Audio Engineering Society

Conference Paper

Presented at the Conference on
Audio for Virtual and Augmented Reality
2016 Sept 30 – Oct 1, Los Angeles, CA, USA

Soundfield Navigation using an Array of Higher-Order Ambisonics Microphones

Joseph G. Tylka and Edgar Y. Choueiri

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

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

ABSTRACT

A method is presented for soundfield navigation through estimation of the spherical harmonic coefficients (i.e., the higher-order ambisonics signals) of a soundfield at a position within an array of two or more ambisonics microphones. An existing method based on blind source separation is known to suffer from audible artifacts, while an alternative method, in which a weighted average of the ambisonics signals from each microphone is computed, is shown to necessarily introduce comb-filtering and degrade localization for off-center sources. The proposed method entails computing a regularized least-squares estimate of the soundfield at the listening position using the signals from the nearest microphones, excluding those that are nearer to a source than to the listening position. Simulated frequency responses and predicted localization errors suggest that, for interpolation between a pair of microphones, the proposed method achieves both accurate localization and minimal spectral coloration when the product of angular wavenumber and microphone spacing is less than twice the input expansion order. It is also demonstrated that failure to exclude from the calculation those microphones that are nearer to a source than to the listening position can significantly degrade localization accuracy.

1 Introduction

Virtual navigation of three-dimensional higher-order ambisonics soundfields (i.e., soundfields that have been decomposed into spherical harmonics) enables a listener to explore an acoustic space and experience a spatially-accurate perception of the soundfield. Applications of soundfield navigation may be found in virtual-reality reproductions of real-world spaces. For example, to reproduce an orchestral performance in virtual reality, navigation of an acoustic recording of the performance is likely to yield superior spatial and tonal fidelity compared to that produced through acoustic simulation of the performance. Navigation of acoustic recordings may also be preferable when reproducing real-world spaces for which computer modeling of complex wave-phenomena and room characteristics may be too computationally intensive for real-time playback and interaction.

A well-known limitation of the higher-order ambisonics (HOA) framework is that a finite-order expansion of a soundfield yields only an approximation to that

soundfield, the accuracy of which decreases with increasing frequency and distance from the expansion center [1], so the prospect of navigating such a soundfield is inherently limited. Indeed, existing techniques for soundfield navigation using a single HOA microphone[1] have been shown to introduce coloration [2] and degrade localization [3, 4] as the listener navigates farther away from the expansion center. Furthermore, the region over which the expansion is valid is limited by the nearest sound source to the expansion center, so near-field sources pose a particularly limiting problem to navigation.

However, employing an array of HOA microphones (which are themselves arrays of microphone capsules) throughout the soundfield (or, equivalently, sampling a synthetic soundfield at discrete positions) not only provides a more accurate description of the soundfield at any intermediate position between the microphones, but also enables navigation near to and around sound sources. In the following section, we review previous methods for soundfield navigation that employ multiple HOA microphones and discuss their deficiencies.

1.1 Previous Work

Zheng [5] developed a method of soundfield navigation in which discrete sound sources are first identified, localized, and isolated through a method known as "collaborative blind source separation" [5, section 3.3]. These discrete sources are then treated as sound "objects," which may be artificially moved relative to the listener to emulate navigation. However, this method is only ideal for soundfields 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). Furthermore, even in ideal situations, the source separation technique employed in the time-frequency domain (i.e., short-time Fourier transform domain) often results in a degradation of sound quality due to the introduction of audible artifacts [5, section 5.3].

An alternative technique is to compute a weighted average of the ambisonics signals, which has been implemented in the context of interpolating ambisonics room impulse responses (RIRs) to enable real-time navigable auralizations of acoustic spaces [6]. However, if a sound source is nearer to one microphone than to another, this technique will necessarily produce two copies of that source's signal, separated by a finite time delay, yielding a comb-filtering-like effect and potentially skewing localization towards the direction of the source from the position of the nearest microphone due to the precedence effect. Furthermore, in this method, even those microphones that are nearer to a source than to the desired listening position would be included in the calculation, likely degrading the accuracy of the estimated soundfield at the listening position.

[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 capture ambisonics signals.

1.2 Objectives and Approach

It is the objective of the present work to develop a method of interpolating between spatially-distinct recordings of a given soundfield with minimal spectral coloration and without the introduction of audible artifacts. To these ends, we develop a method of interpolation which employs a matrix of linear filters that are designed with frequency-dependent regularization to limit spectral coloration.

Additionally, the proposed method should achieve accurate sound localization even in the vicinity of near-field sources. To this end, we propose a method of excluding any microphones which are nearer to a sound source than to the desired listening position, thereby ensuring that all microphones used in the calculation provide valid descriptions of the soundfield at the listening position, and are therefore suitable for interpolation.

