Performance of Linear Extrapolation Methods for Virtual Sound Field Navigation

JOSEPH G. TYLKA AND EDGAR Y. CHOUEIRI

(josephgt@alumni.princeton.edu)

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

Performance errors are characterized for two representative linear extrapolation methods for virtual navigation of higher-order ambisonics sound fields. For such methods, navigation is theoretically restricted to within the so-called region of validity, which spherically extends from the recording ambisonics microphone to its nearest source, but the precise consequences of violating that restriction have not been previously established. To that end, the errors introduced by each method are objectively evaluated, in terms of metrics for sound level, spectral coloration, source localization, and diffuseness, through numerical simulations over a range of valid and invalid conditions. Under valid conditions, results show that the first method, based on translating along plane-waves, accurately reproduces both the level and localization of a source, whereas the second method, based on ambisonics translation coefficients, incurs significant errors in both level and spectral content that increase steadily with translation distance. Under invalid conditions, two common features of the performance of both methods are identified: significant localization errors are introduced and the reproduced level is too low. It is argued that these penalties are inherent to all methods that are bound by the region of validity restriction, including all linear extrapolation methods.

0 INTRODUCTION

Virtual navigation of three-dimensional higher-order ambisonics sound fields (i.e., sound fields that have been decomposed into spherical harmonics) enables a listener to explore an acoustic space and, ideally, experience a spatially and tonally accurate perception of the sound field. Applications of this type of virtual sound field navigation (hereafter just "virtual navigation") may be found in virtual-reality (VR) reproductions of real-world spaces. For example, given an acoustic recording of an orchestral performance, a listener can virtually navigate that recording in order to experience that performance from different vantage points. Another application of virtual navigation can be found in VR games, in which synthetic spatial room impulse responses (RIRs) are used to produce spatial audio. Calculating these spatial RIRs on the fly can be computationally intensive, so it may be preferable to pre-render spatial RIRs on a fixed grid of points and then, in real time, navigate between them to generate spatial RIRs at intermediate positions.

According to ambisonics theory, a higher-order ambisonics (HOA) recording of a sound field contains information about the spatial and temporal distribution of sound over the entire spherical free-field region surrounding the microphone (i.e., the so-called *region of validity*) [1–3]. However, it is well-established that a finite-order expansion of a sound field yields only an approximation to that sound field, the accuracy of which decreases with increasing frequency and distance from the expansion center [4, 5], so the prospect of navigating such a sound field is inherently limited. In particular, a sound field is only accurately represented by an Lth-order expansion up to a distance r, provided that

$$kr \leq L, \tag{1}$$

where k is the angular wavenumber [5]. Fortunately, this limitation is primarily a technical one, since synthetic sound fields can be generated to arbitrarily high orders and microphone array technology[1] is rapidly advancing such that it may soon be practical to capture very high-order expansions of real sound fields.

The more fundamental limitation of the spherical-harmonic description of sound fields is that the region over which the HOA expansion is mathematically *valid* is limited by the nearest sound source (or scattering body) to the expansion center (see Sec. 1 for more details). Consequently,

---

[1] See, for example: the Zylia ZM-1 [6], mh acoustics Eigenmike [7], and VisiSonics Audio Camera [8].

from a mathematical standpoint, near-field sources pose a particularly limiting problem to navigation. However, any degradations in sound quality (e.g., in terms of spatial or spectral fidelity) that might result from violating this region of validity restriction have not been clearly determined and are a primary focus of this work.

Several previous studies have developed extrapolation-based (i.e., using only a single HOA microphone[2]) navigational methods, some of which are more susceptible to violating the region of validity restriction than others (as discussed below). In the following section, we provide a critical review of these existing methods and identify the main challenges they face.

0.1 Previous Work

Perhaps the most intuitive navigational method is the *virtual ambisonics* method, as described by Tylka and Choueiri [9, Sec. 3.1]. In this method, the ambisonics signals are first decoded for a given loudspeaker array, which is then simulated in a virtual space with the listener at any desired position within the array. The combined signals arriving at the virtual listener from the virtual loudspeaker array are then computed numerically and played back to the real listener. Some advantages of this method are that we are free of the practical limitations (such as space, floors/walls, cost, etc.) common to physical loudspeaker arrays and we can leverage the wealth of experience with real HOA systems that have been accumulated by researchers in this field (e.g., studies have shown that, at off-center listening positions: adding near-field compensation yields a more accurately reconstructed sound field [10], max-r_E decoding improves localization accuracy compared to standard pseudoinverse decoding [11], etc.).

In a recent study, Fernandez-Grande [12] proposed a similar method, referred to as the *spherical equivalent source method*, in which the sound field is captured with one (or more) HOA microphone(s) and subsequently fitted, in a least-squares sense, to that created by a predefined array of virtual monopole sources. This yields a virtual sound field consisting of a finite set of known monopole sources, which can then be rendered for an arbitrary listening position elsewhere in the space. Unless these virtual source positions can be determined based on *a priori* knowledge of the real source position(s), the performance of this method is unlikely to differ substantially from that of the virtual ambisonics method (especially if a uniformly distributed spherical array of virtual sources is chosen). Additionally, as they have been defined, neither of these methods have any knowledge of any source (or obstacle) positions; consequently, depending on the positions of the real sources in the sound field, the listener may inadvertently navigate too far and violate the region of validity restriction.

The *plane-wave translation* method, described here in Sec. 2.1, entails computing a plane-wave expansion of the

sound field and translating along each plane-wave term. Menzies and Al-Akaidi [13] first derived the mathematical operations required for this method, 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. Schultz and Spors [14] later formulated the plane-wave translation method for the purpose of sound field navigation and examined the time- and frequency-domain consequences of the translation operation. Similar to the methods described above, this method is also prone to violating the region of validity restriction in the presence of near-field sources.

Hahn and Spors [15] evaluated spectral coloration induced by this method by visually examining impulse and frequency responses and found that the induced coloration is often mitigated by matching Q, the number of plane-wave terms, to N, the number of ambisonics signals (i.e., the so-called "critically-sampled" condition, when $Q = N$) [15, Sec. 5]. In the same study, the authors found that when $Q > N$ (in the so-called "oversampled" condition), navigating parallel to the direction of a source introduced less coloration than navigating perpendicularly. The localization properties of this method were explored by Winter et al. [16], who showed that the range over which accurate localization is achieved increases with HOA expansion order and that increasing the number of plane-wave-expansion terms beyond critically-sampled does not improve localization. More recently, Tylka and Choueiri [9] evaluated the localization errors using the velocity and energy localization vectors developed by Gerzon [17], although the perceptual relevance of the findings of this study are limited since the analysis does not take into account the precedence effect.

Another navigational method, referred to here as *ambisonics translation* and described in Sec. 2.2, entails translating the HOA expansion center by re-expanding the sound field about the desired listening position.[3] To enable fast computation of such re-expansions, Gumerov and Duraiswami [2] derived a set of recurrence formulae, which Zotter [3] later rederived for real-valued spherical harmonics and which Tylka and Choueiri [20] recently distilled and replicated for ease of implementation. Menzies and Al-Akaidi [21] in particular described how this method can be used to allow a listener to virtually navigate an HOA sound field, although a detailed analysis was not performed. More recently, Baumgartner and Zotter [22] discussed time-domain implementation issues of the translation filters and proposed a discrete-time realization with improved stability.

Alternatively, one can use the same ambisonics translation filters and perform an *inverse ambisonics translation*

---

[2] Here, we use the term "HOA microphone" to refer to any array of microphone capsules (typically arranged on the surface of a sphere or tetrahedron) that is used to capture ambisonics signals.

[3] Relatedly, Ahrens and Spors [18] proposed a method that uses very similar mathematical operations (Bessel function translations) in order to analytically move the "sweet-spot" in two dimensions within a circular loudspeaker array, and Hannemann and Donohue [19] proposed a multipole-matching method for similar purposes but which requires a synthetic sound field that can be expanded for an arbitrary desired listener position.

by taking a matrix inverse (or pseudoinverse) of the filters for translating from the listener position to the microphone. Wang and Chen [23] recently proposed a modification to this approach, in which the spherical-harmonic translation coefficients are approximated via a finite-term discrete plane-wave decomposition. In that study, the authors showed that their method tends to improve the stability of the matrix inversion compared to using the traditional spherical-harmonic translation coefficients [23, Sec. 4]. However, both the forward and inverse ambisonics translation methods are prone to violating the region of validity restriction if a listener navigates beyond the distance of the nearest source to the microphone.

In order to overcome the region of validity restriction, Wakayama et al. [24] proposed an extrapolation method that is based on spherical-harmonic translation filters but which requires *a priori* knowledge of the source position. This method, referred to here as *multipole ambisonics translation*, was shown to not only enable navigation beyond a near-field source but also accurately estimate the directivity of the source using a multipole expansion. It is not clear, however, whether or how this method can be extended to accommodate multiple sources.

More recently, Plinge et al. [25] developed a parametric method that relies on a time-frequency (i.e., short-time Fourier transform) analysis of the sound field from a single first-order ambisonics microphone as well as a previously measured "distance map" of the environment. This distance map is essentially a source-distance lookup table, consisting of the measured (e.g., optically) distance to the nearest obstacle in each direction (azimuth and elevation) from the microphone. The recorded sound field is first decomposed via directional audio coding (DirAC) [26] in the time-frequency domain into diffuse and directional sound components. Each directional component is then treated as a virtual point-source, with the direction of the source determined via an acoustic intensity vector calculation (cf. Merimaa and Pulkki [27, Eq. (11)]) and its distance determined via the distance map for that direction. The signals from these virtual sources are then "re-recorded" by a virtual microphone at an arbitrary position and with arbitrary directivity. By construction, this method, referred to here as *distance-mapped sound field modeling*, is free of the region of validity restriction as the listener is navigating within a well-defined model of the sound field. However, it is unclear if this method can accurately capture and reproduce the directivities of the real sources, as the method exclusively uses omnidirectional point sources to model the sound field.

A few years earlier, Pihlajamaki and Pulkki [28] developed a conceptually similar method that also leverages a DirAC-based energetic analysis of the sound field and subsequently maps the directional components of the sound field to a predefined so-called "convex hull," in order to create a set of fixed virtual sources with known positions relative to the listener. The more recent method of Allen [29] is also quite similar, except a recursive approach (rather than DirAC) is taken to perform the directional decomposition, and a similarity analysis is performed to distinguish direct

and reflected sounds. This method too requires a predefined set of possible source positions distributed around a convex hull in order to map the extracted directional signals to virtual sources with known positions. Collectively, we refer to these methods as *directional analysis and convex-hull modeling*.

In light of the above discussion we identify the following main issues that existing extrapolation-based navigational methods can face:

1. the region of validity restriction can be violated,
2. localization information can be degraded,
3. spectral colorations or other audible processing artifacts can be introduced,
4. geometric information about the sound field (e.g., source locations) must be known or inferred,
5. arbitrary (e.g., dense or reverberant) sound fields cannot be reproduced, and/or
6. source directivity cannot be captured or reproduced.

The methods discussed above are listed in Table 1. It is worth emphasizing that all of the linear methods are bound by the region of validity restriction (this is actually true more generally for any linear extrapolation method, as discussed in Sec. 7.1), whereas the parametric methods are able to circumvent that restriction either by employing *a priori* knowledge (or *a posteriori* estimates) of source positions or by generating a well-defined model of the incident sound field (or both).

Several other authors have developed linear and parametric interpolation-based navigational methods [30–34, and references therein], which employ an array of HOA microphones and are expected to yield significant improvements compared to extrapolation methods in terms of spectral coloration and source localization. While it is outside the scope of this work to discuss these methods in detail, a critical review of such methods is given by Tylka and Choueiri [35].

0.2 Objectives and Approach

The objectives of the present work are to 1. determine the penalties incurred by violating the region of validity restriction and 2. characterize and compare the performance of the existing linear extrapolation methods. Here, however, we omit the virtual ambisonics and spherical equivalent source methods, since, conceptually, they are not significantly different from the plane-wave translation method (especially for far and uniformly distributed virtual source arrays). We also omit the inverse ambisonics translation method, which is conceptually very similar to the forward ambisonics translation method. Consequently, we take the plane-wave and ambisonics translation methods to be representative of the entire class of linear extrapolation-based navigational methods.

We characterize the performance of each method through numerical simulations over a range of valid and invalid conditions and use objective metrics to evaluate the errors introduced by each method in terms of sound level, spectral

Table 1. Summary of published extrapolation-based navigational methods.

