OPEN ACCESS Freely available online

L. McCormack, S. Delikaris-Manias, A. Politis, D. Pavlidi, A. Farina, D. Pinardi and V. Pulkki, "Applications of Spatially Localized Active-Intensity Vectors for Sound-Field Visualization" *J. Audio Eng. Soc.*, vol. 67, no. 11, pp. 840–854, (2019 November.).

DOI: https://doi.org/10.17743/jaes.2019.0041

Applications of Spatially Localized Active-Intensity Vectors for Sound-Field Visualization

LEO MCCORMACK[1], *AES Student Member*, **SYMEON DELIKARIS-MANIAS**[1],
(leo.mccormack@aalto.fi)

ARCHONTIS POLITIS[1], *AES Associate Member*, **DESPOINA PAVLIDI**[2],

ANGELO FARINA[3], *AES Fellow*, **DANIEL PINARDI**[3], *AES Student Member*, **AND**

VILLE PULKKI[1], *AES Fellow*

[1]*Aalto University, Espoo, Finland*
[2]*University of Crete, Heraklion, Greece*
[3]*University of Parma, Parma, Italy*

The purpose of this article is to detail and evaluate three alternative approaches to sound-field visualization, which all employ the use of spatially localized active-intensity (SLAI) vectors. These SLAI vectors are of particular interest, as they allow direction-of-arrival (DoA) estimates to be extracted in multiple spatially localized sectors, such that a sound source present in one sector has reduced influence on the DoA estimate made in another sector. These DoA estimates may be used to visualize the sound-field by either: (I) directly depicting the estimates as icons, with their relative size dictated by the corresponding energy of each sector; (II) generating traditional activity maps via histogram analysis of the DoA estimates; or (III) by using the DoA estimates to reassign energy and subsequently sharpen traditional beamformer-based activity maps. Since the SLAI-based DoA estimates are continuous, these approaches are inherently computationally efficient, as they forego the need for dense scanning grids to attain high-resolution imaging. Simulation results also show that these SLAI-based alternatives outperform traditional active-intensity and beamformer-based approaches, for the majority of cases.

0 INTRODUCTION

Capturing a sound-field utilizing an array of microphones, offers many advantages over using a single microphone. These advantages largely stem from the fact that certain spatial attributes of the sound-field are encoded within the array signals, whereby the spatial resolution of this encoding is predominately influenced by the number of microphones and their orientation. Essentially, while a single sensor can be used to establish whether a sound source is present in a sound scene, a microphone array can be used to generate spatially selective filters (commonly referred to as beamformers) or employed to extract meaningful spatial parameters, such as the direction-of-arrival (DoA) of sound sources. Beamformers are traditionally employed to enhance a signal (e.g., speech) in the presence of interferers and noise [1, 2], and may yield improved performance when informed of spatial parameters and subjected to certain constraints [3, 4]. However, beamformers and spatial parameters may also be utilized for parametric sound-field reproduction or sound-field visualization; the latter of which is the primary focus of this work.

Many established sound-field visualization approaches rely on evaluating a suitable localization function over a dense scanning grid, which is selected to sample a particular spatial area of interest from the perspective of the microphone array. For convenience, when utilizing spherical microphone arrays, these localization functions are often formulated in the spherical harmonic domain (SHD) [1, 5]. Many of these functions rely on determining the sound-field energy, or the likelihood of a sound source being present, at each grid point. However, these approaches have an inherent limitation; they require the use of dense scanning grids for high resolution imaging. In this

article, alternative sound-field visualization approaches are explored, which all utilize SLAI vectors. Essentially, these are active-intensity (AI) vectors [6], which are estimated after the sound-field has been subjected to a direction-dependent weighting function. This approach allows for DoA estimation within multiple spatially localized sectors, whereby an estimate in one sector has reduced susceptibility to interferers present in other sectors. Additionally, these DoA estimates are continuous, and as such, they may be quantized to a grid with much higher resolution than that of a practical scanning grid. Once the DoA estimates have been extracted from multiple sectors, sound-field visualization may be subsequently established using one of the following three approaches, which are detailed and evaluated in this work:

I) Directly depicting the estimates as icons, with their relative size dictated by the corresponding energy of each sector [7].
II) Generating traditional activity maps via histogram analysis of the DoA estimates [8, 9].
III) Using the DoA estimates to reassign energy and subsequently sharpen traditional beamformer-based activity maps [10].

This article provides background on SHD processing, traditional sound-field visualization approaches, and SLAI vectors (Sec. 1). It goes on to explain how array signals may be spatially encoded into the SHD (Sec. 2). Following this, Secs. 3 and 4 detail how the AI and SLAI vectors may be obtained in the SHD, respectively. Sec. 5 explores the three aforementioned approaches to sound-field visualization, employing the DoA estimates extracted from multiple SLAI vectors. This is followed by an evaluation of the approaches in Sec. 6 and conclusions, which are provided in Sec. 7.

1 BACKGROUND

1.1 Spatially Encoding Microphone Array Signals into Spherical Harmonic Signals

It is possible to derive spatial parameters and generate beamformers using microphone array signals directly. However, this approach tends to lead to algorithms being tailored for a specific array, as the array specifications become an integral part of the algorithms [11]. Therefore, a more flexible approach is to first transform the microphone array signals into an array-independent domain. This is popularly achieved by decomposing the sound-field into orthonormal basis functions around the unit sphere, via a spherical harmonic transform (SHT) [1]. Essentially, by spatially encoding the microphone array signals into spherical harmonic (SH) signals (also known as Ambisonic [12] or B-Format signals), the array specifications are largely abstracted away from the algorithms that utilize them.

There are two main approaches for performing this transform. In the case of spherical and cylindrical arrays with phase-matched sensors, analytical solutions are available [13, 14] that describe how plane waves interact with the array at specific frequencies. These formulae take into account certain physical aspects of the array design, such as its radius and whether it has an open or rigid baffle construction. However, for atypical microphone array shapes and/or mismatched sensors, for which no analytical solution can be easily derived, the more generalized measurement-based approach is more appropriate. This method requires the measurement of impulse responses (IRs) for many directions on the surface of the array, in order to derive suitable encoding filters [15–19].

1.2 Sound-Field Visualization in the SHD

Today, there are a multitude of algorithms formulated in the SHD, with many spatial audio systems relying on beamforming as a means of reproducing or visualizing a captured sound-field. Regarding the latter function, perhaps the most trivial approach is to steer a beamformer and determine its energy in multiple directions that surround a particular spatial area of interest; an approach commonly referred to as scanning beamforming. This direction-dependent beamformer energy may then be presented as a 2-dimensional image with a suitable color gradient, often described as an activity map, which is a convenient representation for depicting the relative acoustical energy for multiple directions. Bright spots in the activity map infer directions with potential sound sources and/or reflections, which may be extracted numerically via peak-finding. Naturally, this approach is only precise to the degree of separation between the scanning grid directions, and thus, the computational requirements can be significant when high precision is required.

One popular beamformer is based on a Plane-wave Decomposition (PWD) approach, whereby spherical harmonics for certain directions on the sphere are used as the beamforming weights [20, 21]. However, this approach is inherently limited by the spatial resolution of the input signals, which can often be insufficient at lower orders and result in blurred activity maps. Alternatively, signal-dependent beamformers, such as the Minimum-variance Distortionless Response (MVDR) algorithm, can yield improved spatial selectivity by adaptively placing nulls towards interferers. However, in general, these beamformers require increased computation and can often be susceptible to coherent interferers [22].

