OPEN ACCESS Freely available online
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, (2020 March.).
DOI: https://doi.org/10.17743/jaes.2019.0055

Fundamentals of a Parametric Method for Virtual Navigation Within an Array of Ambisonics Microphones*

JOSEPH G. TYLKA AND EDGAR Y. CHOUEIRI
(josephgt@alumni.princeton.edu)

Princeton University, Princeton, New Jersey

Fundamental aspects of a method for virtual navigation of a sound field within an array of ambisonics microphones, wherein the subset of microphones to be used for interpolation is determined parametrically, are presented. An existing, weighted-average–based navigational method serves as a benchmark due to its simplicity and its applicability to arbitrary sound fields but introduces comb-filtering and, for near-field sources, degrades localization. A critical review of existing methods is presented, through which a number of issues are identified. In the proposed method, those microphones that are nearer to the desired listening position than to any source are determined based on the known or inferred positions of sources. The signals from only those microphones are then interpolated using a regularized least-squares matrix of filters. Spectral distortions and source localization errors are characterized for the benchmark and proposed methods via numerical simulations of a two-microphone array, and an experimental validation of these simulations is presented. Results show that, for near-field sources, the proposed method significantly outperforms the benchmark in both spectral and localization accuracy due to the exclusion of the second microphone. For far-field sources, the proposed method achieves slightly decreased spectral distortions due to the flattened response of the interpolation filters.

0 INTRODUCTION

Virtual navigation of 3D ambisonics-encoded sound fields (i.e., sound fields that have been decomposed into spherical harmonics) enables a listener to explore (with 6 degrees of freedom, i.e., translation and rotation) an acoustic space and, ideally, experience a spatially and tonally accurate perception of the sound field. Applications of this type of 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 prerender spatial RIRs on a fixed grid of points and then, during playback, navigate between them to generate spatial RIRs at intermediate positions.

A well-known limitation of the higher-order ambisonics (HOA) 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]. In particular, a well-established rule of thumb states that a sound field is accurately represented by an Lth-order expansion up to a distance r provided that $kr \leqslant L$, where k is the angular wavenumber [2]. Consequently, the navigable region of such a sound field is inherently restricted. Indeed, existing techniques for sound field navigation using a single HOA microphone[1] have been shown to introduce spectral

*This article presents a revised formulation and comprehensive characterization of the method originally presented at the 2016 AES International Conference on Audio for Virtual and Augmented Reality in Los Angeles, California, on September 30th, 2016.

[1] 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 obtain ambisonics signals via a transformation of the raw microphone signals.

distortions [3, 4] and degrade localization [5, 6] as the listener navigates farther away from the expansion center.[2]

Furthermore, according to theory, the HOA 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) [8, 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 previous studies have developed navigational methods that employ an array of HOA microphones distributed throughout the sound field (or, equivalently, encode a synthetic sound field into HOA at multiple discrete positions). In the following section, we provide a critical review of these existing methods and identify the main challenges they face.

0.1 Critical Review of Previous Work

Existing navigational methods may be categorized based on the type of processing employed. For the present discussion, we define the following three types of methods:

- *Linear*: The processing to be applied to the input signals is determined solely based on the geometric arrangement of the recording microphones and the desired listener position. Furthermore, the operations applied to the signals comprise only linear filtering operations (e.g., scalar gains, summing, delays, etc.).
- *Nonlinear*: The construction is similar to linear methods, except the operations applied to the microphone signals cannot be represented as linear filtering operations.
- *Parametric*: The specific processing (either linear or nonlinear) to be applied to the microphone signals is dependent on (i.e., parameterized by) some additional information about the recorded sound field (e.g., source positions), which may be supplied as an additional input or derived from the microphone signals.

In the following sections, we review existing methods of each of these types.

0.1.1 Linear Methods

For all linear methods, since the rendered signals are, by definition, given by linear combinations of the measured signals, such methods are prone to violating the region of validity restriction mentioned above if those microphones that are nearer to a source than to the desired listening position are included in the calculation. As we will show in the present article, such a violation can lead to a degradation of the estimated sound field at the listening position. However, an advantage of linear methods is that they are widely applicable, in that they do not require any additional information about the sound field other than the measured signals, nor do they require any assumptions to be made regarding the types of incident signals or the spatial characteristics of the sound field.

Perhaps the simplest interpolation-based navigational method is to compute a weighted average of the ambisonics signals from each microphone, where the interpolation weights are related to the distances from the listener to each microphone. This approach has been implemented by Mariette and Katz [9] using virtual first-order ambisonics (FOA) microphones spaced 20 m apart and arranged in both linear (two microphones) and triangular (three microphones) configurations. Similarly, Southern et al. [10] employed this method in order to interpolate ambisonics RIRs to enable real-time navigable auralizations of acoustic spaces.

One fundamental limitation of this method is that it necessarily confines the listener to the region interior to the microphone array, since it is purely an interpolation method with no means of extrapolation. Furthermore, objective and perceptual investigations have shown that this method suffers from significant localization errors, in particular when the source distance (from the center of the microphone array) is small compared to the microphone spacing [9, 11] (here, we refer to this condition as having an "interior source"). This effect is consistent with findings from a previous study of ours [12, Fig. 6]. In that study, we also showed that if a sound source is nearer to one microphone than to another, this method will necessarily induce comb-filtering (as it produces at least two copies of that source's signal, separated by a finite time delay) [12, Fig. 4(a)]. In Sec. 2 below, we revise and expand these analyses.

Recently, Patricio et al. [13] 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. The method also employs a low-pass filter on each microphone's signals to mimic the atmospheric absorption of high frequencies. Fundamentally, this method is subject to the same issues and limitations as the weighted average method, since it is essentially the same calculation but with a particular (order and frequency-dependent) prescription for the interpolation weights. Nevertheless, the authors experimentally demonstrated that the proposed distance-biasing approach achieves plausible source localization and perception of listener movement.

Fernandez-Grande [14] proposed an equivalent-source method for representing and reconstructing a measured sound field. In this method, the sound field is captured with one or more HOA microphones and subsequently fitted, in a least-squares sense, to that sound field created by a predefined grid of virtual monopole sources. This yields a virtual sound field consisting of a finite set of known monopole sources, which can then be rendered at an

[2] Similar results were found by Walther and Faller [7], although their aim was not to navigate the sound field but instead to simulate a spaced-microphone recording using a single first-order ambisonics microphone.

arbitrary position elsewhere in the space, such that the listener is not confined to a strictly interior region. However, without a priori knowledge of the real sound source positions, the performance of the method may degrade. Consequently, in order to better accommodate arbitrary source positions, this method might be improved by a parametric implementation that incorporates some basic source localization algorithm to estimate source positions directly from the HOA signals.

Samarasinghe et al. [15] developed a regularized least-squares inverse interpolation approach based on spherical harmonic translation coefficients using an array of HOA microphones. Several subsequent studies have demonstrated or improved upon this method [16–18]. In particular, Ueno et al. [19] developed a similar method that takes a Bayesian inference approach, which was shown to achieve improved performance (in terms of reconstruction errors) at high frequencies.

In a previous publication, we implemented a similar matrix of regularized least-squares interpolation filters [12, Sec. 3.2] and showed that neglecting to account for the region of validity for each microphone can lead to significant localization errors [12, Fig. 8(b)]. Additionally, a qualitative analysis of spectral distortions suggested that these methods may induce significant spectral coloration (i.e., the perception of those distortions) at high frequencies [12, Fig. 4(b)]. It was also shown that, at large microphone spacings (compared to source distance), this method suffers from significant localization errors [12, Fig. 6]. At the time of publication, the objective localization model used to quantify localization errors had not yet been subjectively validated, but we have since refined and validated it (albeit over a limited range of conditions) [20].

More recently, Wang and Chen [21] proposed a modification to this inverse interpolation method 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 [21, Sec. IV]. Fundamentally, this method is subject to the same issues and limitations as the original method of Samarasinghe et al. [15], although its performance has not been evaluated in terms of localization and coloration.

0.1.2 Nonlinear Methods

In the context of HOA RIR interpolation, spectral distortions and localization errors may be mitigated by taking a dynamic time warping approach, similar to that proposed by Masterson et al. [22]. Although this approach has not been implemented to interpolate HOA RIRs specifically, a previous study by the same authors has suggested incorporating arbitrary microphone directivity [23], which would enable extending this method to HOA.[3] One limitation of this method is that it requires knowledge not only of the microphone positions, but also of the source position, which, for an arbitrary sound field, would not be known a priori. Additionally, by its nature, this method can only be applied to RIRs and is therefore unsuitable for interpolating sound fields consisting of arbitrary signals.

Emura [26] recently proposed a more general method which combines the measured signals from two HOA microphones in order to estimate coefficients for a single global plane-wave decomposition of the sound field. In this method, a so-called "dictionary" matrix is precomputed for a high-resolution grid of plane-waves incident on the microphones, and the plane-wave signals that best explain the measured pressures on the microphones are determined by minimizing the ℓ_1 norm of an error signal [26, see Eq. (20)]. This calculation, based on compressed sensing techniques (which have been reviewed by Epain et al. [27] in the context of spatial sound field analysis and synthesis), is consequently nonlinear with respect to the measured signals. (If a least-squares, i.e., ℓ_2-norm minimization approach, were taken, this method could then be implemented linearly.) An evaluation of this method presented by the authors shows that the performance of the method (in terms of low-frequency rms errors) is improved in the vicinity of the second microphone compared to if only a single microphone is used. However, due to the plane-wave–based nature of the proposed method, it can be expected that near-field and interior sources will be problematic.

0.1.3 Parametric Methods

Recently, Bates et al. [28] developed a perceptually motivated method for sound field navigation using a 50-cm × 50-cm square arrangement of four FOA microphones.[4] In this method, each ambisonics microphone is used to create a virtual directional microphone, the placement and directivity of which are varied as a function of listener position and are parameterized based on the source positions (which therefore must be known a priori). The signals from these virtual directional microphones are then encoded into a single HOA stream. Based on its published formulation, this method appears well suited for applications with a small, predefined navigable region but would be difficult to extend to cover a larger region. Additionally, in an objective analysis of perceived source distance and direction, this method achieved promising performance when navigating towards the source but yielded significant directional errors and diminished distance performance when navigating away [28, Sec. 3].

Other parametric methods have been developed which rely on a time-frequency (i.e., short-time Fourier transform)

[3] More recently, Garcia-Gomez and Lopez [24] extended this method to interpolate binaural RIRs. For additional recent progress on the subject of binaural RIR interpolation, the interested reader is referred to the work of Brandenburg et al. [25] and references therein.

