**OPEN ACCESS** Freely available online

J. G. Tylka and E. Y. Choueiri, "Domains of Practical Applicability for Parametric Interpolation Methods for Virtual Sound Field Navigation"
*J. Audio Eng. Soc.*, vol. 67, no. 11, pp. 882–893, (2019 November.).
DOI: https://doi.org/10.17743/jaes.2019.0038



# Domains of Practical Applicability for Parametric Interpolation Methods for Virtual Sound Field Navigation

**JOSEPH G. TYLKA,** *AES Student Member,* **AND EDGAR Y. CHOUEIRI,** *AES Associate Member*
(josephgt@alumni.princeton.edu)

*Princeton University, Princeton, NJ, USA*

Suitable domains are established for the practical application of two state-of-the-art parametric interpolation methods for virtual navigation of ambisonics-encoded sound fields. Although several navigational methods have been developed, existing studies rarely include comparisons between methods and, significantly, practical assessments of such methods have been limited. To that end, the errors introduced by both methods are objectively evaluated, in terms of metrics for sound level, spectral coloration, source localization, and diffuseness, through numerical simulations. Various practical domains are subsequently identified, and guidelines are established with which to choose between these methods based on their intended application. Results show that the first method, which entails a time-frequency analysis of the sound field, is preferable for large-area recordings and when spatial localization accuracy is critical, as this method achieves superior localization performance (compared to the second method) with sparsely distributed microphones. However, the second method, which parametrically excludes from the interpolation any microphones that are farther from the listening position than is any source, is shown to be more suitable for applications in which sound quality attributes such as coloration and diffuseness are critical, since this method achieves smaller spectral errors with sparsely distributed microphones and smaller diffuseness errors under all conditions.

## 0 INTRODUCTION

Given an ambisonics-encoded sound field (i.e., a sound field that has been decomposed into spherical harmonics), *virtual navigation* enables a listener to explore the recorded space and, ideally, experience a spatially and tonally accurate perception of the sound field. One application of such a procedure is to reproduce (e.g., over headphones), from an arbitrary vantage point, an acoustically recorded scene. This would allow, for example, a listener to experience a recording of an orchestral performance from elsewhere in the original venue.

A well-known limitation of the ambisonics framework is 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 [1]. Consequently, the navigable region of such a sound field is inherently restricted. Indeed, existing techniques for virtual navigation using a single ambisonics microphone[1] have been shown to introduce spectral distortions [2] and degrade localization [3, 4] as the listener navigates farther away from the expansion center.

Furthermore, according to theory, the ambisonics expansion provides a mathematically *valid* description of the sound field only in the free field, thereby effectively creating a spherical *region of validity* (also known as the region of convergence), which is centered on the recording microphone and extends up to the nearest sound source (or scattering body) [5, Sec. 6.8]. Consequently, near-field sources may pose a significantly limiting problem to navigation, although the particular degradations in sound quality (e.g., in terms of spatial or tonal fidelity) that might result from violating this region of validity restriction are unclear.

In an effort to overcome these challenges, several authors have developed both *parametric* navigational methods, which leverage additional information about the sound field (e.g., known or inferred source positions) beyond the recorded ambisonics signals, and *interpolation-based* navigational methods, which employ an array of ambisonics microphones (of first-order or higher) distributed throughout the sound field. As it is outside the scope of this work to review all of these methods in detail, the interested reader is referred to Refs. [6–8] and the works cited below.

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

### 0.1 Previous Work and Remaining Problems

Several of the recently developed linear interpolation-based navigational methods have been evaluated experimentally and have shown promising results. Patricio et al. [9] proposed a modified linear interpolation method in which the directional components of the microphone nearest to the listener are emphasized over those of the farther microphones. In their study, the authors experimentally demonstrated that the proposed distance-biasing approach achieves plausible source localization and perception of listener movement.

Grosche et al. [10] recently developed a method that employs multiple distinct virtual loudspeaker arrays (VLAs), each corresponding to a different ambisonics (or other) recording microphone and reproducing the sound field as captured by that microphone. The superposition of the reproduced signals from all of the VLAs is then rendered for the listener at an arbitrary desired position in the virtual sound field. This method has been evaluated via localization tests [11], which showed a significant improvement in performance over a basic weighted-average interpolation approach, and Deppisch and Sontacchi [12] have developed a browser-based implementation of the method. Comparisons between these linear methods under comparable conditions, however, have not been conducted. As described subsequently, in the present article we comprehensively compare two parametric methods, the procedure for which may serve as a template for conducting similar comparisons in the future.

Several of the parametric methods that have been developed entail a directional analysis of the recorded signals (often carried out in the time-frequency domain) in order to construct a navigable model of the incident sound field [13–15]. While such methods might yield accurate source localization, some have been shown to introduce minor degradations of sound quality [14, Sec. 5.3] and they may not be suitable for dense or highly reverberant environments [13, Sec. II]. In a recent publication [16], we developed an alternative method that parametrically selects for the interpolation a suitable subset of microphones to ensure that the region of validity restriction for each included microphone is not violated. The differences in performance between these two parametric methods under the same conditions have not been established, so we seek guidance for choosing between them in various domains of practical application.

To that end, we evaluate and compare the performance of these two state-of-the-art parametric interpolation methods, referred to here as the time-frequency analysis (TFA) method [13] and the valid microphone interpolation (VMI) method [16]. First, in Section 1, we review the formulation of each of these methods. We then describe, in Section 2, the numerical simulations conducted in this study and the objective metrics used to evaluate the errors introduced by each method in terms of sound level, spectral coloration, source localization, and diffuseness. We then present and discuss in Section 3 the results of these simulations, from which we identify in Section 4 considerations regarding the suitability of each method to various applications. Relevant ambisonics theory is reviewed in Appendix A.

The present article is intended to complement the findings of our previous study [16], in which the VMI method is described in detail and fundamental aspects of the method are demonstrated through numerical proof-of-concept analyses. Compared to its original implementation [16], the VMI method is largely unchanged (with the exception of the two-band approach; see Section 1.2.2). However, compared to previous analyses, the present analyses are significantly more comprehensive, several additional performance metrics are employed, and the comparative analysis of the two parametric methods is entirely new.

## 1 REVIEW OF NAVIGATIONAL METHODS

As is common in ambisonics, 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 (non-negative) 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 $(\theta, \phi) = (0, 0)$ 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 $\hat{r} \equiv \vec{r}/r$.

Generally, we seek the ambisonics signals, up to a given order $L_{\text{out}}$, that describe the sound field in the vicinity of a listener located at $\vec{r}_0$. Here, we use real-valued orthonormal (N3D) spherical harmonics, as given by Zotter [17, Sec. 2.2], and we adopt the ambisonics channel number (ACN) convention [18] 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 $Y_n$.

For a finite-order ambisonics expansion about $\vec{r}_0$, the acoustic potential field (defined as the Fourier transform of the acoustic pressure field) is given by

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

where $k$ is the angular wavenumber, $A_n$ is the $n^{\text{th}}$ desired ambisonics signal, $j_l$ is the spherical Bessel function of order $l$, and $N_{\text{out}} = (L_{\text{out}} + 1)^2$ is the number of terms in the expansion (cf., Eq. (A.31) in Appendix A).