Method	Processing	Additional Inputs
Virtual ambisonics [9, Sec. 3.1]	Linear	
Spherical equivalent source method [12]	Linear	
Plane-wave translation [14]	Linear	
Ambisonics translation [2, 21, 3]	Linear	
Inverse ambisonics translation [23]	Linear	
Multipole ambisonics translation [24]	Parametric	Source position
Distance-mapped sound field modeling [25]	Parametric	Distance map of environment
Directional analysis and convex-hull modeling [28, 29]	Parametric	Predefined convex hull of source positions

coloration, source localization, and diffuseness. In order to identify the penalties associated with violating the region of validity restriction, we look for common features of the performance of both methods under invalid conditions (this is discussed in more detail in Sec. 7.1).

First, in Sec. 1, we review relevant acoustical and ambisonics theory; in Sec. 2, we review the formulation of each navigational method; and in Sec. 3, we describe the numerical simulations and objective metrics employed here. We then present, in Sec. 4, simulations of a typical far-field scenario and objectively determine (in terms of these metrics) suitable parameters for the plane-wave decomposition calculation required for the plane-wave translation method. Next, in Sec. 5, we explore basic properties of each method by computing the effective frequency responses induced by translation for various source azimuths. We then present and discuss in Sec. 6 the results of simulations characterizing and comparing the performance of both methods. Finally, in Sec. 7, we summarize the findings from the presented analyses, discuss the practical viability of each method, and draw broader conclusions regarding the penalties associated with violating the region of validity restriction.

1 RELEVANT AMBISONICS THEORY

As is common in HOA, we adopt Cartesian and spherical coordinate systems in which, for a listener positioned at the origin, the $+x$-axis points forward, the $+y$-axis points to the left, and the $+z$-axis points upward. Correspondingly, r is the (nonnegative) radial distance from the origin, $\theta \in [-\pi/2, \pi/2]$ is the elevation angle above the horizontal $(x\text{-}y)$ plane, and $\phi \in [0, 2\pi)$ is the azimuthal angle around the vertical (z) axis, with $(\theta, \phi) = (0, 0)$ corresponding to the $+x$ direction and $(0, \pi/2)$ to the $+y$ direction. For a position vector $\vec{r} = (x, y, z)$, we denote the corresponding unit vector by $\hat{r} \equiv \vec{r}/r$.

Here, we use real-valued orthonormal (N3D) spherical harmonics, as given by Zotter [3, Sec. 2.2], and we adopt the ambisonics channel number (ACN) convention [36] such that, for a spherical harmonic function of degree $l \in [0, \infty)$ and order $m \in [-l, l]$, the ACN index n is given by $n = l(l+1) + m$ and the spherical harmonic function is denoted by Y_n.

In the free field (i.e., in a region free of sources and scattering bodies), the *acoustic potential field*, ψ (defined as the Fourier transform of the acoustic pressure field), satisfies the homogeneous Helmholtz equation and can therefore
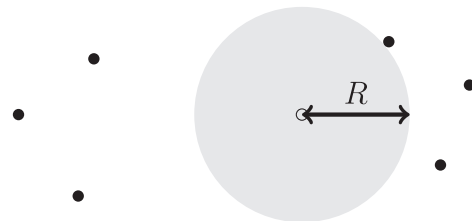

Fig. 1. Diagram of the region of validity for a free-field spherical Fourier-Bessel expansion. The empty circle indicates the expansion center, the filled circles indicate source positions, and the shaded disk indicates the region of validity. The radius of the region of validity is equal to R, the distance of the nearest source from the expansion center.

be expressed as an infinite sum of regular (i.e., not singular) basis solutions. In ambisonics, these basis solutions are given by $j_l(kr)Y_n(\hat{r})$, where j_l is the spherical Bessel function of order l and k is the angular wavenumber. The sum, also known as a spherical Fourier-Bessel series expansion, is given by [2, Ch. 2]

$$\psi(k, \vec{r}) = \sum_{n=0}^{\infty} 4\pi(-i)^l A_n(k) j_l(kr) Y_n(\hat{r}), \qquad (2)$$

where A_n are the corresponding (frequency-dependent) expansion coefficients and we have, without loss of generality, factored out $(-i)^l$ to ensure conjugate-symmetry in each A_n, thereby making each ambisonics signal (i.e., the inverse Fourier transform of A_n) real-valued for a real pressure field. In practice, this expansion is truncated to a finite order L (i.e., $l \in [0, L]$), yielding $N = (L + 1)^2$ terms.

The series solution in Eq. (2) can be used to describe the acoustic potential in an interior region, that is, for $r < R$, where R is a finite distance [1, Sec. 6.8]. So that the region remains source- and obstacle-free, R is typically taken to be the distance of the nearest source or scattering body to the origin. This defines the so-called *region of validity* around a given expansion center, which is illustrated in Fig. 1.

The ambisonics encoding filters for a point source located at $\vec{s}_0$ and expanded about the origin are given in the frequency domain by [4, Eq. (10)]

$$A_n(k) = i^{l+1} k h_l(k s_0) Y_n(\hat{s}_0), \qquad (3)$$

where h_l is the (outgoing) spherical Hankel function of order l. Similarly, the ambisonics encoding filters for a
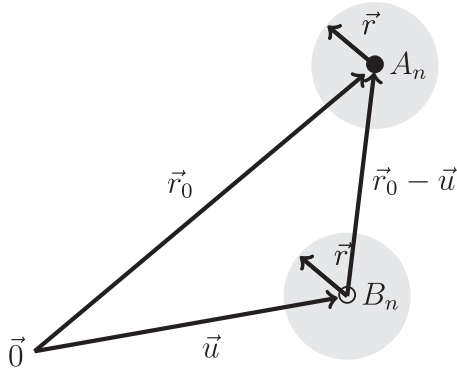
Fig. 2. Diagram of microphone and listener positions. The empty circle indicates the microphone position, the filled circle indicates the listener position, and the shaded disks indicate the regions of the sound field represented by the corresponding ambisonics expansions.

plane-wave source originating from $\hat{s}_0$ are given by [4, Eq. (9)]

$$A_n(k) = Y_n(\hat{s}_0). \qquad (4)$$

2 NAVIGATIONAL METHODS

Generally, we seek the ambisonics signals, A_n, up to order L_{out}, that describe the sound field in the vicinity of the listener, located at $\vec{r}_0$. Using an order-truncated version of Eq. (2), this potential field is given by

$$\psi(k, \vec{r} + \vec{r}_0) = \sum_{n=0}^{N_{\text{out}}-1} 4\pi(-i)^l A_n(k) j_l(kr) Y_n(\hat{r}), \qquad (5)$$

where $N_{\text{out}} = (L_{\text{out}} + 1)^2$, A_n now denotes the ambisonics signals for an expansion about $\vec{r}_0$, and $\vec{r}$ corresponds to the displacement away from that expansion center.

We take as inputs to each navigational method a set of measured (or synthesized) ambisonics signals, B_n, up to order L_{in}, that describe the sound field in the vicinity of the microphone (either real or virtual), which is located at $\vec{u}$.[4] Again using Eq. (2), this potential field is given by

$$\psi(k, \vec{r} + \vec{u}) = \sum_{n=0}^{N_{\text{in}}-1} 4\pi(-i)^l B_n(k) j_l(kr) Y_n(\hat{r}), \qquad (6)$$

where $N_{\text{in}} = (L_{\text{in}} + 1)^2$ and B_n denotes the ambisonics signals for an expansion about $\vec{u}$. Thus, estimation of A_n from B_n amounts to a translation from the position of the microphone, $\vec{u}$, along the vector $\vec{r}_0 - \vec{u}$, as illustrated in Fig. 2.

_____
[4] The order of the ambisonics microphone is generally determined by the number of capsules in the assembly.

For later use, we define vectors of ambisonics signals, given by

$$\mathbf{a} = \begin{bmatrix} A_0 \\ A_1 \\ \vdots \\ A_{N_{\text{out}}-1} \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} B_0 \\ B_1 \\ \vdots \\ B_{N_{\text{in}}-1} \end{bmatrix}. \qquad (7)$$

Below, we review the derivations of the plane-wave and ambisonics translation methods.

2.1 Plane-Wave Translation

Given the measured ambisonics signals, B_n, we first compute the so-called *signature function*, μ, which represents the coefficients of a plane-wave decomposition of the sound field, such that the potential field can be approximately reconstructed using a finite number of plane-waves. For a given plane-wave source direction $\hat{v}$, the signature function is given by [2, Sec. 2.3.3]

$$\mu(k, \hat{v}; \vec{u}) = \sum_{n=0}^{N_{\text{in}}-1} B_n(k) Y_n(\hat{v}_q), \qquad (8)$$

where the $\vec{u}$ parameter of μ denotes the position of the expansion center, i.e., the position of the microphone. Equivalently, written as a matrix equation, we have

$$\begin{bmatrix} \mu(k, \hat{v}_1; \vec{u}) \\ \mu(k, \hat{v}_2; \vec{u}) \\ \vdots \\ \mu(k, \hat{v}_Q; \vec{u}) \end{bmatrix} = \mathbf{Y}^{\text{T}} \cdot \mathbf{b}, \qquad (9)$$

where Q is the number of plane-wave terms; $\hat{v}_q$ is the source direction of the qth term, $\forall q \in [1, Q]$; $(\cdot)^{\text{T}}$ denotes the transpose of the argument; and $\mathbf{Y}$ is a matrix of spherical harmonics, given by

$$\mathbf{Y} = \begin{bmatrix} Y_0(\hat{v}_1) & Y_0(\hat{v}_2) & \cdots & Y_0(\hat{v}_Q) \\ Y_1(\hat{v}_1) & Y_1(\hat{v}_2) & \cdots & Y_1(\hat{v}_Q) \\ \vdots & \vdots & \ddots & \vdots \\ Y_{N_{\text{in}}-1}(\hat{v}_1) & Y_{N_{\text{in}}-1}(\hat{v}_2) & \cdots & Y_{N_{\text{in}}-1}(\hat{v}_Q) \end{bmatrix}. \qquad (10)$$

Due to its direct application of spherical harmonics, this approach for computing μ via Eq. (9) is known as *modal beamforming* (hereafter just "beamforming"), cf. Hahn and Spors [15, Sec. 3].

Relative to the expansion center, the incident potential field is approximately reconstructed by

$$\psi(k, \vec{r} + \vec{u}) \approx \sum_{q=1}^{Q} w_q \mu(k, \hat{v}_q; \vec{u}) e^{-ik\hat{v}_q \cdot \vec{r}}, \qquad (11)$$

where w_q is the quadrature weight of the qth term, which is dependent on the chosen grid of directions. By convention, we assume $\sum_{q=1}^{Q} w_q = 4\pi$ and typically require that $w_q \neq 0$, $\forall q \in [1, Q]$.