Traditional Wiener post-filters, formulated in the SHD [23], may be employed to minimize the incoherent noise present in the beamformer signal that can often result in cleaner activity maps. Other post-filters, such as the Cross-Pattern Coherence (CroPaC) algorithms [24, 25], are also well suited to the task of improving beamformer performance, especially in reverberant and/or noisy environments. These latter post-filters are also generally straightforward to implement, as they do not impose any assumptions regarding the source or interferers, nor do they require any prior knowledge of the signal or noise statistics. Additionally, one may even directly utilize the energy of CroPaC post-filters for visualization purposes, provided that an

additional iterative side-lobe suppression operation is applied [26].

On the other hand, subspace methods, such as MUltiple SIgnal Classification (MUSIC), may also produce high-resolution activity maps. Rather than determining beamformer signal energies, these methods operate by ascertaining parameters that correspond to the likelihood of sound sources being located at each scanning direction. However, these approaches typically require source number estimation [27, 28] and can also be sensitive to coherent interferers and reverberation. There also exist other subspace based [29, 14] and maximum likelihood DoA estimators [30, 31], which do not require scanning to estimate the DoAs of multiple sources; however, these approaches generally require increased implementation and computational complexity and are often unsuitable for real-time systems.

1.3 DoA Estimation Based on SLAI Vectors

One alternative approach to DoA estimation is to utilize the active-intensity (AI) vector [6], which describes the flow of acoustical energy. By making the assumption that the direction opposite to this energy flow is the DoA of a sound source, the approach represents a more computationally efficient option when compared to the traditional activity map and peak-finding based alternatives [32–34, 20, 35]. Furthermore, these DoA estimates are continuous, and thus they may be quantized to a grid of much higher point density than that of a practical beamforming or subspace scanning grid.

Due to the relation between the AI vector and the perceived DoA during sound-field reproduction, it became the predominant spatial parameter for the spatial audio coding and reproduction framework, Directional Audio Coding (DirAC) [36, 37]. However, it was found that for scenarios where multiple sound sources overlap in both time and frequency, the flow of energy would propagate from a direction in-between the sources. Therefore, DirAC has since been generalized to take advantage of the increased spatial resolution of higher-order SH signals, introducing the concept of SLAI vectors in [38, 39]. This higher-order DirAC formulation extracts a separate DoA estimate at spatially selective sectors surrounding multiple directions on the sphere, through beamforming operations in the SHD. Provided that sound sources are located in their own individual sector, this approach has been shown to be more robust for reproduction purposes, subsequently resulting in higher perceived spatial accuracy. The theoretical properties of SLAI for reference sound-field conditions were further analyzed in [40].

1.4 Sound-Field Visualization Utilizing SLAI-Based DoA Estimates

More recently, the SLAI approach has been explored for computationally efficient sound-field visualization, which is the main point of interest for this work. SLAI-based sound-field visualization can be attained in a variety of ways. One approach being to directly represent the sector DoA estimates at multiple frequencies as different-colored icons, with their relative size determined by the corresponding sector energies [7]. This approach relies on fewer computational resources, when compared to many traditional activity map alternatives, and has the added benefit of being able to depict estimates for multiple frequencies, simultaneously. The main downside being that the physical extent of a sound source becomes more ambiguous, as one may infer that a cluster of icons corresponds to a single large source, rather than several point-like sources, or vice versa.

Alternatively, by performing histogram analysis on the estimated DoAs, with a sufficiently long temporal window, one may also generate more traditional activity maps [8, 9]; an approach which still retains much of the computational benefits granted by the SLAI approach.

The SLAI approach may also be employed to sharpen traditional beamformer-based activity maps by reassigning the beamformer energies based on the corresponding DoA estimates for each grid point [10]. This beamformer energy may then be accumulated and quantized to a grid of much higher resolution than that of the scanning grid, resulting in a sharper image. In the trivial case of a single point source in a free-field, the beamformer energy for all scanning grid directions would be reassigned to the true source direction.

2 SPATIAL ENCODING

Before expanding upon the SLAI formulations, and the subsequent steps required for sound-field visualization, the microphone array signals must first be spatially encoded into the SHD. Since this encoding is frequency-dependent, the microphone array signals $\mathbf{x}(t)$ must first be transformed into the time-frequency domain $\hat{\mathbf{x}}(t, f)$ by means of either a short-time Fourier transform (STFT) or perfect reconstruction filterbank, where t and f refer to the down-sampled time and frequency indices, respectively.

2.1 Encoding Based on Analytical Solutions

Consider a spherical array, denoted with Q microphones at $\Omega_q = (\theta, \phi, r)$ locations, where $\theta \in [-\pi/2, \pi/2]$ denotes the elevation angle, $\phi \in [-\pi, \pi]$ the azimuthal angle and r the radius. A SHT may then be employed to convert the array signals, $\hat{\mathbf{x}} \in \mathbb{C}^{Q\times 1}$, into a set of SH signals for each frequency band. The accuracy of this conversion depends on the microphone distribution on the surface of the array, the type of the array, and the radius [1]. The total number of microphones defines the highest order of SH signals L that can be estimated. Please note that the frequency and time indices are omitted for the brevity of notation.

The SH signals can be estimated as

$$\mathbf{s} = \mathbf{W}\hat{\mathbf{x}}, \tag{1}$$

where

$$\mathbf{s} = [s_{00}, s_{1(-1)}, s_{10}, \ldots, s_{L(L-1)}, s_{LL}]^{\mathrm{T}} \in \mathbb{C}^{(L+1)^2\times 1}, \tag{2}$$

are the SH signals and $\mathbf{W} \in \mathbb{C}^{(L+1)^2\times Q}$ is a frequency-dependent spatial encoding matrix defined as

$$\mathbf{W} = \mathbf{W}_l\mathbf{Y}_e, \tag{3}$$

where $\mathbf{Y}_e \in \mathbb{R}^{(L+1)^2 \times Q}$ is the SH encoding matrix, which estimates the pressure on the surface of the sphere and $\mathbf{W}_l \in \mathbb{C}^{(L+1)^2 \times (L+1)^2}$ is an equalization matrix that eliminates the effect of the sphere itself.

The frequency-independent SH encoding matrix can be calculated as

$$\mathbf{Y}_e = \begin{cases} \frac{1}{Q}\mathbf{Y}^{\mathrm{T}} & \text{uniform} \\ \mathbf{Y}^{\dagger} = (\mathbf{Y}^{\mathrm{T}}\mathbf{Y})^{-1}\mathbf{Y}^{\mathrm{T}} & \text{non-uniform,} \end{cases} \tag{4}$$

where † denotes the Moore-Penrose pseudo-inverse operation and $\mathbf{Y} \in \mathbb{R}^{Q \times (L+1)^2}$ is a matrix containing SHs weights for each microphone direction

$$\mathbf{Y}(\Omega_q) = \begin{bmatrix} Y_{00}(\Omega_1) & Y_{1(-1)}(\Omega_1) & \dots & Y_{LL}(\Omega_1) \\ Y_{00}(\Omega_2) & Y_{1(-1)}(\Omega_2) & \dots & Y_{LL}(\Omega_2) \\ Y_{00}(\Omega_3) & Y_{1(-1)}(\Omega_3) & \dots & Y_{LL}(\Omega_3) \\ \vdots & \vdots & \vdots & \vdots \\ Y_{00}(\Omega_Q) & Y_{1(-1)}(\Omega_Q) & \dots & Y_{LL}(\Omega_Q) \end{bmatrix}, \tag{5}$$

where Y_{lm} are the individual real-valued SHs of order $l \geqslant 0$, degree $m \in [-l, l]$ and microphone direction Ω_q. This double SH indexing (l, m) may also be expressed by the single index $k = l^2 + l + m + 1$, which is referred to as the Ambisonic Channel Number (ACN) in Ambisonics literature.