We take as inputs to each navigational method $P$ sets of measured ambisonics signals, $B_n^{[p]}$, up to order $L_{\text{in}}$, that describe the sound field in the vicinity of the $p^{\text{th}}$ microphone, which is located at $\vec{u}_p$, $\forall p \in [1, P]$. This local potential field is given by

$$\psi(k, \vec{r} + \vec{u}_p) = \sum_{n=0}^{N_{\text{in}}-1} 4\pi(-i)^l B_n^{[p]}(k)\, j_l(kr) Y_n(\hat{r}), \qquad (2)$$

where $N_{\text{in}} = (L_{\text{in}} + 1)^2$.

### 1.1 Time-Frequency Analysis (TFA) Method

In the method proposed by Thiergart et al. [13], the sound field is first analyzed in the time-frequency domain

and subsequently modeled as a finite set of monochromatic omnidirectional point sources. As will become clear below, this method can only use the first-order ambisonics signals, so we must have $L_{\text{in}} = 1$. The output signals, however, can be computed to an arbitrary order $L_{\text{out}}$. Below, we describe our implementation of this method.

#### 1.1.1 Time-Frequency Sound Field Analysis

For each microphone, we first compute the short-time Fourier transform (STFT) of each of the first four ambisonics signals, which gives $B_n^{[p]}(\xi, \kappa)$ for $n \in [0, 3]$, where $\xi$ and $\kappa$ are the time and frequency indices, respectively. Typically, we take an overlap fraction of $R = 0.5$, and set the FFT length to be

$$N_{\text{FFT}} = 2^{\lceil \log_2(\frac{F_s}{1-R}\frac{\Delta}{c}) \rceil}, \tag{3}$$

where $\lceil \cdot \rceil$ denotes rounding up to the nearest integer, $F_s$ is the sampling rate of the system, $\Delta$ is the distance between microphones (defined later in Section 2), and $c \approx 343$ m/s is the speed of sound. For the STFT analysis window, we choose a Hamming window [19] of length $N_{\text{FFT}}$.

Using the transformed signals from each microphone, we then compute the acoustic intensity vector, $\vec{v}_{\text{I}}^{[p]}(\xi, \kappa)$, given by [20, Eq. (11)]

$$\vec{v}_{\text{I}}(\xi, \kappa) = \frac{\sqrt{2}}{\rho_0 c} \text{Re}\left\{ \overline{W}(\xi, \kappa) \vec{X}(\xi, \kappa) \right\}, \tag{4}$$

where $\overline{(\cdot)}$ and $\text{Re}\{ \cdot \}$ denote taking the complex conjugate and the real part of the argument, respectively; $W$ and $\vec{X} = \begin{bmatrix} X & Y & Z \end{bmatrix}$ are defined below in Eq. (28); and $\rho_0 \approx 1.225$ kg/m$^3$ is the density of air.

At each time-frequency bin, we then triangulate a single "effective" source (which may or may not coincide with a real source). For two microphones and with sources restricted to the horizontal plane, triangulation is computed as follows:

$$\begin{aligned} \vec{s}_0(\xi, \kappa) = \vec{u}_1 + c_1 \hat{\text{v}}_{\text{I}}^{[1]}(\xi, \kappa) = \vec{u}_2 + c_2 \hat{\text{v}}_{\text{I}}^{[2]}(\xi, \kappa), \\ \implies \vec{u}_2 - \vec{u}_1 = c_1 \hat{\text{v}}_{\text{I}}^{[1]}(\xi, \kappa) - c_2 \hat{\text{v}}_{\text{I}}^{[2]}(\xi, \kappa), \end{aligned} \tag{5}$$

where $\vec{s}_0$ is the triangulated source position and $c_1$ and $c_2$ are scalars found for each time-frequency bin. These scalars are computed by

$$\begin{bmatrix} c_1 \\ c_2 \end{bmatrix} = \begin{bmatrix} \cos \phi_{\text{I}}^{[1]} & -\cos \phi_{\text{I}}^{[2]} \\ \sin \phi_{\text{I}}^{[1]} & -\sin \phi_{\text{I}}^{[2]} \end{bmatrix}^{-1} \cdot \begin{bmatrix} x_2 - x_1 \\ y_2 - y_1 \end{bmatrix}, \tag{6}$$

where $\phi_{\text{I}}^{[p]}$ denotes the azimuth of $\vec{v}_{\text{I}}^{[p]}$ and $x_p$ and $y_p$ denote the $x$ and $y$ components of $\vec{u}_p$, respectively. Note that the matrix inversion in Eq. (6) fails when $\phi_{\text{I}}^{[1]} = \phi_{\text{I}}^{[2]}$, i.e., when the intensity vectors are parallel. A more general approach for source triangulation, either in three dimensions or for $P > 2$ microphones (or both), is described by Thiergart et al. [13, Sec. IV.A]. It is worth noting that this triangulation step assumes that the sound field consists of a finite number of discrete sources that can be easily separated (i.e., sources that are far enough apart or not emitting sound simultaneously) [13, Sec. II]; consequently, the performance of this method may suffer in dense or highly reverberant environments.

Finally, we compute the acoustic potential, given by

$$\psi^{[p]}(\xi, \kappa) = \sqrt{4\pi} B_0^{[p]}(\xi, \kappa), \tag{7}$$

and the diffuseness parameter, $\Psi^{[p]}(\xi, \kappa)$, as given below in Eq. (29) and described in Section 2.2.4. (Note that, if we were not using orthonormal N3D spherical harmonics, an additional normalization coefficient would be needed in the above equation.)

#### 1.1.2 Sound Field Modeling and Synthesis

The estimated ambisonics output signals are assembled in the time-frequency domain as follows. For a given listener position $\vec{r}_0$, we let $\vec{s}_0{}' = \vec{s}_0 - \vec{r}_0$ denote the position of the triangulated source relative to the listener for each time-frequency bin. Additionally, we choose a reference microphone with index $p = p_{\text{ref}}$, such that the position of the triangulated source relative to the reference microphone is given by $\vec{s}_{p_{\text{ref}}} = \vec{s}_0 - \vec{u}_{p_{\text{ref}}}$. By default, we choose as the reference the nearest microphone to the listener. We further define a direct-to-diffuse sound ratio parameter, given by

$$\Gamma(\xi, \kappa) = \frac{1}{\Psi^{[p_{\text{ref}}]}(\xi, \kappa)} - 1, \tag{8}$$

as well as direct and diffuse components of the sound field, given by

$$S_{\text{dir}} = \sqrt{\frac{\Gamma(\xi, \kappa)}{1 + \Gamma(\xi, \kappa)}} \frac{\psi^{[p_{\text{ref}}]}}{ikh_0(ks_{p_{\text{ref}}})}, \tag{9}$$

$$S_{\text{diff}} = \sqrt{\frac{1}{1 + \Gamma(\xi, \kappa)}} \psi^{[p_{\text{ref}}]}, \tag{10}$$

respectively, where $h_0$ is the zeroth-order outgoing spherical Hankel function.

From these, we compute the ambisonics output signals up to order $L_{\text{out}}$ by

$$\begin{aligned} \tilde{A}_n(\xi, \kappa) = i^{l+1} k h_l(k s_0{}') Y_n(\hat{s}_0{}') S_{\text{dir}}(\xi, \kappa) \\ + \frac{1}{\sqrt{4\pi}} S_{\text{diff}}(\xi, \kappa), \end{aligned} \tag{11}$$

where we have used Eq. (A.32) to encode the direct point-source components and applied the factor of $1/\sqrt{4\pi}$ to encode the diffuse sound components, thereby compensating for the "directivity" of each ambisonics channel. (This factor is only independent of channel for orthonormal spherical harmonics.) Each of these signals is finally converted into the time domain via an inverse STFT.