From this equation, we can estimate the potential field in the vicinity of an arbitrary listener position, $\vec{r}_0$, by substituting $\vec{r} + \vec{r}_0 - \vec{u}$ for $\vec{r}$ in Eq. (11). For each term in that summation, the potential field at $\vec{r} + \vec{r}_0$ then differs only by a phase-factor $e^{-ik\hat{v}_q \cdot (\vec{r}_0 - \vec{u})}$, so we combine this factor into

(a) Varying $\|\vec{r}_0 - \vec{u}\|$; fixed $L_{\text{in}} = 4$

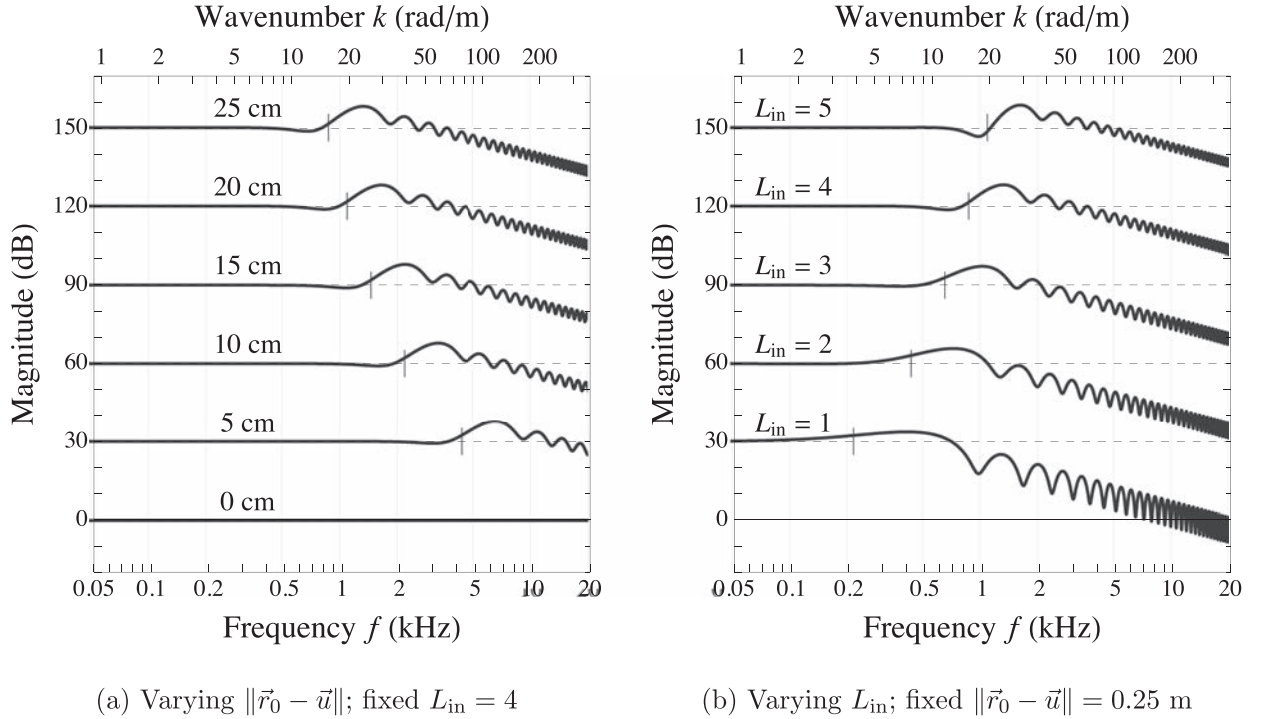(b) Varying L_{in}; fixed $\|\vec{r}_0 - \vec{u}\| = 0.25$ m

Fig. 3. Magnitude responses caused by the ambisonics translation method for (a) various translation distances and (b) input orders. The bottom axes show frequency in kHz while the top axes show the angular wavenumber k. The short vertical lines indicate the nondimensional frequency $k\|\vec{r}_0 - \vec{u}\| = L_{\text{in}}$. For legibility, each frequency response is offset by 30 dB.

the signature function and define the *translated signature function*, μ', given by [13]

$$\mu'(k, \hat{v}_q; \vec{r}_0) = \mu(k, \hat{v}_q; \vec{u})e^{-ik\hat{v}_q \cdot (\vec{r}_0 - \vec{u})}. \tag{12}$$

Effectively, this operation translates the expansion center from $\vec{u}$ to $\vec{r}_0$.

The translated signature function can be converted back into ambisonics signals, up to an arbitrary order, L_{out}, yielding ambisonics signals for the listener, A_n, given by

$$A_n(k) = \sum_{q=1}^{Q} w_q \mu'(k, \hat{v}_q; \vec{r}_0) Y_n(\hat{v}_q), \tag{13}$$

where we have essentially encoded into ambisonics the sum of Q plane-wave signals via Eq. (4).

As will become relevant in Sec. 4, a plane-wave decomposition can alternatively be computed via a *pseudoinversion* approach: instead of using Eq. (9), μ is given by

$$\begin{bmatrix} \mu(k, \hat{v}_1; \vec{u}) \\ \mu(k, \hat{v}_2; \vec{u}) \\ \vdots \\ \mu(k, \hat{v}_Q; \vec{u}) \end{bmatrix} = \mathbf{W}^{-1} \cdot \mathbf{Y}^{+} \cdot \mathbf{b}, \tag{14}$$

where $(\cdot)^{+}$ denotes pseudoinversion and

$$\mathbf{W} = \text{diag}\left\{ \begin{bmatrix} w_1 & w_2 & \cdots & w_Q \end{bmatrix} \right\}. \tag{15}$$

2.2 Ambisonics Translation

Given ambisonics signals B_n for an expansion about $\vec{u}$, translated ambisonics signals for an expansion about $\vec{r}_0$ are given by [2, Ch. 3]

$$A_{n'}(k) = \sum_{n=0}^{N_{\text{in}}-1} T_{n,n'}(k, \vec{r}_0 - \vec{u}) B_n(k), \tag{16}$$

where $T_{n,n'}$ are the so-called *translation coefficients*, which we compute using the recurrence formulae given by Gumerov and Duraiswami [2, Sec. 3.2], Zotter [3, Ch. 3], and Tylka and Choueiri [20]. Note that the translated expansion coefficients $A_{n'}$ can be computed to an arbitrary output order L_{out}. In matrix form, we can write Eq. (16) as

$$\mathbf{a}(k) = (\mathbf{T}(k, \vec{r}_0 - \vec{u}))^{\text{T}} \cdot \mathbf{b}(k), \tag{17}$$

where, omitting dependencies,

$$\mathbf{T} = \begin{bmatrix} T_{0,0} & T_{0,1} & \cdots & T_{0,N_{\text{out}}-1} \\ T_{1,0} & T_{1,1} & \cdots & T_{1,N_{\text{out}}-1} \\ \vdots & \vdots & \ddots & \vdots \\ T_{N_{\text{in}}-1,0} & T_{N_{\text{in}}-1,1} & \cdots & T_{N_{\text{in}}-1,N_{\text{out}}-1} \end{bmatrix}. \tag{18}$$

As the full derivation for computing these translation coefficients is not easily presented in a concise manner, the interested reader is referred to another publication of ours, in which we have distilled and replicated the necessary formulae for this computation [20]. Essentially, the translation coefficients are computed by decomposing the translation operation into three steps: first, a rotation of the ambisonics signals to orient the z-axis in the direction of desired
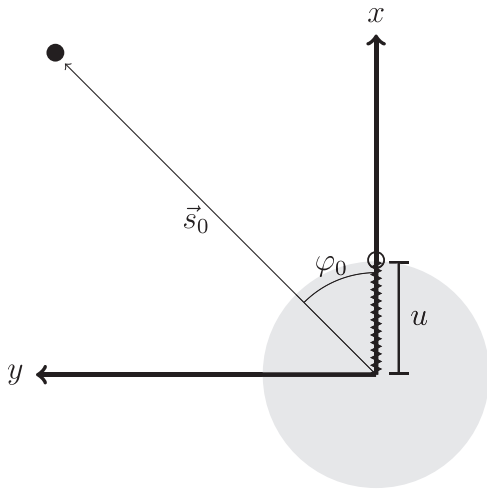
Fig. 4. Diagram of a microphone (empty circle) and a single source (filled circle). The shaded gray disk indicates the interior region, where $r < u$. The jagged line segment indicates the navigable region, where $x \in [0, u]$ and $y = z = 0$.

translation; followed by a translation along that new z-axis; and finally a rotation back to the original (or any other desired) orientation. Due to the axial symmetry of many of the spherical harmonics, the recurrence formulae are significantly simplified by translation along the z-axis. Rotation of the listener (or, equivalently, of the sound field) in ambisonics is straightforward and has been well-established in the literature [2, 3] and involves the application of rotation matrices to change coordinates and a frequency-independent mixing of the ambisonics signals.

Using Eq. (17) to translate along $\vec{r}_0 - \vec{u}$, the potential field at $\vec{r} + \vec{r}_0$ is then given by Eq. (5). It is important to note that the re-expanded field is still limited in accuracy and region of validity by the original expansion. In other words, with increasing L_{out}, the re-expanded field approaches the original *order-limited* field, not the incident field.

2.2.1 Frequency-Domain Behavior

According to Eq. (1), the accuracy of the translated ambisonics signals will degrade with increasing frequency and distance away from the microphone [5]. To explore this behavior in the frequency domain, we compute the frequency response induced by translation away from the microphone and plot, in Fig. 3, the magnitude responses corresponding to various translation distances and input orders. Simulations of this type are described in more detail in Sec. 3.1.1. From these plots, we see that translation effectively acts as a low-pass filter, the corner frequency of which approximately corresponds to $k \propto L_{in}/\|\vec{r}_0 - \vec{u}\|$, where $\|\cdot\|$ denotes the ℓ^2 norm (Euclidean distance) of a vector. (Although not shown here, it can be verified that this behavior also holds across all source and listener positions. See, for example, Fig. 6(b), which shows this behavior over multiple source azimuths.)

It is worth noting that the inverse ambisonics translation method (discussed in Sec. 0.1) simply entails taking an

inverse (or pseudoinverse if $L_{in} \neq L_{out}$) of the translation coefficients matrix, such that

$$\mathbf{a}(k) = \left((\mathbf{T}(k, \vec{u} - \vec{r}_0))^{\mathrm{T}} \right)^{-1} \cdot \mathbf{b}(k). \tag{19}$$

However, as the forward translation operation (given by Eq. (17)) leads to an attenuation (i.e., a "roll-off") of high frequencies, the inverse operation will yield excessive high-frequency amplification unless steps (such as regularization) are taken to mitigate such excessive gains.

3 SIMULATIONS AND METRICS

Consider a single-microphone array geometry, illustrated in Fig. 4, in which a single microphone is separated from the origin by a distance u and placed along the longitudinal x-axis, such that its position is given by $\vec{u} = (u, 0, 0)$. In this configuration, we define the *navigable region* as the segment of the x-axis connecting the origin and microphone position, i.e., all listener positions $\vec{r}_0 = (x_0, 0, 0)$ where $x_0 \in [0, u]$.

We further define a nondimensional geometrical parameter $\gamma = r/u$ and refer to the region with $\gamma > 1$ as the *exterior region* and that with $\gamma < 1$ as the *interior region* (see Fig. 4).

Note that, by these definitions, interior sources (where $s_0/u = \gamma < 1$) tend to yield invalid conditions, as most of the listening positions lie outside of the region of validity of the microphone, whereas exterior sources ($\gamma > 1$) tend to yield valid conditions. Strictly speaking, all listener positions are valid when $\gamma > 2$ and become invalid as $\gamma \to 0$; otherwise, the precise validity of the situation depends on the azimuth of the source as well as its distance, since a listening position is only valid when $\|\vec{r}_0 - \vec{u}\| < \|\vec{s}_0 - \vec{u}\|$.

3.1 Simulation Parameters

We simulate recording of the sound field depicted in Fig. 4 for a range of microphone positions, $u \in [0.1, 10]$ m, and all source distances $s_0 = \gamma u$ for $\gamma \in [0.1, 10]$. In each simulation, we vary the source azimuth from $\varphi_0 = 0°$ to $180°$ in increments of $5°$ and generate, using Eq. (3), an artificial HOA impulse response at the microphone. We then estimate, via translation using each method, the HOA impulse responses at listener positions from $x_0 = 0$ to u, taken in 20 equal increments.

In all simulations, unless stated otherwise, we choose $L_{in} = 4$ and $L_{out} = 1$.[5] The sampling rate is 48 kHz and all impulse responses are calculated with 16,384 samples ($\approx$ 341 ms). Additionally, we filter all point-source HOA impulse responses with order-dependent near-field compen-

---

[5] Note that, for the metrics listed in Sec. 3.2, only the localization model (described in Sec. 3.2.3) depends on L_{out}; all of the other metrics, by construction, use only the zeroth- and first-order signals.

sation high-pass filters, given for the lth-order ambisonics signals by

$$H_l(f) = 1 - \frac{1}{\sqrt{1 + \left(\frac{f}{f_l}\right)^l}}, \tag{20}$$

where f_l is the corner frequency of the lth filter, which we choose to be $f_l = (200 \times l)$ Hz.

3.1.1 Source Azimuth Dependence

In order to explore the basic properties of a given navigational method, we consider a representative far-field scenario and compute the effective frequency response induced by translation. For these simulations, we pick $s_0 = 2.5$ m and $\gamma = 10$ (so $u = 0.25$ m) and translate to $\vec{r}_0 = (0, 0, 0)$. Then, for each source azimuth, we compute the induced frequency response, which is given by the ratio of the zeroth-order translated ambisonics signal, $A_0(k)$, to the zeroth-order reference ambisonics signal, $B_0(k)$, that would have been measured at $\vec{r}_0$.