[4] This work extends the so-called "perspective control microphone array" method, which creates the perception of navigation by varying the mixing of the signals from 5 coincident pairs of microphones (cardioid and hypercardioid), spaced ∼2 m apart [29, 30].

analysis of the sound field using two (or more) FOA microphones. One such method is known as "collaborative blind source separation" [31, Sec. 3.3], in which discrete sound sources are first identified, localized, and isolated and are subsequently treated as virtual sources, which may be artificially moved relative to the listener to emulate navigation. Similarly, Thiergart et al. [32] developed a method of sound field navigation in which the sound field is decomposed into diffuse sound components and multiple discrete sources, which are triangulated in the time-frequency domain via acoustic intensity vector calculations (cf. Merimaa and Pulkki [33, Eq. (11)]) from each microphone. The signals from these virtual sources are then "re-recorded" by a virtual microphone at an arbitrary position and with arbitrary directivity.[5] Taking a similar approach, Schörkhuber et al. [34] developed for sound field analysis a wireless system consisting of an array of FOA microphones, which the authors showed to be able to accurately localize multiple sources [34, 35].

While clearly promising, these methods are only ideal for sound fields consisting 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) [32, Sec. II]. An advantage of these methods is that, using the virtual model of the captured sound field, the listener is free to navigate anywhere in 3D space rather than being confined to the region interior to the microphone array. However, even in ideal situations, such parametric processing methods employed in the time-frequency domain often result in a minor degradation of sound quality [31, Sec. 5.3]. Furthermore, it is unclear if these methods can accurately capture and reproduce the directivities of the real sources, as this issue has not been addressed in the literature. Indeed, both Zheng [31] and Thiergart et al. [32] exclusively use omnidirectional point sources to model the sound field. We speculate, however, that for the method of Thiergart et al. [32], perhaps source directivity information can be implicitly contained in the modeled sound field by the spatial distribution of the virtual point sources.

Recently, Wakayama et al. [36] proposed an extrapolation method (i.e., a navigational method using a single HOA microphone) based on spherical-harmonic translation filters, which was shown to enable navigation beyond a near-field source and to estimate its directivity using a multipole expansion but requires a priori knowledge of the source's position. It is not clear, however, whether or how this method can be extended to accommodate multiple sources.

In our previous publication, we presented a parametric method of excluding any microphones that are nearer to any sound source than to the desired listening position, which also requires either a priori knowledge of or a means of estimating the positions of any near-field sources [12, Sec. 3.3]. However, this method ensures that all microphones used in the calculation provide valid descriptions of the sound field at the listening position, and the ambisonics signals from those "valid" microphones are then combined using a matrix of regularized least-squares interpolation filters in order to obtain an estimate of the sound field at the listening position [12, Sec. 3.2]. The spectral distortions and localization errors induced by this method, however, have not been fully characterized.

[5] Although not specifically addressed by the authors, the generalization of the method to a virtual HOA microphone appears straightforward.

0.2 Objectives and Approach

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

1. the method restricts navigation to a finite region (e.g., a strictly interior region, the horizontal plane, etc.),
2. the method violates the region of validity restriction,
3. the method degrades localization information,
4. the method introduces spectral coloration or other audible processing artifacts,
5. the method requires additional geometric information about the sound field (e.g., source locations),
6. the method cannot accommodate arbitrary signals,
7. the method cannot accommodate arbitrary (e.g., dense or reverberant) sound fields,
8. the method cannot reproduce source directivities, and/or
9. the method cannot reproduce moving sources.

The issues suffered by and addressed by each method discussed in the previous section are summarized in Table 1.

In this work, we take the weighted average method as a benchmark as it is both simple to implement and broadly applicable to arbitrary sound fields consisting of arbitrary signals and with an arbitrary placement of sources (i.e., this method does not suffer from issues 5, 6, or 7). A fundamental aspect of our proposed navigational method is the parametric exclusion of microphones, which ensures that we do not violate the region of validity restriction for any microphone (issue 2) but requires some means of obtaining information about the distances of sources to each microphone (issue 5). We aim to demonstrate the benefits of our method over the benchmark in terms of improvements in spectral and localization accuracy (issues 3 and 4). Potentially, our method might also enable navigation beyond the strict interpolation-only navigable region (issue 1), whereas the benchmark method cannot, but here we only evaluate these methods in this region. Both methods may also preserve source directivity or moving-source information (issues 8 and 9), but we do not explore these issues here.

The objectives of the present work are as follows: 1) to demonstrate the fundamental problems inherent to the weighted average method, 2) to revise our previously proposed parametric navigational method in order to mitigate its induced spectral coloration, 3) to establish fundamental aspects of the proposed method and provide an experimentally validated proof-of-concept demonstration of its advantages over the weighted average method, and 4) to

Table 1. Summary of published interpolation methods and the corresponding issues (numbers refer to the list given in Sec. 0.2) suffered by or addressed by each method. Issues for which each method has yet to be tested are omitted from this table; issues marked with an asterisk (*) indicate subjects demonstrated in the present article.

Method	Processing	Suffers	Addresses
Weighted-average interpolation (Sec. 1.2) [10, 9]	Linear	1,2,3*,4*	5,6,7
Distance-biased linear interpolation [13]	Linear	1,2	3,5,6,7
Spherical equivalent source method [14]	Linear	2,5	1,6,7
Regularized inverse interpolation (Sec. 3.2) [15, 12, 19]	Linear	2,3,4	1,5,6,7
Inverse ambisonics translation via plane-waves [21]	Linear	2	1,5,6,7
Dynamic time warping room impulse response (RIR) interpolation [22, 23]	Nonlinear	1,5,6,7	3,4
Sparse plane-wave estimation and translation [26]	Nonlinear	7	1,5,6
Perspective control ambisonics microphone array [28]	Parametric	1,3,5	6
Collaborative blind-source separation [31]	Parametric	4,6,7,8	1,2,3,5
Time-frequency analysis and modeling [32]	Parametric	4,6,7	1,2,3,5
Singular ambisonics translation (extrapolation) [36]	Parametric	5,7	1,2,6,8
Proposed: valid-only interpolation (Sec. 3.1) [12]	Parametric	5	2,3*,4*

characterize and compare the performances of both methods.

To these ends, in Sec. 1, we formulate the general problem of virtual navigation of ambisonics and review the weighted average interpolation method. We then evaluate, in Sec. 2, the fundamental problems (spectral distortions due to comb-filtering and localization errors due to the precedence effect) inherent to this method through numerical analyses of frequency response and perceived localization. In Sec. 3, we describe our parametric navigational method and the regularized least-squares interpolation filters. We then describe, in Sec. 4, numerical simulations of simple incident sound fields (consisting of two microphones and a single source), which we conduct in order to characterize and compare the performance, in terms of objective metrics for perceived spectral coloration and localization, of each of the two methods. We then present, in Sec. 6, an experimental validation of the numerical simulations through a comparison with physical measurements of spectral distortions and source localization, taken over a subset of the simulated conditions. Finally, in Sec. 7, we summarize our findings and conclude.

1 VIRTUAL NAVIGATION OF AMBISONICS

As is common in higher-order 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 unit vectors by $\hat{r} \equiv \vec{r}/r$.

1.1 Problem Formulation

Consider an array of P HOA microphones, where the pth microphone is located at $\vec{u}_p$ for $p \in [1, P]$. For microphones of order L_{in}, each microphone "captures" $N_{\text{in}} = (L_{\text{in}} + 1)^2$ ambisonics signals (computed from the raw microphone capsule signals), which we represent with a vector, $\mathbf{b}_p$. (See Appendix A for a review of the relevant ambisonics conventions and theory.) In general, techniques for virtual navigation of ambisonics aim to approximate, up to order L_{out} and with $N_{\text{out}} = (L_{\text{out}} + 1)^2$ terms, the exact ambisonics signals, $\mathbf{a}$, of the sound field at a listening position $\vec{r}_0$.

1.2 Weighted Average Interpolation

In the navigation method proposed by Mariette and Katz [9] and Southern et al. [10] (which we take to be a benchmark), a weighted sum of the captured ambisonics signals is computed to obtain an estimate of the ambisonics signals at the listening position, given by

$$\tilde{\mathbf{a}} = \sum_{p=1}^{P} w_p \mathbf{b}_p. \tag{1}$$

Here, we normalize the weights such that

$$\sum_{p=1}^{P} w_p = 1. \tag{2}$$

Note that as the sum in Eq. (1) is computed term by term, we must have an output order $L_{\text{out}} \leqslant L_{\text{in}}$, where the inequality arises if one chooses to discard higher-order terms captured by the microphones. Depending on the placement of the microphones, the weights w_p may be computed using standard linear or bilinear schemes, for example.

2 FUNDAMENTAL PROBLEMS

In this section, we evaluate two fundamental problems inherent to the weighted average interpolation method: 1) comb-filtering introduced by summing two very similar signals separated by a time delay and 2) localization errors due to the precedence effect.

2.1 Array Geometry

Consider a linear microphone array geometry, illustrated in Fig. 1, in which a pair of microphones ($P = 2$) are separated by a distance Δ, equidistant from the origin and placed along the lateral y-axis, such that their positions are given

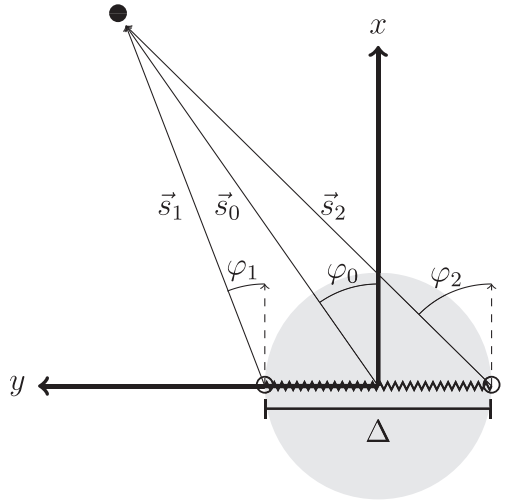


Fig. 1. Diagram of a two-microphone array (empty circles) with a single source (filled circle). The shaded gray disk indicates the interior region, where $r < \Delta/2$. (Note that this region is actually spherical, but here we only consider the horizontal plane.) The jagged line segment indicates the navigable region, where $y \in [-\Delta/2, \Delta/2]$ and $x = z = 0$.