### 1.2 Valid Microphone Interpolation (VMI) Method

Below, we describe our previously proposed parametric navigational method, which comprises two components

(1) A parametric method for excluding "invalid" microphones from the interpolation calculation based on estimated source positions, and

(2) A two-band implementation of regularized least-squares interpolation filters.

The method (as originally implemented, without the two-band analysis) and its performance are discussed in more detail by Tylka and Choueiri [16].

### 1.2.1 Source Position and Microphone Validity

According to theory (see Appendix A), ambisonics signals provide a valid description of the captured sound field only in a spherical free-field region around the ambisonics microphone that extends up to the nearest source or obstacle. Consequently, in order to determine the set of microphones for which the listening position is valid, we first localize any near-field sources. Several methods for acoustically localizing sources using signals from two or more ambisonics microphones are discussed by Zheng [14, Ch. 3]; such methods often involve triangulation via acoustic intensity vectors, as described in Section 1.1.1.

Once the locations of any near-field sources are determined, we compute the distances from each microphone to its nearest source and to the listening position. Only those microphones that are nearer to the listening position than to any near-field source are included in the interpolation calculation (i.e., all microphones such that $r_p = \|\vec{r}_0 - \vec{u}_p\| < \|\vec{s}_0 - \vec{u}_p\| = s_p$). As described in the following section, a matrix of interpolation filters is then computed and applied to the signals from the remaining valid microphones.[2]

### 1.2.2 Regularized Least-Squares Interpolation

We first define vectors of ambisonics signals, given by

$$\mathbf{a} = \begin{bmatrix} A_0 \\ A_1 \\ \vdots \\ A_{N_{\text{out}}-1} \end{bmatrix}, \quad \mathbf{b}_p = \begin{bmatrix} B_0^{[p]} \\ B_1^{[p]} \\ \vdots \\ B_{N_{\text{in}}-1}^{[p]} \end{bmatrix}. \tag{12}$$

We also define a matrix of interpolation weights, given by

$$\mathbf{W} = \begin{bmatrix} \sqrt{w_1}\,\mathbf{I} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \sqrt{w_2}\,\mathbf{I} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & \sqrt{w_P}\,\mathbf{I} \end{bmatrix}, \tag{13}$$

where $w_p$ is the interpolation weight for the $p^{\text{th}}$ microphone; $\mathbf{I}$ is the $N_{\text{in}} \times N_{\text{in}}$ identity matrix; and $\mathbf{0}$ is an $N_{\text{in}} \times N_{\text{in}}$ matrix of zeros. Here, we compute the weights $w_p$ using a standard linear interpolation scheme.

We then pose interpolation as an inverse problem, in which we consider the ambisonics signals at the listening position and, using matrices of ambisonics translation coefficients, we write a system of equations simultaneously describing the ambisonics signals at all $P$ valid (first- or higher-order) ambisonics microphones.[3] That is, for each frequency, we write

$$\mathbf{W} \cdot \mathbf{M} \cdot \mathbf{a} = \mathbf{W} \cdot \mathbf{b}, \tag{14}$$

where, omitting frequency dependencies, we let

$$\mathbf{M} = \begin{bmatrix} \mathbf{T}(-\vec{r}_1) \\ \mathbf{T}(-\vec{r}_2) \\ \vdots \\ \mathbf{T}(-\vec{r}_P) \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \\ \vdots \\ \mathbf{b}_P \end{bmatrix}, \tag{15}$$

where $\vec{r}_p$ is the vector from the $p^{\text{th}}$ microphone to the listening position, given by $\vec{r}_p = \vec{r}_0 - \vec{u}_p$, and $\mathbf{T}$ is a matrix of ambisonics translation coefficients, which we compute using the recurrence formulae given by Gumerov and Duraiswami [22, Sec. 3.2], Zotter [17, Ch. 3], and Tylka and Choueiri [23]. Essentially, this formulation allows us to compute the ambisonics signals at the desired listening position that "best explain" (in a least-squares sense) the measured signals, while weighting most heavily those signals from the microphone nearest to the listening position.

Next, we compute the singular value decomposition of $\mathbf{M}_w = (\mathbf{W} \cdot \mathbf{M})$, such that $\mathbf{M}_w = \mathbf{U}\Sigma\mathbf{V}^*$, where $(\cdot)^*$ represents conjugate-transposition. This allows us to compute a regularized pseudoinverse of $\mathbf{M}_w$, given by [24, Sec. 5.1]

$$\mathbf{L} = \mathbf{V}\Theta\Sigma^{+}\mathbf{U}^*, \tag{16}$$

where $(\cdot)^+$ represents pseudoinversion, and $\Theta$ is a square, diagonal matrix whose elements are given by

$$\Theta_{nn} = \frac{\sigma_n^2}{\sigma_n^2 + \beta}. \tag{17}$$

Here, $\sigma_n$ is the $n^{\text{th}}$ singular value of $\mathbf{M}_w$ and $\beta$ is a frequency-dependent regularization parameter, for which we choose a high-shelf filter profile (cf. Tylka and Choueiri [16, Eq. (17)]).

Rearranging Eq. (14) yields an estimate of $\mathbf{a}$, given by

$$\tilde{\mathbf{a}} = \mathbf{L} \cdot \mathbf{W} \cdot \mathbf{b}, \tag{18}$$

which we apply below some critical wavenumber $k_0$. Ideally, as $L_{\text{in}} \to \infty$, we should find $\tilde{\mathbf{a}} \to \mathbf{a}$ (the exact ambisonics signals of the sound field at $\vec{r}_0$). Above $k_0$, we compute a weighted average,[4] given by

$$\tilde{\mathbf{a}} = \begin{bmatrix} w_1\mathbf{I} & w_2\mathbf{I} & \cdots & w_P\mathbf{I} \end{bmatrix} \cdot \mathbf{b}, \tag{19}$$

where now $\mathbf{I}$ is the $N_{\text{out}} \times N_{\text{in}}$ identity matrix.

Finally, the combined interpolated signals are given by

$$\tilde{\mathbf{a}} = \begin{cases} \mathbf{L} \cdot \mathbf{W} \cdot \mathbf{b} & \text{for } k < k_0, \\ \begin{bmatrix} w_1\mathbf{I} & w_2\mathbf{I} & \cdots & w_P\mathbf{I} \end{bmatrix} \cdot \mathbf{b}, & \text{for } k \geq k_0, \end{cases} \tag{20}$$

[2] In practice, as the listener traverses the navigable region, the number of valid microphones may change. Consequently, one should crossfade between the filters for successive audio frames to prevent any audible discontinuities.

[3] Samarasinghe et al. perform a similar derivation for a two-dimensional sound field [22, Sec. III.A].

[4] Future refinements to this method might instead employ one of the state-of-the-art linear interpolation methods [10, 9] (provided that we have $P > 1$ valid microphones), which have been shown to outperform the basic weighted-average interpolation approach adopted here.