1.3 Paper Overview

In Section 2, we review relevant concepts from acoustics and ambisonics theory. Next, in Section 3, we review an existing method of interpolation between HOA microphones and present the proposed method. In Section 4, we describe the numerical simulations conducted to evaluate these methods. We then present and discuss the results of these simulations in Section 5, and draw conclusions in Section 6.

2 Review of Acoustical Theory

Here, we adopt a spherical coordinate system commonly used in higher-order ambisonics, in which r is the (nonnegative) radial distance from the origin, $\theta \in [-\pi/2, \pi/2]$ is the elevation angle above the horizontal (x-y) plane, and $\phi \in [0, 2\pi)$ is the azimuthal

angle around the vertical (z) axis, with $\phi = 0$ corresponding to the $+x$-axis and $\phi = \pi/2$ to the $+y$ axis. For a position vector $\vec{r} = (x, y, z)$, we denote unit vectors by $\hat{r} = \vec{r}/r$.

We define the *acoustic potential field* ψ as the Fourier transform of the acoustic pressure field, such that, in a source-free region (i.e., under free-field conditions), the acoustic potential field satisfies the homogeneous Helmholtz equation,

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

where ∇^2 is the Laplace operator and k is the angular wavenumber.

Here, we use real-valued spherical harmonics as given by Zotter [7, section 2.2] and we adopt the ambisonics channel number (ACN) convention [8] 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. Regular (i.e., not singular) solutions to the Helmholtz equation are given by $j_l(kr)Y_n(\hat{r})$, where j_l is the spherical Bessel function of order l. These solutions are only valid under free-field conditions, and can be used to describe the acoustic potential in an interior region, that is, for $r < R$, where R is a finite distance. So that the region remains source-free, R is typically taken to be the distance of the nearest source to the origin.

Provided these restrictions are met, any acoustic potential field can be written as an infinite sum of regular solutions, known as a spherical Fourier-Bessel series expansion, given by [9, chapter 2]

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

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

2.1 Plane-Wave Decomposition

Given spherical Fourier-Bessel expansion coefficients, A_n, the so-called *signature function*, μ, in the direction $\hat{v}_q$ is given by [9, section 2.3.3]

$$\mu(k, \hat{v}_q) = \sum_{n=0}^{N-1} A_n(k) Y_n(\hat{v}_q). \tag{3}$$

The signature function represents the coefficients of a plane-wave decomposition of the soundfield, such that the potential field can be reconstructed using a finite number of plane-waves, given by

$$\psi(k, \vec{r}) = \sum_{q=1}^{Q} w_q \mu(k, \hat{v}_q) e^{ik\hat{v}_q \cdot \vec{r}}, \tag{4}$$

where w_q is the quadrature weight of the q^{th} plane-wave term and is dependent on the chosen grid of directions.

2.2 Ambisonics Translation

It can be shown that, given spherical Fourier-Bessel expansion coefficients, A_n, for an expansion about the origin, translated expansion coefficients for an expansion about $\vec{d}$ are given by [9, chapter 3]

$$B_{n'}(k; \vec{d}) = \sum_{n=0}^{N-1} \Gamma_{n',n}(k, \vec{d}) A_n(k), \tag{5}$$

where $\Gamma_{n',n}$ 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 [9, section 3.2] and Zotter [7, chapter 3]. 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

$$\mathbf{b}(k) = \mathbf{T}(k; \vec{d}) \cdot \mathbf{a}(k), \tag{6}$$

where, omitting dependencies,

$$\mathbf{b} = \begin{bmatrix} B_0 \\ B_1 \\ \vdots \\ B_{N'-1} \end{bmatrix}, \quad \mathbf{a} = \begin{bmatrix} A_0 \\ A_1 \\ \vdots \\ A_{N-1} \end{bmatrix}, \tag{7}$$

and

$$\mathbf{T} = \begin{bmatrix} \Gamma_{0,0} & \Gamma_{0,1} & \cdots & \Gamma_{0,N-1} \\ \Gamma_{1,0} & \Gamma_{1,1} & \cdots & \Gamma_{1,N-1} \\ \vdots & \vdots & \ddots & \vdots \\ \Gamma_{N'-1,0} & \Gamma_{N'-1,1} & \cdots & \Gamma_{N'-1,N-1} \end{bmatrix}. \tag{8}$$

2.3 The Energy Vector

To predict the localization of sound in multichannel playback systems, Gerzon [10] defines two localization metrics: the velocity and energy vectors. The velocity vector is used to predict localization due to interaural time differences (ITD) at low frequencies (< 700 Hz), while the energy vector is used to predict localization due to interaural level differences (ILD) at higher frequencies (500 Hz – 5 kHz) and is given by [10]

$$\vec{r}_{\mathrm{E}}(k) = \frac{\sum_{q=1}^{Q} |G_q(k)|^2 \hat{v}_q}{\sum_{q=1}^{Q} |G_q(k)|^2}, \tag{9}$$

where G_q is the complex-valued, frequency-dependent "gain" of the q^{th} source, and $\hat{v}_q$ points in the direction of that source from the origin. The direction of the resulting vector indicates the expected localization direction and its magnitude indicates the quality of the localization. Ideally, the vector should have a magnitude equal to unity and point in the direction of the virtual source.