The equalization matrix can be defined as

$$\mathbf{W}_l = \mathrm{diag}\{[w_0, w_1, w_1, w_1, \dots, w_L]\}, \tag{6}$$

where w_l are order-dependent and frequency-dependent equalization weights, which are the inverse of the theoretical modal coefficients b_l

$$w_l = \frac{1}{b_l}, \tag{7}$$

which take into account whether the array construction is open or rigid and the directivity of the sensors (e.g., cardioid or omnidirectional). In the case of a rigid baffle construction, these modal coefficients can also take into consideration the frequency-dependent acoustical admittance of the surface of the array and also the distance between the surface of the array and the microphones in order to cater for cases in which the microphones protrude from the surface of the array. For more details on calculating these modal coefficients, the reader is directed to [13, 14].

However, the problem with directly inverting the modal coefficients in Eq. (7) is that the internal noise of the sensors can be amplified significantly in practice, especially at low frequencies and higher orders. Therefore, a regularized inversion is often more preferable, such as the Tikhonov approach [15]

$$w_l = \frac{1}{b_l}\frac{|b_l|^2}{|b_l|^2 + \lambda^2}, \tag{8}$$

where λ is a regularization parameter that influences the sensor noise amplification, allowing the user to make a compromise between the accuracy of the transformation and the increase in sensor noise. For details on some alternative options for regularization and for calculating the equalization matrix $\mathbf{W}_l$, the reader is referred to [15, 41–43, 21, 44].

2.2 Arbitrary Encoding via Impulse Response Measurements

The theoretical encoding approach is a convenient means of extracting SH signals from microphone arrays, provided that they employ phase-matched sensors in a spherical or cylindrical arrangement. However, analytical scattering solutions for other geometries do not exist [13], and the assumption that the array utilizes phase-matched sensors may not always be met. Therefore, in these cases, array response measurements in free-field environments for several directions around the array must be obtained in order to derive a suitable spatial encoding matrix. Such proposed solutions are based on a regularized least-squares inversion of the array responses in the space domain [15, 18] or the SHD [16], based on a weighted singular value decomposition [17], or a quadratically constrained least-squares solution [19].

3 ACTIVE-INTENSITY VECTOR ESTIMATION

The AI vector can be estimated from the zeroth and first-order SH signals, which represent the pressure and pressure-gradient signals, respectively [6]. The DoA and the diffuseness estimates can then be derived utilizing the AI vector and the energy density at the recording point. Estimating these parameters in such a manner represents a computationally efficient and convenient means of acquiring information, utilized historically by DirAC for parametric reproduction purposes [36].

The AI vector can be estimated as [6]

$$\mathbf{i}_{\mathrm{a}} = \Re[p\mathbf{u}^*], \tag{9}$$

where $\Re$ denotes the real operator; * denotes the complex conjugate operator; p is the sound pressure, estimated as the omnidirectional signal $p \simeq s_{00}$; and $\mathbf{u}$ is the particle velocity, which can be estimated using the pressure gradient signals [45], assuming that the sound sources are received as plane-waves, as[1]

$$\mathbf{u} \simeq -\frac{1}{\rho_0 c\sqrt{3}} \begin{bmatrix} \mathbf{s}_{11} \\ \mathbf{s}_{1(-1)} \\ \mathbf{s}_{10} \end{bmatrix}, \tag{10}$$

where ρ_0 is the mean density of the medium and c is the speed of sound.

The DoA estimate γ is assumed to be in the opposite direction to the AI vector

$$\gamma_{\mathrm{DoA}}(\theta, \phi) = -\frac{\mathbf{i}_{\mathrm{a}}}{||\mathbf{i}_{\mathrm{a}}||}. \tag{11}$$

This approach provides one estimate of the DoA per time and frequency index, which becomes problematic when multiple sound sources have similar spectral content and are active simultaneously. This is because in these scenarios,

[1] Assuming ortho-normalized (N3D) real SHs with ACN indexing. Omit the $1/\sqrt{3}$ term if using the semi-normalized (SN3D) convention.

the DoA estimate will oscillate between the different sound source directions. Furthermore, in stationary sound-fields, and with sufficiently long temporal observation windows, the AI DoA Eq. (11) is stable but points to a spatial weighted average of the individual source DoAs, with the weights dependent on the cross-correlation between the source signals [46].

4 SPATIALLY LOCALIZED ACTIVE-INTENSITY

The above relations show that sound-field analysis based on the AI vector does not require signals of order higher than one. The SLAI-based analysis exploits the fact that if SH signals of order $L > 1$ are available, then the sound-field may be weighted directionally before the AI vector is estimated. The simple principle of the SLAI is to compute the normal AI vector after a directional weighting has been applied to the sound-field. This approach preserves contributions from the region of interest, hereby referred to as a *sector*, while attenuating contributions from outside that region. This directional weighting $w(\Omega)$ is in essence a beamformer with its main lobe oriented towards the center of the sector, and with its beamwidth approximately covering the sector area. To compute the AI vector for this sector, one requires the pressure p_w and velocity $\mathbf{u}_w$ due to the sound-field weighted by $w(\Omega)$, with the SLAI vector calculated in a similar way to Eq. (9) as

$$\mathbf{i}_{\mathrm{a,w}} = \Re[p_w \mathbf{u}_w^*]. \tag{12}$$

The sector pressure signal is given simply as the sector beamformer $w(\Omega)$ output, while the sector velocity signals correspond to beampatterns of the following form, as given in [38, 40],

$$\begin{bmatrix} w_{\mathrm{x}}(\Omega) \\ w_{\mathrm{y}}(\Omega) \\ w_{\mathrm{z}}(\Omega) \end{bmatrix} = \begin{bmatrix} w(\theta, \phi)\cos\phi\cos\theta \\ w(\theta, \phi)\sin\phi\cos\theta \\ w(\theta, \phi)\sin\theta \end{bmatrix}. \tag{13}$$

This last relationship reveals that the sector velocity beampatterns correspond to the sector beampattern weighted with three orthogonal dipoles.

Assuming that there are higher-order SH signals available $L > 1$, the next point of interest is to obtain the sector pressure-velocity signals directly from them. This is performed by means of beamforming matrices computed directly in the SHD and applied to the SH signals, in order to determine the sector signals and, subsequently, the SLAI vector

$$\begin{bmatrix} p_w \\ \mathbf{u}_w \end{bmatrix} = \mathbf{W}_{\mathrm{p,u}}^{\mathrm{H}} \mathbf{s}. \tag{14}$$

The following subsections describe various means of calculating this beamforming matrix $\mathbf{W}_{\mathrm{p,u}} \in \mathbb{R}^{(L+1)^2 \times 4}$ and also provide some insights into the direction-dependent performance of the DoA estimation.