3.2 Metrics

We quantify the errors incurred through navigation by each method using the following metrics:

1. the level error, e_λ, of the mean audible energy (MAE), as given in Sec. 3.2.1;
2. the range, ρ_η, of the auditory band spectral error (ABSE), as given in Sec. 3.2.2;
3. the localization error, e_ν, as given in Eq. (26), for the localization vector model described in Sec. 3.2.3; and
4. the error, e_Ψ, in the diffuseness parameter, as given in Sec. 3.2.4.

For each simulation presented here, we average these error metrics over the entire navigable region (as defined above) and all source azimuths, for specified combinations of source distance s_0 and microphone distance u.

3.2.1 Mean Audible Energy (λ)

We define the *mean audible energy* (MAE), λ, of an ambisonics signal as the average energy of the zeroth-order term across a set of critical bands,[6] i.e.,

$$\lambda = 10 \log_{10} \left(\frac{1}{N_b} \sum_{c=1}^{N_b} \frac{\int |H_\Gamma(f; f_c)||A_0(f)|^2 df}{\int |H_\Gamma(f; f_c)| df} \right), \tag{21}$$

where $H_\Gamma(f; f_c)$ is the transfer function of a gammatone filter[7] (which approximates critical bands), with center fre-

quency f_c for $c \in [1, N_b]$, and integration is taken over all frequencies f. Here, we choose f_c to be a set of equivalent rectangular bandwidth (ERB)-spaced center frequencies [39] spanning the range $f_c \in [50$ Hz, 21 kHz]. The summand in Eq. (21) approximately represents the power spectrum of a signal reaching the cochlea, as defined by Salomons [40, Eq. (5.12)]. We further define the level error, given in dB by

$$e_\lambda = \tilde{\lambda} - \lambda, \tag{22}$$

where λ is the MAE for a reference signal and $\tilde{\lambda}$ is that for a translated signal.

3.2.2 Auditory Band Spectral Error (η)

The *auditory band spectral error* (ABSE), adapted from Schärer and Lindau [41, Eq. (9)], is given by

$$\eta(f_c) = 10 \log_{10} \left(\frac{\int |H_\Gamma(f; f_c)||\tilde{A}_0(f)|^2 df}{\int |H_\Gamma(f; f_c)||A_0(f)|^2 df} \right), \tag{23}$$

where A_0 and $\tilde{A}_0$ are the zeroth-order terms of the reference and translated HOA transfer functions, respectively; each integration is taken over all frequencies f; and we again choose ERB-spaced $f_c \in [50$ Hz, 21 kHz], as recommended by Boren et al. [42]. We further define the *spectral error*, given by

$$\rho_\eta = \max_c \eta(f_c) - \min_c \eta(f_c), \tag{24}$$

which we found in a previous study to be a strong predictor of perceptible colorations induced through navigation [43].

To provide some intuition for relevant magnitudes of this spectral error, we refer to that prior subjective validation study, in which we showed that the perceived coloration $C \sim 2.88\rho_\eta$ [43, Table 4], where C is a percentage indicating perceived coloration relative to a reference ($C = 0\%$ coloration) and a 3.5 kHz low-pass filter (100%). On this scale, $\rho_\eta \approx 3.5$ dB corresponds to a 10% relative coloration, which we take to suggest that errors on the order of 3 dB or greater are likely significant and easily perceptible, whereas errors smaller than 3 dB may be quite tolerable in some applications.

3.2.3 Precedence-Effect Localization Vector ($\tilde{\nu}$)

Localization is predicted using a recently developed precedence-effect-based localization model, the details of which are provided in an earlier publication [43, Sec. 2.A.i]. This model extends the precedence-effect-based energy vector model of Stitt et al. [44] in order to compute a predicted perceived source localization vector, $\vec{\nu}$, from an ambisonics impulse response.[8] Briefly, this model entails decomposing the ambisonics impulse response into a finite set of plane-wave impulse responses, which are further divided into wavelets with distinct arrival times. The signal amplitudes, plane-wave directions, and times-of-arrival for all wavelets are fed into the original energy vector model of

---

[6] Roughly speaking, a *critical band* refers to the bandwidth of the effective auditory filter created by the cochlea, within which a stronger second tone will mask the perception of a weaker first tone [37, Ch. 3].

[7] In this work, we used the gammatone filters implemented in the large time-frequency analysis toolbox (LTFAT) for MATLAB [38].

---

[8] In contrast, the original model of Stitt et al. only considered loudspeaker gains and positions relative to the listener.

Stitt et al. to produce a single predicted source localization vector.[9]

In their model, Stitt et al. predict localization as a weighted average of position vectors for each of a finite set of discrete sources. Perceptual weights are first computed for each source based on its time-of-arrival to and amplitude at the listening position. The calculation of these weights is not trivial, but briefly it involves weighting earlier signals more heavily but adjusting the weights of later signals based on their amplitudes and directions-of-arrival (see the original paper for details). The predicted localization is then given as a vector by

$$\vec{v} = \frac{\sum_{q=1}^{Q} |w_q(\alpha) G_q / v_q|^2 \, \hat{v}_q}{\sum_{q=1}^{Q} |w_q(\alpha) G_q / v_q|^2}, \tag{25}$$

where w_q is the perceptual weight for the qth source, G_q is the amplitude of that source, $\vec{v}_q$ is its position, and $\alpha \in [0, 1]$ is a free parameter that specifies the relative importance of stationary (i.e., time-averaged) to transient information in the stimulus signal.[10] Here, we use the optimal parameter values (e.g., for α) as previously determined through subjective experimental validation by Tylka and Choueiri [43, Table 2].

Given this predicted localization direction, the corresponding localization error e_v is then computed by

$$e_v = \cos^{-1} \left(\hat{v} \cdot \hat{s}_0' \right), \tag{26}$$

where $\hat{s}_0'$ is the direction of the source relative to the listener, found by normalizing the vector $\vec{s}_0' = \vec{s}_0 - \vec{r}_0$.

3.2.4 Diffuseness Parameter (Ψ)

According to Merimaa and Pulkki [27], the acoustic intensity vector and a diffuseness parameter can be computed using the four standard "B-format" signals. These signals are related to the first four ACN/N3D ambisonics signals by

$$W = \frac{A_0}{\sqrt{2}}, \quad Y = \frac{A_1}{\sqrt{3}}, \quad Z = \frac{A_2}{\sqrt{3}}, \quad X = \frac{A_3}{\sqrt{3}}, \tag{27}$$

where this relationship can be derived by comparing Eq. (4) and the standard B-format encoding equations, as specified by Malham and Myatt [45, p. 62].

---

[9] More precisely, this operation is performed in several frequency bands, and a weighted average localization direction is computed using signal energies in each band as weights [43, Sec. 2.A.i].

[10] For example, a highly transient signal is expected to require a low value of α, while a more stationary signal would require a higher value [44].

We construct a frequency-dependent Cartesian row vector, $\vec{X} = \begin{bmatrix} X & Y & Z \end{bmatrix}$, and the diffuseness parameter Ψ is then given by [27, Eq. (12)]

$$\Psi(f) = 1 - \sqrt{2} \frac{\left\| \text{Re} \left\{ \overline{W}(f) \vec{X}(f) \right\} \right\|}{|W(f)|^2 + \left\| \vec{X}(f) \right\|^2 / 2}, \tag{28}$$

where $\overline{(\cdot)}$ denotes taking the complex conjugate of the argument, and, for a complex-valued vector, $\| \cdot \|$ denotes the norm given by $\|\vec{x}\| = \sqrt{\langle \vec{x}, \vec{x} \rangle} \equiv \sqrt{\vec{x} \, \vec{x}^{\text{H}}}$, where $(\cdot)^{\text{H}}$ denotes the conjugate (Hermitian) transpose of the argument. Thus, Ψ is a real-valued scalar which takes on values between $\Psi \in [0, 1]$, where $\Psi = 0$ corresponds to a purely directional incident sound field and $\Psi = 1$ corresponds to a purely diffuse incident sound field.

We then compute the logarithmically weighted mean of the difference between the diffuseness spectra for the test and reference samples, given by

$$e_\Psi = \frac{\int_{f_L}^{f_U} \frac{1}{f} \left(\tilde{\Psi}(f) - \Psi(f) \right) df}{\log(f_U) - \log(f_L)}, \tag{29}$$

where $f_L = 50$ Hz and $f_U = 21$ kHz.

## 4 PLANE-WAVE DECOMPOSITION APPROACH

In this section, we compare the two methods for performing the plane-wave decomposition introduced in Sec. 2.1:

1. beamforming, as given by Eq. (9); and
2. pseudoinversion, as given by Eq. (14).

In particular, we explore the performance of each method across a range of plane-wave grid densities and input ambisonics orders, L_{in}. For all plane-wave expansions, we compute Q plane-wave terms arranged on Fliege nodes and use the corresponding quadrature weights (available online [46]). Additionally, for ease of comparison with L_{in}, we define the "order" of a given plane-wave expansion as $\sqrt{Q} - 1$. As mentioned in Sec. 3.2, all errors are computed averaging over all source azimuths and over all listener positions in the navigable region (see Fig. 4). For these simulations, we examine a typical far-field scenario by choosing $u = 0.25$ m and $s_0 = 2.5$ m (so $\gamma = 10$).

In Figs. 5(a) and 5(b), we plot the level errors for each plane-wave decomposition method as a function of Q and L_{in}. From Fig. 5(a), we note a region of minimal errors that follows the diagonal line $L_{\text{in}} \approx \sqrt{Q} - 1$. This suggests that, for the beamforming method, it is advantageous to match the number of plane-wave terms to the number of ambisonics signals, i.e., $Q \approx N_{\text{in}}$. Following Hahn and Spors [15], we refer to this condition as the "critically-sampled" condition, since $Q = N_{\text{in}}$, while we refer to $Q > N_{\text{in}}$ as oversampled and $Q < N_{\text{in}}$ as undersampled.
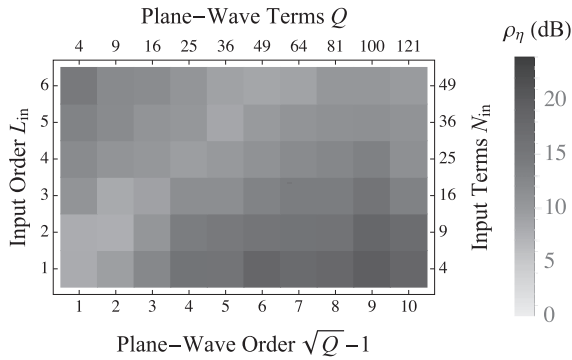
From Fig. 5(b), we see a similar region of small errors, but it is less pronounced, since for $Q > N_{\text{in}}$ (bottom right corner of the plot), the errors are also small. However, we
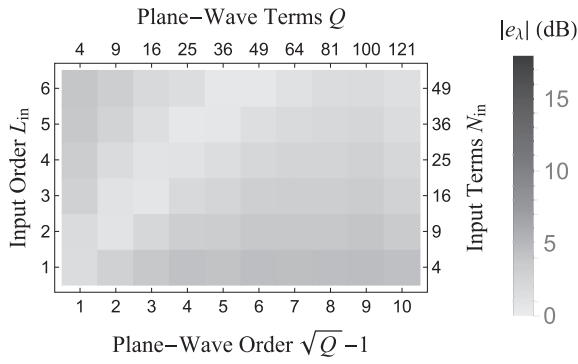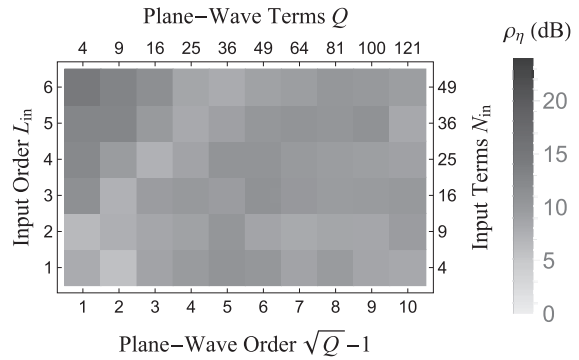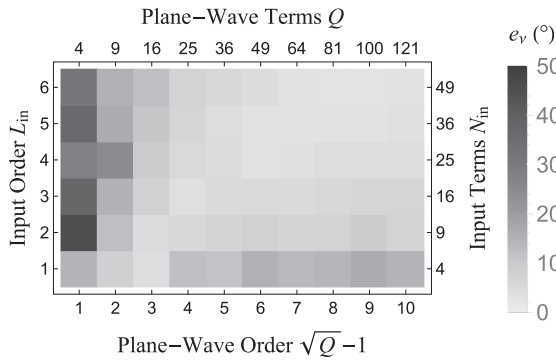
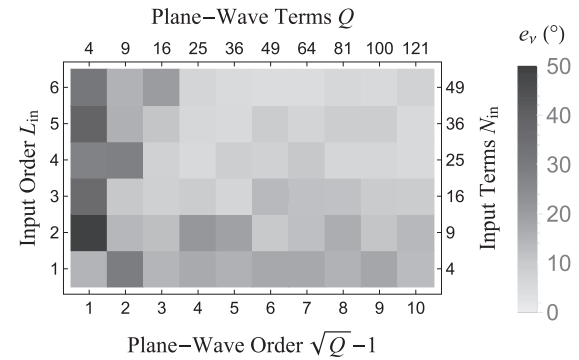(a) Level error e_λ – beamforming
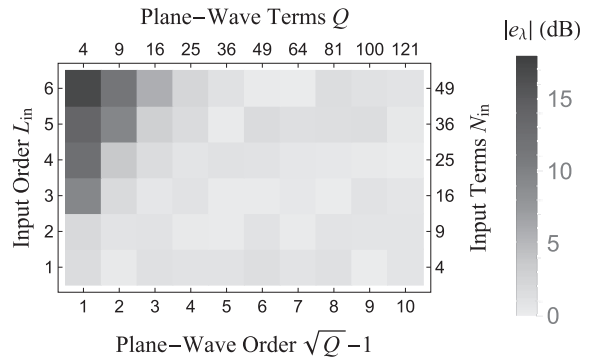
(b) Level error e_λ – pseudoinversion