3 Interpolation Methods

In this section, we generally describe two interpolation methods, considering an array of P HOA microphones, where the p^{th} microphone is located at $\vec{u}_p$ and its captured ambisonics signals are denoted $\mathbf{b}_p$. Both techniques aim to approximate the exact expansion coefficients, $\mathbf{a}$, at the listening position $\vec{r}_0$.

3.1 Weighted Average

In the interpolation method proposed by Southern et al. [6], a weighted sum of the captured ambisonics signals is computed to obtain an estimate of the ambisonics signals at the listening position, given by

$$\tilde{\mathbf{a}} = \sum_{p=1}^{P} w_p \mathbf{b}_p, \tag{10}$$

where the weights are normalized such that

$$\sum_{p=1}^{P} w_p = 1. \tag{11}$$

Depending on the placement of the microphones in the soundfield, the weights w_p may be computed using standard linear or bilinear schemes, for example.

3.2 Regularized Least-Squares

We pose the proposed interpolation method as an inverse problem, where we consider the expansion coefficients at the listening position and, using the translation coefficient matrices given by Eq. (8), we write a system of equations simultaneously describing the ambisonics signals at all HOA microphones in the array. For each frequency, we write

$$\mathbf{M} \cdot \mathbf{x} = \mathbf{y}, \tag{12}$$

where, omitting frequency dependencies,

$$\mathbf{M} = \begin{bmatrix} \sqrt{w_1}\mathbf{T}(-\vec{d}_1) \\ \sqrt{w_2}\mathbf{T}(-\vec{d}_2) \\ \vdots \\ \sqrt{w_P}\mathbf{T}(-\vec{d}_P) \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} \sqrt{w_1}\mathbf{b}_1 \\ \sqrt{w_2}\mathbf{b}_2 \\ \vdots \\ \sqrt{w_P}\mathbf{b}_P \end{bmatrix}, \tag{13}$$

and $\vec{d}_p$ is the vector from the p^{th} microphone to the listening position, given by $\vec{d}_p = \vec{r}_0 - \vec{u}_p$. Ideally, for infinite-order expansions, $\mathbf{x} = \mathbf{a}$. In practice, each microphone captures ambisonics signals up to order L_{in}, so for all microphones, $\mathbf{b}_p$ is a column-vector of length N_{in} and, consequently, $\mathbf{y}$ is a column-vector of length $P \cdot N_{\text{in}}$.

In order to ensure that the system in Eq. (12) is not under-determined, we compute the maximum order for $\mathbf{x}$, given by

$$L_{\text{max}} = \left\lfloor \sqrt{P \cdot N_{\text{in}}} \right\rfloor - 1, \tag{14}$$

where $\lfloor \cdot \rfloor$ denotes rounding down to the nearest integer (i.e., taking the floor of the argument). Therefore, $\mathbf{x}$ is a column-vector of length N_{max} and each translation coefficient matrix $\mathbf{T}$ in Eq. (13) will have dimensions $N_{\text{in}} \times N_{\text{max}}$ (rows $\times$ columns).

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 the regularized pseudoinverse of $\mathbf{M}$, given by

$$\mathbf{L} = \mathbf{V}\Theta\Sigma^{+}\mathbf{U}^*, \tag{15}$$

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

$$\Theta_{ii} = \frac{\sigma_i^2}{\sigma_i^2 + \beta}, \tag{16}$$

where σ_i is the i^{th} singular value of $\mathbf{M}$. 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

$$\beta(k) = \beta_0 \left| \frac{G_\pi ik\Delta + 1}{ik\Delta + G_\pi} \right|, \tag{17}$$

where G_π determines the amplitude of the high-shelf filter, Δ is the microphone spacing in meters, and

$$\beta_0 = \max_i \frac{\sigma_i}{\gamma}, \tag{18}$$

with $\gamma \gg 1$. Note that the singular values, σ_i, 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 $\gamma = 1000$.

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

$$\tilde{\mathbf{a}} = \mathbf{L} \cdot \mathbf{y}. \tag{19}$$

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

3.3 Microphone Validity

As discussed previously, the spherical Fourier-Bessel expansion is a valid description of the captured soundfield 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 localize any near-field sources. Several existing methods for localizing near-field sources using ambisonics signals from one or more HOA microphones are discussed by Zheng [5, chapter 3], and require only knowledge of the microphones' positions and orientations.