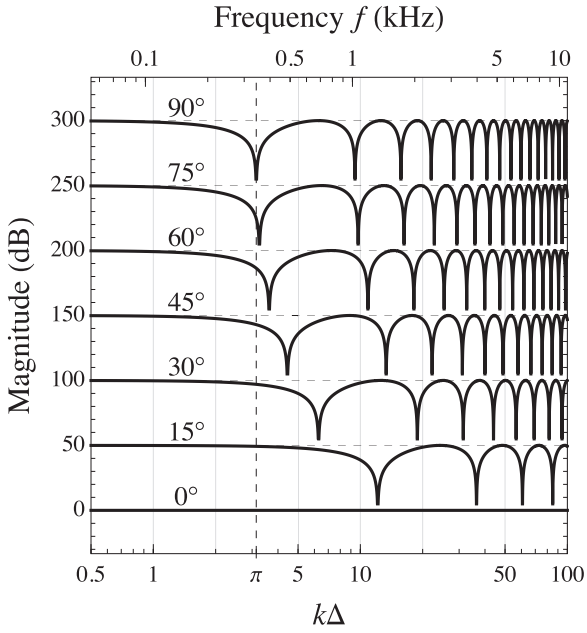


Fig. 2. Comb-filter magnitude responses caused by the weighted average method for various source azimuths. The bottom axis shows the nondimensional frequency $k\Delta$ while the top axis shows frequency in kHz assuming a microphone spacing of $\Delta = 0.5$ m. For legibility, each frequency response is offset by 50 dB and notch depths have been artificially truncated to not exceed −45 dB.

by $\vec{u}_1 = (0, \Delta/2, 0)$ and $\vec{u}_2 = (0, -\Delta/2, 0)$. In this configuration, 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 $y_0 \in [-\Delta/2, \Delta/2]$. In this configuration, linear interpolation weights for Eq. (1) are given by $w_1 = 0.5 + y_0/\Delta$ and $w_2 = 0.5 - y_0/\Delta$ for $y_0 \in [-\Delta/2, \Delta/2]$.

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 pth 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 φ_p and the relative source distance from that microphone is s_p.

For later use, we further define a nondimensional geometrical parameter $\gamma = r/(\Delta/2)$. Here we refer to the region with $\gamma > 1$ as the *exterior region* and that with $\gamma < 1$ as the *interior region* (see Fig. 1).

2.2 Spectral Distortion: Comb-Filtering

For a plane-wave source (i.e., $s_0 \to \infty$), the path-length difference from the source to each microphone is $\Delta \sin|\varphi_0|$, and the corresponding time-of-arrival delay is $(\Delta/c)\sin|\varphi_0|$, where c is the speed of sound. For a listener at the origin, the interpolation weights are given by $w_1 = w_2 = 0.5$. Thus, the weighted-average impulse response for the zeroth-order (i.e., omnidirectional) ambisonics signal[6] is given by

$$a_0(t) = 0.5\delta(t) + 0.5\delta\left(t - \frac{\Delta}{c}\sin|\varphi_0|\right), \tag{3}$$

where $\delta(t)$ is the Dirac delta function, and the corresponding frequency response is given by

$$A_0(k) = 0.5 + 0.5e^{-ik\Delta\sin|\varphi_0|}, \tag{4}$$

where $k = 2\pi f/c$ is the angular wavenumber for frequency f. As this frequency response depends only on $|\varphi_0|$ and the nondimensional frequency $k\Delta$, we plot, in Fig. 2, magnitude responses of A_0 for several source azimuths. Note that due to the lateral symmetry of the geometry, these responses hold for negative azimuths ($\varphi_0' = -\varphi_0$), and due to the front-back symmetry, they also hold for rear azimuths ($\varphi_0' = 180° \pm \varphi_0$).

From this plot, we see that only sources at 0° (or 180°) azimuth are interpolated without comb-filtering. For all other azimuths, comb-filtering is introduced due to the time-of-arrival delay between microphones. Note that for a point source (i.e., $s_0 < \infty$), the structure of the induced spectral distortions are very similar to those shown in Fig. 2 (cf. Tylka and Choueiri [12, Fig. 4(a)]). The primary differences are 1) that the notches are shallower due to an inexact cancellation of the summed signals and 2) that the positions of the notches are rescaled based on the new time delays.

2.3 Localization: Precedence-Effect Errors

For a plane-wave source, the localization information received by each microphone will be identical, so localization of the interpolated signal is likely unchanged. However, for a finite-distance source, the apparent source direction will differ between the perspectives of each microphone. In such cases, interpolation between the microphones

[6] Note that the subscript of a_n refers to the ambisonics channel number (ACN), as described in Appendix A.

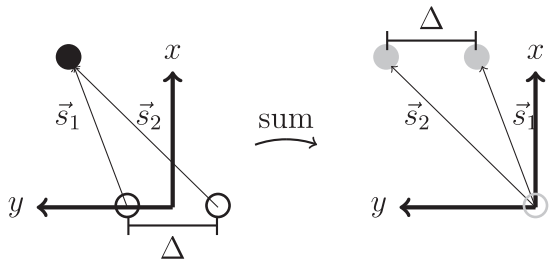


Fig. 3. Diagram of virtual sources (light gray filled circles) effectively created by the weighted average method for a pair of microphones (empty circles) in a sound field with a single source (black filled circle).

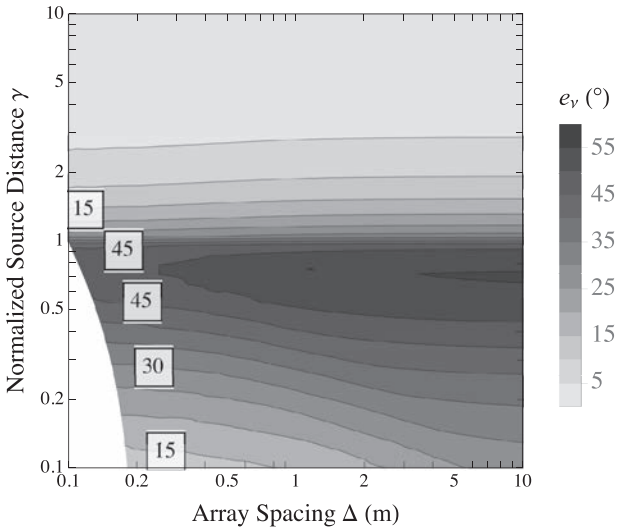


Fig. 4. Predicted localization errors e_v incurred by the weighted average method for various combinations of microphone spacing Δ and normalized source distance $\gamma = s_0/(\Delta/2)$. The plotted errors have been averaged over the entire navigable region (as defined in Sec. 2.1) and all source azimuths ($\varphi_0 \in [0, 90°]$ in increments of 5°). Contour lines are drawn every 5°.

effectively leads to the creation of distinct *virtual sources*, as illustrated in Fig. 3.

To explore the potential localization errors created by these virtual sources, we employ the precedence-effect–based localization model of Stitt et al. [37]. This model takes as inputs the source positions (in this case, $\vec{s}_1$ and $\vec{s}_2$) and signal amplitudes (the interpolation weights w_1 and w_2 from Eq. (1)) and computes a predicted localization vector, $\vec{v}$. Here, we let the free parameter, α, as defined by Stitt et al., take a value of $\alpha = 0.6$, which is somewhat typical for a stimulus signal consisting of both transient and stationary components [37]. We then compute, using the real source position ($\vec{s}_0$) and the desired listener position ($\vec{r}_0$), a localization error, e_v, given by

$$e_v = \cos^{-1}\left(\hat{v} \cdot \hat{s}_0{}'\right), \tag{5}$$

where $\hat{s}_0{}'$ is the direction of the source relative to the listener, found by normalizing the vector $\vec{s}_0{}' = \vec{s}_0 - \vec{r}_0$, and $\| \cdot \|$ denotes the ℓ^2 norm (Euclidean distance) of a vector.

In Fig. 4, we plot these predicted localization errors, averaged over the entire navigable region (as defined in Sec. 2.1) and all source azimuths, for various combinations of source distance s_0 and microphone spacing Δ. Note that we exclude from this contour plot the region in which $s_0 + \Delta/2 < 0.1$ m (i.e., the bottom left corner of Fig. 4), 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 this plot, we first note that localization errors appear primarily dependent on γ but tend to increase with increasing Δ. We also see that, at large microphone spacings ($\Delta > 0.5$ m), localization errors for "slightly" interior sources ($0.5 < \gamma < 1$) become extreme ($e_v > 50°$). Localization errors for distant exterior sources ($\gamma > 2$), however, are uniformly small ($e_v < 10°$). Although not shown here, it can be verified that these localization errors tend to decrease with increasing $\alpha \in [0, 1]$, as $\alpha = 1$ corresponds to purely energy-based localization, with no effect from time-of-arrival delays [37]. Consequently, as $\alpha \to 1$, the dependence of e_v on Δ disappears.

3 PROPOSED NAVIGATION METHOD

In this section, we describe our proposed parametric method for virtual navigation of ambisonics sound fields. Compared to the originally presented version (cf. Tylka and Choueiri [12, Sec. 3]), the method described below has been revised to mitigate spectral coloration (see Sec. 3.3, in particular), but the two are otherwise conceptually and mathematically very similar. Nevertheless, we review the entire method for completeness.

3.1 Source Localization and Microphone Validity

As discussed previously, the ambisonics signals provide a valid description of the captured sound field only in a spherical region around the HOA 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 must first locate any near-field sources. Several existing methods for acoustically localizing near-field sources using ambisonics signals from one or more HOA microphones are discussed by Zheng [31, Ch. 3] and require only knowledge of the positions and orientations of the microphones.

Briefly, such methods often involve taking a short-time Fourier transform of the first-order ambisonics signals and, for each time-frequency bin, calculating the acoustic intensity vector, as given in Merimaa and Pulkki [33, Eq. (11)]. For each HOA microphone, a histogram is generated using the direction of the intensity vector at each time and frequency. The peaks of the histogram indicate source directions, and source positions are determined through triangulation with multiple HOA microphones.