4.1 Sector Design Using SHD Beamformer Patterns

Beamforming in the SHD is equivalent to applying the SH coefficients $\mathbf{w}$ of the beampattern $w(\Omega)$ as a weight-and-sum operation on the SH signals

$$p_w = \mathbf{w}^{\mathrm{H}} \mathbf{s}. \tag{15}$$

The coefficients for certain beampatterns may be derived analytically and, although the beampatterns do not need to be axisymmetric, this section focuses on axisymmetric beampatterns due to their design simplicity. Axisymmetric patterns may be factorized into a pattern-dependent component and an orientation-dependent component, such that the kth term of the coefficient vector $\mathbf{w}$ is

$$[\mathbf{w}(\Omega_0)]_k = w_{lm}(\Omega_0) = w_l Y_{lm}(\Omega_0) \tag{16}$$

where Ω_0 is the main lobe's orientation. Some useful axisymmetric patterns that are suitable for the analysis are

$$w_l = \frac{L!L!}{(L+l+1)!(L-l)!} \qquad \text{cardioid} \tag{17}$$

$$w_l = \frac{1}{(L+1)^2} \qquad \text{hypercardioid} \tag{18}$$

$$w_l = \frac{P_l(\cos\kappa_L)}{\sum_{l=0}^{L}(2l+1)P_l(\cos\kappa_L)} \qquad \max - \mathrm{rE} \tag{19}$$

where P_l are the Legendre polynomials and $\kappa_L = \cos(2.407/(L + 1.51))$ as given by [47]. Higher-order cardioids are defined here as patterns with a single rear null; hypercardioids are normalized PWD beamformers that attain the maximum directivity factor for a given order; and max-rE are patterns that maximize the length of the intensity vector in a diffuse field.

After the coefficients of the jth sector pattern $\mathbf{w}(\Omega_j)$ have been obtained, the sector velocity patterns should be computed. Since they correspond to multiplication of the sector pattern with dipole patterns, their coefficients can be expressed as a linear combination of the coefficients of the sector pattern, as has been shown in [40]. Due to the fixed nature of the dipole patterns, the sector velocity patterns are given by the following relation

$$\mathbf{w}_{\mathrm{i}}(\Omega_j) = \mathbf{A}_{\mathrm{i}} \mathbf{w}(\Omega_j), \qquad \text{with i} = \{\mathrm{x, y, z}\}. \tag{20}$$

where $\mathbf{A}_{\mathrm{i}} \in \mathbb{R}^{(L+2)^2 \times (L+1)^2}$ are matrices that linearly map the coefficients of the sector and sector velocity patterns. Since the sector velocity patterns are based on the products of both sector and dipoles, their SH order is larger than each of the original patterns and equal to the sum of their orders. Since dipoles are of order $L = 1$, the sector velocity patterns are one order higher $L + 1$ than that of the sector pattern.

The matrices $\mathbf{A}_{\mathrm{i}}$ are independent of the sector pattern and can be pre-computed up to some maximum order of interest [38, 40]. Their detailed derivation and structure is given in [40]. Finally, the beamforming matrix for a single sector oriented at Ω_j is

$$\mathbf{W}_{\mathrm{p,u}}(\Omega_j) = [\hat{\mathbf{w}}(\Omega_j), \mathbf{w}_{\mathrm{x}}(\Omega_j), \mathbf{w}_{\mathrm{y}}(\Omega_j), \mathbf{w}_{\mathrm{z}}(\Omega_j)], \tag{21}$$

where $\hat{\mathbf{w}}(\Omega_j)$ are the coefficients of the sector pattern when zero-padded to order $L + 1$.

4.2 Sector Design Based on Arbitrary Directional Functions

The above SHD-based sector design formulation represents an explicit solution with known properties, and as such, the formulation relies on analytical descriptions of the target sector patterns. One alternative approach, however, is to approximate the sector patterns such that they mimic arbitrary directional functions in a least-squares (LS) sense [7]. This approach may accommodate a wider range of sector designs.

From henceforth, the Vector-Base Amplitude Panning (VBAP) directional function [48] is employed as an example; however, the principles may also be extended to arbitrary functions or other panning methods. As before, the look directions of the sectors are defined at the following points: $\Omega_j = (\theta_j, \phi_j)$, for $j = 1, \ldots, J$, where J is the total number of sectors. The VBAP gains $\mathbf{g}(\Omega_j) \in \mathbb{R}^{1\times\Upsilon}$ are then computed for a dense grid of Y points surrounding the sphere, for each of the sector look-directions. These VBAP gain patterns are then imposed onto zeroth and first-order SH patterns $\mathbf{Y}_{\mathrm{p,u}}^{(norm)} \in \mathbb{R}^{4\times\Upsilon}$, which are evaluated using the same dense grid and normalized such that each component $\max(|\hat{Y}|) = 1$. The target pressure and velocity patterns $\mathbf{T}_{\mathrm{p,u}}(\Omega_j) \in \mathbb{R}^{4\times\Upsilon}$ are then derived as

$$\mathbf{T}_{\mathrm{p,u}}(\Omega_j) = \mathbf{G}(\Omega_j) \odot \mathbf{Y}_{\mathrm{p,u}}^{(norm)}. \tag{22}$$

where $\odot$ denotes the Hadamard product, and $\mathbf{G}(\Omega_j) = [\mathbf{g}(\Omega_j); \ldots; \mathbf{g}(\Omega_j);] \in \mathbb{R}^{4\times\Upsilon}$ is a matrix that consists of the VBAP gains replicated for each of the SH components.

The sector beamforming weights may then be approximated in the SHD via a LS regression as

$$\mathbf{W}_{\mathrm{p,u}}^{(\mathrm{VBAP})}(\Omega_j) = \mathbf{T}_{\mathrm{p,u}}(\Omega_j)\mathbf{Y}_{\mathrm{LS}}^{\dagger}, \tag{23}$$

where $\mathbf{Y}_{\mathrm{LS}} \in \mathbb{R}^{(L+1)^2\times\Upsilon}$ are SHs of order $L > 1$. Note that the more intricate the target pattern, the higher the order of SH components are required to sufficiently approximate them.

4.3 Directional Bias in the Presence of a Diffuse Field

In the case of a single source in a free field, a sector's directional analysis will produce the correct DoA, even in situations where the DoA is far outside the look-direction of the sector. The only exception is when the source direction coincides with a null of the sector pattern, in this case the DoA is undefined. However, the presence of interferers and diffuse noise will affect the estimated DoA even for a dominant source. For the fundamental case of a sound-field mixture, consisting of one plane wave source of power P_{pw} and a diffuse-field of power P_{d}, a simple expression exists for the directional bias, which is detailed in [40]. First, an assumption is made that the plane-wave signal is incident from the DoA Ω_0, with the sector directed at Ω_j; the angle between the two is given by α. The direct-to-diffuse ratio $\Gamma = P_{\mathrm{pw}}/P_{\mathrm{d}}$, and the magnitude of the directional energy at the centroid of the sector pattern $\mathbb{K} = ||\hat{\mathbf{w}}^{\mathrm{T}}(0, \pi/2)\mathbf{A}_{\mathrm{z}}\mathbf{w}(0, \pi/2)||$, are calculated for the sector oriented at the top, in order to exploit its axisymmetry and simplify the expression. The directional bias will always tend towards the center of the sector pattern and is given by

$$\delta(\Gamma, \alpha) = \arcsin\left(\frac{\mathbb{K}\sin\alpha}{4\pi\Delta(\Gamma, \alpha)}\right) \tag{24}$$

where Δ is defined as

$$\Delta(\Gamma, \alpha) = \sqrt{\Gamma^2 w^4(\alpha) + (\mathbb{K}/4\pi)^2 + 2\Gamma(\mathbb{K}/4\pi)w^2(\alpha)\cos\alpha}. \tag{25}$$

This directional bias is illustrated in Sec. 6.1.