(c) Spectral error ρ_η – beamforming

(d) Spectral error ρ_η – pseudoinversion
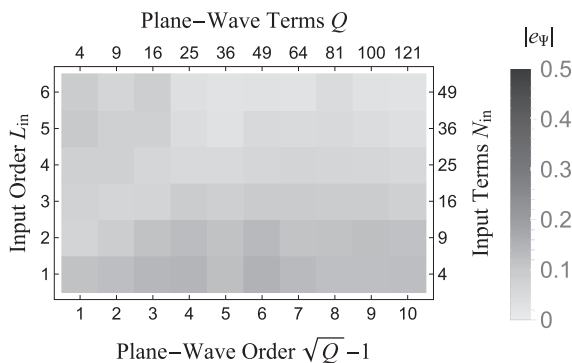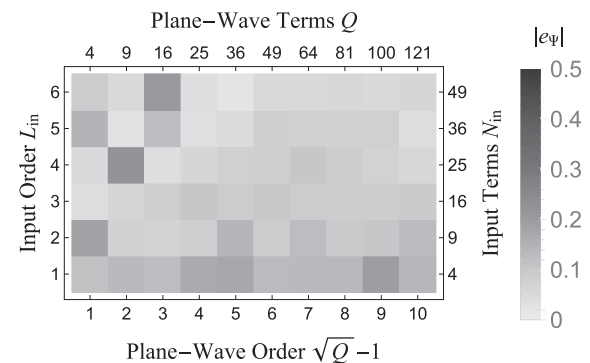

(e) Localization error e_ν – beamforming

(f) Localization error e_ν – pseudoinversion

(g) Diffuseness error e_Ψ – beamforming

(h) Diffuseness error e_Ψ – pseudoinversion

Fig. 5. Dependence of errors on Q and ambisonics input order L_in for the (a,c,e,g) beamforming and (b,d,f,h) pseudoinversion plane-wave decomposition methods.

(a) Plane-wave translation
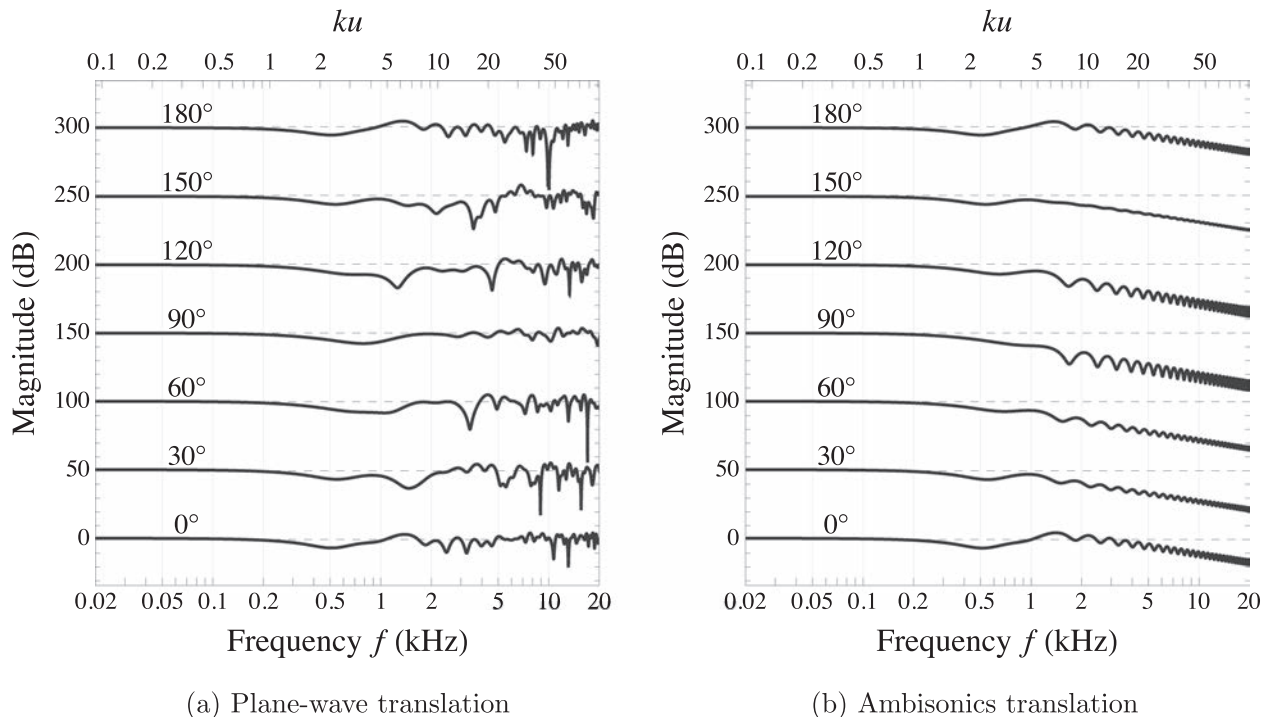
(b) Ambisonics translation

Fig. 6. Magnitude responses caused by the plane-wave (a) and ambisonics translation (b) methods for various source azimuths. The bottom axes show frequency in kHz while the top axes show the nondimensional frequency ku for a microphone distance of $u = 0.25$ m. For legibility, each frequency response is offset by 50 dB and the responses have been artificially truncated (where needed) to not exceed -45 dB.

note that a region of very large errors exists where $Q < N_{in}$ (top left corner). This suggests that, for the pseudoinversion method, it is advantageous to have $Q \geqslant N_{in}$; i.e., to have at least as many plane-wave terms as ambisonics signals.

The spectral errors plotted in Figs. 5(c) and 5(d) show trends similar to those shown by the level errors. From these plots, it is clear that only the critically-sampled condition is ideal for the beamforming method. This is in agreement with the finding of Hahn and Spors [15, cf. Fig. 7]: that the critical-sampling condition yields decreased coloration (compared to oversampling) when using the beamforming method.[11] From Fig. 5(d), we see that both the critical-sampling and oversampling conditions yield small errors for the pseudoinversion method. Oversampling and under-sampling for the beamforming method, as well as under-sampling for the pseudoinversion method, all yield significantly larger errors than other conditions.

From the localization errors plotted in Figs. 5(e) and 5(f), we see that both methods yield large errors for undersampled conditions. For the beamforming method, the errors are otherwise uniformly small, except at very small ambisonics orders ($L_{in} = 1$). That these errors do not improve with increasing Q corroborates the finding of Winter et al. [16, cf. Fig. 4]: that increasing the number of plane-waves beyond critically-sampled does not improve localization when using the beamforming method. From Fig. 5(f), we

see that the pseudoinversion method appears sensitive to mismatches between Q and N_{in} given the presence of large, sporadic errors that do not follow any obvious pattern.

From the diffuseness errors plotted in Fig. 5(g), we see that the beamforming method only yields relatively large errors at low ambisonics orders, e.g., $L_{in} = 1, 2$. For the pseudoinversion method, however, we again see a sensitivity to mismatched sampling conditions, yielding relatively large errors without any discernible pattern. Consequently, unless the plane-wave terms can be carefully chosen ahead of time for a given ambisonics order, beamforming is likely the safer method.

5 SOURCE AZIMUTH DEPENDENCE

In this section, we examine the effective frequency response induced by translation via each method as a function of source azimuth. As described in Sec. 3.1.1, for these simulations, we let $L_{in} = 4$, pick $u = 0.25$ m and $s_0 = 2.5$ m (so $\gamma = 10$), and translate to $\vec{r}_0 = (0, 0, 0)$. Based on the findings discussed above, here we use beamforming with $Q = N_{in} = 25$ for the plane-wave translation method.

The induced frequency responses are plotted in Fig. 6. From Fig. 6(a), we see that the plane-wave translation method introduces sporadic notches in the frequency response of the signal, which do not appear to follow any obvious pattern. Otherwise, however, the frequency responses are largely flat for all source azimuths. This is in agreement with the finding of Hahn and Spors [15, cf. Figs. 7(c) and 7(f)]: that the induced frequency responses are largely flat

---

[11] Similar phenomena related to the over- and undersampling of ambisonics have been explored in detail by Solvang [47] and Bernschütz et al. [48].
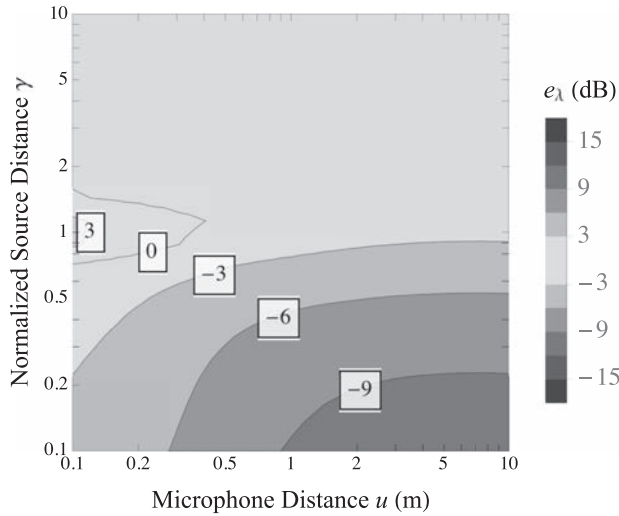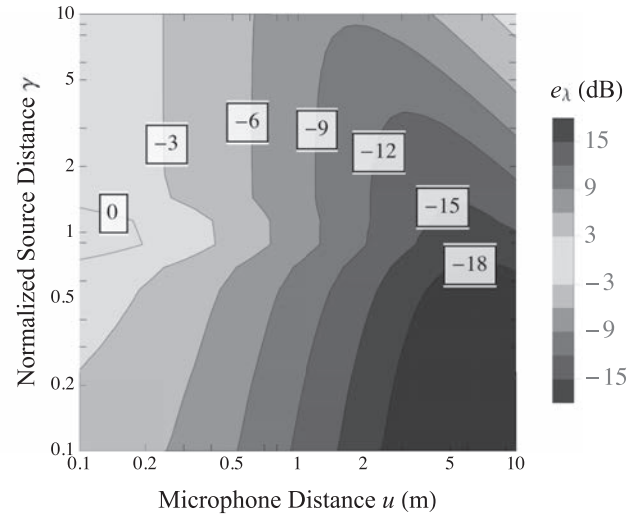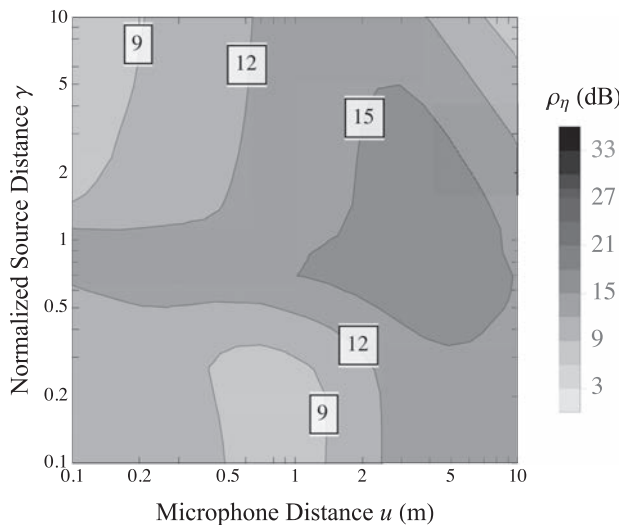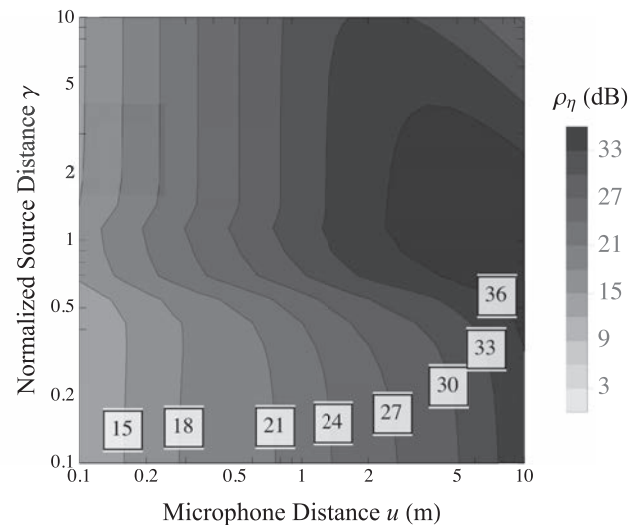
(a) Level errors e_λ – plane-wave translation