Once the locations of the near-field sources are determined, we compare the distances from each microphone to its nearest source and the distance of that microphone to the desired listening position. Only the

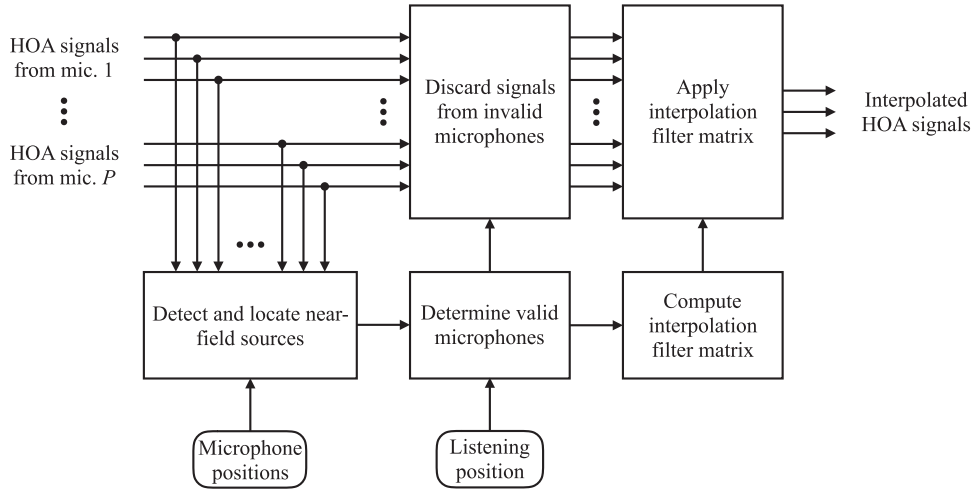


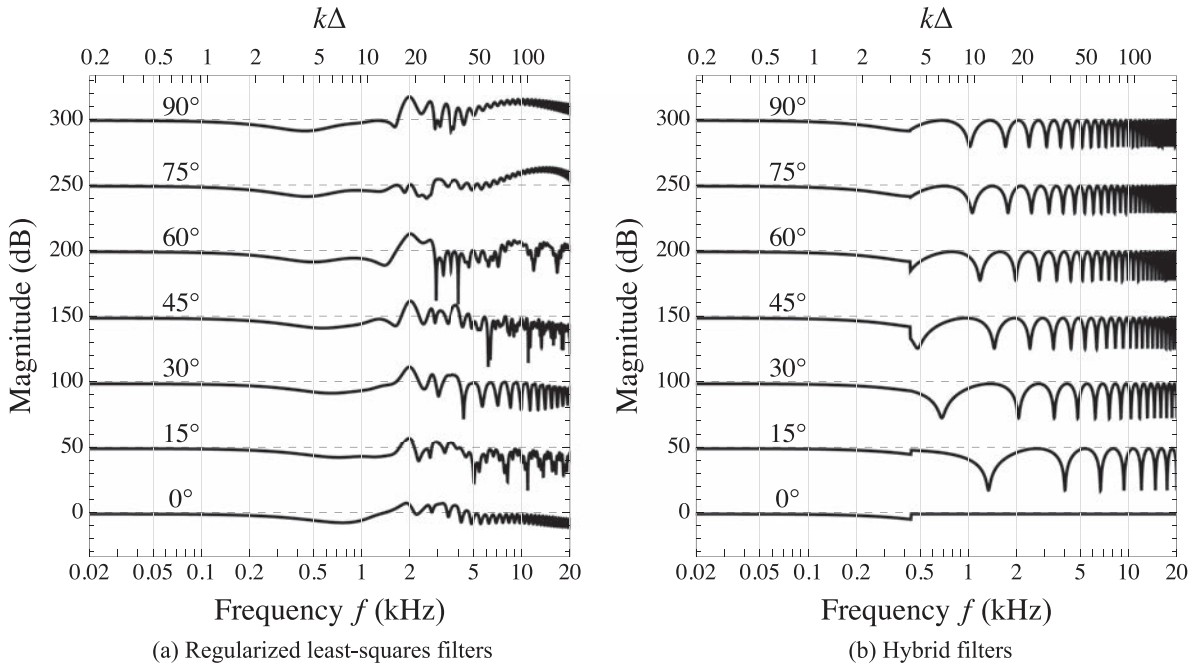
Fig. 5. Flowchart of the proposed method for virtual sound field navigation excluding invalid microphones.

signals from those microphones that are nearer to the listening position than to any near-field source are included in the navigation calculation (i.e., all microphones such that $r_p = \|\vec{r}_0 - \vec{u}_p\| < \|\vec{s}_0 - \vec{u}_p\| = s_p$). A matrix of interpolation filters, described in the following sections, is then computed for and applied to the signals from the remaining "valid" microphones. This procedure is illustrated by the flowchart in Fig. 5.

In this work, we assume that any near-field sound sources can be located accurately and we choose to focus on characterizing the performance of the proposed navigational method under that assumption. Accordingly, we do not concern ourselves with the sensitivity of the proposed method to inaccuracies in the estimated positions of near-field sources. This assumption will be valid in scenarios where the positions of the nearest sound sources are either known a priori or can be accurately obtained (e.g., through physical distance measurements) a posteriori. In other practical applications, however, the validity of this assumption may need to be established experimentally.

3.2 Regularized Least-Squares Interpolation Filters

To compute the interpolation filters, we first pose interpolation as an inverse problem, in which we consider the ambisonics signals at the listening position and, using the



(a) Regularized least-squares filters (b) Hybrid filters

Fig. 6. Magnitude responses caused by the regularized least-squares and hybrid interpolation filters for various source azimuths. The bottom axis shows frequency in kHz while the top axis shows the nondimensional frequency $k\Delta$ for a microphone spacing of $\Delta = 0.5$ 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.

translation coefficient matrices, $\mathbf{T}(k;\vec{r})$, given by Eq. (27) in Appendix B, we write a system of equations simultaneously describing the ambisonics signals at all P-valid HOA microphones (cf. Samarasinghe et al. [15, Sec. III.A], who perform a similar derivation for a 2D sound field). That is, for each frequency, we write

$$\mathbf{M}\cdot\mathbf{x}=\mathbf{y}, \tag{6}$$

where, omitting frequency dependencies,

$$\mathbf{M}(\vec{r}_0)=\begin{bmatrix}\sqrt{w_1}\mathbf{T}(-\vec{r}_1)\\ \sqrt{w_2}\mathbf{T}(-\vec{r}_2)\\ \vdots\\ \sqrt{w_P}\mathbf{T}(-\vec{r}_P)\end{bmatrix},\quad \mathbf{y}=\begin{bmatrix}\sqrt{w_1}\mathbf{b}_1\\ \sqrt{w_2}\mathbf{b}_2\\ \vdots\\ \sqrt{w_P}\mathbf{b}_P\end{bmatrix}, \tag{7}$$

where w_p is the interpolation weight for the p^{th} microphone and $\vec{r}_p$ is the vector from the pth microphone to the listening position, given by $\vec{r}_p=\vec{r}_0-\vec{u}_p$. Again, the interpolation weights may be computed using standard linear or bilinear schemes (see Sec. 2.1 for explicit expressions in the case of a linear configuration of $P=2$ microphones), and their inclusion ensures that the signals from the microphones nearest to the desired listening position are weighted most heavily during the interpolation. Ideally, as $L_{\text{in}}\to\infty$, we should find $\mathbf{x}\to\mathbf{a}$ (the exact ambisonics signals of the sound field at $\vec{r}_0$). In practice, each microphone captures only N_{in} ambisonics signals, so each $\mathbf{b}_p$ is a column vector of length N_{in} and $\mathbf{y}$ is a column vector of length $P\cdot N_{\text{in}}$.

In order to ensure that the system in Eq. (6) is not underdetermined, we define the maximum order for $\mathbf{x}$, given by

$$L_{\max}=\left\lfloor\sqrt{P\cdot N_{\text{in}}}\right\rfloor-1, \tag{8}$$

where $\lfloor\cdot\rfloor$ denotes rounding down to the nearest integer. Therefore, $\mathbf{x}$ is a column vector of length $N_{\max}=(L_{\max}+1)^2$ and each matrix $\mathbf{T}$ in Eq. (7) will have dimensions $N_{\text{in}}\times N_{\max}$ (rows $\times$ columns). Note that by this definition, we will always have $L_{\max}\geqslant L_{\text{in}}$ irrespective of the number of microphones.

Next, we compute the singular value decomposition of $\mathbf{M}$, such that $\mathbf{M}=\mathbf{U}\Sigma\mathbf{V}^*$, where $(\cdot)^*$ represents conjugate-transposition. This allows us to compute a regularized pseudoinverse of $\mathbf{M}$, given by [38, Sec. 5.1]

$$\mathbf{L}=\mathbf{V}\Theta\Sigma^{+}\mathbf{U}^*, \tag{9}$$

where $(\cdot)^+$ represents pseudoinversion (recall that by definition, Σ is diagonal), and Θ is a square, diagonal matrix whose elements are given by

$$\Theta_{nn}=\frac{\sigma_n^2}{\sigma_n^2+\beta}, \tag{10}$$

where σ_n is the n^{th} singular value of $\mathbf{M}$, with $n\in[1,N_{\max}]$. In general, the regularization parameter β may be a function of frequency. Here, we choose the magnitude of a high-shelf filter as the regularization function, given by [39, Sec. 5.2]

$$\beta(k)=\beta_0\left|\frac{G_\pi i\frac{k}{k_0}+1}{i\frac{k}{k_0}+G_\pi}\right|, \tag{11}$$

where G_π is the amplitude of the high-shelf filter and k_0 is its zero-dB crossing, which, for convenience, we take to be the same as that given below in Eq. (16). This choice of regularization function helps to prevent excessive high-frequency amplification in the resulting filter matrix, $\mathbf{L}$, due to the inversion of the translation coefficient matrices, which can become very small at high frequencies. (In our original publication, we had chosen $k_0=1/\Delta$ for simplicity [12, Eq. (17)], but we did not consider cases where $P\neq 2$.) We then let

$$\beta_0=\frac{1}{\sigma_0}\max_n\sigma_n, \tag{12}$$

with some constant $\sigma_0\gg 1$. Note that the singular values (σ_n) of $\mathbf{M}$ are calculated for each frequency, so, in general, β_0 is also frequency-dependent. Here, we choose $G_\pi=10^{1.5}$ (i.e., 30 dB) and $\sigma_0=1{,}000$.

Finally, we obtain an estimate of $\mathbf{a}$, given by

$$\tilde{\mathbf{a}}=\mathbf{L}\cdot\mathbf{y}. \tag{13}$$

Note that, as with the weighted average method, we may choose to drop the higher-order terms in $\tilde{\mathbf{a}}$ such that we keep only up to order L_{out}, where $L_{\text{out}}\leqslant L_{\max}$.