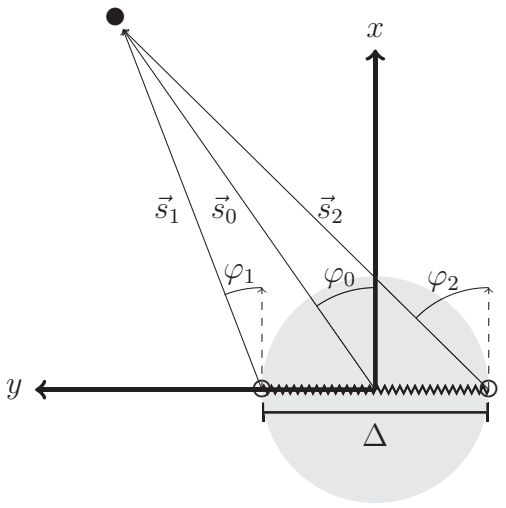



Fig. 1. Diagram of a two-microphone array (empty circles) with a single source (filled circle). The shaded gray disk indicates the interior region, where $r < \Delta/2$. The jagged line segment indicates the navigable region, where $y \in [-\Delta/2, \Delta/2]$ and $x = z = 0$.

where we have chosen

$$k_0 = \begin{cases} 1/r_1, & \text{for } P = 1, \\ \Delta/(r_1 r_2), & \text{for } P = 2, \\ 1/\max_{p\in[1,P]} r_p, & \text{otherwise,} \end{cases} \tag{21}$$

which we found empirically to perform well in terms of spectral errors.

It is worth noting that, for large distances, this critical wavenumber may correspond to very low frequencies, such that the above computation in Eq. (20) simplifies to just a weighted-average interpolation. For example, consider a pair of microphones separated by a distance of $\Delta = 2$ m and a desired listener position at the midpoint, such that $r_1 = r_2 = 1$ m. In this configuration, if only one microphone is valid (i.e., $P = 1$), the frequency corresponding to $k_0$ is given by $c/(2\pi r_1) \approx 54.6$ Hz; if both microphones are valid (i.e., $P = 2$), this frequency is given by $c\Delta/(2\pi r_1 r_2) \approx 109.2$ Hz.

## 2 SIMULATIONS AND METRICS

Consider a linear microphone array geometry, illustrated in Fig. 1, in which a pair of microphones ($P = 2$) are separated by a distance $\Delta$, equidistant from the origin and placed along the lateral $y$-axis, such that their positions are given by $\vec{u}_1 = (0, \Delta/2, 0)$ and $\vec{u}_2 = (0, -\Delta/2, 0)$. We define the *navigable region* as the segment of the $y$-axis connecting the two microphone positions, i.e., all listener positions $\vec{r}_0 = (0, y_0, 0)$ where $y_0 \in [-\Delta/2, \Delta/2]$. We further define a nondimensional geometrical parameter $\gamma = r/(\Delta/2)$, and refer to the region with $\gamma > 1$ as the *exterior region* and that with $\gamma < 1$ as the *interior region* (see Fig. 1).

A single point source is placed on the horizontal plane at $\vec{s}_0 = (s_0 \cos\varphi_0, s_0 \sin\varphi_0, 0)$. From the position of the $p^{\text{th}}$ microphone, the apparent source position is given by $\vec{s}_p = \vec{s}_0 - \vec{u}_p = (s_p \cos\varphi_p, s_p \sin\varphi_p, 0)$, such that the apparent source azimuth is $\varphi_p$ and the relative source distance from the microphone is $s_p$.

### 2.1 Simulation Parameters

We simulate recording of the sound field depicted in Fig. 1 for a range of microphone spacings, $\Delta \in [0.1, 10]$ m, and all source distances $s_0 = \gamma\Delta/2$ for $\gamma \in [0.1, 10]$. In each simulation, we vary the source azimuth from $\varphi_0 = 0^\circ$ to $90^\circ$ in increments of $5^\circ$ and generate, using Eq. (A.32), an artificial ambisonics impulse response at the microphone. We then compute, using each method, the ambisonics impulse responses at listener positions from $y_0 = -\Delta/2$ to $+\Delta/2$, taken in 20 equal increments. Note that for the TFA method only, we intentionally omit source azimuths of $90^\circ$ since, for $\varphi_0 = \pm 90^\circ$, the source becomes collinear with the microphones and consequently the triangulation calculation (see Eq. (6)) can no longer produce a unique solution.

In all simulations, we choose $L_{\text{in}} = L_{\text{out}} = 1$.[5] Although the VMI method has been derived for an arbitrary $L_{\text{in}}$ (whereas the TFA method has only been derived for first-order ambisonics input signals; see Section 1.1), it can be verified that the performance of that method does not vary significantly with input order due to our order-independent choice of critical wavenumber (see Eq. (21)). The sampling rate is 48 kHz and all impulse responses are calculated with 16,384 samples (≈341 ms). Additionally, we filter all point-source ambisonics impulse responses with order-dependent near-field compensation high-pass filters, given for the $l^{\text{th}}$-order ambisonics signals by

$$H_l(f) = 1 - \frac{1}{\sqrt{1 + \left(\frac{f}{f_l}\right)^l}}, \tag{22}$$

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

### 2.2 Metrics

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

1. The level error, $e_\lambda$, of the mean audible energy (MAE), as given in Section 2.2.1,
2. The range, $\rho_\eta$, of the auditory band spectral error (ABSE), as given in Section 2.2.2,
3. The localization error, $e_\nu$, for the precedence-effect localization vector, as given in Section 2.2.3, and
4. The error, $e_\Psi$, in the diffuseness parameter, as given in Section 2.2.4.

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 spacing $\Delta$.

[5] Note that, for the metrics listed in Section 2.2, only the localization model (described in Section 2.2.3) depends on $L_{\text{out}}$; all of the other metrics, by construction, use only the zeroth and first order signals.

#### 2.2.1 Mean Audible Energy ($\lambda$)

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

$$\lambda = 10\log_{10}\left(\frac{1}{N_b}\sum_{c=1}^{N_b}\frac{\int|H_\Gamma(f;f_c)||A_0(f)|^2df}{\int|H_\Gamma(f;f_c)|df}\right), \quad (23)$$

where $H_\Gamma(f;f_c)$ is the transfer function of a gammatone filter[6] (which approximates critical bands), with center frequency $f_c$ for $c \in [1, N_b]$, and integration is taken over all frequencies $f$. Here, we choose $f_c$ to be a set of ERB-spaced (equivalent rectangular bandwidth) center frequencies [26] spanning the range $f_c \in [50\text{ Hz}, 21\text{ kHz}]$. We further define the level error, given in dB by

$$e_\lambda = \tilde{\lambda} - \lambda, \quad (24)$$

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

#### 2.2.2 Auditory Band Spectral Error ($\eta$)

The *auditory band spectral error* (ABSE), adapted from Schärer and Lindau [27, Eq. (9)], is given by

$$\eta(f_c) = 10\log_{10}\left(\frac{\int|H_\Gamma(f;f_c)||\tilde{A}_0(f)|^2df}{\int|H_\Gamma(f;f_c)||A_0(f)|^2df}\right), \quad (25)$$

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

$$\rho_\eta = \max_c \eta(f_c) - \min_c \eta(f_c), \quad (26)$$

which we found through a previous subjective validation study to be a strong predictor of perceptible colorations induced through navigation [28].

#### 2.2.3 Precedence-Effect Localization Vector ($\vec{v}$)

Localization is predicted using a recently developed and subjectively validated precedence-effect-based localization model, the details of which are provided in an earlier publication [28, Sec. 2.A.i]. 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 precedence-effect-based energy vector model of Stitt et al. [29] to produce a single predicted source localization vector, $\vec{v}$. The corresponding localization error is then computed by

$$e_v = \cos^{-1}\left(\hat{v}\cdot\hat{s}_0{}'\right), \quad (27)$$

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

[6] Here, we used the gammatone filters implemented in the large time-frequency analysis toolbox (LTFAT) for MATLAB [26].