(b) Level errors e_λ – ambisonics translation

(c) Spectral errors ρ_η – plane-wave translation

(d) Spectral errors ρ_η – ambisonics translation

Fig. 7. Level errors e_λ (a,b) and spectral errors ρ_η (c,d) for microphone distance u and normalized source distance γ. Contour lines are drawn every 3 dB.

under critically-sampled conditions, i.e., when the number of plane-wave terms matches the number of ambisonics signals.

We also note from Fig. 6(a) that the frequency response for a source azimuth of 90° is particularly flat, which suggests that translation perpendicular to the direction of the source introduces very little spectral coloration. This appears to contradict another finding of Hahn and Spors [15]: that translation parallel to the source direction introduces less coloration than translation perpendicular. However, their finding was only shown for oversampled conditions, where $Q > N_{\text{in}}$ [15, cf. Fig. 7]; for critically-sampled conditions (where $Q = N_{\text{in}}$), as considered here, the induced coloration is apparently smaller for perpendicular translations. Although not shown here, it can be verified that oversampling the plane-wave decomposition leads to similar behavior (albeit slightly more complicated due to the

spherical, rather than circular, plane-wave grid) as reported by those authors.

As shown in Fig. 6(b), the ambisonics translation method introduces a high-frequency roll-off, which does not depend on source azimuth. Recall from Sec. 2.2.1, however, that the corner frequency of this roll-off does vary proportionally to $L_{\text{in}}/\|\vec{r}_0 - \vec{u}\|$. Consequently, we expect the spectral coloration induced by this method to increase steadily with increasing translation distance, while the overall level will steadily decrease, although both of these errors should improve with increasing L_{in}.

6 CHARACTERIZATION AND DISCUSSION

Based on our findings in Sec. 4, we use beamforming with $Q = N_{\text{in}}$ for the plane-wave translation method in all simulations discussed below.

Level error contour plots are shown in the top panels of Fig. 7. For the plane-wave translation method (see Fig. 7(a)), we see that exterior sources ($\gamma > 1$) experience negligible level error; interior sources ($\gamma < 1$), however, are reproduced approximately 6 dB too quietly for most microphone distances. For the ambisonics translation method (see Fig. 7(b)), we see a general trend of increasing level error with microphone distance at all source distances, although exterior sources consistently experience less severe level errors than interior ones for the same u, and at very large u and γ (top right corner of Fig. 7(b)), the errors become less severe. The degradation in performance with increasing microphone distance is due to the high-frequency roll-off induced by the ambisonics translation filters, as discussed in Sec. 5, which yields a decrease in overall level. Taken together, these results imply that a violation of the region of validity restriction (i.e., for interior sources) tends to yield a reproduced level that is several dB too low.

In the bottom panels of Fig. 7, we plot the spectral errors incurred by both methods. The plane-wave translation method does not appear to exhibit any clear trend (see Fig. 7(c)), although we do see that, for any given microphone distance, the greatest spectral errors occur for source distances around $\gamma = 1$. Furthermore, this method tends to experience the largest spectral error ($\rho_\eta > 15$ dB) at large microphone distances ($u > 1$ m) with $\gamma \sim 1$. Additionally, there are two regions of relatively low error: 1. at very small microphone distances with very far-field sources (top left corner of Fig. 7(c)) and 2. for microphone distances around $u \approx 1$ m with far interior sources ($\gamma < 0.3$).

As shown in Fig. 7(d), the ambisonics translation method exhibits similar behavior in terms of spectral error as it does for level errors: in this plot, we again see a clear trend of increasing error with microphone distance (again due to the high-frequency roll-off), with the exception of a region of slightly less severe errors at very large u and γ. However, in this case, interior sources experience approximately 6 dB less spectral error than corresponding exterior ones (for a fixed u).

From Fig. 8(a), we see that the plane-wave translation method introduces very small localization errors for exterior sources. This is somewhat intuitive, since as γ increases, the source appears more like a plane-wave source, which should be natural to reproduce using the plane-wave translation method. The ambisonics translation method, however, is only accurate at very small microphone distances with very far exterior sources (as shown in the top left corner of Fig. 8(b)). Otherwise, for exterior sources, the errors incurred by the ambisonics translation method increase steadily with increasing microphone distance. Additionally, both methods yield large errors at all microphone distances for interior sources with approximately $0.3 < \gamma < 1$, as well as at very small u and γ (bottom left corners of both Figs. 8(a) and 8(b)). That this behavior is common to both methods implies that it is the violation of the region of validity restriction which causes these extremely large localization errors.

As shown in the bottom panels of Fig. 8, both methods achieve nearly zero diffuseness errors, although the

ambisonics translation method exhibits a region of minor errors ($e_\Psi \approx -0.1$) for interior sources at large microphone distances (bottom right corner of Fig. 8(d)).

6.1 Order Dependence

In Fig. 9, we plot errors for each metric and for both methods as functions of microphone distance u for a fixed source distance of $s_0 = 1$ m and for several input ambisonics orders L_{in} (with matching $Q = N_{in} = (L_{in} + 1)^2$ for the plane-wave translation method).

From Fig. 9(a), we see that, in terms of level errors, increasing the ambisonics input order tends to improve performance, albeit only marginally. Additionally, the performance of both methods degrades with increasing microphone distance, with exterior sources (where $u < s_0 = 1$ m) exhibiting significantly less severe errors than interior ones (where $u > s_0$). Furthermore, all of these curves exhibit a "knee" at $u = 1$ m, where the curves experience a qualitative change in behavior. In particular, the plane-wave translation method exhibits two distinct regimes on either side of $u = 1$ m: for exterior sources, the method incurs constant level errors, whereas for interior sources, the errors grow more extreme with increasing u. For the ambisonics translation method, increasing u tends to produce more extreme errors overall, and the errors grow more rapidly in the interior source regime than in the exterior source regime. Only the plane-wave translation method with exterior sources exhibits negligible level errors (~ 1 dB), which is a consequence of our choice to match the number of plane-wave terms, Q, to the number of ambisonics signals, $N_{in} = (L_{in} + 1)^2$.

From Fig. 9(b), we again see that increasing the ambisonics order L_{in} yields an improvement in performance (i.e., a decrease in spectral errors) for the ambisonics translation method. Here again we note two distinct regimes for this method on either side of $u = 1$ m with a transition region where the behavior changes. Additionally, within each regime, the errors again increase monotonically with increasing u. However, unlike with level errors, we now see a more rapid (but still monotonic) increase in error for *exterior* sources, while the more gradual performance degradation occurs for *interior* sources. For the plane-wave translation method, we see that increasing L_{in} yields an *increase* in spectral error and that this increase in error is more severe for interior sources than exterior ones. This suggests a penalty from violating the region of validity restriction that is particular to the plane-wave translation method: that a larger ambisonics input order incurs more spectral coloration.

As shown in Fig. 9(c), the localization errors for both methods tend to improve with increasing L_{in}, and the existence of two regimes on either side of $u = 1$ m is evident. For exterior sources, the errors improve with decreasing u, whereas for interior sources, the opposite is true: the errors improve with *increasing* u. Evidently, a microphone distance of $u \approx s_0$ (i.e., $\gamma \approx 1$) yields the most extreme errors. Overall, the ambisonics translation method tends to perform worse than the plane-wave translation method (except
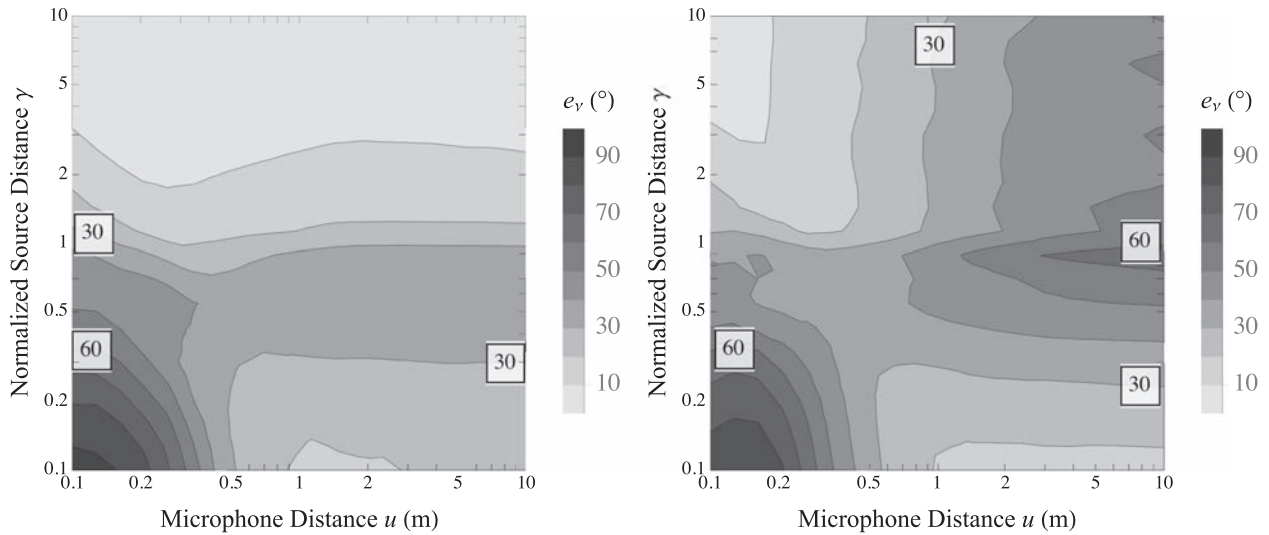
(a) Localization errors e_ν – plane-wave translation

(b) Localization errors e_ν – ambisonics translation
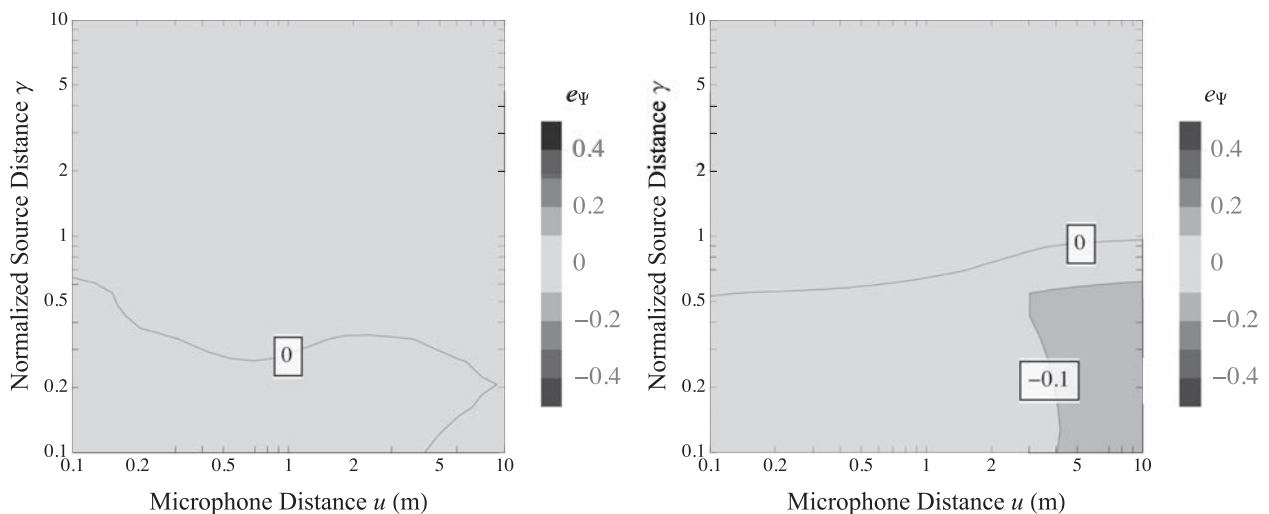
(c) Diffuseness errors e_Ψ – plane-wave translation

(d) Diffuseness errors e_Ψ – ambisonics translation

Fig. 8. Localization errors e_ν (a,b) and diffuseness errors e_Ψ (c,d) for microphone distance u and normalized source distance γ. Localization error contour lines are drawn every $10°$; diffuseness error contour lines are drawn in increments of 0.1.

at very large u and $L_{in} = 1$), and the errors for both methods are typically smaller for exterior sources than interior ones.