Also note that we can factor out the interpolation weights into a diagonal matrix, such that

$$\tilde{\mathbf{a}}=\left(\mathbf{L}\cdot\begin{bmatrix}\sqrt{w_1}\mathbf{I} & \mathbf{0} & \cdots & \mathbf{0}\\ \mathbf{0} & \sqrt{w_2}\mathbf{I} & \ddots & \vdots\\ \vdots & \ddots & \ddots & \mathbf{0}\\ \mathbf{0} & \cdots & \mathbf{0} & \sqrt{w_P}\mathbf{I}\end{bmatrix}\right)\cdot\begin{bmatrix}\mathbf{b}_1\\ \mathbf{b}_2\\ \vdots\\ \mathbf{b}_P\end{bmatrix}, \tag{14}$$

where $\mathbf{0}$ is an $N_{\text{in}}\times N_{\text{in}}$ matrix of zeros, and $\mathbf{I}$ is the $N_{\text{in}}\times N_{\text{in}}$ identity matrix. For compactness, we let

$$\mathbf{L}_w=\mathbf{L}\cdot\begin{bmatrix}\sqrt{w_1}\mathbf{I} & \mathbf{0} & \cdots & \mathbf{0}\\ \mathbf{0} & \sqrt{w_2}\mathbf{I} & \ddots & \vdots\\ \vdots & \ddots & \ddots & \mathbf{0}\\ \mathbf{0} & \cdots & \mathbf{0} & \sqrt{w_P}\mathbf{I}\end{bmatrix}. \tag{15}$$

3.3 Two-Band Interpolation Filters

As found in a previous analysis [12, Fig. 4(b)], the regularized least-squares interpolation filters derived in the previous section can induce significant spectral distortions at high frequencies, whereas, below some critical frequency, they induce negligible distortions. Consequently, we propose a two-band approach which applies the regularized least-squares interpolation filters at low frequencies ($k<k_0$) and employs the weighted average method at higher frequencies ($k\geqslant k_0$). Here, we let

$$k_0=\begin{cases}\dfrac{1}{r_1}, & \text{for } P=1,\\ \dfrac{\Delta}{r_1 r_2}, & \text{for } P=2,\\ \dfrac{1}{\max_{p\in[1,P]} r_p}, & \text{otherwise,}\end{cases} \tag{16}$$

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

We then rewrite the weighted average calculation, given in Eq. (1), as a matrix equation, such that

$$\tilde{\mathbf{a}} = \begin{bmatrix} w_1\mathbf{I} & w_2\mathbf{I} & \cdots & w_P\mathbf{I} \end{bmatrix} \cdot \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \\ \vdots \\ \mathbf{b}_P \end{bmatrix}, \tag{17}$$

where now $\mathbf{I}$ is the $N_{\text{out}} \times N_{\text{in}}$ identity matrix. Thus, we define the combined interpolation filter matrix as

$$\mathbf{H} = \begin{cases} \mathbf{L}_w, & \text{for } k < k_0, \\ \begin{bmatrix} w_1\mathbf{I} & w_2\mathbf{I} & \cdots & w_P\mathbf{I} \end{bmatrix}, & \text{for } k \geq k_0. \end{cases} \tag{18}$$

Note that if we have only $P = 1$ valid microphone, the high-frequency filter matrix becomes the identity matrix (since $w_1 = 1$ after normalization).

Given the well-established rule of thumb that a sound field is accurately represented up to a distance r provided that $kr \leqslant L_{\text{in}}$ [2], we expect that for $P = 1$, the spectral distortions should be negligible below $kr_1 \approx L_{\text{in}}$. Additionally, for $P = 2$ microphones, we previously found that spectral distortions appeared negligible below $k\Delta \approx 2L_{\text{in}}$ for a listener equidistant from the microphones [12, cf. Fig. 5]. However, these approximations were found to break down due to the near-field compensation filters needed in practice (described in Sec. 4.3). Consequently, here we take a more conservative critical frequency, as in Eq. (16).[7]

3.4 Source Azimuth Dependence

To examine the effective frequency responses induced by translation via the regularized least-squares and hybrid interpolation filters, we simulate a representative far-field scenario in which we let $L_{\text{in}} = 4$, pick $\Delta = 0.5$ m and $s_0 = 2.5$ m (so $\gamma = 10$), vary $\varphi_0 \in [0, 90^\circ]$, and interpolate to $\vec{r}_0 = (0, 0, 0)$. For each source azimuth, we compute the induced frequency response, i.e., the ratio of the zeroth-order interpolated ambisonics signal, $A_0(k)$, to the zeroth-order reference ambisonics signal, $B_0(k)$, that would have been measured at $\vec{r}_0$.

The induced frequency responses are plotted in Fig. 6. For the regularized least-squares interpolation filters (see Fig. 6(a)), we see that the frequency response is largely flat below a critical frequency (cf. Tylka and Choueiri [12, Fig. 4(b)]), whereas above that frequency, the response exhibits significant broadband deviations (e.g., for $\varphi_0 = 75^\circ, 90^\circ$) as well as sporadic notches. For the hybrid filters (see Fig. 6(b)), we see a comb-filtering frequency response above the critical frequency, k_0, given in Eq. (16). Below this frequency, however, the hybrid filter responses exhibit a wide, flat region, rather than the continued comb-filtering response exhibited by the weighted average method (as shown in Fig. 2).

3.5 Practical Implementation

In practice, as the listener traverses the navigable region, the number of valid microphones may change. Consequently, one should crossfade between audio frames to prevent any audible discontinuities caused by a sudden change in the filters. Additionally, it is likely preferable to implement a "crossover" between the low and high-frequency ranges of the combined filter matrix, thereby blending the two filter matrices. Here, however, we take a simple frequency-domain concatenation approach, as indicated in Eq. (18).

4 NUMERICAL SIMULATIONS

Here we again consider the linear array geometry described in Sec. 2.1 and illustrated in Fig. 1. We simulate recording of this sound field for a range of microphone spacings, $\Delta \in [0.1, 10]$ m, and all source positions $s_0 = \gamma\Delta/2$ for $\gamma \in [0.1, 10]$. In each simulation, we vary the source azimuth from $\varphi_0 = 0^\circ$ to 90° in increments of 5° and generate an artificial HOA impulse response at each microphone. We then estimate, using both the proposed and weighted average methods, the microphone's impulse responses at intermediate listener positions from $y_0 = -\Delta/2$ to $\Delta/2$, taken in 20 equal increments. We ultimately compute, for each method, metrics quantifying the spectral coloration (described in Sec. 4.1) and predicted localization error (described in Sec. 4.2) incurred through navigation.

In all simulations, unless stated otherwise, we choose $L_{\text{in}} = 4$ and $L_{\text{out}} = 1$. Linear interpolation weights (explicit expressions for which are given in Sec. 2.1) are attributed to each microphone for each intermediate position. The sampling rate is 48 kHz, and all impulse responses are calculated with 16,384 samples (≈341 ms).

4.1 Coloration Metric

To quantify induced spectral distortions, we first compute the *auditory band spectral error* (ABSE), adapted from Schärer and Lindau [40, Eq. (9)] and given by

$$\eta(f_c) = 10\log_{10}\left(\frac{\int |\Gamma(f; f_c)||\tilde{A}_0(f)|^2 df}{\int |\Gamma(f; f_c)||A_0(f)|^2 df}\right), \tag{19}$$

where A_0 and $\tilde{A}_0$ are the zeroth-order terms of the exact and estimated (respectively) HOA transfer functions, integration is taken over all frequencies, and $\Gamma(f; f_c)$ is a gammatone filter[8] with center frequency f_c for $c \in [1, N_b]$, for a set of ERB-spaced (equivalent rectangular bandwidth) center frequencies [42] spanning the range $f \in [50$ Hz,

[7] Note that in this work we do not explore the case of $P > 2$, so a superior critical frequency likely exists. It may also still be worth pursuing order-dependent critical frequencies for $P = 1, 2$, since, in principle, increasing the expansion order should improve accuracy, but we do not do so here.

[8] In this work, we used the gammatone filters implemented in the large time-frequency analysis toolbox (LTFAT) for MATLAB [41].

21 kHz], as recommended by Boren et al. [43]. We further define the *spectral error*, given by

$$\rho_\eta = \max_c \eta(f_c) - \min_c \eta(f_c), \tag{20}$$

which we found in a previous study to be a strong predictor of the perceived coloration induced through navigation [20].

4.2 Localization Model

Localization is predicted using a recently developed precedence-effect–based localization model, the details of which are provided in an earlier publication [20, Sec. 2.A.i]. This model extends the precedence-effect–based energy vector model of Stitt et al. [37] in order to compute a predicted perceived source localization vector, $\vec{v}$, from an ambisonics impulse response. In contrast, the original model of Stitt et al. only considered loudspeaker gains and positions relative to the listener. We have previously shown our extended model to outperform, in terms of agreement with subjective listening test results, the binaural localization model of Dietz et al. [44]. Furthermore, an advantage of our model over a binaural localization model is that we do not need to first render to binaural, a process which will likely introduce errors that depend on the choice of renderer and head-related transfer function. Consequently, we expect this model to give a reasonable prediction of perceived localization. However, it is worth noting that the performance of the model was optimized for speech signals over a limited range of source positions [20, Sec. 4.A].

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. This information (signal amplitude, plane-wave direction, and time of arrival) for all wavelets is fed into the original energy vector model of Stitt et al. to produce a predicted source localization vector. This operation is performed in ERB-spaced frequency bands, and a weighted average localization direction is computed using signal energies in each band as weights [20, Sec. 2.A.i]. The localization error e_v is then computed using Eq. (5).

4.3 Point-Source Encoding

The ambisonics encoding filters for a point source are given by Eq. (23) in Appendix A. To limit excessive low-frequency amplification when encoding near-field point sources into ambisonics, we apply, to each microphone and at all ambisonics orders $l \in [1, L_{\text{in}}]$, an order-dependent, zero-phase, high-pass Butterworth filter. The frequency response of this filter is given by

$$H_l(f) = 1 - \frac{1}{\sqrt{1 + \left(\frac{f}{f_l}\right)^l}}, \tag{21}$$

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

Note that the frequency at which this near-field amplification occurs increases as the distance between the source and microphone decreases, and the magnitude of the amplification increases with increasing order l [45, Sec. 2.1]. However, the compensation filters given in Eq. (21) are independent of source position, which will lead to excessive low-frequency amplification (due to insufficient compensation) at small source distances and excessive low-frequency attenuation at large source distances. This approach, while inexact, is representative of practical reality since, in general, the source position(s) may be unknown. Indeed, the Eigenmike by mh acoustics uses fixed-frequency compensation filters [46, Sec. 4.3].

5 RESULTS AND DISCUSSION

Spectral errors for each method are shown in Fig. 7. From these plots, we see that, at large spacings ($\Delta > 0.5$ m) with an interior source ($\gamma < 1$), the proposed method achieves significantly (∼4 dB) smaller spectral errors than the weighted average method. This is primarily due to the exclusion of the invalid microphone from the navigation calculation, which prevents any comb-filtering in the proposed method.

Additionally, at small microphone spacings ($\Delta < 0.5$ m) with an exterior source ($\gamma > 1$), the proposed method achieves slightly (∼1 dB) smaller spectral errors than the weighted average method. This is due to the widening (as Δ decreases) of the frequency range over which the regularized least-squares interpolation filters achieve a nearly flat frequency response (see Fig. 6(a)). As specified in Eq. (16), for $P = 2$ microphones, the crossover frequency increases with decreasing Δ, since $r_1 r_2 \propto \Delta^2$. At large microphone spacings ($\Delta > 0.5$ m) with an exterior source ($\gamma > 1$), the proposed and weighted average methods perform comparably.

Localization errors for each method are shown in Fig. 8. From these plots, we see that, at large spacings ($\Delta > 0.5$ m) with an interior source ($\gamma < 1$), the proposed method achieves significantly (i.e., at least 10°) smaller localization errors than the weighted average method. This too is primarily due to the exclusion of the invalid microphone from the navigation calculation, which prevents any corruption of the localization information by the invalid microphone.

