

With the exception of the plane-wave expansion and translation technique, the errors generated by the navigational techniques described above and their effects on localization have not been investigated.

1.2. Objectives and Approach

The objective of this work is to compare various existing techniques for binaural navigation of higher-order ambisonic soundfields. To that end, we perform numerical simulations of each navigational technique and use objective metrics to evaluate the errors introduced by each technique and their effects on localization.

1.3. Paper Overview

In Section 2, we briefly review the theory of 3D soundfields which will be used to formulate the navigational techniques presented in Section 3. Then, in Section 4, we present objective metrics with which we compare the performances of these navigational techniques and describe the numerical simulations conducted. The results of these simulations are presented and discussed in Section 5 and conclusions indicated by these results are summarized in Section 6.

2. ACOUSTICAL THEORY

In this section we review the 3D acoustical theory that will be used in Section 3 to formulate navigational techniques.

2.1. Definitions and Conventions

Here, we adopt a spherical coordinate system commonly used in HOA, in which r is the (nonnegative) radial distance from the origin, $\theta \in [-\pi/2, \pi/2]$ is the elevation angle above the horizontal (x-y) plane, and $\phi \in [0, 2\pi)$ is the azimuthal angle around the vertical (z) axis, with $\phi = 0$ corresponding to the $+x$-axis and $\phi = \pi/2$ to the $+y$ axis. For a position vector $\mathbf{r} = (x, y, z)$, we denote unit vectors with a "hat," such that $\hat{\mathbf{r}} = \mathbf{r}/r$.

Also common in HOA, we use real-valued spherical harmonics of degree $n \geq 0$ and order $m \in [-n, n]$ given by

$$Y_n^m(\theta, \phi) = N_n^m P_n^{|m|}(\sin\theta) \times \begin{cases} \cos|m|\phi & \text{for } m \geq 0, \\ \sin|m|\phi & \text{for } m < 0, \end{cases}$$

where N_n^m is a normalization term and $P_n^{|m|}$ is the associated Legendre polynomial of degree n and order $|m|$. For the orthonormal (N3D) spherical harmonics, the normalization term is given by [13]

$$N_n^m = \sqrt{\frac{(2n+1)(2-\delta_m)}{4\pi} \frac{(n-|m|)!}{(n+|m|)!}},$$

where δ_m is the Kronecker delta.

2.2. Description of 3D Soundfields

We define the *acoustic potential field* ψ as the Fourier transform of the acoustic pressure field, such that, in

a source-free region (i.e., under free-field conditions), the acoustic potential field satisfies the homogeneous Helmholtz equation,

$$\left(\nabla^2 + k^2\right) \psi(k, \mathbf{r}) = 0, \tag{1}$$

where ∇^2 is the Laplace operator and k is the angular wavenumber. Regular (i.e., not singular) solutions to the Helmholtz equation are given by [9]

$$R_n^m(k, \mathbf{r}) \equiv j_n(kr) Y_n^m(\theta, \phi), \tag{2}$$

where j_n is the spherical Bessel function of order n. These solutions are only valid under free-field conditions, and can be used to describe the acoustic potential in an interior region, that is, for $r < r_0$, where r_0 is a finite distance. So that the region remains source-free, r_0 is typically taken to be the distance of the nearest source to the origin.

Provided these restrictions are met, any acoustic potential can be written as an infinite sum of regular solutions, known as a spherical Fourier-Bessel series expansion, given by [9]

$$\psi(k, \mathbf{r}) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} A_n^m(k) R_n^m(k, \mathbf{r}), \tag{3}$$

where A_n^m are the corresponding (frequency-dependent) expansion coefficients.

3. NAVIGATIONAL TECHNIQUES

In this section, we review three techniques for binaural navigation of higher-order ambisonic soundfields. Each technique operates using the same finite spherical Fourier-Bessel series expansion of the soundfield, so each technique is constrained by the same fundamental limitations regarding the accuracy and the region of validity of the original expansion. We refer to this expansion as the "band-limited" (BL) potential field, given by

$$\psi_{\mathrm{BL}}(k, \mathbf{r}) = \sum_{n=0}^{N_S} \sum_{m=-n}^{n} A_n^m(k) R_n^m(k, \mathbf{r}), \tag{4}$$

where N_S is the order-limit of the expansion and A_n^m are the complex-valued, frequency-dependent expansion coefficients, when the expansion is taken about the origin.

3.1. Virtual Higher-Order Ambisonics

The first navigational technique we consider involves simulating HOA playback over a virtual array of loudspeakers (hereafter called "virtual HOA"). In this case, binaural navigation requires only that the HRTFs applied to each loudspeaker signal (appropriately attenuated and delayed based on distance) be updated based on the position of the listener relative to that loudspeaker.

The process of decoding HOA to N_L loudspeakers is typically expressed as a (frequency-domain) matrix multiplication between the so-called *decoding matrix* and ambisonic signals, which yields the appropriate loudspeaker signals. Methods of calculating the decoding matrix have been extensively researched (see, for example, Heller *et al.* [17]) and it is outside of the scope of this work to discuss them in detail. Briefly, the decoding matrix attempts to use all available loudspeakers to create a perceptually-accurate reproduction the recorded soundfield [17].

Here we model the virtual loudspeakers as point-sources,[1] such that a single loudspeaker at $\mathbf{r}_0$, driven with a (Fourier-transformed) signal V_0, produces a potential field given by [18]

$$\psi_0(k, \mathbf{r}) = \frac{e^{ik|\mathbf{r}_0 - \mathbf{r}|}}{|\mathbf{r}_0 - \mathbf{r}|} V_0(k). \tag{5}$$

The total potential field produced by virtual HOA playback (denoted by the subscript "VA") is then given by

$$\psi_{\mathrm{VA}}(k, \mathbf{r}) = \sum_{l=0}^{N_L - 1} \frac{e^{ik|\mathbf{r}_l - \mathbf{r}|}}{|\mathbf{r}_l - \mathbf{r}|} V_l(k), \tag{6}$$

where V_l is the Fourier transform of the signal sent to the l^{th} loudspeaker and $\mathbf{r}_l$ is the position of that loudspeaker.