#### 2.2.4 Diffuseness Parameter ($\Psi$)

According to Merimaa and Pulkki [20], the acoustic intensity vector and a diffuseness parameter can be computed using the four standard "B-format" signals, which are related to the first four ACN/N3D ambisonics signals by

$$W = \frac{A_0}{\sqrt{2}}, \quad Y = \frac{A_1}{\sqrt{3}}, \quad Z = \frac{A_2}{\sqrt{3}}, \quad X = \frac{A_3}{\sqrt{3}}. \quad (28)$$

The diffuseness parameter is a nondimensional measure of the fraction of the total acoustic energy that is not directional. To compute it, we first construct a frequency-dependent Cartesian row vector, $\vec{X} = \begin{bmatrix} X & Y & Z \end{bmatrix}$. The diffuseness parameter $\Psi$ is then given by [20, Eq. (12)]

$$\Psi(f) = 1 - \sqrt{2}\frac{\left\|\mathrm{Re}\left\{\overline{W}(f)\vec{X}(f)\right\}\right\|}{|W(f)|^2 + \left\|\vec{X}(f)\right\|^2/2}, \quad (29)$$

where $\overline{(\cdot)}$ denotes taking the complex conjugate of the argument and, for a complex-valued vector, $\|\cdot\|$ denotes the norm given by $\|\vec{x}\| = \sqrt{\langle\vec{x},\vec{x}\rangle} \equiv \sqrt{\vec{x}\,\vec{x}^{\,\mathrm{H}}}$, where $(\cdot)^{\mathrm{H}}$ denotes the conjugate (Hermitian) transpose of the argument. Thus, $\Psi$ 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 translated and reference signals, given by

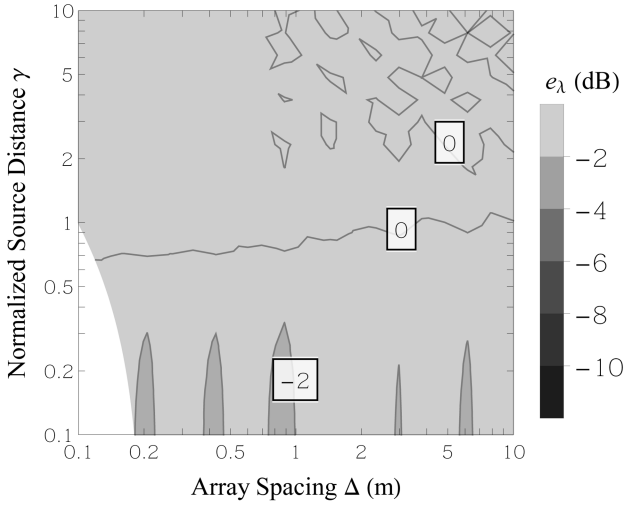
$$e_\Psi = \frac{\int_{f_\mathrm{L}}^{f_\mathrm{U}}\frac{1}{f}\left(\tilde{\Psi}(f) - \Psi(f)\right)df}{\log(f_\mathrm{U}) - \log(f_\mathrm{L})}, \quad (30)$$

where $f_\mathrm{L} = 50$ Hz and $f_\mathrm{U} = 21$ kHz.

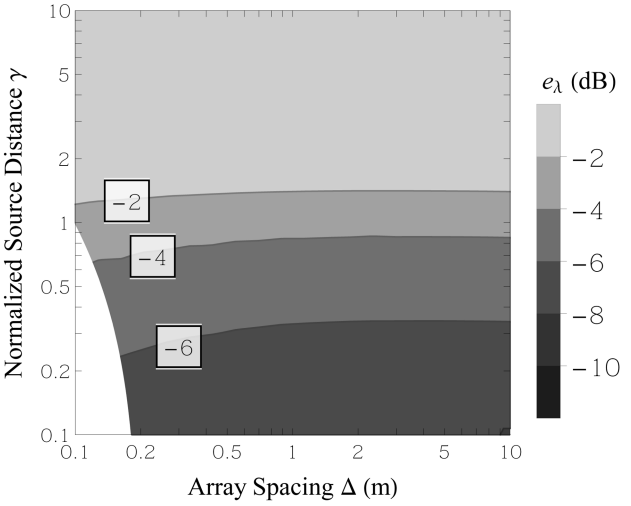
## 3 CHARACTERIZATION AND DISCUSSION

In Fig. 2(a), we plot the level errors incurred by the TFA method as a function of array spacing $\Delta$ and normalized source distance $\gamma$. Note that we exclude from these plots the region in which $s_0 + \Delta/2 < 0.1$ m (i.e., the bottom left corner of each panel in Figs. 2 and 3), as this corresponds to geometries for which the source is "inside the head" (for an approximate head radius of 10 cm) at all positions within the navigable region. From Fig. 2(a), we see that the TFA method is able to achieve approximately zero error almost everywhere, with the exception of far interior sources ($\gamma < 1$). This yields an improvement over the VMI method, shown in Fig. 2(b), which is only able to accurately reconstruct the sound level for exterior sources ($\gamma > 1$) and is otherwise several dB too quiet for interior sources.

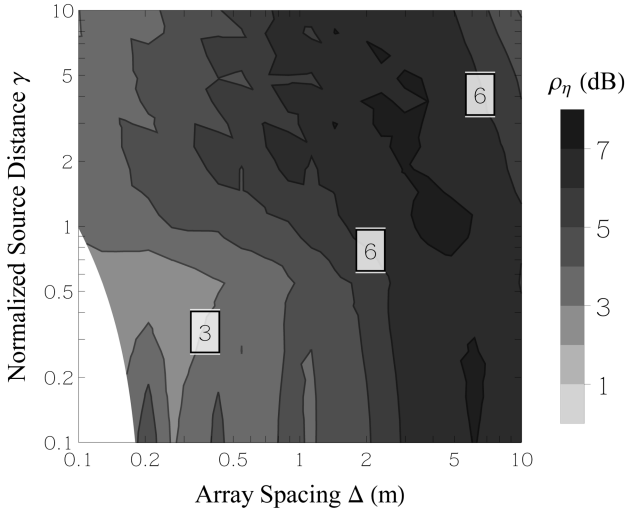
That the reconstructed level is too low may be particularly detrimental to a listener's perception of source proximity, since one of the primary distance cues humans expect is an increase in level [30, Sec. 3.1.1]. Consequently, the impact of these errors on a listener's perception of distance should be investigated, although it is outside the scope of this work to do so.