From Fig. 8(b), we see that, compared to the weighted average method (see Fig. 8(a)), the proposed method incurs larger localization errors ($e_v > 20°$) for $\Delta < 0.5$ m and $\gamma < 1$ (bottom left corner of the plot). This penalty may not be too limiting, however, as these values of Δ and γ correspond to sources very near to the origin ($s_0 < 0.25$ m), and in any case, microphone spacings of less than 0.5 m may be impractical.

5.1 Effect of Input Order

To further explore the performance of these methods, we plot, in Fig. 9, spectral and localization errors for a source distance of $s_0 = 1$ m and for input ambisonics orders $L_{\text{in}} = 1, 2, \ldots, 6$. Since $L_{\text{out}} = 1$ in all cases, the weighted average method is only plotted for $L_{\text{in}} = L_{\text{out}} = 1$.

From these figures, we immediately see that the errors incurred by the proposed method are virtually identical at all orders. This result is somewhat surprising, as one might

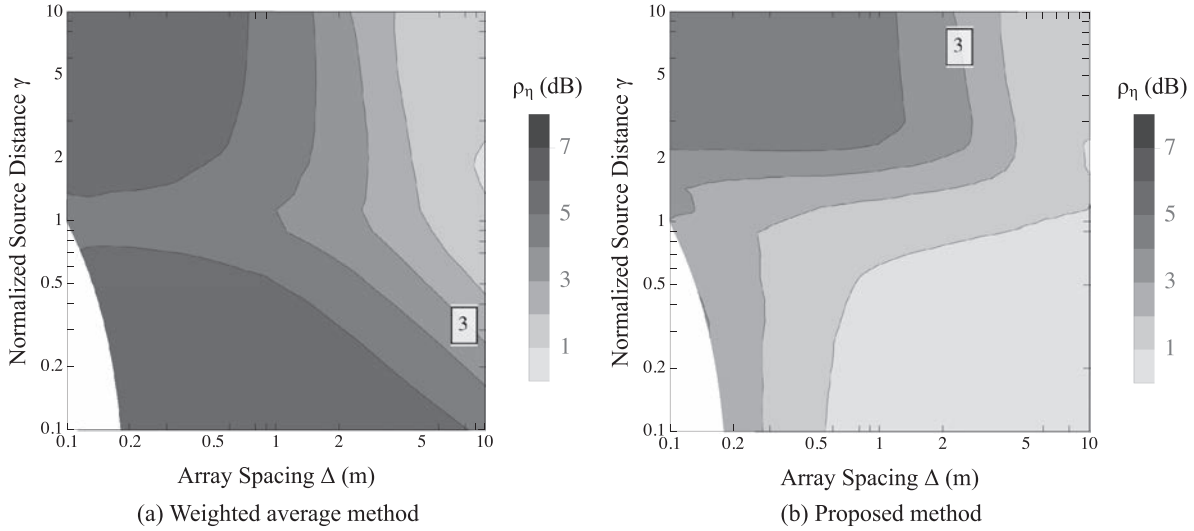


Fig. 7. Spectral errors ρ_η for microphone spacing Δ and normalized source distance γ. Contour lines are drawn every 1 dB.

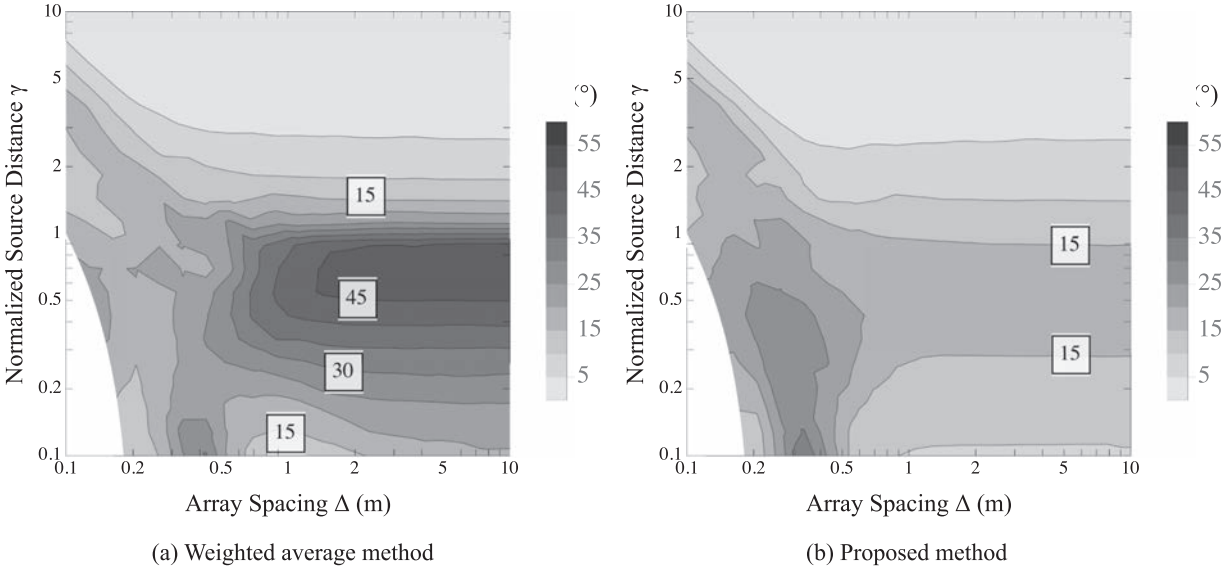


Fig. 8. Predicted localization errors e_v for microphone spacing Δ and normalized source distance γ. Contour lines are drawn every 5°.

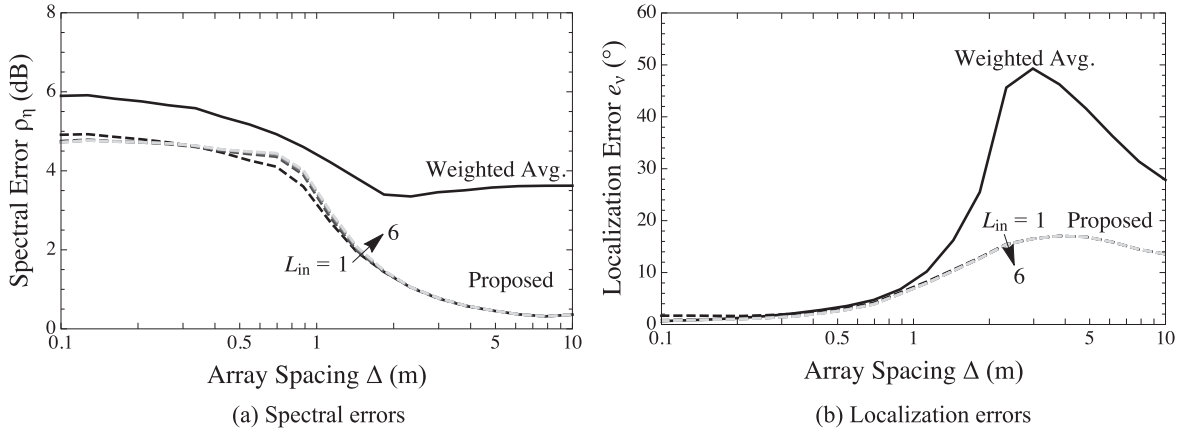


Fig. 9. Spectral errors ρ_η (a) and predicted localization errors e_v (b) for various microphone spacings Δ with a fixed source distance $s_0 = 1$ m. Errors are plotted for the weighted average method (solid curves) and the proposed method (dashed curves). For the proposed method only, six input ambisonics orders are shown: $L_{in} = 1$ (black) to $L_{in} = 6$ (lightest gray).

expect higher-order recordings to produce a more accurate interpolated sound field. Evidently, however, the input ambisonics order does not significantly affect the performance of the proposed method. This insensitivity to input order is largely due to the order-independent choice of critical frequency, k_0, as given in Eq. (16). Potentially, this critical frequency could be optimized to yield improved accuracy as input order increases, at least over a range of array spacings or source positions. This is a topic for further development.

In Fig. 9(a), we see that, for microphone spacings $0.4 \leqslant \Delta \leqslant 1.5$ m, increasing the input order yields increased spectral errors. This behavior can be attributed to the mismatch between the near-field compensation filters (given in Eq. (21)) and the actual low-frequency amplification caused by the near-field effect, as the magnitude of this mismatch increases with order (cf. Daniel [45, Fig. 6]). However, at smaller microphone spacings ($\Delta < 0.3$ m), the opposite effect is observed: increasing the order yields a slight improvement in spectral errors. This is due to the higher-order terms yielding a more accurate estimate of the sound field at the listening position, although this effect is evidently very minor.

In terms of localization, we see that, for $\Delta < 1$ m, the performance of the proposed method is already nearly optimal with $L_{\text{in}} = 1$. This implies that the regularized least-squares interpolation filters do not improve the performance of the proposed method by this metric beyond that achieved by the weighted average method; evidently, the dominant effect of these filters is to decrease spectral errors. For $\Delta > 1$ m, on the other hand, the performance of the proposed method is improved significantly compared to the weighted average method. This demonstrates that any improvement seen by the proposed method in terms of localization must be primarily due to the exclusion of invalid microphones from the interpolation calculation.

Overall, the results shown in Fig. 9 reaffirm our previous finding that the proposed method tends to outperform the weighted average method for interior sources ($\Delta/2 > s_0 = 1$ m) since, for both metrics, the proposed method outperforms the weighted average method at all spacings $\Delta \geqslant 1$ m.

6 EXPERIMENTAL VALIDATION

In order to validate the simulations described in Sec. 4, we replicate a subset of the simulations through acoustical measurements taken in a 3.6 × 2.35 × 2.55-m (length × width × height) anechoic chamber with 8-inch deep (equal to 1/4 wavelength at ∼425 Hz) anechoic foam wedges. We consider three source positions: $\vec{s}_{\text{A}} = (0.35, 0, 0)$ m, $\vec{s}_{\text{B}} = (0.35, 0.35, 0)$ m, and $\vec{s}_{\text{C}} = (0.35, 0.7, 0)$ m. For each source, we measure, up to order $L_{\text{in}} = 4$, ambisonics impulse responses for all microphone positions $\vec{u} = (0, u_y, 0)$ with $u_y = [-0.5, 0.5]$ m in increments of 0.05 m. Here, we use Genelec 8010A [47] loudspeakers for the sources, and the HOA impulse responses are recorded using the Eigenmike by mh Acoustics [48], which comprises an array of 32 capsules flush-mounted on a 4.2-cm–radius rigid sphere. These microphone and source positions are illustrated in Fig. 10.