To binaurally render the soundfield described by Eq. (6) for a listener at position $\mathbf{d}$, the left and right binaural potentials (indicated by the superscripts "L" and "R," respectively) are computed by

$$\psi_{\mathrm{VA}}^{L,R}(k, \mathbf{d}) = \sum_{l=0}^{N_L - 1} \frac{e^{ik|\mathbf{r}_l - \mathbf{d}|}}{|\mathbf{r}_l - \mathbf{d}|} V_l(k) H^{L,R}(k, \hat{\mathbf{s}}_l(\mathbf{d})), \tag{7}$$

where

$$\hat{\mathbf{s}}_l(\mathbf{d}) = \frac{\mathbf{r}_l - \mathbf{d}}{|\mathbf{r}_l - \mathbf{d}|} \tag{8}$$

is a unit vector pointing from the translated position of the listener to the l^{th} loudspeaker and $H^{L,R}(k, \hat{\mathbf{s}})$ is the far-field HRTF for a source in the direction $\hat{\mathbf{s}}$.

[1] Note that we could have instead modeled the loudspeakers as plane-wave sources, infinitely far away from the listener. However, we chose to use finite-distance virtual loudspeakers so that this technique will differ more significantly from the plane-wave expansion technique described in Section 3.2.

3.2. Plane-Wave Expansion and Translation

The second navigational technique we consider uses a plane-wave expansion of the band-limited potential field. It was shown by Schultz and Spors that, given a plane-wave expansion of a soundfield, translation can be achieved by applying a frequency-domain phase-factor (or group delay in the time domain) to each plane-wave term, based on the direction of travel of the listener relative to the propagation direction of each plane-wave [14].

It can be shown that any free-field potential field can be written as an infinite sum of plane-waves, given by [1]

$$\psi(k,\mathbf{r}) = \frac{1}{4\pi}\int_{S_u} \mu(k,\hat{\mathbf{s}}) e^{ik\hat{\mathbf{s}}\cdot\mathbf{r}} dS(\hat{\mathbf{s}}), \tag{9}$$

where μ is the complex-valued, frequency-dependent amplitude of each plane-wave, called the *far-field signature function*, $\hat{\mathbf{s}}$ is the propagation direction of each plane-wave, and the integration is taken over the surface of the unit sphere. Given the spherical Fourier-Bessel expansion coefficients of a soundfield, the signature function is given by [1]

$$\mu(k,\hat{\mathbf{s}}) = \sum_{n=0}^{\infty}\sum_{m=-n}^{n} i^{-n} A_n^m(k) Y_n^m(\hat{\mathbf{s}}). \tag{10}$$

In practice, we approximate integrating over the unit sphere by numerical quadrature, so the plane-wave (PW) potential field is given by [1]

$$\psi_{\mathrm{PW}}(k,\mathbf{r}) = \sum_{p=0}^{N_P-1} \mu(k,\hat{\mathbf{s}}_p) e^{ik\hat{\mathbf{s}}_p\cdot\mathbf{r}} w_p, \tag{11}$$

where N_P is the total number of plane-waves, $\hat{\mathbf{s}}_p$ is the propagation direction of the p^{th} plane-wave, and w_p is the corresponding quadrature weight. For each term in this summation, the potential field at $\mathbf{r}+\mathbf{d}$ differs only by a phase-factor $e^{ik\hat{\mathbf{s}}_p\cdot\mathbf{d}}$, so we combine this factor into the signature function and define the translated signature function μ', given by [6]

$$\mu'(k,\hat{\mathbf{s}}_p;\mathbf{d}) = \mu(k,\hat{\mathbf{s}}_p) e^{ik\hat{\mathbf{s}}_p\cdot\mathbf{d}}. \tag{12}$$

The left and right binaural potentials for a listener at $\mathbf{d}$ are then given by [14]

$$\psi_{\mathrm{PW}}^{L,R}(k,\mathbf{d}) = \sum_{p=0}^{N_P-1} \mu'(k,\hat{\mathbf{s}}_p;\mathbf{d}) H^{L,R}(k,-\hat{\mathbf{s}}_p) w_p. \tag{13}$$

3.3. Soundfield Re-Expansion

The final navigational technique we consider is to compute a new set of ambisonic signals by re-expanding the soundfield about a translated expansion point using frequency-domain translation coefficients.

It can be shown that any individual spherical Fourier-Bessel term can be expressed as an infinite spherical Fourier-Bessel series centered about a translated expansion point, $\mathbf{d}$, as given by [9]

$$R_n^m(k,\mathbf{r}+\mathbf{d}) = \sum_{n'=0}^{\infty}\sum_{m'=-n'}^{n'} \Gamma_{n',n}^{m',m}(k,\mathbf{d}) R_{n'}^{m'}(k,\mathbf{r}), \tag{14}$$

where $\Gamma_{n',n}^{m',m}$ are the so-called *translation coefficients*. Integral forms of these translation coefficients as well as fast recurrence relations for computing them are given by Gumerov and Duraiswami [9] and Zotter [10].

Therefore, the potential field at $\mathbf{r}+\mathbf{d}$ obtained through soundfield re-expansion (SR) about $\mathbf{d}$ is given by

$$\psi_{\mathrm{SR}}(k,\mathbf{r};\mathbf{d}) = \sum_{n'=0}^{N_S'}\sum_{m'=-n'}^{n'} C_{n'}^{m'}(k;\mathbf{d}) R_{n'}^{m'}(k,\mathbf{r}), \tag{15}$$

where N_S' is the order-limit of the new expansion and $C_{n'}^{m'}$ are the complex-valued, frequency-dependent expansion coefficients, when the expansion is taken about $\mathbf{d}$. Complementary to Eq. (14), the translated expansion coefficients are given by

$$C_{n'}^{m'}(k;\mathbf{d}) = \sum_{n=0}^{N_S}\sum_{m=-n}^{n} \Gamma_{n',n}^{m',m}(k,\mathbf{d}) A_n^m(k). \tag{16}$$

It is important to note that these translated expansion coefficients can be computed to arbitrarily high orders, although the re-expanded field is still limited in accuracy and region of validity by the original expansion. In other words, with increasing N_S', the re-expanded field approaches the original *band-limited* field, not the incident field.