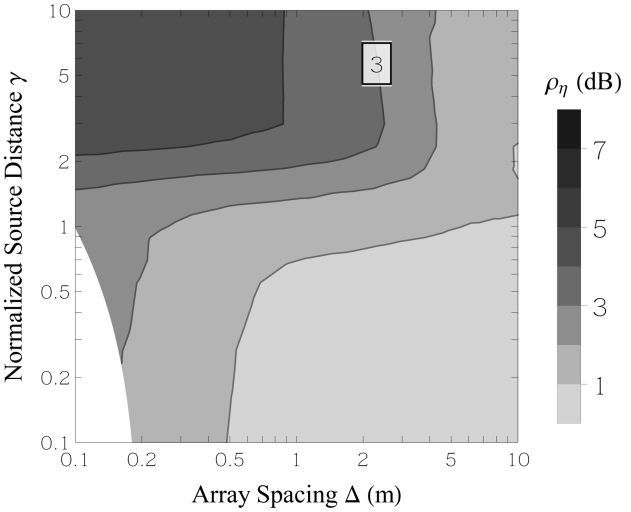

(a) Level errors – TFA method




(b) Level errors – VMI method




(c) Spectral errors – TFA method

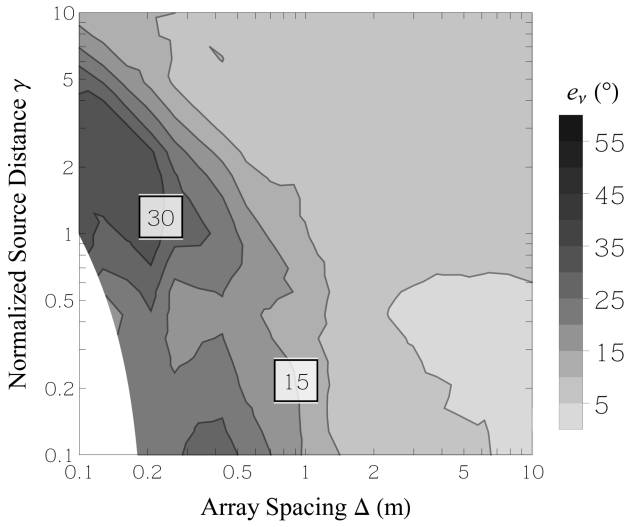



(d) Spectral errors – VMI method

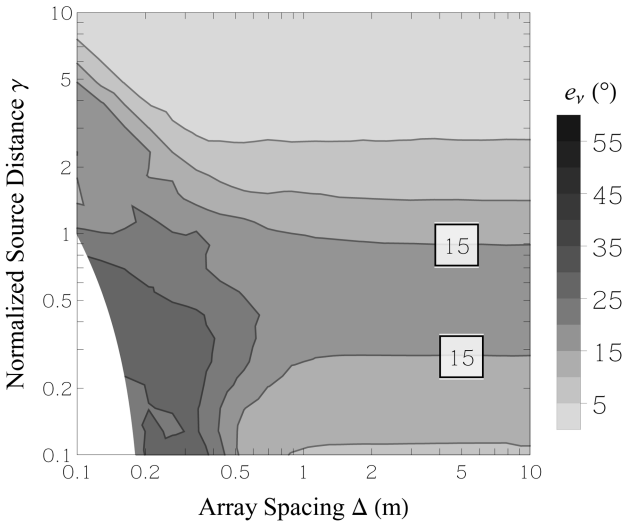
Fig. 2. Level errors $e_\lambda$ (top panels) and spectral errors $\rho_\eta$ (bottom) for microphone spacing $\Delta$ and normalized source distance $\gamma$. Level error contour lines are drawn every 2 dB; spectral error contour lines are drawn every 1 dB.

For spectral coloration, we see in Figs. 2(c) and 2(d) that the TFA method yields larger errors than the VMI method for all microphone spacings larger than approximately 0.5 m. In particular, the VMI method yields significantly smaller errors for interior sources with large microphone spacings ($\gamma < 1$ and $\Delta > 0.5$ m). Only for exterior sources with microphone spacings smaller than approximately 0.25 m does the TFA method achieve smaller spectral errors than the VMI method. Future investigations should attempt to determine the source of, and correct for, the spectral coloration induced by the TFA method at large microphone spacings.

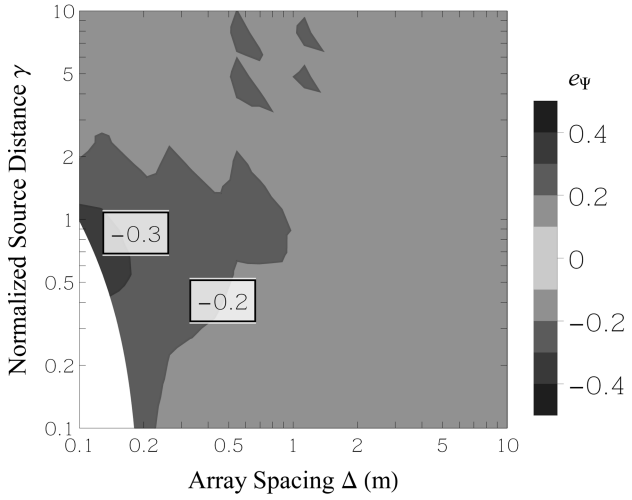
Localization errors for the TFA method are shown in Fig. 3(a). Contrary to that method's coloration performance (see Fig. 2(c)), which is most accurate at small microphone spacings (e.g., $\Delta < 0.5$ m), the localization errors incurred by this method are largest at those small microphone spacings. In particular, for exterior sources with microphone spacings smaller than approximately 0.3 m, the VMI method yields a significant improvement ($\sim$15°) over the TFA method. Additionally, for far exterior sources ($\gamma > 3$) and at all microphone spacings, the VMI method yields a marked improvement ($\sim$5°) over the TFA method. For microphone spacings larger than approximately 0.5 m, the errors incurred by the VMI method are relatively constant with spacing, whereas those incurred by the TFA method improve with increasing spacing and even become very small ($\epsilon_v <$ 5°) at large $\Delta$ and $\gamma < 1$. Accordingly, the TFA method yields a significant improvement over the VMI method for interior sources with microphone spacings larger than approximately 1 m.



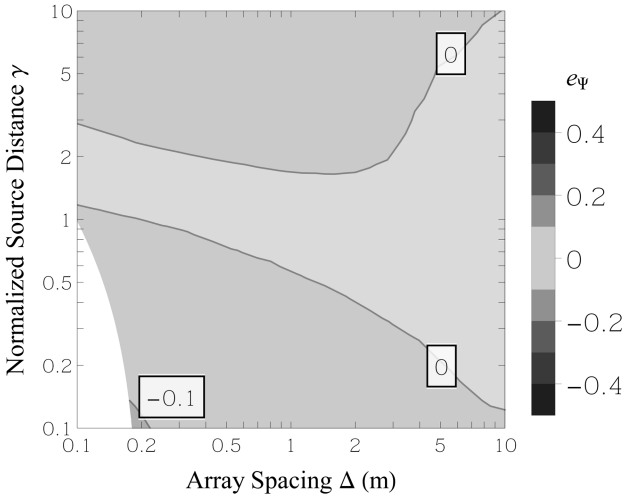

(a) Localization errors – TFA method




(b) Localization errors – VMI method




(c) Diffuseness errors – TFA method




(d) Diffuseness errors – VMI method

Fig. 3. Predicted localization errors $e_\nu$ (panels (a) and (b)) and diffuseness errors $e_\Psi$ (panels (c) and (d)) for microphone spacing $\Delta$ and normalized source distance $\gamma$. Localization error contour lines are drawn every 5°; diffuseness error contour lines are drawn in increments of 0.1.

Under such conditions, although the absolute localization performance of the VMI method ($\epsilon_\nu \sim 15°$) may be tolerable for some applications, it is unclear to what extent the method can reproduce *dynamic* localization cues as the listener moves. At large translation distances, the interpolation calculation effectively reduces to a "switching" between different microphones (since $k_0$ will be very low; see Eq. (21)) to the one that is valid for a given listener position, typically with a small region of overlap wherein multiple microphones can be used. Thus, to prevent abrupt changes in perspective, the audio streams may require some form of crossfading between frames. Alternatively, more sophisticated extrapolation methods [6, 15, 8] could potentially be employed for cases with only $P = 1$ valid microphone. This is a topic for exploration and development in future practical implementations.