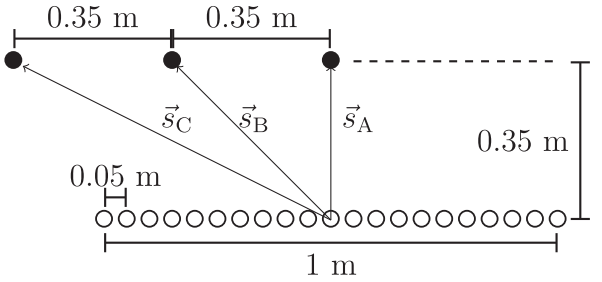


Fig. 10. Diagram of microphone positions (empty circles) and source positions (filled circles).

The ambisonics impulse responses are then equalized by the frequency response of the same source measured by an omnidirectional reference microphone at the same position, thereby compensating for the directivity of the source.

For each microphone spacing $\Delta \in [0.1, 1]$ m (taken in increments of 0.1 m) and each source position, we estimate, using both the weighted average and proposed navigation methods, the ambisonics impulse response each intermediate microphone position. In all cases and at each intermediate position, we compute the "measured" ABSE and localization vector, $\eta'(f_c)$ and $\vec{v}'$, respectively. To better match the measurements, which were taken using the Eigenmike, we modify the near-field compensation filters given by Eq. (21) to use the following corner frequencies: $f_2 = 400$ Hz, $f_3 = 1$ kHz, and $f_4 = 1.8$ kHz (no filters are applied for orders $l = 0, 1$), as specified in the EigenUnits user manual [46, Sec. 4.3].

6.1 Results

Given the simulated and measured ABSE spectra, we first compute the discrepancy, $d_\eta(f_c) = |\eta(f_c) - \eta'(f_c)|$, for each navigation method, source position, microphone spacing, and intermediate microphone position (a total of $1 + (\Delta/0.05)$ distinct positions for each microphone spacing). We then average, in a single operation, these discrepancies over every combination of microphone spacing and *strictly interior* (i.e., $|u_y| < \Delta/2$) intermediate microphone position (note that this is only $(\Delta/0.05) - 1$ positions per spacing).

In Fig. 11, we plot, as a function of frequency, these average discrepancies in ABSE between the simulations and measurements for each navigation method and source position. From this plot, we see that the simulations consistently match, within ∼1 dB, the physical measurements over a frequency range of approximately 150 Hz to 10 kHz. The sharp increase in discrepancy at high frequencies can be explained by spatial aliasing, a well-known effect which we do not currently account for in our simulations (see Rafaely [49], for example) but which could potentially be incorporated in the future with a simple model based on the geometrical arrangement of capsules on the spherical microphone array used in the measurements. The gradual increase in discrepancy at low frequencies, however, is explained by a combination of 1) mismatches between the near-field compensation filters, 2) nonanechoic conditions

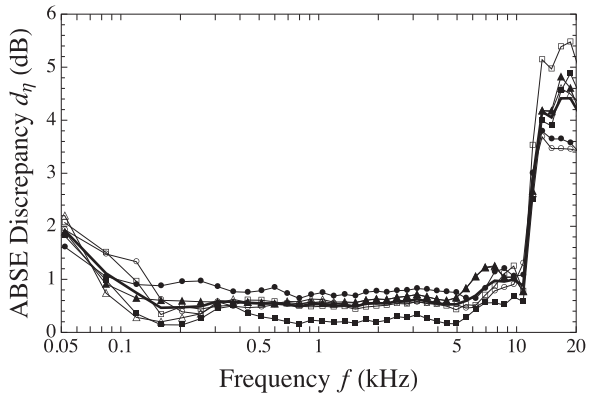


Fig. 11. Average discrepancies in auditory band spectral error (ABSE) spectra between simulations and measurements. Discrepancies are plotted for both the weighted average method (denoted by filled symbols) and the proposed method (empty symbols), as well as for each source: A (△), B (□), and C (◯). For each source and method combination, a thin black line connects the data points, while a thick black line indicates the average over all six curves.

below ∼425 Hz, and 3) low-frequency ambient noise in the measurements.

In Fig. 12(a), we plot, now as a function of microphone spacing, the average discrepancies in ABSE, where two averages are taken: first over all frequencies $f_c \in [0, 10]$ kHz and subsequently over all strictly interior intermediate microphone positions. From this plot, we see that the discrepancy between simulation and measurement is consistently smaller than 1 dB, with a very slight and gradual increase with increasing microphone spacing.

Given the simulated and measured localization directions, we next compute the discrepancy, $d_v = \cos^{-1}\left(\hat{v} \cdot \hat{v}'\right)$, for each navigation method, source position, microphone spacing, and intermediate microphone position. In Fig. 12(b), we plot, as a function of microphone spacing, averages of these discrepancies over all strictly interior intermediate microphone positions. From this plot, we see that the discrepancy between simulation and measurement is consistently smaller than 5°, with an average value of approximately 3.5°, and does not vary significantly with microphone spacing.

Taken together, Figs. 11–12(b) further suggest that the discrepancies between simulations and measurements do not depend significantly on navigational method or source position.

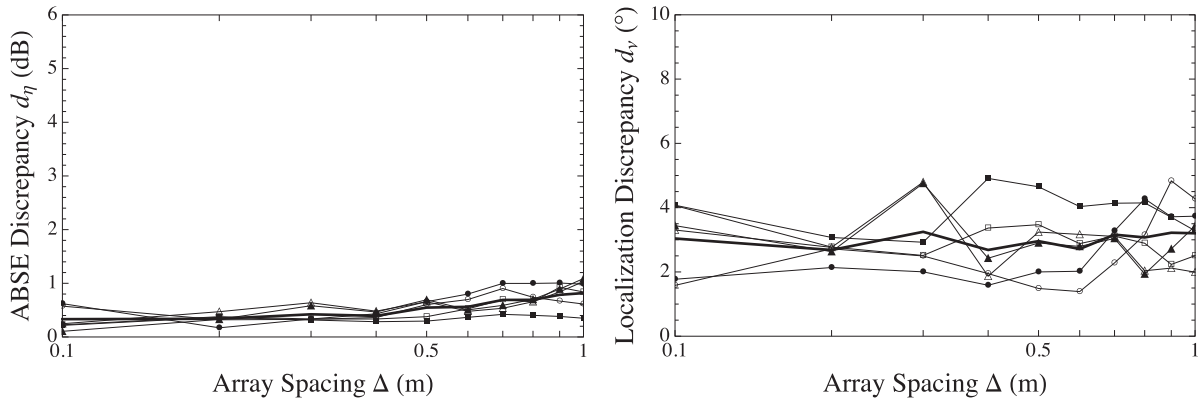
7 SUMMARY AND CONCLUSIONS

In this work, we proposed and characterized an interpolation-based method for virtual navigation, wherein the subset of microphones to be used is parametrically determined to ensure that the region of validity restriction (defined in Sec. 0) is not violated. An existing alternative method, in which navigation is performed by computing a weighted average of the higher-order ambisonics (HOA) signals from each microphone, was shown in Sec. 2 to incur spectral distortions due to comb-filtering and localization errors due to the precedence effect. The proposed method, described in Sec. 3, employs knowledge of the locations of any near-field sources in order to determine which HOA microphones are valid for use in the navigation calculation as a function of the desired listening position. Additionally, at low frequencies, the proposed method applies a matrix of regularized least-squares inverse filters to estimate the ambisonics signals at the listening position, while at high frequencies, the weighted average method is employed.

As described in Sec. 4, we compared, through numerical simulations of simple incident sound fields, the proposed method to the weighted average method. These two methods were evaluated for a linear array geometry (illustrated in Fig. 1) in terms of induced spectral distortions (see Sec. 4.1) and predicted localization errors (see Sec. 4.2). Results show that, for interior sources, the proposed method achieves a significant improvement (in terms of spectral and localization accuracy) over the existing method. In particular, the proposed method yields significantly improved localization errors over the existing method for large microphone spacings (larger than 0.5 m). These improvements primarily result from excluding the invalid microphone, which would otherwise add spectral distortions and corrupt the localization information in the reproduced signals. Additionally, for small microphone spacings (smaller than 0.5 m) and exterior sources, the proposed method achieves slightly smaller spectral errors than does the existing method. This is due to the widening (as microphone spacing decreases) of the frequency range over which the regularized least-squares interpolation filters achieve a nearly flat frequency response (see Fig. 6a).

Results also show that the performance of the proposed method is largely independent of the input ambisonics order (see Sec. 5.1). As this is primarily a consequence of our order-independent choice of critical frequency for the hybrid interpolation filters (see Eq. (16)), future refinements to the proposed method should explore the use of order-dependent critical frequencies in an effort to better leverage the additional information about the sound field contained in the higher-order signals. Ideally, this information could be used to further improve localization accuracy for interior sources and/or mitigate the spectral distortions induced by the proposed method for exterior sources.

Finally, in order to validate our numerical simulations, we conducted a set of acoustical measurements, as described in Sec. 6, taken over a subset of the simulated conditions. Results of these measurements are in good agreement with those of the simulations, indicating that our simulations are indeed representative of reality. In particular, spectral error discrepancies are consistently smaller than 1 dB across all frequencies within approximately 150 Hz to 10 kHz. Additionally, localization direction discrepancies are consistently within 5° (3.5° on average) across all microphone spacings. A more comprehensive validation of our simulation framework could consider alternative navigational methods and span wider ranges of microphone spacings and source positions. However, as expected, the present results suggest that the observed discrepancies (and therefore



(a) ABSE discrepancies (b) Localization discrepancies

Fig. 12. Average discrepancies in ABSE over $f_c \in [0, 10]$ kHz (top panel) and in localization direction (bottom) for each microphone spacing. See Fig. 11 for a description of the lines and symbols used.

the fidelity of the simulations) do not depend significantly on the navigational method, microphone spacing, or source position.

7.1 Future Work

As the present article is primarily a proof of concept of the proposed navigational method to demonstrate its fundamental aspects, future work should include a complementary practical implementation of the method and further experimental validation (e.g., through subjective listening tests) of the results presented here. Such a practical demonstration might also explore the performance of the method under more complicated and realistic conditions, e.g., in sound fields consisting of multiple sources, moving sources, and/or diffuse sound. Additionally, future work might extend the present investigations to consider 2D array configurations, such as an equilateral triangle or a square (cf. Mariette et al. [11] and Bates et al. [28], respectively). It may also be insightful to characterize the performance of these methods in terms of other perceptually relevant attributes (e.g., perceived source width).

8 ACKNOWLEDGMENT

This work was sponsored by the Sony Corporation of America. The authors wish to thank R. Sridhar for fruitful discussions throughout the work, P. Stitt for providing the MATLAB code for the precedence-effect–based energy vector model (available online [50]), and the anonymous reviewers for their feedback.