Briefly, such methods often involve calculating the frequency-dependent acoustic intensity vector from the first-order ambisonics signals. For each HOA microphone, a histogram is generated using the direction of the intensity vector at each frequency. The peaks of the histogram indicate source directions, and source positions are determined through triangulation with multiple HOA microphones. Note that rather than localizing sources in order to isolate their emitted signals, the present method only requires determining their locations.

Once any near-field sources are identified, we compare the distances from each microphone to its nearest source and the distance of that microphone to the desired listening position. Only those microphones that are nearer to the listening position than to any near-field source are included in the interpolation calculation. The interpolation weights of those microphones that are excluded from a calculation are then set to zero and the remaining weights are re-normalized such that Eq. (11) holds. This procedure is illustrated by the flowchart in Fig. 1.

4 Objective Evaluation

The first simulation we conduct is for the simplest case of the soundfield radiated by a point-source and recorded by two HOA microphones. The microphones are placed on the y-axis equidistant from the origin and separated by a distance Δ. The source is placed in the x-y plane at $\vec{r}_{\text{S}} = (r_{\text{S}} \cos\phi, r_{\text{S}} \sin\phi, 0)$ m, for various azimuths ϕ and a fixed source distance of $r_{\text{S}} = 1$ m. This geometry is depicted in Fig. 2. The ambisonics signals are captured up to order L_{in} by each microphone and then interpolated, using both the proposed method and the weighted average method, to the midpoint (i.e., the origin), yielding an estimate of the soundfield up to order L_{out}. We evaluate each method in terms of frequency response and predicted localization accuracy (to be discussed in Section 4.1) at the interpolated position.

The second simulation we conduct is the soundfield radiated by a point-source and recorded by four HOA microphones. The microphones are placed equidistant from the origin and arranged in a square at the coordinates $(\pm 0.5\Delta, \pm 0.5\Delta, 0)$. We consider two source positions: (a) the source placed at $(0.75\Delta, 0, 0)$, and (b) at $(0.75\Delta, 0.75\Delta, 0)$. This geometry is depicted in Fig. 3. The ambisonics signals captured by each microphone are then interpolated, using the proposed method, to the midpoint (i.e., the origin) under two conditions: 1) using all four microphones for interpolation, and 2) using only the valid microphones (see Section 3.3). For source position (a), the two microphones at $(0.5\Delta, \pm 0.5\Delta, 0)$ are invalid, while for source position (b), only the microphone at $(0.5\Delta, 0.5\Delta, 0)$ is invalid. We evaluate the proposed method for each source position and under each condition in terms of predicted localization accuracy at the interpolated position.

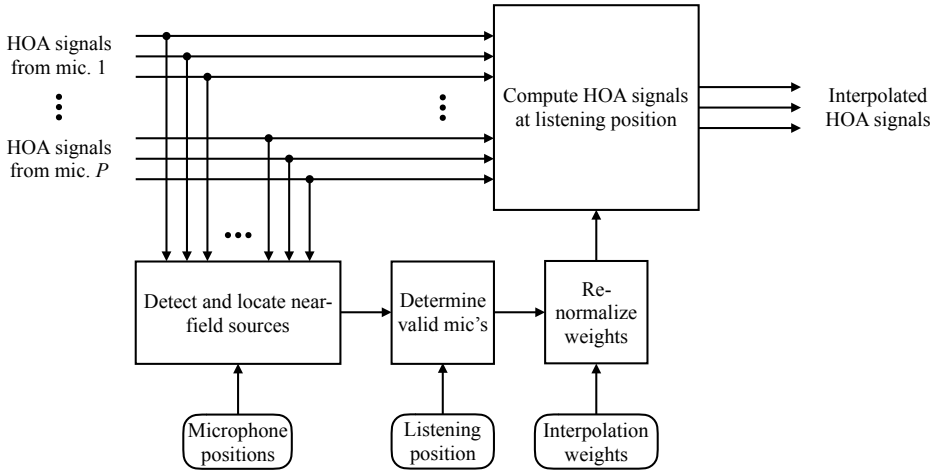


Fig. 1: Flowchart of the proposed method for excluding invalid microphones.

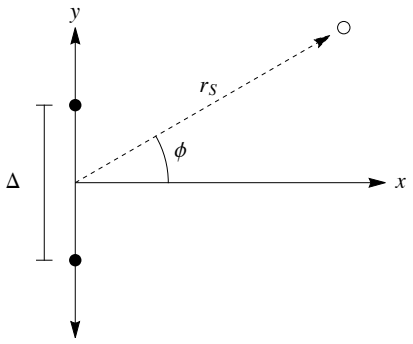


Fig. 2: Diagram of source and microphone positions used in the first simulation. Microphone positions are indicated by filled circles, while the source position is indicated by the empty circle.