It is worth noting that by construction, since s_0 is fixed, as u increases beyond 1 m, more of the navigable region (see Fig. 4) becomes valid for translation. For instance, with $u = 10$ m and $s_0 = 1$ m, the listener can navigate approximately 9 m away from the microphone and still remain inside its region of validity. This explains in part the improvement with increasing u seen for $u > 1$ m, since, on average, more of the navigable region is valid. That is, when averaging errors over the entire navigable region, a smaller fraction of that region will actually be in violation of the region of validity restriction.

Finally, as shown in Fig. 9(d), both methods achieve small diffuseness errors, and generally, increasing L_{in} im-

proves the performance for exterior sources. Again we see a transition between regimes occurring at $u \approx 1$ m.

7 SUMMARY AND CONCLUSIONS

In this work, we presented the results of numerical simulations conducted in order to demonstrate the penalties incurred by violating the region of validity restriction and characterize and compare the performance of existing linear extrapolation-based navigational methods. Although we restricted our analysis to the plane-wave and ambisonics translation methods (both of which are reviewed in Sec. 2), we take these methods to be representative of the entire class of linear extrapolation methods (as discussed
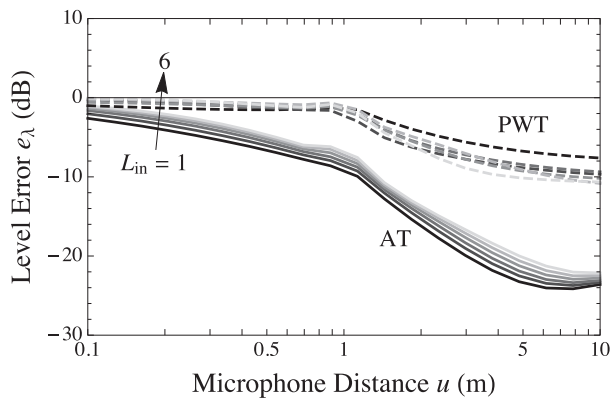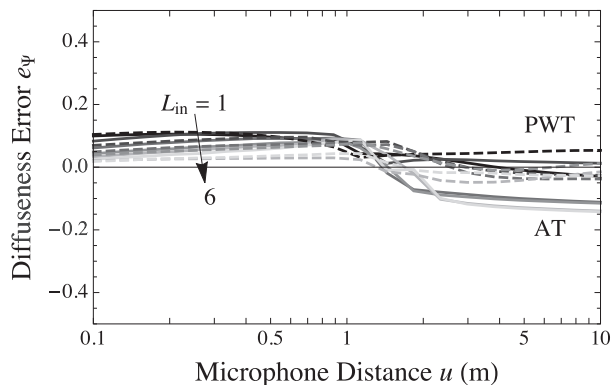
(a) Level errors e_λ
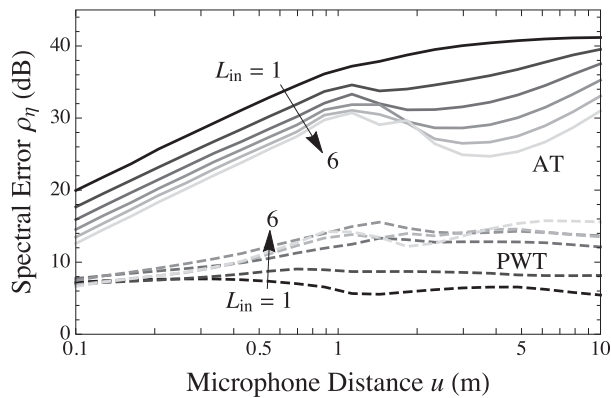
(b) Spectral errors ρ_η
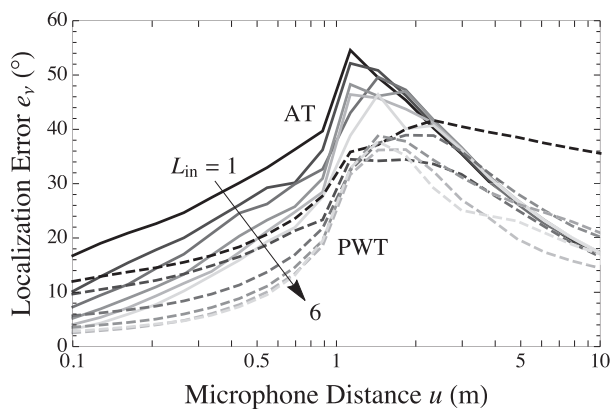
(c) Localization errors e_ν

(d) Diffuseness errors e_Ψ

Fig. 9. Errors for various microphone distances u with a fixed source distance $s_0 = 1$ m. Errors are plotted for the ambisonics translation method (solid curves, labeled "AT") and the plane-wave translation method (dashed curves, labeled "PWT"). For each method, six input ambisonics orders are shown: $L_{in} = 1$ (black) to $L_{in} = 6$ (lightest gray).

in Sec. 0.2). A critical review of all existing extrapolation-based methods is given in Sec. 0.1.

Following the simulation framework laid out in Sec. 3, we simulated simple incident sound fields consisting of a single microphone and single point-source and varied source distance and azimuth as well as microphone distance and listener position. We first determined suitable parameters for the plane-wave translation method, comparing the beamforming and pseudoinversion methods for computing the plane-wave decomposition and varying both the ambisonics input order and number of plane-wave terms. We then explored basic properties of each method by computing the effective frequency responses induced by the plane-wave and ambisonics translation methods across source azimuths. Finally, we conducted a more comprehensive analysis of both methods in terms of the metrics enumerated in Sec. 3.2 for sound level, spectral coloration, source localization, and diffuseness.

The analyses presented in this work yielded the following major findings:

- for the plane-wave translation method, a clear advantage exists to using the beamforming plane-wave

decomposition method (see Eq. (8)) and matching the number of plane-wave terms, Q, to the number of ambisonics signals, N_{in};[12]

- the frequency responses induced by the plane-wave translation method are largely flat but with sporadic notches while those induced by the ambisonics translation method exhibit a consistent low-pass-like roll-off of high-frequency energy, but, qualitatively, both methods appear largely insensitive to source azimuth;

- the ambisonics translation method incurs significant errors in both level and coloration at all source distances which, overall, increase steadily with translation distance;

- for exterior (viz., far-field) sources, the plane-wave translation method achieves a high degree of accuracy in both level and localization;

- for interior (near-field) sources, where the region of validity restriction is violated, both methods incur significant errors in both level and localization;

---

[12] All subsequent findings relate to this implementation of the plane-wave translation method: beamforming with matched $Q = N$.

- increasing the ambisonics order tends to uniformly improve the performance of both methods, with the exception of the spectral errors incurred by the plane-wave translation method; and
- more generally, both methods tend to exhibit two distinct regimes of behavior for exterior and interior sources, with a transition region between the two, and the performance for interior sources is often degraded compared to that for exterior sources.

Due to the extremely large level and coloration errors incurred by the ambisonics translation method, the plane-wave translation method is the better choice for most applications with practical translation distances (e.g., $u > 0.5$ m). This method is particularly well-suited for exterior sources, but its performance degrades significantly, in terms of localization in particular, once the region of validity restriction is violated. The remaining challenge for this method is the introduction of significant spectral errors ($\rho_\eta > 10$ dB) over all microphone and source distances.

7.1 Violating the Region of Validity Restriction

In this work, we identified common features of the performance of the plane-wave and ambisonics translation methods under a range of invalid conditions, i.e., conditions in which the region of validity restriction was violated. Under such conditions, both methods exhibit extremely large localization errors ($e_v > 30°$) for all translation distances (see Figs. 8(a) and 8(b)). This behavior can be explained geometrically: as the listener traverses the navigable region (as defined in Sec. 3), the intended source direction can change drastically, in particular when comparing listener positions inside and outside of the region of validity. Evidently, neither of these methods can adequately compensate for the changing intended direction of the source as the listener moves.

The other common error introduced under invalid conditions is that the reproduced sound level is several dB too low (see Figs. 7(a) and 7(b)). This too is largely a geometric effect; neither of the methods considered here are able to adequately increase the reproduced level as the listener navigates very close to a source (or decrease it as the listener navigates away from close). Since these penalties relate primarily to the drastic changes in the apparent geometry of the sound field as the listener navigates, they are not particular to the chosen methods but are rather inherently associated with violating the region of validity restriction.

As indicated in Sec. 0.1, all linear extrapolation methods are bound by the region of validity restriction and, consequently, will necessarily incur the associated penalties. This is a direct consequence of ambisonics theory: all linear methods operate by constructing an alternative representation of the recorded sound field (e.g., with plane-waves, point-sources, etc.) that approximates the input order-limited series expansion of the sound field (see Eq. (6)). As this expansion is subject to its region of validity, all subsequent approximations are subject to the same

restriction.[13] Therefore, the results of this work imply that all linear extrapolation methods will incur the penalties in level and localization described above.

7.2 Future Directions

As the evaluations presented in this article have been purely numerical, one avenue for further research is to assess the perceptibility and/or perceptual severity of the errors discussed here through listening experiments (following the approaches demonstrated by Rudrich et al. [49], for example). Future work might also seek to establish perceptual thresholds for the metrics employed here, e.g., through subjective evaluations of stimuli with varying diffuseness in order to determine an audibility threshold for diffuseness errors. Additionally, although errors in the reproduced sound level are likely sufficient to produce errors in distance perception [50, Sec. 3.1.1], it may be useful to develop a dedicated model for perceived source distance, perhaps based on the known formula for distance coding of point-sources (see Eq. (3)).

Another avenue for further research is to evaluate the capability of each navigational method to capture and reproduce source directivities (issue 6 in Sec. 0.1), which may be important in many applications (e.g., in immersive applications with intimate sources) but is largely unaddressed in the literature (with the exception of the study by Wakayama et al. [24]). Similarly, the performance of these should be evaluated under more realistic scenarios (e.g., with multiple sources or under reverberant conditions; issue 5 in Sec. 0.1). Intuitively, we might expect that their performance will not change due to the linearity of the methods; however, it might be the case that superimposed direct and reflected sounds, for example, will serve to "average out" the incurred spectral distortions, e.g., for the plane-wave translation method, as shown in Fig. 6(a). Future investigations might also characterize the performance of alternative methods, such as the parametric extrapolation methods of Wakayama et al. [24] and Plinge et al. [25], which should theoretically circumvent the region of validity restriction.

8 ACKNOWLEDGMENT

---

[13] It is worth noting, however, that some methods may perform more *accurately* outside of the region of validity than others by choosing a better alternative representation for the incident sound field. For example, with the spherical equivalent source method [12], if the virtual sources are positioned to mimic the positions of the real sources, the translated sound field will likely provide an accurate estimate of the recorded sound field, even outside of the region of validity. However, this can only be achieved reliably using some *a priori* information about the incident sound field.

9 REFERENCES

[1] E. G. Williams, *Fourier Acoustics: Sound Radiation and Nearfield Acoustical Holography* (Academic Press, 1999).