A.1 RELEVANT AMBISONICS THEORY

Here, we use real-valued orthonormal (N3D) spherical harmonics as given by Zotter [51, Sec. 2.2], and we adopt the ambisonics channel number (ACN) convention [52] 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 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 the sum, also known as a spherical Fourier–Bessel series expansion, is given by [53, Ch. 2]

$$\psi(k, \vec{r}) = \sum_{n=0}^{\infty} 4\pi(-i)^l A_n(k) j_l(kr) Y_n(\hat{r}), \tag{A.22}$$

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, 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 $\vec{s}_0$ are given in the frequency domain by [1, Eq. (10)]

$$A_n(k) = i^{l+1} k h_l(ks_0) Y_n(\hat{s}_0), \tag{A.23}$$

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

B.1 AMBISONICS TRANSLATION

It can be shown that, given ambisonics signals (or, more generally, any spherical Fourier–Bessel expansion coefficients), A_n, for an expansion about the origin, translated ambisonics signals for an expansion about $\vec{r}$ are given by [53, Ch. 3]

$$B_{n'}(k; \vec{r}) = \sum_{n=0}^{N-1} T_{n',n}(k, \vec{r}) A_n(k), \tag{B.24}$$

where $T_{n',n}$ are the so-called *translation coefficients*. Integral forms of these translation coefficients as well as fast recurrence relations for computing them are given by Gumerov and Duraiswami [53, Sec. 3.2] and Zotter [51,

Ch. 3], and were recently distilled and replicated by Tylka and Choueiri [54]. Note that the translated expansion coefficients $B_{n'}$ can be computed to an arbitrary order L', with $N' = (L' + 1)^2$ terms. In matrix form, we can write

$$\mathbf{b}(k) = \mathbf{T}(k; \vec{r}) \cdot \mathbf{a}(k), \tag{B.25}$$

where, omitting dependencies,

$$\mathbf{b} = \begin{bmatrix} B_0 \\ B_1 \\ \vdots \\ B_{N'-1} \end{bmatrix}, \quad \mathbf{a} = \begin{bmatrix} A_0 \\ A_1 \\ \vdots \\ A_{N-1} \end{bmatrix}, \tag{B.26}$$

and

$$\mathbf{T} = \begin{bmatrix} T_{0,0} & T_{0,1} & \cdots & T_{0,N-1} \\ T_{1,0} & T_{1,1} & \cdots & T_{1,N-1} \\ \vdots & \vdots & \ddots & \vdots \\ T_{N'-1,0} & T_{N'-1,1} & \cdots & T_{N'-1,N-1} \end{bmatrix}. \tag{B.27}$$

9 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] D. B. Ward and T. D. Abhayapala, "Reproduction of a Plane-Wave Sound Field Using an Array of Loudspeakers," *IEEE Transactions on Speech and Audio Processing*, vol. 9, no. 6, pp. 697–707 (2001 Sep.), doi: 10.1109/89.943347.

[3] A. Kuntz and R. Rabenstein, "Limitations in the Extrapolation of Wave Fields From Circular Measurements," *Proc. 15th European Signal Processing Conference*, pp. 2331–2335 (2007 Sep.).

[4] 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.

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

[6] 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.

[7] A. Walther and C. Faller, "Linear Simulation of Spaced Microphone Arrays Using B-Format Recordings," presented at the *128th Convention of the Audio Engineering Society* (2010 May), conference paper 7987.

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

[9] N. Mariette and B. F. G. Katz, "SOUNDDELTA—Large Scale, Multi-User Audio Augmented Reality," *Proc. EAA Symposium on Auralization*, pp. 37–42 (2009 Jan.).

[10] A. Southern, J. Wells, and D. Murphy, "Rendering Walk-Through Auralisations Using Wave-Based Acoustical Models," *17th European Signal Processing Conference*, pp. 715–719 (2009 Aug.).

[11] N. Mariette, B. F. G. Katz, K. Boussetta, and O. Guillerminet, "SoundDelta: A Study of Audio Augmented Reality Using WiFi-Distributed Ambisonic Cell Rendering," presented at the *128th Convention of the Audio Engineering Society* (2010 May), convention paper 8123.

[12] 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.

[13] 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 Audio Engineering Society Convention* (2019 Mar.), convention paper 10141.

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

[15] 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, no. 3, pp. 647–658 (2014 Jan.), doi:10.1109/TASLP.2014.2300341.

[16] C. Fan, S. M. A. Salehin, and T. D. Abhayapala, "Practical Implementation and Analysis of Spatial Soundfield Capture by Higher Order Microphones," *Proc. 2014 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA)*, pp. 1–8 (2014 Dec.).

[17] H. Chen, T. D. Abhayapala, and W. Zhang, "3D Sound Field Analysis Using Circular Higher-Order Microphone Array," *Proc. 23rd European Signal Processing Conference (EUSIPCO)*, pp. 1153–1157 (2015 Aug. -Sep.).

[18] P. Samarasinghe, T. Abhayapala, M. Poletti, and T. Betlehem, "An Efficient Parameterization of the Room Transfer Function," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 23, no. 12, pp. 2217–2227 (2015 Aug.), doi:10.1109/TASLP.2015.2475173.

[19] N. Ueno, S. Koyama, and H. Saruwatari, "Sound Field Recording Using Distributed Microphones Based on Harmonic Analysis of Infinite Order," *IEEE Signal Processing Letters*, vol. 25, no. 1, pp. 135–139 (2018 Jan.), doi:10.1109/LSP.2017.2775242.

[20] J. G. Tylka and E. Y. Choueiri, "Models for Evaluating Navigational Techniques for Higher-Order Ambison-

ics," *Proc. Mtgs. Acoust.*, vol. 30, no. 1, p. 050009 (2017 Oct.), doi:10.1121/2.0000625.

[21] 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 Jun.), doi:10.1121/1.5041742.

[22] C. Masterson, G. Kearney, and F. Boland, "Acoustic Impulse Response Interpolation for Multichannel Systems Using Dynamic Time Warping," presented at the *AES 35th International Conference: Audio for Games* (2009 Feb.), conference paper 34.

[23] G. Kearney, C. Masterson, S. Adams, and F. Boland, "Dynamic Time Warping for Acoustic Response Interpolation: Possibilities and Limitations," *Proc. 17th European Signal Processing Conference*, pp. 705–709 (2009 Aug.).

[24] V. Garcia-Gomez and J. J. Lopez, "Binaural Room Impulse Responses Interpolation for Multimedia Real-Time Applications," presented at the *144th Convention of the Audio Engineering Society* (2018 May), convention paper 9962.

[25] K. Brandenburg, E. Cano, F. Klein, T. Kóllmer, H. Lukashevich, A. Neidhardt, U. Sloma, and S. Werner, "Plausible Augmentation of Auditory Scenes Using Dynamic Binaural Synthesis for Personalized Auditory Realities," presented at the *2018 AES International Conference on Audio for Virtual and Augmented Reality* (2018 Aug.), conference paper P8-3.

[26] S. Emura, "Sound Field Estimation Using Two Spherical Microphone Arrays," presented at the *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 101–105 (2017 Mar.), doi:10.1109/ICASSP.2017.7952126.

[27] N. Epain, C. Jin, and A. Van Schaik, "The Application of Compressive Sampling to the Analysis and Synthesis of Spatial Sound Fields," presented at the *127th Convention of the Audio Engineering Society* (2009 Oct.), convention paper 7857.

[28] E. Bates, H. O'Dwyer, K.-P. Flachsbarth, and F. M. Boland, "A Recording Technique for 6 Degrees of Freedom VR," presented at the *144th Convention of the Audio Engineering Society* (2018 May), convention paper 10022.

[29] H. Lee, "A New Multichannel Microphone Technique for Effective Perspective Control," presented at the *130th Convention of the Audio Engineering Society* (2011 May), convention paper 8337.

[30] H. Lee, "Subjective Evaluations of Perspective Control Microphone Array (PCMA)," presented at the *132nd Convention of the Audio Engineering Society* (2012 Apr.), convention paper 8625.

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

[32] 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.), doi:10.1109/TASL.2013.2280210.

[33] 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 Dec.).

[34] C. Schörkhuber, M. Zaunschirm, and I. M. Zmölnig, "WiLMA—A Wireless Large-Scale Microphone Array," presented at the *Linux Audio Conference 2014* (2014 Jan.).

[35] C. Schörkhuber, P. Hack, M. Zaunschirm, F. Zotter, and A. Sontacchi, "Localization of Multiple Acoustic Sources with a Distributed Array of Unsynchronized First-Order Ambisonics Microphones," presented at the *6th Congress of Alps-Adria Acoustics Assosiation* (2014 Oct.).

[36] K. Wakayama, J. Trevino, H. Takada, S. Sakamoto, and Y. Suzuki, "Extended Sound Field Recording Using Position Information of Directional Sound Sources," *Proc. 2017 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pp. 185–189 (2017 Oct.).

[37] 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), doi:10.17743/jaes.2016.0008.

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

[39] U. Zölzer, *Digital Audio Signal Processing*, 2nd ed. (Wiley, Hoboken, New Jersey, 2008).

[40] 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.

[41] 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].

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

[43] B. Boren, M. Geronazzo, F. Brinkmann, and E. Choueiri, "Coloration Metrics for Headphone Equalization," *Proc. 21st International Conference on Auditory Display*, pp. 29–34 (2015 Jul.).

[44] M. Dietz, S. D. Ewert, and V. Hohmann, "Auditory Model Based Direction Estimation of Concurrent Speakers From Binaural Signals," *Speech Communication*, vol. 53, no. 5, pp. 592–605 (2011 May), doi:10.1016/j.specom.2010.05.006.

[45] 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 Process-*

ing in Audio Recording and Reproduction (2003 May), conference paper 16.

[46] *EigenUnits VST Plugins for macOS and Windows*, mh acoustics, LLC, version 2, Rev. A (Summit, New Jersey, 2018).

[47] Genelec, Inc., "8010A Studio Monitor," https://www.genelec.com/8010, 2014 [Online; accessed 5-June-2019].

[48] mh acoustics, LLC, "Eigenmike® Microphone," https://www.mhacoustics.com/products#eigenmike1, 2014 [Online; accessed 16-February-2019].

[49] B. Rafaely, "Analysis and Design of Spherical Microphone Arrays," *IEEE Transactions on Speech and Audio Processing*, vol. 13, no. 1, pp. 135–143 (2005 Jan.), doi:10.1109/TSA.2004.839244.

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

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

[52] C. Nachbar, F. Zotter, E. Deleflie, and A. Sontacchi, "ambiX—A Suggested Ambisonics Format," *Proc. 3rd Ambisonics Symposium* (2011 Jun.).

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

[54] 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 Mar.).

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.