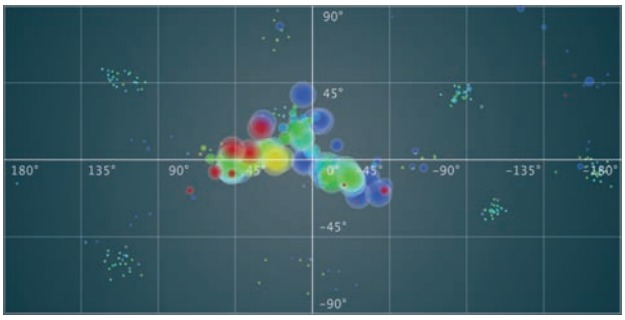

Fig. 1 An example of how one might depict the direction-of-arrival (DoA) estimates for multiple frequency bands and sectors utilizing the software implementation described in [7]. The icons for different frequencies are distinguishable by their color, whereas their relative energetic contributions to the sound-field are indicated by their size.

5 SOUND-FIELD VISUALIZATION UTILIZING DOA ESTIMATES DERIVED FROM SLAI VECTORS

5.1 Approach I: Sound-Field Visualization Based on Directly Depicting the DoA Estimates

In [7], sound-field visualization based on directly depicting multiple DoA estimates was explored, by employing the generalized SLAI formulation of Eq. (23). The DoA estimates for multiple frequency bands and sectors were depicted using icons of different color, in a similar manner as in [49]. The relative size of the icons were also influenced by the sector energy, such that the DoA estimate that corresponded with the highest sector energy, would be represented by a larger icon; an example is shown in Fig. 1.

This approach to sound-field visualization has a number of benefits when compared to its traditional activity map counterparts. First, it allows the DoA estimates for multiple frequencies to be depicted simultaneously, whereas activity maps are presented over only one specific frequency range at a time. Furthermore, since the DoA estimates are continuous, the method does not rely on dense scanning grids; rather, the estimates can be quantized to the nearest pixel, thus demanding fewer computational resources. The primary downside, however, is that the physical extent of

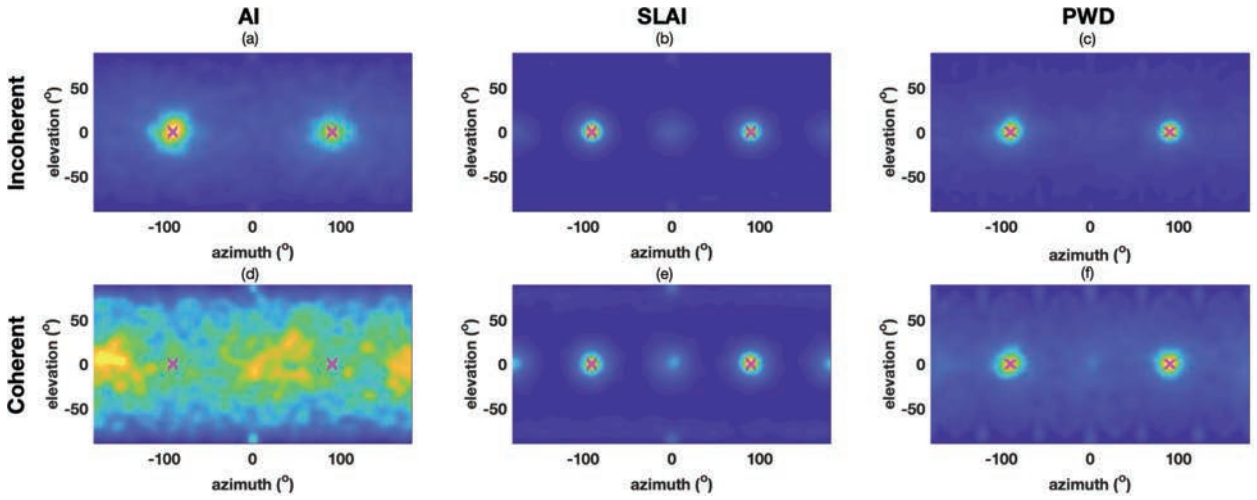

Fig. 2 Histograms of direction-of-arrival (DoA) estimates for a time window of 1 s for two incoherent (top) and two coherent sources (bottom): (a), (d) using first-order signals to obtain the standard active-intensity (AI) vector and subsequent DoA estimate; (b), (e) using Vector-Base Amplitude Panning (VBAP) patterns to generate six uniformly distributed spatially localized active-intensity (SLAI) vectors; (c), (f) using Plane-wave Decomposition (PWD) beamforming uniformly around the array. The true source directions are denoted with crosses.

sound sources becomes more ambiguous. The DoA estimation performance of the generalized formulation of Eq. (23) is evaluated in Secs. 6.1 and 6.2.

5.2 Approach II: Generating Activity Maps via Histogram Analysis of DoA Estimates

The DoA estimates derived from SLAI vectors may also be utilized to generate more traditional activity maps by applying histogram analysis on the data [50]; provided that the data is analyzed over a sufficiently long time window. This approach retains much of the computational benefits described in Sec. 5.1, while foregoing the ability to simultaneously depict the DoA estimates at different frequencies. In this approach, at each time and frequency index only the DoA of the sector with the highest energy is used in the histogram formulation. The selected estimates of a fixed time window length, which slides forward one frame at a time, form a 2-dimensional histogram and several examples are shown in Fig. 2. The histograms are smoothed by applying a circularly-symmetric Gaussian window as

$$\mathbf{h}_\mathrm{s}(\theta, \phi) = \sum_i \sum_j \mathbf{h}(i, j)\mathbf{w}_\mathrm{s}(\theta - i, \phi - j), \tag{26}$$

where $\mathbf{w}_\mathrm{s}$ is the Gaussian window of zero mean and standard deviation (std) σ_s, $\mathbf{h}(\theta, \phi)$ is the original 2-dimensional histogram and $\mathbf{h}_\mathrm{s}(\theta, \phi)$ is the smoothed variant.

In order to extract the DoA estimates, and assuming a known number of sources, G, the histograms are processed in an iterative fashion; such that for each iteration, g, the highest peak of the smoothed histogram $\mathbf{h}_s^g(\theta, \phi)$ is detected and its index is identified as the DoA of a source

$$[\theta_g, \phi_g] = \arg\max_{\theta,\phi} \mathbf{h}_s^g(\theta, \phi). \tag{27}$$

The contribution of this source to the histogram is then estimated as

$$\delta_g = \mathbf{h}_s(\theta, \phi) \odot \mathbf{w}_\mathrm{C}(\theta - \theta_g, \phi - \phi_g), \tag{28}$$

where $\mathbf{w}_\mathrm{C}(\theta, \phi)$ is a second Gaussian window of zero mean and std equal to σ_C. The contribution of the detected source is then removed from the histogram, which is then subjected to the next iteration

$$\mathbf{h}_s^{g+1}(\theta, \phi) = \mathbf{h}_s^g(\theta, \phi) - \delta_g. \tag{29}$$

The DoA estimation performance of this approach, when subjected to multiple speech sources, is evaluated in Sec. 6.3, alongside the traditional AI approach with the same histogram analysis and the PWD activity map approach.