4. METRICS AND SIMULATIONS

In this section, we introduce the various metrics by which we evaluate and compare the navigational techniques described above, and describe the numerical simulations conducted. Results are presented and discussed in Section 5.

4.1. Reconstruction Errors

Any practical implementation of any of the navigational techniques described above will necessarily introduce errors in the reconstructed soundfield. We classify these errors into two types: truncation error and rendering error. Truncation error is introduced by using a finite-order approximation of an infinite-order potential field. For example, truncation error is introduced when computing the band-limited potential field from the incident soundfield. This type of error has been well-documented in the literature, and is known to create a finite "sweet-spot" (i.e., a region of space in which the expansion accurately represents the incident soundfield) in the band-limited field [5]. However, since all of the navigational techniques operate using the band-limited potential field, this error will be present in all cases (for the same original expansion order N_S). Consequently, we exclude this truncation error from our analyses, to better compare only the errors generated by each navigational technique. Note, however, that soundfield re-expansion introduces a similar truncation error, as we are computing a new, finite-order expansion of the band-limited potential field, as shown in Eqs. (15) and (16).

Rendering (or aliasing) error is introduced by converting a finite-order soundfield expansion into a finite sum of discrete sources, e.g., through decoding to virtual loudspeakers or by converting to plane-waves. This type of error not only occurs when rendering the original soundfield expansion, but also when rendering a soundfield that has been re-expanded about a translated origin. As is well-established in the literature, this type of error also creates a sweet-spot in which the rendering error is small [5].

Combined, these separate sources of error in each navigational technique result in an overall reconstruction error, which is a measure of the total discrepancy between the reconstructed (i.e., translated and rendered) potential field and the original band-limited field. We quantify this error as the *normalized reconstruction error*, given by

$$\varepsilon_R(k,\mathbf{r};\mathbf{d}) = \frac{|\psi_{\mathrm{BL}}(k,\mathbf{r}+\mathbf{d}) - \psi'(k,\mathbf{r};\mathbf{d})|^2}{|\psi_{\mathrm{BL}}(k,\mathbf{r}+\mathbf{d})|^2}, \qquad (17)$$

where $\psi'(k,\mathbf{r};\mathbf{d})$ is the reconstructed potential field obtained through any of the navigational techniques described above. For the virtual-HOA and plane-wave translation techniques, ψ' is the rendered field evaluated at $\mathbf{r}+\mathbf{d}$, as given by Eqs. (6) and (11), respectively. In the case of soundfield re-expansion, however, ψ' is the soundfield re-expanded about $\mathbf{d}$ and evaluated at $\mathbf{r}$, as given by Eq. (15).

It is readily verified that for both the virtual-HOA and plane-wave translation techniques, the process of translation introduces no new errors. Instead, the reconstruction error for these techniques consists only of a rendering error that creates a static sweet-spot in the rendered field. Soundfield re-expansion, on the other hand, introduces truncation errors that change with translation position. Furthermore, to facilitate conversion to binaural (and to apply the localization metrics defined in Section 4.2), the re-expanded soundfield must be rendered as a finite sum of discrete sources, which introduces an additional rendering error. Thus, the reconstruction error for the soundfield re-expansion technique consists of both truncation and rendering errors. For all simulations in this work, we compute a plane-wave expansion of the re-expanded soundfield.

We expect that the most relevant errors to the generation of binaural signals are those in the vicinity of the listener's head. Consequently, we also define a *volumetric reconstruction error*, given by

$$\varepsilon_V(k;\mathbf{d}) = \frac{\iiint_V |\psi_{\mathrm{BL}}(k,\mathbf{r}+\mathbf{d}) - \psi'(k,\mathbf{r};\mathbf{d})|^2 dV}{\iiint_V |\psi_{\mathrm{BL}}(k,\mathbf{r}+\mathbf{d})|^2 dV}, \quad (18)$$

where the volume integral is performed over a spherical region centered at $\mathbf{d}$, i.e., surrounding the head. In our simulations of each navigational technique, we evaluate volumetric reconstruction errors for a point-source on the x-axis as a function of translation distance along the y-axis, and compute each integral over a sphere with a radius of 9 cm.

4.2. Localization Vectors

To predict the localization of sound in multichannel playback systems, Gerzon [19] defines two localization metrics: the velocity and energy vectors. The velocity vector is used to predict localization due to interaural time differences at low frequencies (< 700 Hz) and is given by [19]

$$\mathbf{r}_{\mathrm{V}}(k) = \mathrm{Re}\left[\frac{\sum_n G_n(k)\hat{\mathbf{r}}_n}{\sum_n G_n(k)}\right], \qquad (19)$$

where G_n is the complex-valued, frequency-dependent "gain" of the n^{th} source, and $\hat{\mathbf{r}}_n$ points in the direction of that source from the origin. The energy vector is used to predict localization due to interaural level differences at

higher frequencies (500 Hz – 5 kHz) and is given by [19]

$$\mathbf{r}_{\mathrm{E}}(k) = \frac{\sum_n |G_n(k)|^2 \hat{\mathbf{r}}_n}{\sum_n |G_n(k)|^2}. \tag{20}$$

The directions of these vectors indicate the expected localization direction and their magnitudes indicate the quality of the localization. Ideally, the vectors should have a magnitude equal to unity and point in the direction of the virtual source.