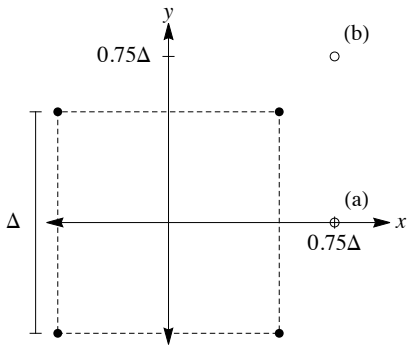


Fig. 3: Diagram of source and microphone positions (indicated as in Fig. 2) used in the second simulation.

In all simulations, unless stated otherwise, the expansion order for all HOA microphones is $L_{\text{in}} = 4$, and the expansion order of the interpolated ambisonics signals is $L_{\text{out}} = 1$. Equal interpolation weights are attributed to each microphone, since in all simulations we are interpolating to the exact midpoint of the array. The sampling rate is 48 kHz and all impulse responses are calculated with 2048 samples (≈ 43 ms).

4.1 Localization Prediction

Localization is predicted using a precedence-effect based localization model [11] that was derived as an extension to the energy vector (see Section 2.3). In order to employ this model, we first convert, via Eq. (3), the interpolated ambisonics signals into a set of Q impulse responses for a specified grid of plane-wave directions. Here, we use $Q = 25$ terms arranged on Fliege nodes [12] and use the corresponding quadrature weights.[2] As this discrete plane-wave soundfield would generally be rendered according to Eq. (4), the impulse response for each plane-wave term is given by the inverse Fourier transform of $w_q \mu(k, \hat{v}_q)$.

We then identify and isolate temporally-distinct impulse response "wavelets." To do this, we first apply a 4$^{\text{th}}$-order Butterworth high-pass filter with a cut-off

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

frequency of 500 Hz to all impulse responses in the set and compute the largest peak (i.e., the maximum of the absolute value) in the set. Then, for each impulse response, we take the absolute value and search for peaks whose amplitude is at least 12.6% (-18 dB) of the largest peak in the set. If no peaks are found in a given impulse response, then that response in its entirety is treated as a wavelet. If at least one peak is found, then, for each peak, we apply a Tukey window beginning 1 ms before the peak and ending either 1 ms after the peak, or at the position of the following peak, whichever yields a larger window length. Both the cosine fade-in and fade-out of the Tukey window are 1 ms in duration. In this way, a single impulse response may be split into several wavelets. For each wavelet, we then apply a -18 dB (now relative to the peak of the wavelet) threshold to determine the time-delay of the onset.

Each wavelet is then treated as a distinct source in the precedence-effect-based energy vector model, wherein wavelets extracted from the same impulse response originate from the same direction, but at different times, given by their onset times. Each wavelet is then transformed to the frequency domain via FFT, yielding frequency-dependent gains for each source. These gains are then fed into the model, yielding a frequency-dependent predicted localization vector $\vec{r}_{\mathrm{PE}}(k)$, and the localization error ε is given by

$$\varepsilon(k) = \cos^{-1}\left(\hat{r}_{\mathrm{PE}}(k) \cdot \hat{r}_{\mathrm{S}}\right). \tag{20}$$

5 Results

Figure 4 shows frequency responses at the listening position for various source azimuths, with a fixed microphone spacing of $\Delta = 0.5$ m. Frequency responses for the weighted average method are shown in Fig. 4a, while those for the proposed method are shown in Fig. 4b. We observe that the proposed method exhibits very little spectral coloration for low frequencies ($\leq$ 1 kHz), corresponding to $k\Delta \sim 10$. The weighted average method, however, introduces comb-filtering for sources at non-zero azimuths, whose lowest-frequency notch occurs at $k\Delta \approx \pi/\sin\phi$, where equality is achieved when the source distance is far compared to the microphone spacing, $r_{\mathrm{S}} \gg \Delta$.