From the plots of diffuseness errors shown in Figs. 3(c) and 3(d), we immediately see that the VMI method achieves more accurate performance than the TFA method over all conditions. The TFA method consistently yields a diffuseness parameter which is too small, whereas the VMI method achieves nearly exact diffuseness (except at very small $\gamma$ and $\Delta$). This apparent deficiency in diffuseness of the TFA method suggests that the diffuse sound term in the sound field re-synthesis equation, Eq. (11), is underestimated by the method. Consequently, it may be relatively straightforward to modify the TFA method to correct for this behavior. This too is a topic for further development.

## 4 CONCLUSIONS

In this work we conducted numerical simulations in order to characterize and compare the performance of the time-frequency analysis (TFA) interpolation method of Thiergart et al. [13] to our recently proposed parametric valid microphone interpolation (VMI) method [16]. Following the simulation framework laid out in Section 2, we simulated simple incident sound fields consisting of a two-microphone array and a single point-source and varied source distance and azimuth, microphone spacing, and listener position. We conducted a comprehensive analysis of the methods by computing, over a wide range of conditions, the metrics enumerated in Section 2.2 for sound level, spectral coloration, source localization, and diffuseness. These analyses yielded the following findings:

- The TFA method yields virtually exact sound levels for all conditions and is particularly superior to the VMI method for interior sources;
- The TFA method yields significantly larger spectral errors than the VMI method for microphone spacings larger than approximately 0.5 m;
- The TFA method yields significantly smaller localization errors than the VMI method for interior sources with microphone spacings larger than approximately 1 m; and
- The TFA method does not sufficiently reproduce the diffuseness of a sound field, whereas the VMI method yields nearly exact diffuseness for almost all conditions.

### 4.1 Practical Implications

Taken together, these findings suggest that the TFA and VMI methods may each be more suitable in different practical domains. For the present discussion, we define the following practically relevant "axes":

(1) The *sparsity* of the microphone array (i.e., the size of the desired navigable region relative to the number of available microphones),

(2) The *intimacy* of the sound sources (i.e., the proximity of the sources to the navigable region), and

(3) The *complexity* of the sound field (i.e., the total number of sources and/or the reverberance of the recording environment).

While the first two of these axes can be easily related to microphone spacing and normalized source distance, respectively, the third axis has not been directly explored here. However, based on the construction of the TFA method, we speculate that this method may have difficulties accommodating multiple sources since, at each time-frequency bin, only a single point-source is created (see Section 1.1). Consequently, the capability of this method to accurately reproduce multiple sources warrants further study.

Nevertheless, below, we identify several general principles with which to choose between the two methods in various applications spanning these axes:

(1a) With a sparse microphone array (e.g., when covering a large room with only a few microphones), the TFA method will generally yield superior localization accuracy, whereas the VMI method will incur less spectral coloration.

(1b) With a dense microphone array, the methods perform comparably to each other (i.e., neither method is particularly superior) in terms of localization accuracy and spectral coloration.

(2a) When recording primarily intimate sources (e.g., an immersive recording of a small group of musicians), the TFA method will yield superior localization accuracy and will likely better convey source distance information (due to its accurate reproduction of sound level), whereas the VMI method will again incur less spectral coloration.

(2b) When recording primarily distant sources (e.g., when covering the audience section only of a concert hall), the VMI method will yield smaller spectral and localization errors.

(3a) For an acoustically complex sound field (e.g., in a room with highly reflective surfaces and/or many scattering bodies), the VMI method is likely more suitable as it more accurately reproduces diffuseness and, potentially, the TFA method will fail to adequately reproduce many sources.

(3b) For an acoustically simple sound field (e.g., an outdoor recording of a park with sparsely distributed sources), the TFA method will likely yield superior localization accuracy, and its deficiency in diffuseness will be less problematic.

While these principles specify the superior method for any given practical domain, another way of summarizing the present results is to determine the domains, in terms of these practical axes, over which each method yields *accurate and superior* performance. As we did not explore the complexity axis explicitly, here we omit that axis and focus only on the sparsity of the microphone array and the intimacy of the sources. Additionally, since the level and diffuseness results are relatively straightforward (see the top panels of Fig. 2 and the bottom panels of Fig. 3, respectively), we omit those metrics as well.

Using the spectral errors plotted in the bottom panels of Fig. 2 and the localization errors plotted in the top panels of Fig. 3, we first identify, for each method, regions of low coloration ($\rho_\eta < 3$ dB) and regions of accurate localization ($e_\nu < 10°$). We then determine the regions in which each method performs a) more accurately than that error limit and b) more accurately than, or at least comparably to, the other method. These regions are sketched in Fig. 4.

From these plots, we see that, in applications with distant sources and with a sparse microphone array, the VMI method yields accurate and superior performance in both coloration and localization. Furthermore, for most applications with a sparse microphone array or with intimate sources, the VMI method yields accurate and often superior spectral coloration performance, and for most applications with distant sources, the VMI method yields accurate and

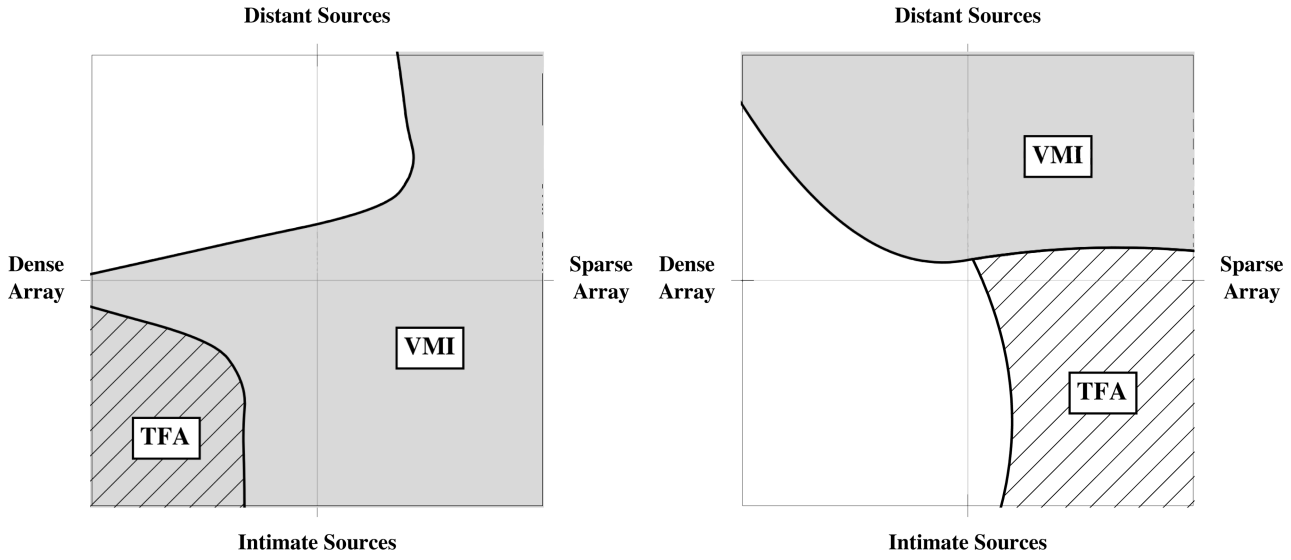



Fig. 4. Region plots illustrating accurate and superior methods in terms of spectral coloration (panel (a)) and localization accuracy (panel (b)) across practical applications with varying microphone array sparsity (horizontal axis) and sound source intimacy (vertical). The gray filled regions correspond to the valid microphone interpolation (VMI) method and the hatched regions correspond to the time-frequency analysis (TFA) method. Regions that are both filled and hatched indicate that the methods perform comparably; empty regions indicate that neither method satisfies the specified error limit.

superior localization performance. The TFA method, however, yields accurate spectral coloration performance only for applications with a dense microphone array and with intimate sources, and accurate and superior localization performance only for applications with a sparse microphone array and with intimate sources. Consequently, in such applications (with both a sparse microphone array and with intimate sources), the TFA method yields improved localization but degraded coloration performance compared to the VMI method.