5.3 Approach III: Sharpening Activity Maps Based on Directional Reassignment

Another option for employing SLAI vectors was explored in [10], where the approach was used to sharpen traditional PWD activity maps. The approach is, in essence, greatly inspired by the time-frequency reassignment principle; an operation that can attain high-resolution time-frequency spectrograms, without having to resort to computing Fourier transforms of much higher order [51]. In a similar manner to the time-frequency reassignment approach, this method first generates a PWD activity map but then reassigns the energies or amplitudes, for each of the scanning grid points, to new directions. These new directions correspond to the continuous SLAI-based DoA estimates, which are quantized to a much higher-resolution display grid. The approach can be realized through the following steps:

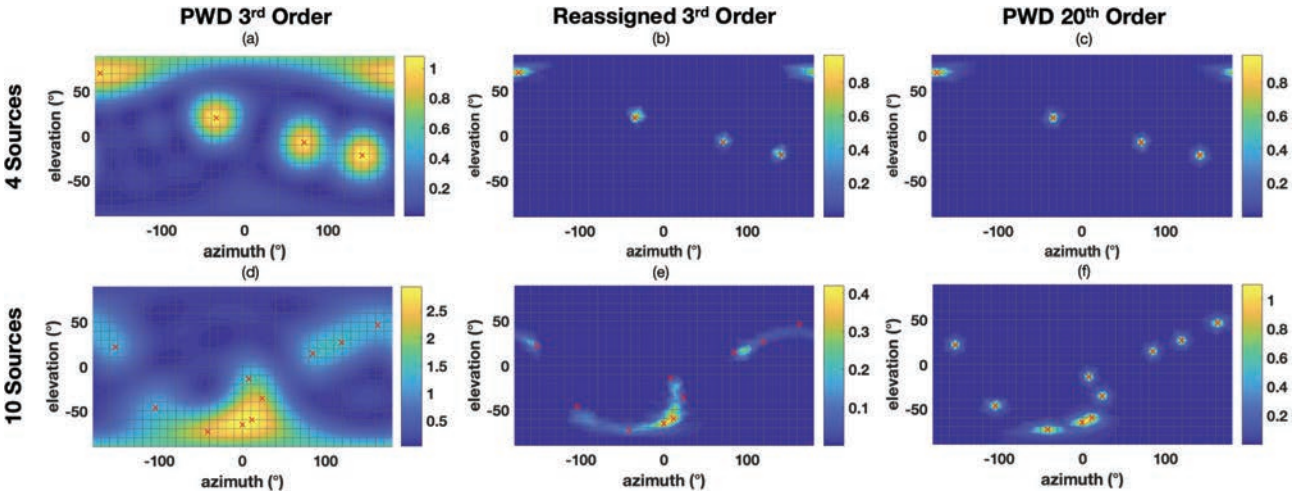

Fig. 3 Third-order activity maps utilizing Plane-wave Decomposition (PWD) (left) and the directional reassignment approach upscaled to 20th order (middle); both using a uniform scanning grid of 900 points. A 20th order PWD activity map is provided as reference (right). Two free-field sound scenes are depicted with 4 (top) and 10 (bottom) sources, respectively. The true source directions are denoted with crosses.

(a) Compute the spatially localized pressure and velocity signals as in Eq. (14), for every direction Ω_j of a low-density scanning grid.

(b) Derive the energy of the spatially localized pressure signals to generate a base-line activity map; note that when using the hypercardioid weights of Eq. (19), this base-line is equivalent to a traditional PWD activity map.

(c) Compute the SLAI vectors as in Eq. (12), and derive the DoAs Φ_j for each scanning grid direction.

(d) Reassign the sector energy based on the corresponding DoAs.

This reassignment operation (d) may be conducted in a variety of ways. The most computationally efficient approach is to quantize the DoA estimates to the nearest display grid points and subsequently accumulate the energy of the spatially localized pressure signals for each point. In the trivial case of a single plane wave source in a free-field, the DoA estimates for all scanning grid directions will be that of the true source direction regardless of the beamformer look-directions. Therefore, the energy of all the spatially localized pressure signals will be accumulated at the same point [38]. In order to preserve the energy in this case, i.e., for the peak in the reassigned activity map to have the same energy as that of the PWD activity map, and assuming a uniformly-distributed scanning grid, the sector beamformers must be normalized to preserve energy as

$$\beta_{\mathrm{EP}} \sum_{j=1}^{J} |w(\Omega_j, \Omega)|^2 = 1 \quad \text{with} \quad \beta_{\mathrm{EP}} = \frac{(L+1)^2}{J}. \tag{30}$$

Alternatively, one may also reassign by re-encoding the PWD signals $a_j = \mathbf{w}^{\mathrm{H}}(\Omega_j)\mathbf{s}$ into a higher-order $\hat{L} > L$, and subsequently compute the activity map using the traditional PWD approach on the resulting sharpened SH signals $\mathbf{s}_{\hat{L}}$

$$\mathbf{s}_{\hat{L}} = \sum_{j=1}^{J} a_j \mathbf{y}_{\hat{L}}(\Phi_j), \tag{31}$$

where $\mathbf{y}_{\hat{L}}(\Phi_j)$ is the vector of SHs for the analyzed DoA. This ultimately results in a smoother activity map, at the cost of increased computation; although, it should be noted that the SH weights for encoding may be generated off-line in an initialization stage. Furthermore, in this case, the sector beamformers should meet an amplitude-preserving condition, such as

$$\beta_{\mathrm{AP}} \sum_{j=1}^{J} w(\Omega_j, \Omega) = 1 \quad \text{with} \quad \beta_{\mathrm{AP}} = \frac{4\pi}{J}. \tag{32}$$

Some third-order examples of the PWD approach and the reassignment method up-scaled to 20th order are presented in Fig. 3, which utilized a 900-point uniformly distributed scanning grid. It can be observed that for sources that are in close proximity to each other, the energy tends to migrate in-between them. However, the reassigned activity maps are still generally closer to a PWD reference of much higher order, as is further demonstrated in Sec. 6.4.

6 EVALUATION

6.1 DoA Estimation of a Single Source in a Diffuse Field

To illustrate the direction-dependent performance of the SLAI-based DoA estimation approach, as described in Sec. 4.3, a white noise source was placed in a simulated diffuse-field (SNR = 6 dB). This field was approximated by 1442 uncorrelated white noise sources uniformly distributed around the sphere. The single white noise source was then panned in each of the 1442 grid directions and

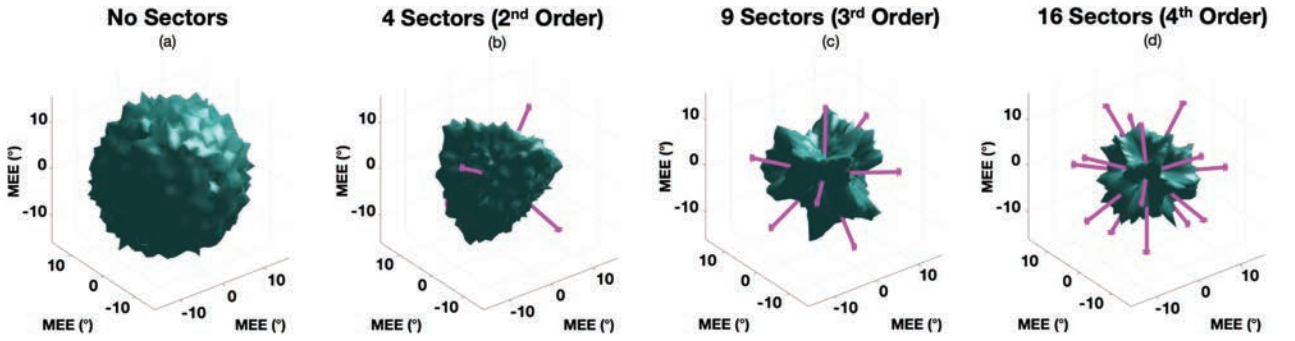

Fig. 4 The Mean Estimation Error (MEE) averaged over 0.5 s when panning a single source around a sphere using a uniform grid of 1442 points in a diffuse-field (SNR = 6 dB), including the mean and standard deviations (a) when using the standard active-intensity (AI) vector of Eq. (10) and (b), (c), (d) when employing 4, 9, and 16 uniformly distributed SLAI vectors using Eq. (23), with second, third, and fourth-order signals, respectively. The sector directions are depicted with lines protruding from the origin. The mean and standard deviations were as follows: no sectors: [14.61°, 0.79], 4 sectors: [9.88°, 2.02], 9 sectors: [7.37°, 2.28], and 16 sectors: [6.42°, 2.11].