In this work, we apply the above definitions to predict localization both in the case of virtual HOA, where each virtual loudspeaker is a source and the signals sent to those loudspeakers are the source gains, and in the case of plane-wave expansions, where each plane-wave term is a source and the signature function and quadrature weights determine the source gains. Moreover, we extend the definitions of these localization vectors to off-center listening positions, similar to the work of Moore and Wakefield [20]. We are then able to evaluate these localization vectors as a function of translation position and assess the ability of each navigational technique to accurately reproduce localization cues at off-center listening positions.

For virtual HOA, the source gains at the origin are simply the loudspeaker signals, i.e., $G_n(k) = V_n(k)$, and the source directions are those of each loudspeaker, $\hat{\mathbf{r}}_n$. However, at a translated position $\mathbf{d}$, the *effective* source gains (accounting for point-source radiation) become

$$G_n(k;\mathbf{d}) = \frac{e^{ik(|\mathbf{r}_n - \mathbf{d}| - |\mathbf{r}_n|)}}{|\mathbf{r}_n - \mathbf{d}| \,/\, |\mathbf{r}_n|} V_n(k), \tag{21}$$

and the source directions are given by $\hat{\mathbf{r}}_n = \hat{\mathbf{s}}_n(\mathbf{d})$, as defined in Eq. (8).

For plane-wave expansions, the source gains at the origin are given by the product of the signature function and the quadrature weight for each plane-wave term, i.e., $G_n(k) = w_n \mu(k, \hat{\mathbf{s}}_n)$, but the source directions are given by $\hat{\mathbf{r}}_n = -\hat{\mathbf{s}}_n$, since $\hat{\mathbf{s}}_n$ is the direction in which the plane-wave *propagates*. With translation, the source gains become

$$G_n(k;\mathbf{d}) = w_n \mu'(k, \hat{\mathbf{s}}_n; \mathbf{d}), \tag{22}$$

but the source directions do not change.

To compute localization vectors for the soundfield re-expansion technique, we compute a plane-wave expansion of the re-expanded soundfield and use that plane-wave expansion to compute the localization vectors.

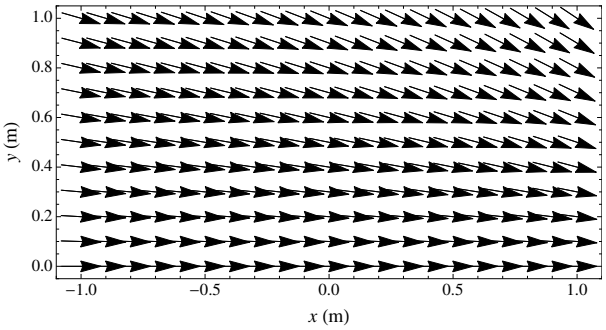


Fig. 1: Ideal localization vectors plotted on a rectangular 10 cm × 10 cm grid in the *x*-*y* plane for a point-source at $\mathbf{r}_s = (2.5, 0, 0)$ m.

To qualitatively evaluate the localization performance of each navigational technique, we plot localization vectors for a given frequency on a rectangular 10 cm × 10 cm grid of translation positions in the *x*-*y* plane. Figure 1 shows the ideal localization vectors (for all frequencies) on this grid, for a point-source at $\mathbf{r}_s = (2.5, 0, 0)$ m.

To assess each navigational technique's ability to preserve localization information throughout the translation process, we evaluate the deviation of these localization vectors from ideal with translation in terms of a *directional error*. For a point-source at $\mathbf{r}_s$ and translation position $\mathbf{d}$, the directional error δ is given by[2]

$$\delta_{\mathrm{V,E}}(k;\mathbf{d}) = \left| \hat{\mathbf{r}}_{\mathrm{V,E}}(k;\mathbf{d}) - \frac{\mathbf{r}_s - \mathbf{d}}{|\mathbf{r}_s - \mathbf{d}|} \right|. \tag{23}$$

In our simulations, we compute, as a function of frequency, the average (RMS) directional errors in a plane over a polar grid of translation positions with radial increments of 10 cm and azimuthal increments of 15°.

4.3. Simulation Parameters

To evaluate the navigational techniques described in Section 3, we simulated navigation of a soundfield containing a single point-source at 2.5 m in front of the listener, $\mathbf{r}_s = (2.5, 0, 0)$ m. Due to the axial symmetry of this source placement, we consider only translation on the half-plane defined by $z = 0$, $y \geq 0$. Unless otherwise noted, the original expansion of this incident soundfield was taken up to fourth order ($N_S = 4$) and all soundfield re-expansions were performed up to fourth order as well

[2] The directional error can be converted to an angular error (i.e., the angle between the two unit vectors in Eq. (23)) by $\cos^{-1}(1 - \delta^2/2)$.

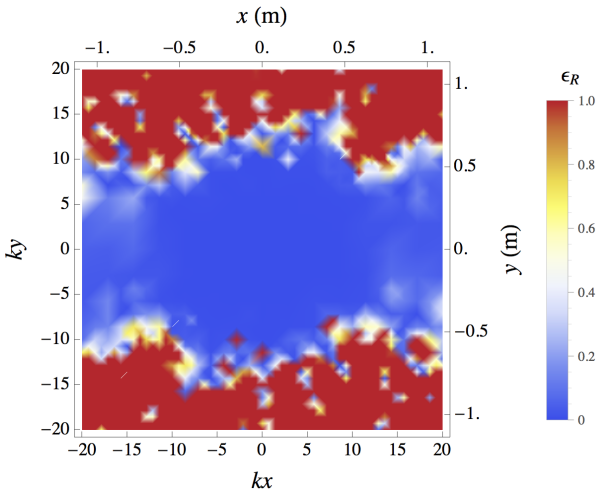


Fig. 2: Normalized reconstruction error (see Eq. (17)) in the x-y plane at 1 kHz generated by plane-wave expansion and translation. Other conditions are given in Section 4.3. Left and bottom axes show nondimensionalized distance at 1 kHz, while right and top axes show distance in meters. Errors have been clipped beyond $\epsilon_R = 1$ to make the regions of low error more readable.