Although the results shown here are for first-order ambisonics input signals only, future modifications to the VMI method might yield improved performance through including higher-order terms (e.g., by choosing a different critical wavenumber, cf. Eq. (21)). At present, however, it is practically advantageous to employ first-order ambisonics microphones (which tend to be significantly less expensive than higher-order ones and require fewer recording channels, preamplifiers, etc.). Also, as mentioned in Section 2.1, the TFA method fails to triangulate sources with azimuths of $|\varphi_0| = 90^\circ$. While, in practice, a source azimuth of exactly $\pm 90^\circ$ is virtually impossible (e.g., due to positioning errors, noise, etc.), this does suggest that the triangulation calculation (see Eq. (6)) may be very sensitive to small changes in azimuth near these extremes. This issue might be easily avoided, however, by using $P > 2$ microphones arranged in a triangular or rectangular configuration, for example.

While the evaluation presented here has been purely numerical, we hope that the practical recommendations enumerated above will facilitate real-world implementations of these navigational methods. Ideally, the conclusions drawn from these analyses will be borne out by future experimental investigations. In particular, subjective listening assessments of these and other methods would be useful to both perceptually validate the present findings and identify other areas for improvement in these methods.

## 5 ACKNOWLEDGMENTS

This work was sponsored by the Sony Corporation of America. The authors wish to thank R. Sridhar for fruitful discussions throughout the work and P. Stitt for providing the MATLAB code for the precedence-effect-based energy vector model [31].

## REFERENCES

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

[2] 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.

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

[4] 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.

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

[6] 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 Aug.).

[7] K. Wakayama, J. Trevino, H. Takada, S. Sakamoto, and Y. Suzuki, "Extended Sound Field Recording Using Position Information of Directional Sound Sources," presented at the *2017 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pp. 185–189 (2017 Oct.), https://doi.org/10.1109/WASPAA.2017.8170020.

[8] A. Allen, "Ambisonics sound field navigation using directional decomposition and path distance estimation" (2019 Jan.), US Patent 10,182,303, https://patents.google.com/patent/US10182303B1/.

[9] 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.

[10] P. Grosche, F. Zotter, C. Schörkhuber, M. Frank, and R. Höldrich, "Method and apparatus for acoustic scene playback" (2018 May), WIPO (PCT) Patent App. WO/2018/077379.

[11] D. Rudrich, F. Zotter, and M. Frank, "Evaluation of Interactive Localization in Virtual Acoustic Scenes," *43. Jahrestagung für Akustik (DAGA 2017)*, pp. 279–282 (2017 Mar.).

[12] T. Deppisch and A. Sontacchi, "Browser Application for Virtual Audio Walkthrough," *Forum Media Technology*, pp. 145–150 (2017).

[13] O. Thiergart, G. Del Galdo, M. Taseska, and E. A. P. Habets, "Geometry-Based Spatial Sound Acquisition Using Distributed Microphone Arrays," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 21, no. 12, pp. 2583–2594 (2013 Dec.), https://doi.org/10.1109/TASL.2013.2280210.

[14] X. Zheng, *Soundfield Navigation: Separation, Compression and Transmission*, Ph.D. thesis, University of Wollongong (2013).

[15] 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.

[16] 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 Sep.), conference paper 4-2.

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

[18] C. Nachbar, F. Zotter, E. Deleflie, and A. Sontacchi, "ambiX - A Suggested Ambisonics Format," *Proceedings of the 3rd Ambisonics Symposium* (2011 June).

[19] The MathWorks, Inc., "Hamming window," https://www.mathworks.com/help/signal/ref/hamming.html (2019), [Online; accessed 16-February-2019].

[20] J. Merimaa and V. Pulkki, "Spatial Impulse Response Rendering I: Analysis and Synthesis," *J. Audio Eng. Soc.*, vol. 53, pp. 1115–1127 (2005 Dec.).

[21] P. Samarasinghe, T. Abhayapala, and M. Poletti, "Wavefield Analysis Over Large Areas Using Distributed Higher Order Microphones," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 22, pp. 647–658 (2014 Mar.), ISSN 2329-9290, https://doi.org/10.1109/TASLP.2014.2300341.

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

[23] J. G. Tylka and E. Y. Choueiri, "Algorithms for Computing Ambisonics Translation Filters," Technical report, 3D Audio and Applied Acoustics Laboratory, Princeton University (2019 Feb.).

[24] P. C. Hansen, *Rank-Deficient and Discrete Ill-Posed Problems: Numerical Aspects of Linear Inversion* (Society for Industrial and Applied Mathematics, 1998).

[25] Z. Průša, P. L. Søndergaard, N. Holighaus, C. Wiesmeyr, and P. Balazs, "The Large Time-Frequency Analysis Toolbox," http://ltfat.github.io (2012) (Online; accessed 16 February, 2019).

[26] B. R. Glasberg and B. C. J. Moore, "Derivation of Auditory Filter Shapes from Notched-Noise Data," *Hearing Research*, vol. 47, no. 1–2, pp. 103–138 (1990 Aug.).

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

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

[29] 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 May).

[30] P. Zahorik, D. S. Brungart, and A. W. Bronkhorst, "Auditory Distance Perception in Humans: A Summary of Past and Present Research," *Acta Acustica united with Acustica*, vol. 91, no. 3, pp. 409–420 (2005 May).

[31] P. Stitt, "Matlab Code," https://circlesounds.wordpress.com/matlab-code/ (2016) (Online; accessed 16 February, 2019).

## A.1 A RELEVANT AMBISONICS THEORY

In the free field (i.e., in a region free of sources and scattering bodies), the acoustic potential field satisfies the homogeneous Helmholtz equation and can therefore be expressed as an infinite sum of regular (i.e., not singular) basis solutions. In ambisonics, these basis solutions are given by $j_l(kr)Y_n(\hat{r})$, and the sum, also known as a spherical Fourier-Bessel series expansion, is given by [23, Ch. 2]

$$\psi(k,\vec{r}) = \sum_{n=0}^{\infty} 4\pi(-i)^l A_n(k) j_l(kr) Y_n(\hat{r}), \tag{A.31}$$

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.

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 [1, Eq. (10)]

$$A_n(k) = i^{l+1} k h_l(ks_0) Y_n(\hat{s}_0), \tag{A.32}$$

where $h_l$ is the (outgoing) spherical Hankel function of order $l$.

## THE AUTHORS

Joseph G. Tylka

Edgar Y. Choueiri

Joe Tylka is a recent Ph.D. recipient from the 3D Audio and Applied Acoustics Laboratory at Princeton University, where his dissertation research explored binaural rendering and virtual navigation of recorded 3D sound fields. Joe's research interests include machine listening and acoustic signal processing.

•

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.