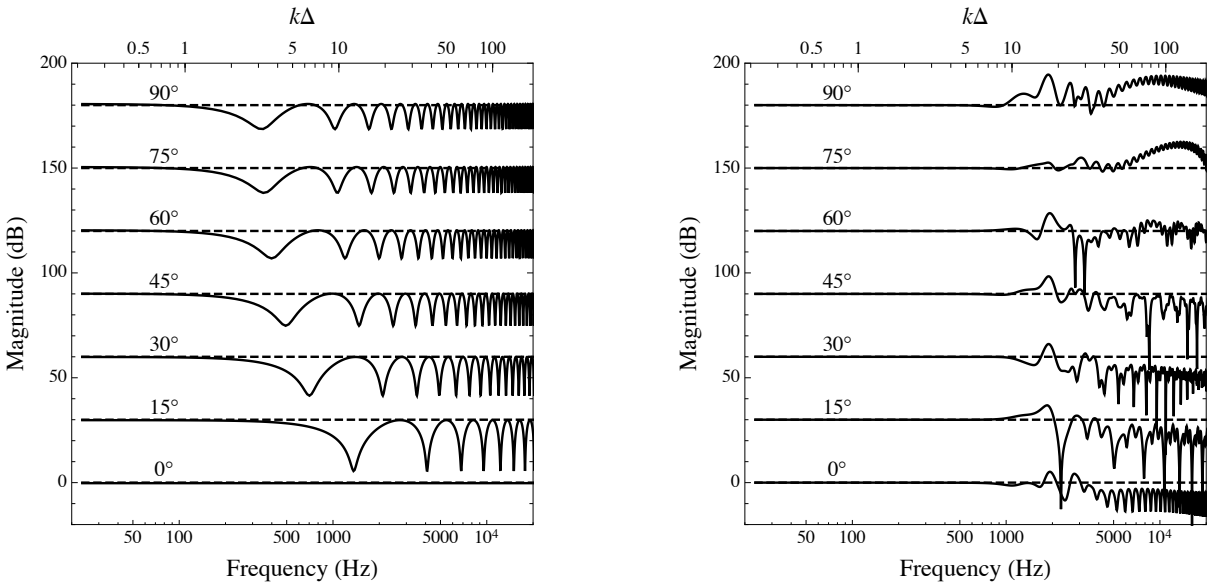
We also observe that for source azimuths $\phi = 45^\circ, 60^\circ, 75^\circ$, the proposed method induces very little spectral coloration even up to $\sim$ 5 kHz, with the exception of two narrow-band notches around 3 kHz for $\phi = 60^\circ$. We expect such spectral coloration may be tolerable, if not imperceptible, although its perceptibility must be determined through subjective listening tests.

To further explore the spectral coloration induced by the proposed method, we plot, in Fig. 5, frequency responses for several input expansion orders, L_{in}, with a fixed source azimuth $\phi = 45^\circ$ and for $\Delta = 0.5$ m. We observe from this plot that spectral coloration appears negligible for frequencies such that $k\Delta \leq 2L_{\mathrm{in}}$, as indicated by the short vertical lines superimposed on the plot.

Figure 6 shows the average localization errors over source azimuths $\phi = 0^\circ, 15^\circ, \ldots, 90^\circ$, and averaged in a fixed 1/3-octave frequency band centered on 1 kHz. From this figure, we see that for small microphone spacings ($\Delta \leq 0.5$ m), corresponding to $k\Delta \sim 10$ for 1 kHz, the proposed method (denoted "Reg-LS" in the plot) achieves smaller localization errors ($\bar{\varepsilon} \approx 3.9^\circ$) than the weighted average method ($\bar{\varepsilon} \approx 7.7^\circ$).

To further explore predicted localization, we plot, in Fig. 7, localization errors, averaged over the same set of source azimuths and frequency band, for several input expansion orders, L_{in}. Similar to the trend observed in the frequency responses shown in Fig. 5, we observe that localization errors remain small for $k\Delta \leq 2L_{\mathrm{in}}$, as indicated by the vertical grid lines drawn on the plot.

Figure 8 shows, for two different source positions, localization errors achieved using all four microphones (labeled "All" in the plots) and those achieved using only the valid microphones ("Valid") as functions of microphone spacing. Localization errors are again averaged in a fixed 1/3-octave frequency band centered on 1 kHz. In the symmetric source configuration, the invalid microphones have little effect on the localization accuracy, as shown in Fig. 8a. In the off-axis source configuration, however, including the invalid microphone in the interpolation significantly increases localization errors across all but the largest microphone spacings, as shown in Fig. 8b. Again, from this plot we see that the proposed method, when microphone validity is considered, achieves small localization errors for $k\Delta \leq 10$.



(a) Weighted average method

(b) Regularized least-squares method

Fig. 4: Frequency responses at the listening position for various source azimuths (offset by +30 dB for each 15° increment in azimuth). The bottom axes show frequency in Hz, while the top axes show nondimensionalized wavenumber $k\Delta$ for $\Delta = 0.5$ m.

6 Conclusions

In this work, we presented a method for soundfield navigation that uses an array of higher-order ambisonics (HOA) microphones and interpolates between them by computing a regularized least-squares estimate of the ambisonics signals at the desired listening position. We compared, through numerical simulations of simple incident soundfields, the proposed method to an existing alternative method in which interpolation is performed by computing a weighted average of the ambisonics signals from each microphone. The methods were evaluated in terms of frequency response and predicted localization error.