the Mean Estimation Error (MEE) between the true source direction and the estimated direction was derived as [52]

$$\text{MEE} = \frac{1}{N_\text{F}} \sum_f \cos^{-1}(\mathbf{v}_\text{f}^\text{T} \hat{\mathbf{v}}_\text{f}) \tag{33}$$

where $\mathbf{v}_\text{f}$ and $\hat{\mathbf{v}}_\text{f}$ are unit vectors for the true and estimated DoAs, respectively; and $N_\text{F} = 128$ is the number of frequency bands used for the analysis.

Fig. 4(a) depicts the direction-dependent MEE for each source direction when utilizing the standard AI vector to derive the DoA estimate. It can be observed that the error is quite uniform, with a mean and standard deviation of 14.61° and 0.79, respectively. Figs. 4(b)–(d) depict the direction-dependent MEE when estimating the DoA using the SLAI vectors of Eq. (23), for orders $L = [2, 3, 4]$ and $S = L^2$ uniformly distributed sectors, respectively. Note that the DoA estimate with the highest sector energy was selected for the plotting. Although the variance is generally higher, the SLAI-based DoA estimates yield lower error on average, with the MEE minimally corresponding to the center of each sector. It can be further observed that the higher the order and the more sound-field segregation, the more accurate the estimates become.

6.2 DoA Estimation of Multiple Sound Sources in Simulated Environments

In order to investigate the DoA estimation performance of the generalized SLAI approach using Eq. (23), when it is presented with multiple sound sources in a more typical environment, several simulations were performed using the MCRoomSim ray-based multichannel acoustical modeler [53]. The *MediumRoom* preset was selected for the first simulation; this configures the room dimensions, absorption coefficients, and scattering coefficients to resemble a typical 10 x 8 x 3 m rectangular office room with broadband reverberation time (RT60) of approximately 0.3 s.

The receiver position was placed directly in the center of the room, with the simulator quantizing and subsequently convolving the received rays with a grid of measured IRs of a real Eigenmike32; a rigid spherical microphone array

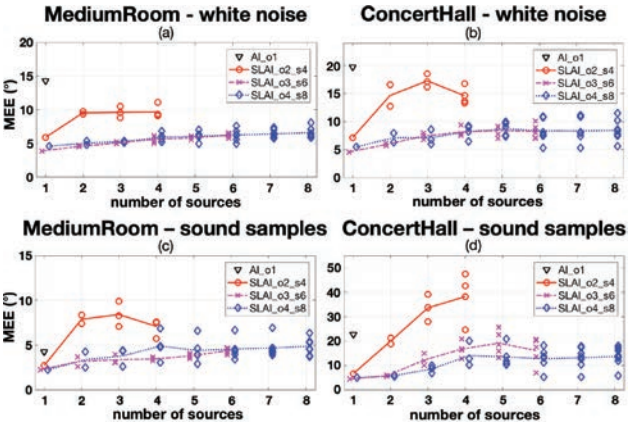

Fig. 5 The Mean Estimation Error (MEE) values when utilizing the spatially localized active-intensity (SLAI) approach with different orders and number of sectors. The markers denote the individual direction-of-arrival (DoA) estimates per source, whereas the lines denote the mean of the DoA estimates for each number of sources (where applicable); adopted from [7].

with 32 sensors. This provided the 32 simulated microphone array signals. Varying numbers of sound sources (between 1 and 8) were then introduced into the simulated room 1 m from the receiver position. The microphone array signals were encoded into SH signals using theoretical filters, as described in Sec. 2, for each test case and were band-passed to retain only frequencies between 1 and 5 kHz, where the spatial performance is high [15].

For the following results, the MEE was averaged using a weighted-sum over a 2 s window, for each DoA estimate extracted from SLAI vectors, the weights for which were derived from the normalized sector energies. Fig. 5(a) depicts the results of a MediumRoom simulation, where uncorrelated white-noise sound sources were introduced into each sector one by one, up to the number of sectors utilized for each processing order $S = 2L$. In Fig. 5(b), the same approach was employed using the 49 x 19 x 18 m *ConcertHall* preset; with broadband RT60 of 2.4 s. Figs. 5(c)–(d) used the same approach as Figs. 5(a)–(b), only the white-noise sound sources were replaced by mono audio files, which were (in the corresponding order of introduction): male

speech, female speech, cello, trumpet, piano, birds with background noise, clapping, and, finally, an acoustic guitar. The sound files were all normalized to have identical peak level.

For a single source, it can be observed in Figs. 5(a)–(d) that the SLAI approach of Eq. (23) yields a reduced MEE for all single source cases. This is because the sector-based approach essentially provides some degree of dereverberation, as the reflections and diffuse sound present in the other sectors have reduced influence on the DoA estimate. As the number of sources increases, it can be observed that the general trend is an increase in the MEE, while on the other hand, the higher the SH order and number of sectors, the lower the MEE becomes. There are some outliers, such as the test cases that employed four sources encoded into second-order SH signals in Figs. 5(b)–(c), which performed better than three sources. This could possibly be rectified by rotating the sound-field, such that the reflections of interfering sound sources for each sector are shifted away from any problematic reflecting surfaces. Furthermore, when comparing the DoA estimation performance between the white-noise sources and the audio sound sources, the silent periods in the mono audio may have led to the observable discrepancy between them, since diffuse-field DoA estimates may have been included into the weighted average.

6.3 DoA Estimation of Multiple Sound Sources Utilizing the Histogram Approach

In the next simulation, the DoA accuracy of the histogram approach described in Sec. 5.2, was evaluated with between one and six simultaneous speech sources. For these results, the SLAI design of Eq. (23) was employed, with a fixed number of six sectors. A 5.5 x 6 x 3 m sized office room with RT60 = 0.3 s was simulated also using MCRoomSim, and a virtual Eigenmike array was placed in the center of the room. The speech sources were then introduced into the simulation, one meter away from the array in the center of each sector. For the formation of the histograms, the fixed length time window was set to one second, and the std of the smoothing and contribution Gaussian windows was $\sigma_S = 10^\circ$ and $\sigma_C = 40^\circ$, respectively. For each number of sources, the MEE was evaluated in a similar manner as in Eq. (33):

$$\text{MEE} = \frac{1}{N_P N_T G} \sum_{p,t,g} \cos^{-1}\left(\mathbf{v}_{ptg}^T \hat{\mathbf{v}}_{ptg}\right), \tag{34}$$

where the mean is taken over all possible combinations of the sources around the array N_P, all time frames N_T, and all simultaneously active speakers G, using estimates that exhibited a DoA estimation error lower than 15°. Given that the sources are at the center of each sector, the value N_P differs for each number of sources, defined as the binomial coefficient $N_P = \binom{6}{G}$.