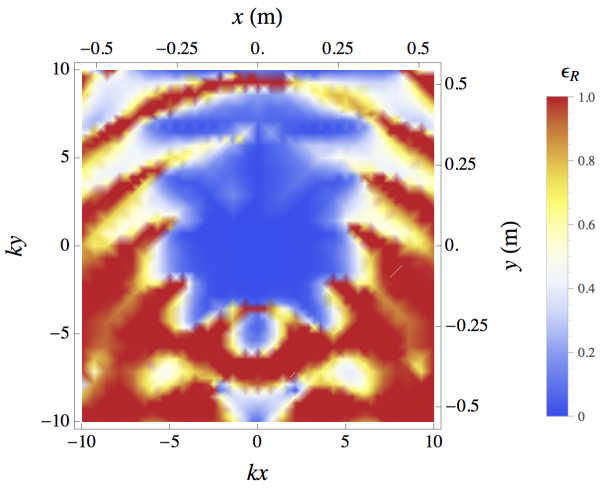


Fig. 3: Normalized reconstruction error (see Eq. (17)) in the x-y at 1 kHz for soundfield re-expansion about $y = 100$ cm followed by plane-wave expansion. Other conditions are given in Section 4.3. Left and bottom axes show nondimensionalized distance at 1 kHz, while right and top axes show distance in meters. Errors have been clipped beyond $\epsilon_R = 1$ to make the regions of low error more readable. In contrast to Fig. 2, the origin of this plot is the translation position, $\mathbf{d} = (0,1,0)$ m.

($N'_S = 4$). For all virtual HOA simulations, we decode to a virtual array of $N_L = 36$ loudspeakers arranged on Fliege nodes [21], with an array radius of $r_l = 5$ m, and use near-field-compensated decoding [13] with both "basic" (or "mode-matching") and max-r_E weighting [17]. For all plane-wave expansions, we compute $N_P = 100$ plane-wave terms also arranged on Fliege nodes and use the corresponding quadrature weights.[3]

5. RESULTS AND DISCUSSION

Now we present and discuss the results of the simulations described in the previous section.

5.1. Reconstruction Errors

Normalized reconstruction errors are plotted in Figs. 2 and 3. Figure 2 shows the reconstruction errors at 1 kHz generated by a plane-wave expansion, plotted in the horizontal plane for approximately ± 1 m in each direction. This plot clearly indicates the existence of a sweet-spot which allows for relatively large back and forth translation but is significantly more restrictive of lateral translations. Nondimensionalized distances (kx and ky) are given on the left and bottom axes, and it can be verified that the reconstruction errors, when plotted over the same range of nondimensional distances, retain the same qualitative structure across a wide range of frequencies. This trend breaks down, however, at low frequencies, since the source distance is finite (2.5 m from the origin). For example, at 100 Hz, $kr_s \approx 4.6$ is on the order of the expansion order, $N_S = 4$, at which point, the near-field effect of the source has a significant effect on the reconstruction error. Although not shown here, the reconstruction errors generated by virtual HOA exhibit very similar behavior in terms of the structure and frequency-dependence of the sweet-spot, although the size of the sweet-spot is more limited than for plane-wave expansions, likely due to the fewer number of discrete sources ($N_L < N_P$).

Figure 3 shows reconstruction errors at 1 kHz generated by re-expanding the soundfield at $\mathbf{d} = (0,1,0)$ m and then converting to plane-waves, plotted in the horizontal plane for approximately ± 0.5 m in each direction. Again we note the existence of a sweet-spot, although now cen-

[3] Node coordinates and corresponding quadrature weights can be found here: `http://www.mathematik.uni-dortmund.de/lsx/research/projects/fliege/nodes/nodes.html`

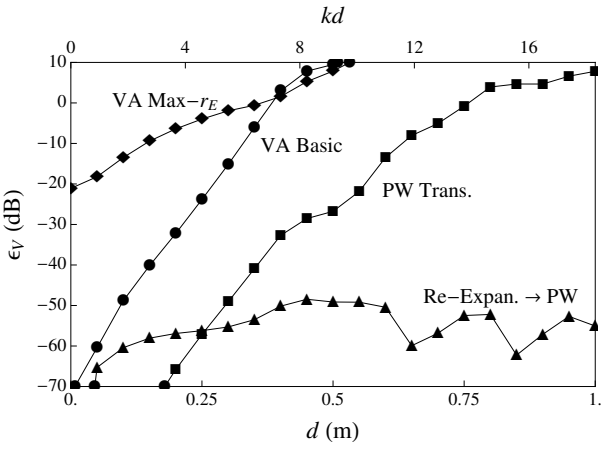


Fig. 4: Volumetric reconstruction errors (see Eq. (18) and subsequent text) at 1 kHz as a function of translation distance along the y-axis. Bottom axis shows distance in meters, while top axis shows nondimensionalized distance at 1 kHz. Other conditions are given in Section 4.3.

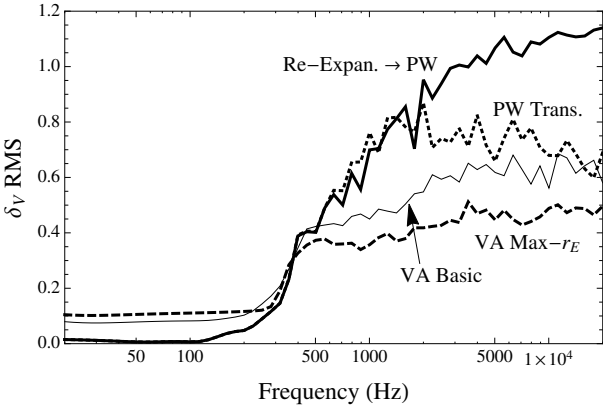


Fig. 5: Average directional errors (see Eq. (23)) of the velocity vectors as functions of frequency, RMS-averaged over a polar grid of translation positions for $0 \le r \le 1$ m in increments of 10 cm and $\phi \in [0, 180°]$ in increments of 15°. Other conditions are given in Section 4.3.