Spectral coloration induced by the proposed method is shown to be negligible for small microphone spacings and/or low frequencies. As a rule of thumb, for interpolating between a pair of microphones, coloration appears negligible for $k\Delta \leq 2L_{\text{in}}$, where k is angular wavenumber, Δ is the microphone spacing in meters, and L_{in} is the maximum expansion order captured by the microphones. On the other hand, the weighted average method introduces comb-filtering for frequencies approximately $k\Delta \geq \pi/\sin\phi$, where ϕ is the azimuth of the source. Consequently, the coloration induced by the weighted average technique is independent of expansion order L_{in}, whereas for the proposed method, increasing expansion order increases the lowest frequency at which coloration becomes significant.

We also showed that for small microphone spacings, the localization errors obtained with the weighted average method are, when averaged over varying source direction, approximately twice as large as those obtained with the proposed method. Again we found a rule of thumb that, for interpolating between a pair of microphones, localization errors are small for $k\Delta \leq 2L_{\text{in}}$.

Finally, we demonstrated the degradation of localization accuracy incurred by including in the calculation those microphones that are nearer to a source than to the listening position (i.e., invalid microphones). It should be noted, however, that inclusion of the invalid microphones did not have a significant effect in the symmetric case (see Fig. 8a), while the effect was significant in the case of an off-axis source (see Fig. 8b).

Future work should involve subjective verification of the objective analyses shown here. In particular, the predicted localization errors obtained in this work rely

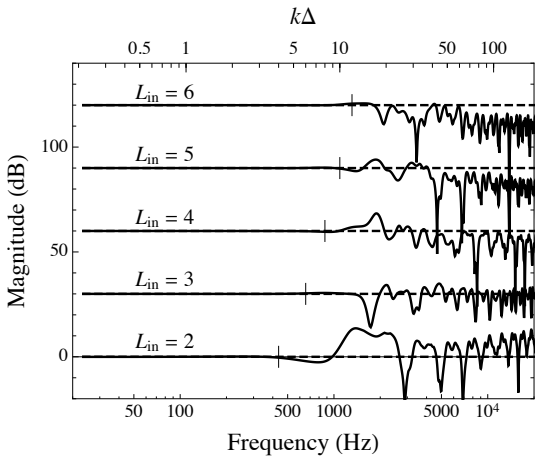


Fig. 5: Frequency responses at the listening position for various input orders (offset by +30 dB) with source azimuth $\phi = 45°$. The bottom axes show frequency in Hz, while the top axes show nondimensionalized wavenumber $k\Delta$ for $\Delta = 0.5$ m. The short vertical lines indicate $k\Delta = 2L_{in}$.

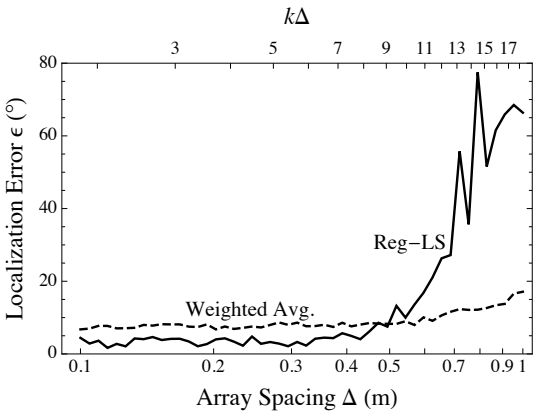


Fig. 6: Predicted localization errors as functions of microphone spacing. Errors are averaged over azimuths and averaged in a 1/3-octave band centered on 1 kHz. The bottom axis shows microphone spacing in meters, while the top axis shows nondimensionalized microphone spacing $k\Delta$ at 1 kHz.

on our procedure for segmenting plane-wave impulse responses into wavelets (see Section 4.1), which has not been validated through subjective testing. Spectral coloration should also be evaluated subjectively, to assess the perceptibility of the frequency response deviations induced by each method.

Additionally, methods to minimize spectral coloration induced by the proposed method at higher frequencies should be investigated. One such method may involve iteratively adjusting the frequency-dependent regularization, by first computing the set of interpolation filters for a given level of regularization, estimating the spectral coloration induced by the filters, and then adjusting the regularization function accordingly.

Acknowledgements

This work was sponsored by the Sony Corporation of America. The authors wish to thank P. Stitt for providing the MATLAB code for the precedence-effect-based energy vector model.[3]

[3]Available here: `https://circlesounds.wordpress.com/matlab-code/`

References

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

[2] Hahn, N. and Spors, S., "Physical Properties of Modal Beamforming in the Context of Data-Based Sound Reproduction," in *Audio Engineering Society Convention 139*, 2015.