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

[3] F. Zotter, *Analysis and Synthesis of Sound-Radiation with Spherical Arrays*, Ph.D. thesis, University of Music and Performing Arts Graz (2009).

[4] M. A. Poletti, "Three-Dimensional Surround Sound Systems Based on Spherical Harmonics," *J. Audio Eng. Soc.*, vol. 53, no. 11, pp. 1004–1025 (2005).

[5] D. B. Ward and T. D. Abhayapala, "Reproduction of a Plane-Wave Sound Field Using an Array of Loudspeakers," *IEEE Trans. Speech Audio Process.*, vol. 9, no. 6, pp. 697–707 (2001), ISSN 1063-6676, doi:10.1109/89.943347.

[6] Zylia Sp. z o. o., "ZYLIA ZM-1 Microphone," http://www.zylia.co/zylia-zm-1-microphone.html, accessed: 2019-02-16.

[7] mh acoustics LLC, "Eigenmike® Microphone," https://www.mhacoustics.com/products#eigenmike1, accessed: 2019-02-16.

[8] VisiSonics Corporation, "VisiSonics 5/64 Audio/Visual Camera," https://visisonics.com/products-2/#camera, accessed: 2019-02-16.

[9] J. G. Tylka and E. Y. Choueiri, "Comparison of Techniques for Binaural Navigation of Higher-Order Ambisonic Soundfields," presented at the *139th Convention of the Audio Engineering Society* (2015 Oct.), convention paper 9421.

[10] 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: Signal Processing in Audio Recording and Reproduction* (2003 May), conference paper 16.

[11] M. Frank, F. Zotter, and A. Sontacchi, "Localization Experiments Using Different 2D Ambisonics Decoders," presented at the *25th Tonmeistertagung – VDT International Convention* (2008).

[12] E. Fernandez-Grande, "Sound Field Reconstruction Using a Spherical Microphone Array," *J. Acoust. Soc. Am.*, vol. 139, no. 3, pp. 1168–1178 (2016), doi:10.1121/1.4943545.

[13] D. Menzies and M. Al-Akaidi, "Nearfield Binaural Synthesis and Ambisonics," *J. Acoust. Soc. Am.*, vol. 121, no. 3, pp. 1559–1563 (2007), doi:10.1121/1.2434761.

[14] F. Schultz and S. Spors, "Data-Based Binaural Synthesis Including Rotational and Translatory Head-Movements," presented at the *AES 52nd International Conference: Sound Field Control - Engineering and Perception* (2013 Sept.), conference paper P-7.

[15] N. Hahn and S. Spors, "Physical Properties of Modal Beamforming in the Context of Data-Based Sound Reproduction," presented at the *139th Convention of the Audio Engineering Society* (2015 Oct.), convention paper 9468.

[16] F. Winter, F. Schultz, and S. Spors, "Localization Properties of Data-Based Binaural Synthesis Including Translatory Head-Movements," presented at the *Forum Acusticum* (2014).

[17] M. A. Gerzon, "General Metatheory of Auditory Localisation," presented at the *92nd Convention of the Audio Engineering SocietyAudio Engineering Society* (1992 Mar.), convention paper 3306.

[18] J. Ahrens and S. Spors, "An Analytical Approach to Sound Field Reproduction with a Movable Sweet Spot Using Circular Distributions of Loudspeakers," *2009 IEEE Int. Conf. Acoust. Speech Signal Process.*, pp. 273–276 (2009), ISSN 1520-6149, doi:10.1109/ICASSP.2009.4959573.

[19] J. Hannemann and K. D. Donohue, "Virtual Sound Source Rendering Using a Multipole-Expansion and Method-of-Moments Approach," *J. Audio Eng. Soc.*, vol. 56, no. 6, pp. 473–481 (2008).

[20] J. G. Tylka and E. Y. Choueiri, *Algorithms for Computing Ambisonics Translation Filters*, Technical Report #2, 3D Audio and Applied Acoustics Laboratory (Princeton University, 2019), https://www.princeton.edu/3D3A/Publications/Tylka_3D3A_HOATranslationFilters.html.

[21] D. Menzies and M. Al-Akaidi, "Ambisonic Synthesis of Complex Sources," *J. Audio Eng. Soc.*, vol. 55, no. 10, pp. 864–876 (2007).

[22] R. Baumgartner and F. Zotter, "Time-Domain Translation Operators for the Fast-Multipole-Method," presented at the *Tagungsband der 38. Deutschen Jahrestagung für Akustik (DAGA 2012)* (2012).

[23] Y. Wang and K. Chen, "Translations of Spherical Harmonics Expansion Coefficients for a Sound Field Using Plane Wave Expansions," *J. Acoust. Soc. Am.*, vol. 143, no. 6, pp. 3474–3478 (2018), doi:10.1121/1.5041742.

[24] K. Wakayama, J. Trevino, H. Takada, S. Sakamoto, and Y. Suzuki, "Extended Sound Field Recording Using Position Information of Directional Sound Sources," *2017 IEEE Workshop Appl. Signal Process. Audio Acoust. (WASPAA)*, pp. 185–189 (2017), doi:10.1109/WASPAA.2017.8170020.

[25] A. Plinge, S. J. Schlecht, O. Thiergart, T. Robotham, O. Rummukainen, and E. A. P. Habets, "Six-Degrees-of-Freedom Binaural Audio Reproduction of First-Order Ambisonics with Distance Information," presented at the *2018 AES International Conference on Audio for Virtual and Augmented Reality* (2018 Aug.), conference paper P6-2.

[26] V. Pulkki, "Spatial Sound Reproduction with Directional Audio Coding," *J. Audio Eng. Soc.*, vol. 55, no. 6, pp. 503–516 (2007).

[27] J. Merimaa and V. Pulkki, "Spatial Impulse Response Rendering I: Analysis and Synthesis," *J. Audio Eng. Soc.*, vol. 53, no. 12, pp. 1115–1127 (2005).

[28] T. Pihlajamaki and V. Pulkki, "Synthesis of Complex Sound Scenes with Transformation of Recorded Spatial Sound in Virtual Reality," *J. Audio Eng. Soc.*, vol. 63, no. 7/8, pp. 542–551 (2015), doi:10.17743/jaes.2015.0059.

[29] A. Allen, Ambisonics Sound Field Navigation Using Directional Decomposition and Path Distance Estimation. US Patent 10,182,303, Filed July 12, 2017; Granted Jan. 15, 2019, https://patents.google.com/patent/US10182303B1.

[30] O. Thiergart, G. Del Galdo, M. Taseska, and E. A. P. Habets, "Geometry-Based Spatial Sound Acquisition Using Distributed Microphone Arrays," *IEEE Trans. Audio Speech Lang. Process.*, vol. 21, no. 12, pp. 2583–2594 (2013), ISSN 1558-7916, doi:10.1109/TASL.2013.2280210.

[31] P. Samarasinghe, T. Abhayapala, and M. Poletti, "Wavefield Analysis Over Large Areas Using Distributed Higher Order Microphones," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 22, no. 3, pp. 647–658 (2014), ISSN 2329-9290, doi:10.1109/TASLP.2014.2300341.

[32] J. G. Tylka and E. Y. Choueiri, "Soundfield Navigation Using an Array of Higher-Order Ambisonics Microphones," presented at the *2016 AES International Conference on Audio for Virtual and Augmented Reality* (2016 Sept.), conference paper 4-2.

[33] P. Grosche, F. Zotter, C. Schörkhuber, M. Frank, and R. Höldrich, Method and Apparatus for Acoustic Scene Playback. US Patent App. US20190253826A1, Filed Apr. 24, 2019; Published Aug. 15, 2019, https://patents.google.com/patent/US20190253826A1.

[34] E. Patricio, A. Rumiński, A. Kuklasiński, Ł. Januszkiewicz, and T. Żernicki, "Toward Six Degrees of Freedom Audio Recording and Playback Using Multiple Ambisonics Sound Fields," presented at the *146th Convention of the Audio Engineering Society* (2019 Mar.), convention paper 10141.

[35] J. G. Tylka and E. Y. Choueiri, "Fundamentals of a Parametric Method for Virtual Navigation Within an Array of Ambisonics Microphones," *J. Audio Eng. Soc.* vol. 68, no. 3, pp. 120–137, doi:10.17743/jaes.2019.0055.

[36] C. Nachbar, F. Zotter, E. Deleflie, and A. Sontacchi, "ambiX - A Suggested Ambisonics Format," presented at the *3rd Ambisonics Symposium* (2011).

[37] B. C. J. Moore, *An Introduction to the Psychology of Hearing* (Koninklijke Brill NV, 2013).

[38] Z. Průša, P. L. Søndergaard, N. Holighaus, C. Wiesmeyr, and P. Balazs, "The Large Time-Frequency Analysis Toolbox," http://ltfat.github.io, accessed: 2019-02-16.

[39] B. R. Glasberg and B. C. J. Moore, "Derivation of Auditory Filter Shapes from Notched-Noise Data," *Hearing Res.*, vol. 47, no. 1, pp. 103–138 (1990).

[40] A. M. Salomons, *Coloration and Binaural Decoloration of Sound Due to Reflections*, Ph.D. thesis, Delft University of Technology (1995).

[41] Z. Schärer and A. Lindau, "Evaluation of Equalization Methods for Binaural Signals," presented at the *126th Convention of the Audio Engineering SocietyAudio Engineering Society* (2009 May), convention paper 7721.

[42] B. Boren, M. Geronazzo, F. Brinkmann, and E. Choueiri, "Coloration Metrics for Headphone Equalization," *Proc. 21st Int. Conf. Audit. Display*, pp. 29–34 (2015).

[43] J. G. Tylka and E. Y. Choueiri, "Models for Evaluating Navigational Techniques for Higher-Order Ambisonics," *Proc. Meetings Acoust.*, vol. 30, no. 1, p. 050009 (2017), doi:10.1121/2.0000625.

[44] P. Stitt, S. Bertet, and M. van Walstijn, "Extended Energy Vector Prediction of Ambisonically Reproduced Image Direction at Off-Center Listening Positions," *J. Audio Eng. Soc.*, vol. 64, no. 5, pp. 299–310 (2016), doi:10.17743/jaes.2016.0008.

[45] D. G. Malham and A. Myatt, "3-D Sound Spatialization Using Ambisonic Techniques," *Comp. Music J.*, vol. 19, no. 4, pp. 58–70 (1995).

[46] J. Fliege, "Integration Nodes for the Sphere," http://www.mathematik.uni-dortmund.de/lsx/research/projects/fliege/nodes/nodes.html, accessed: 2019-02-16.

[47] A. Solvang, "Spectral Impairment of Two-Dimensional Higher Order Ambisonics," *J. Audio Eng. Soc.*, vol. 56, no. 4, pp. 267–279 (2008).

[48] B. Bernschütz, A. V. Giner, C. Pörschmann, and J. Arend, "Binaural Reproduction of Plane Waves with Reduced Modal Order," *Acta Acustica United Acustica*, vol. 100, no. 5, pp. 972–983 (2014), doi:10.3813/AAA.918777.

[49] D. Rudrich, F. Zotter, and M. Frank, "Evaluation of Interactive Localization in Virtual Acoustic Scenes," presented at the *43. Jahrestagung für Akustik (DAGA 2017)* (2017).

[50] P. Zahorik, D. S. Brungart, and A. W. Bronkhorst, "Auditory Distance Perception in Humans: A Summary of Past and Present Research," *Acta Acustica United Acustica*, vol. 91, no. 3, pp. 409–420 (2005), ISSN 1610-1928.

[51] P. Stitt, "Matlab Code," https://circlesounds.wordpress.com/matlab-code/, accessed: 2019-02-16.

THE AUTHORS

Joseph G. Tylka

Edgar Y. Choueiri

Dr. Joseph (Joe) G. Tylka is a research scientist at Siemens Corporate Technology whose expertise lies in multichannel signal processing, intelligent control, and machine learning. He received his Ph.D. in 2019 from Princeton University, where his dissertation research focused on virtual navigation of measured 3D sound fields.

•

Edgar Choueiri is a professor of applied physics in the Department of Mechanical and Aerospace Engineering at Princeton University and associated faculty in the Department of Astrophysical Sciences. He heads Princeton's Electric Propulsion and Plasma Dynamics Lab and the 3D Audio and Applied Acoustics Lab. His research interests are plasma physics, plasma propulsion, acoustics, and 3D audio.