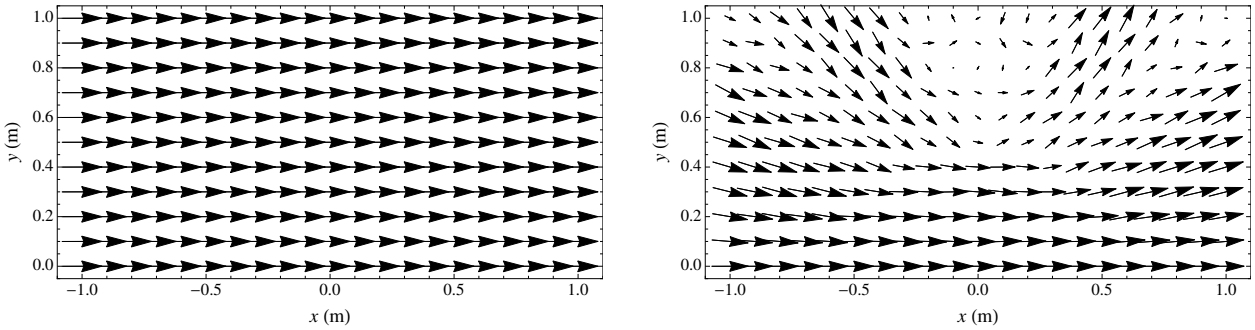
tered around the *translation position*, whereas in Fig. 2, the sweet-spot is centered around the original expansion center. For other translation positions, it can be verified that the size of each sweet-spot is approximately the same for any translation distance, and can be approximated by $kr = N_S'$ ($= 4$ in this case), where r is the distance from the translated expansion center. This result is in very good agreement with the "rule of thumb" found in the literature: that a soundfield expansion up to order N_S is accurate to within $\sim 4\%$ for $kr \le N_S$ [5]. Indeed, it can be seen directly from the plot that for $|kx|, |ky| \le 4$ (~ 22 cm at 1 kHz), the reconstruction error is low.

To further explore this point, we plot in Fig. 4 the volumetric reconstruction errors generated by each technique for translation along the y-axis. As expected, the reconstruction errors grow with translation distance for the virtual-HOA and plane-wave expansion techniques. Soundfield re-expansion, however, is able to maintain a low (approximately -50 dB) volumetric reconstruction error at all translation distances. We note that for small translation distances ($d \le 25$ cm, which again corresponds to $kd \le N_S = 4$), a plane-wave expansion of the original soundfield yields smaller errors than that generated by re-expansion. This suggests that soundfield re-expansion is only beneficial for larger translation distances or, equivalently, at higher frequencies.

As we noted earlier, soundfield re-expansion incurs a truncation error similar to that present in the original expansion of the soundfield. Consequently, with higher-order re-expansions, we expect the sweet-spot to grow both in spatial and frequency extent. In this way, the reconstruction errors generated by soundfield re-expansion, especially those in the vicinity of the listener, can be made arbitrarily small simply by computing higher-order re-expansions. Of course, doing so results in a greater computational load and may not be practical.

5.2. Localization Vectors

Localization vectors reproduced by each navigational technique were plotted over a range of translation positions and qualitatively compared. The velocity vectors were evaluated at 150 Hz and the energy vectors were evaluated at 1 kHz. Although not shown here, we observed that all techniques are able to accurately reproduce 150-Hz velocity vectors for translation distances on the order of 1 m, as the plotted vectors agreed very well with ideal (shown in Fig. 1). The average directional errors for the velocity vectors generated by each technique are plotted in Fig. 5. From this figure, we see that the velocity-vector directional errors at low frequencies (< 300 Hz) are small, but sharply increase around 300 Hz for all techniques. This is explained by the fact that a fourth order expansion is accurate within 1 m of the origin only for frequencies up to ~ 220 Hz (see previous discussions in Section 5.1).



(a) Plane-wave expansion and translation; RMS $\delta_E = 0.248$

(b) Soundfield re-expansion; RMS $\delta_E = 0.609$

Fig. 6: Reproduced energy vectors (see Eq. (20)) at 1 kHz, plotted on the same grid as in Fig. 1. Other conditions are given in Section 4.3.

For the 1-kHz energy vectors, however, the results are varied. Figure 6 (a) shows the reproduced energy vectors for plane-wave translation. This figure clearly shows that the energy vectors are not reproduced correctly with lateral translation, as the directions of the vectors do not adjust to point back towards the source on the x-axis. Although not shown here, the reproduced 1-kHz energy vectors for virtual HOA exhibited very similar behavior.

This result, in the case of a plane-wave expansion, can be explained mathematically by the nature of translation along plane-waves. From Eq. (12), we recall that translation along a plane-wave is achieved through multiplication of the signature function by a complex-exponential phase-factor. Consequently, when the translated signature function is used to compute the energy vector, the phase-factors, which have unity magnitude, have no effect on the result. This fact points to a fundamental limitation of the plane-wave translation technique: that translation phase-factors are unable to spatially-redistribute the energy in the soundfield. This limitation, as well as the structure of the reconstruction errors observed above in Fig. 2, may explain the findings of Winter *et al.*: that localization errors increase with lateral translation distance [15].

However, the same work showed that accurate localization can be achieved for lateral translations, although larger translation distances require higher original expansion orders [15]. It is possible that, since phase-factors manifest themselves as group delays in the time-domain, other psychoacoustic factors such as the dominance of low-frequency interaural time difference cues [22] and/or the precedence effect [23] may compensate for incorrect energy vectors and lead to correct localization, but we do not explore this here.

In the case of virtual HOA, given the relatively large size of the array ($r_l = 5$ m) and the frequency 1 kHz, the point-source loudspeakers behave very much like plane-wave sources over the region considered. For example, if the listener is 4 m from the nearest loudspeaker, at 1 kHz, $kr \approx 73.3 \gg 1$, so the wavefronts will be nearly planar [18]. Consequently, the behavior of the energy vectors for this technique should be very similar over the navigable range to that of the plane-wave technique, which was indeed the case.