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

[4] Tylka, J. G. and Choueiri, E. Y., "Comparison of Techniques for Binaural Navigation of Higher-Order Ambisonic Soundfields," in *Audio Engineering Society Convention 139*, 2015.

[5] Zheng, X., *Soundfield navigation: Separation, compression and transmission*, Ph.D. thesis, University of Wollongong, 2013.

[6] Southern, A., Wells, J., and Murphy, D., "Rendering walk-through auralisations using wave-based acoustical models," in *Signal Processing Conference, 2009 17th European*, pp. 715–719, 2009.

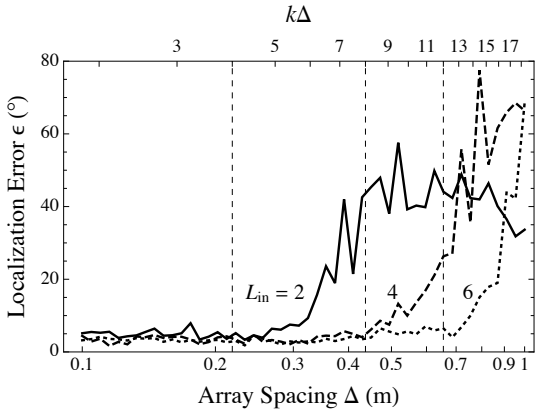


Fig. 7: Predicted localization errors as functions of microphone spacing for various input orders. Errors are averaged over azimuths and averaged in a 1/3-octave band centered on 1 kHz. The bottom axis shows microphone spacing in meters, while the top axis shows nondimensionalized microphone spacing $k\Delta$ at 1 kHz. The vertical grid lines indicate $k\Delta = 2L_{\text{in}}$ for $L_{\text{in}} = 2, 4, 6$.

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

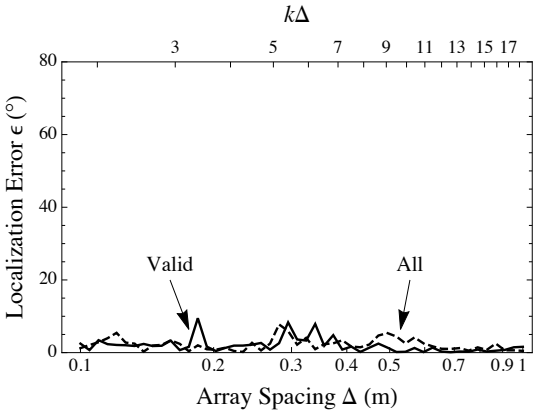
[8] Nachbar, C., Zotter, F., Deleflie, E., and Sontacchi, A., "ambiX - A Suggested Ambisonics Format," in *Proceedings of the 3rd Ambisonics Symposium*, 2011.

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

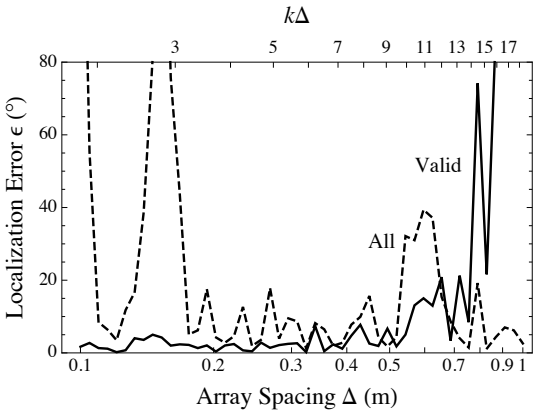
[10] Gerzon, M. A., "General Metatheory of Auditory Localisation," in *Audio Engineering Society Convention 92*, 1992.

[11] Stitt, P., Bertet, S., and van Walstijn, M., "Extended Energy Vector Prediction of Ambisonically Reproduced Image Direction at Off-Center Listening Positions," *J. Audio Eng. Soc.*, 64(5), pp. 299–310, 2016.

[12] Fliege, J. and Maier, U., "The distribution of points on the sphere and corresponding cubature formulae," *IMA Journal of Numerical Analysis*, 19(2), pp. 317–334, 1999, doi:10.1093/imanum/19.2.317.



(a) Source position $\vec{r}_{\text{S}} = (0.75\Delta, 0, 0)$



(b) Source position $\vec{r}_{\text{S}} = (0.75\Delta, 0.75\Delta, 0)$

Fig. 8: Predicted localization errors as functions of microphone spacing. Errors are averaged in a 1/3-octave band centered on 1 kHz. The bottom axes show microphone spacing in meters, while the top axes show nondimensionalized microphone spacing $k\Delta$ at 1 kHz.