The energy vectors produced by soundfield re-expansion are shown in Fig. 6 (b). In this case, the energy vectors show significant errors beyond a certain translation distance and the magnitudes of the vectors become very small, indicating the localization would be ambiguous, if not wholly incorrect. These results suggest that, while plane-wave expansions are unable to reproduce energy vectors which turn towards the source with lateral translation, the energy vectors produced by soundfield re-expansion are much more sensitive to the truncation errors inherent in the band-limited expansion of the soundfield. Consequently, the energy-vector directional error for plane-wave translation increases very gradually with translation distance, whereas the directional error generated by soundfield re-expansion increases very sharply beyond a certain translation distance.

To further explore this point, we plot in Fig. 7 the average directional errors for the energy vectors generated by each technique. We observe that the soundfield re-

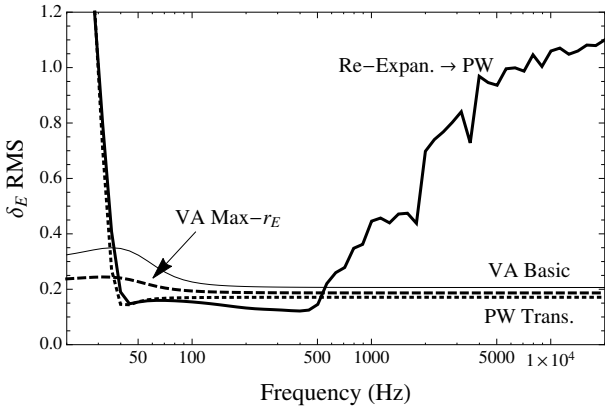


Fig. 7: Average directional errors (see Eq. (23)) of the energy vectors as functions of frequency, RMS-averaged over the same grid as in Fig. 5. Other conditions are given in Section 4.3.

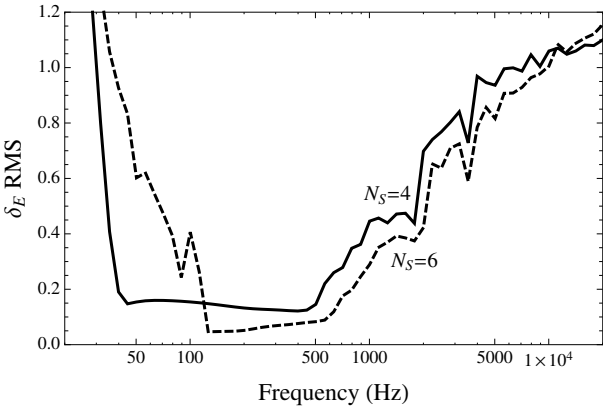


Fig. 8: Directional errors (see Eq. (23)) as a function of frequency for soundfield re-expansion followed by plane-wave expansion, given two different original expansion orders: $N_S = 4$ and 6; RMS-averaged over the same grid as in Fig. 5. Other conditions are given in Section 4.3.

expansion technique is the only technique to exhibit a sharp increase in energy-vector directional error, which occurs at approximately 500 Hz. The virtual-HOA and plane-wave translation techniques do not show any increase in energy-vector directional error with frequency. At very low frequencies, we see large directional errors generated by both plane-wave expansion and soundfield re-expansion, although these errors are attributed to the near-field effect of the point-source at 2.5 m. Similar to the sweet-spot discussion in Section 5.1, at frequencies for which $kr_s \ll N_S$ (i.e., below ~ 87 Hz), the near-field effect becomes significant and leads to directional errors in the energy vectors. A surprising consequence of this trend is that, to achieve directionally-accurate energy vectors at these low frequencies, only the lowest-order terms of the soundfield expansion should be considered. However, this effect may not have significant perceptual consequences since the energy vector predicts interaural level differences and is only applicable for high frequencies (> 500 Hz) [19].

5.2.1. Energy Vectors for Lateral Translations

The localization vector results presented above seem to suggest that none of the considered navigational techniques are capable of accurately reproducing energy vectors for lateral translations. However, by taking a higher-order original expansion of the soundfield, we find that soundfield re-expansion (up to order $N'_S = 4$) is indeed capable of reproducing directionally-accurate energy vectors with lateral translation. Figure 8 shows the average directional errors in the energy vectors generated by soundfield re-expansion, for two original expansion orders: $N_S = 4$ and 6. At low frequencies, we again note the influence of the near-field effect, which becomes more prominent for the higher expansion order. At higher frequencies, however, a higher-order original expansion yields lower average directional errors. For example, given an original expansion of the soundfield up to order $N_S = 6$, the reproduced energy vector grid at 400 Hz matches very well with the ideal and yields an RMS directional error of $\delta_{\mathrm{E}} = 0.124$.

6. CONCLUSIONS

Numerical simulations of three navigational techniques (virtual higher-order ambisonics, plane-wave expansion and translation, and soundfield re-expansion) are conducted to compare these techniques in the context of binaural navigation of higher-order ambisonic soundfields.

In terms of the soundfield reconstruction errors introduced by each technique, results show that virtual-HOA and plane-wave translation techniques necessarily create static sweet-spots that restrict the range of motion of the listener, whereas the sweet-spot created through soundfield re-expansion coincides with the translated position of the listener. We also show that the reconstruction errors generated by soundfield re-expansion can be made arbitrarily small (or, equivalently, the sweet-spot can be

made arbitrarily large) by increasing the re-expansion order. However, the overall discrepancy between any reconstructed soundfield and the exact incident soundfield is still limited by the original expansion order, since all of the navigational techniques operate on a band-limited expansion of the incident soundfield.

As for the effect of each navigational technique on predicted localization cues, results suggest that all of the techniques are capable of reproducing low-frequency interaural time difference cues over a wide range of translation positions. However, the limited high-frequency information provided by the band-limited expansion of the incident soundfield imposes more stringent limitations on the reproduction of high-frequency interaural level difference cues. Our analysis reveals a fundamental limitation of the plane-wave translation technique: that it is unable to spatially-redistribute energy information in the soundfield to update high-frequency localization cues in response to lateral listener translation, although accurate low-frequency temporal cues may or may not compensate for this limitation. For the soundfield re-expansion technique, we observe that the range of translation positions over which localization vectors are accurately reproduced is strictly limited to a finite translation distance, beyond which directional errors increase rapidly. The virtual-HOA and plane-wave translation techniques, however, experience more gradual increases in directional error with translation distance.

Also, as a proof-of-concept, we show that higher-order expansions of the original soundfield enable the soundfield re-expansion technique to more accurately reproduce energy localization cues with lateral translation. As the original expansion order is increased, we expect these results to extend to higher frequencies and larger translation distances.

It is important to keep in mind that all of these techniques are limited by the accuracy and the region of validity of the original expansion. Consequently, for low-order recordings and those containing sources very near to the microphone array, the range of motion allowed by any navigational technique will be significantly limited.

6.1. Future Work

The localization results of this work are encouraging, but need to be further verified. As noted by Gerzon, each localization vector only tells part of the story, and should not be taken in isolation as a predictor of localization [19]. Additionally, when the localization vectors do not agree or are small in magnitude, the localization predicted by these models becomes ambiguous. Consequently, future work should incorporate more sophisticated binaural models (such as those already used by Winter *et al.* for the plane-wave translation technique [15]) and consider additional psychoacoustic effects such as the dominance of low-frequency interaural time difference cues and the precedence effect to more accurately and definitively predict localization. Ultimately, subjective listening tests are needed to both verify the predictions of this work and gain general feedback on the performance of these navigational techniques.

ACKNOWLEDGEMENTS

This work was sponsored by the Sony Corporation of America. The authors wish to thank Rahulram Sridhar for fruitful discussions during this work and the anonymous reviewers for their feedback.

7. REFERENCES

[1] 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". Presented at the AES 119th Convention, October 2005.

[2] S. Berge and N. Barrett. "A New Method for B-Format to Binaural Transcoding". Presented at the AES 40th International Conference, October 2010.

[3] A. McKeag and D. S. McGrath. "Sound Field Format to Binaural Decoder with Head Tracking". Presented at the AES 6th Australian Regional Convention, August 1996.

[4] M. Noisternig, A. Sontacchi, T. Musil, and R. Holdrich. "A 3D Ambisonic Based Binaural Sound Reproduction System". Presented at the AES 24th International Conference, June 2003.

[5] M. A. Poletti. "Three-Dimensional Surround Sound Systems Based on Spherical Harmonics". *J. Audio Eng. Soc.*, 53(11):1004–1025, 2005.

[6] D. Menzies and M. Al-Akaidi. "Nearfield binaural synthesis and ambisonics". *J. Acoust. Soc. Am.*, 121(3):1559–1563, 2007.

[7] B. Rafaely and A. Avni. "Interaural cross correlation in a sound field represented by spherical

harmonics". *J. Acoust. Soc. Am.*, 127(2):823–828, 2010.

[8] B. Bernschütz, A. V. Giner, C. Pörschmann, and J. Arend. "Binaural Reproduction of Plane Waves With Reduced Modal Order". *Acta Acustica united with Acustica*, 100(5):972–983, 2014.

[9] N. A. Gumerov and R. Duraiswami. *Fast Multipole Methods for the Helmholtz Equation in Three Dimensions*. Elsevier Science, 2005.

[10] F. Zotter. *Analysis and Synthesis of Sound-Radiation with Spherical Arrays*. PhD thesis, University of Music and Performing Arts Graz, 2009.

[11] M. Frank, F. Zotter, and A. Sontacchi. "Localization Experiments Using Different 2D Ambisonics Decoders". In *25th Tonmeistertagung – VDT Int. Conv.*, Leipzig, Germany, November 2008.

[12] D. Satongar, C. Dunn, Y. Lam, and F. Li. "Localisation Performance of Higher-Order Ambisonics for Off-Centre Listening". BBC Research & Development White Paper, October 2013.

[13] J. Daniel. "Spatial Sound Encoding Including Near Field Effect: Introducing Distance Coding Filters and a Viable, New Ambisonic Format". Presented at the AES 23rd International Conference, May 2003.

[14] F. Schultz and S. Spors. "Data-Based Binaural Synthesis Including Rotational and Translatory Head-Movements". Presented at the AES 52nd International Conference, September 2013.

[15] F. Winter, F. Schultz, and S. Spors. "Localization Properties of Data-based Binaural Synthesis including Translatory Head-Movements". In *Forum Acusticum*, September 2014.

[16] D. Menzies and M. Al-Akaidi. "Ambisonic Synthesis of Complex Sources". *J. Audio Eng. Soc.*, 55(10):864–876, 2007.

[17] A. J. Heller, E. M. Benjamin, and R. Lee. "A Toolkit for the Design of Ambisonic Decoders". In *Linux Audio Conf.*, April 2012.

[18] P. M. Morse and K. U. Ingard. *Theoretical Acoustics*. Princeton University Press, 1986.

[19] M. A. Gerzon. "General Metatheory of Auditory Localisation". Presented at the AES 92nd Convention, March 1992.

[20] D. Moore and J. Wakefield. "Optimization of the Localization Performance of Irregular Ambisonic Decoders for Multiple Off-Center Listeners". Presented at the AES 128th Convention, May 2010.

[21] J. Fliege and U. Maier. "The distribution of points on the sphere and corresponding cubature formulae". *IMA J. Numer. Anal.*, 19(2):317–334, 1999.

[22] F. L. Wightman and D. J. Kistler. "The dominant role of low-frequency interaural time differences in sound localization". *J. Acoust. Soc. Am.*, 91(3):1648–1661, 1992.

[23] R. Y. Litovsky, H. S. Colburn, W. A. Yost, and S. J. Guzman. "The precedence effect". *J. Acoust. Soc. Am.*, 106(4):1633–1654, 1999.