In Fig. 6, the MEE results for the DoA estimation are depicted when utilizing the SLAI vector approach and the standard AI vector and PWD activity map approaches. The different SNR conditions were attained via appropriate introduction of additive Gaussian white noise into the simulation. Also provided are tables with success rate scores (SRS), which are the percentage of estimates where the error satisfied the threshold of 15°. It is encouraged by the authors that the MEE results should be interpreted alongside these accompanying SRS. It can be observed that the error of the AI approach is always greater than both the SLAI and PWD approaches, as it also exhibits the worst SRS. In the case of coherent sources, the AI method struggles to attain estimates with an error lower than 50%. This inability of the AI approach to properly address coherent sources is also visually depicted in Fig. 2(d), where the majority of the estimates tend to fall in between the true position of the two coherent sources. In the same figure, the PWD approach appears to correctly estimate the DoA of both coherent and incoherent sources; however, it tends to yield more noisy histograms when compared to the SLAI approach. The superiority of the SLAI-based approach is also reflected both in the MEE results, as well as in the SRS values for all the test cases.

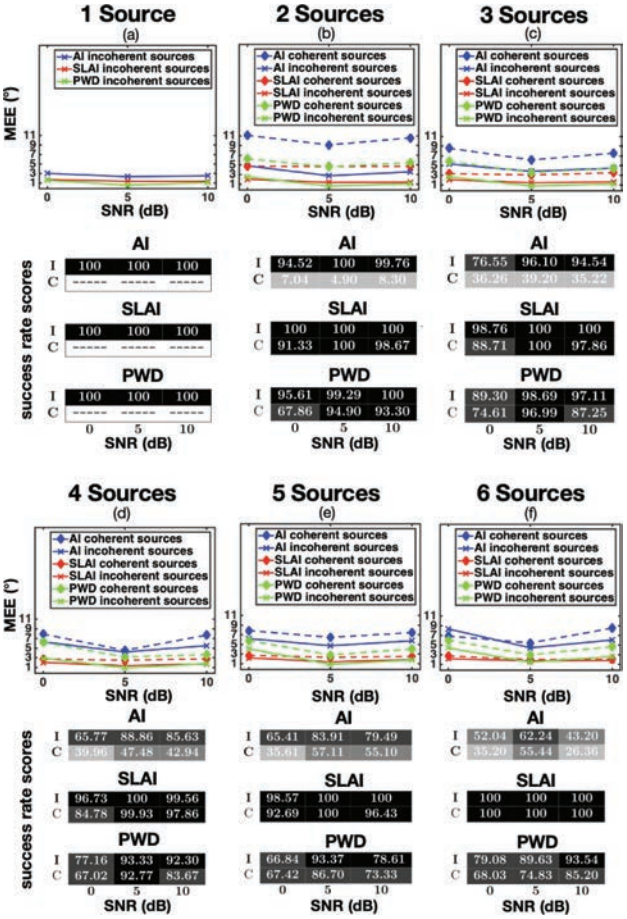

Fig. 6 Mean Estimation Error (MEE) of histogram direction-of-arrival (DoA) estimates for incoherent (I) and coherent (C) speech sources and corresponding success rate scores in a simulated environment of RT60 = 0.3 s.

6.4 Sharpening of Traditional PWD Activity Maps via Directional Reassignment

To evaluate the performance of employing SLAI-based DoA estimates to sharpen traditional PWD activity maps,

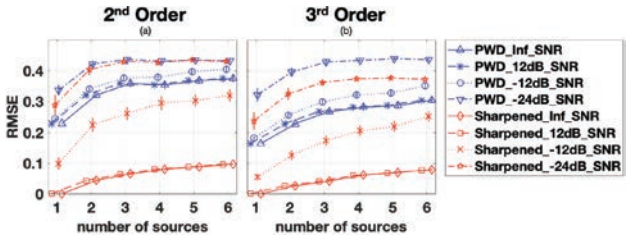

Fig. 7 The means and 95% confidence intervals of the root-mean-square error (RMSE) values derived from the residual, which corresponds to how dissimilar the output activity maps are, when compared to the 20th-order Plane-wave Decomposition (PWD) reference case; adapted from [10].

as described in Sec. 5.3, several sound scenes were synthesized with between 1 and 6 incoherent noise sources at various SNRs. These sound scenes were encoded into both second and third-order. The PWD activity maps, both with and without directional reassignment, were then generated. Note that the reassignment was up-scaled to the 20th order as in Eq. (31). These activity maps were then subtracted from a 20th-order PWD activity map, which served as the high-resolution reference case, and the root-mean-square error (RMSE) of the residual was subsequently computed. This process was repeated 100 times with randomly assigned source directions for each case. The means and 95% confidence intervals for the RMSE at the various SNR values and number of sources are shown in Fig. 7.

It can be observed that the reassignment approach yields a lower RMSE than the traditional PWD for the majority of test cases, where the performance benefit of the reassignment approach is inversely proportional to the SNR of the scene. For negative SNR values, the RMSE increases more drastically and eventually converges to that of the performance of the PWD approach. Therefore, even in the worst case scenario, the reassignment approach yields results that are never inferior to the PWD method.

7 SUMMARY AND CONCLUSION

This article has detailed and evaluated three approaches to sound-field visualization, which all employ the use of spatially localized active-intensity (SLAI) vectors. These SLAI vectors are simply active-intensity (AI) vectors that are derived after the sound-field has been spatially weighted. This subsequently biases the AI vector into favoring energetic contributions made within a certain spatially localized sector. Provided that sound sources and/or reflections can be segregated into their own sector, the approach has been shown to be more robust than the traditional AI approach.

Once a sound-field has been segregated and multiple DoA estimates extracted from the SLAI vectors, the sound-field may be visualized by either: (I) directly depicting the estimates as icons, with their relative size dictated by the corresponding energy of each sector; (II) generating traditional activity maps via histogram analysis of the DoA estimates; or (III) by using the DoA estimates to reassign energy and subsequently sharpen traditional beamformer-based activity maps.

Since the SLAI-based DoA estimates are continuous, they may be quantized to a grid of much higher density than would be practical for scanning-based alternatives. Furthermore, as much of the required computations for these approaches may be performed during an initialization stage, the run-time computational complexity remains relatively low. Additionally, unlike many high-resolution sound-field visualization alternatives, these approaches do not require an estimation of the number of sources and make no assumptions regarding the sound-field conditions. Following multiple evaluations comprising various simulated acoustical environments it was found that the SLAI-based approaches are more robust in identifying the DoA of multiple sound sources when compared to the traditional AI approach. Finally, when compared to the PWD approach of the same order, the reassignment approach yielded activity maps which more closely resembled the selected high-resolution reference.

8 ACKNOWLEDGMENTS

This research has received funding from the Aalto University Doctoral School of Electrical Engineering and the Academy of Finland project no 317341. This work was also supported by the Italian Ministry of Economic Development (MISE), FUND FOR THE SUSTAINABLE GROWTH (F.C.S.) under grant agreement (CUP) B48I15000130008, project VASM (Vehicle Active Sound Management).

9 REFERENCES

[1] B. Rafaely, *Fundamentals of Spherical Array Processing*, vol. 8 (Springer, 2015).

[2] M. Brandstein and D. Ward, *Microphone Arrays: Signal Processing Techniques and Applications* (Springer Science & Business Media, 2013).

[3] S. Doclo and M. Moonen, "GSVD-Based Optimal Filtering for Multi-microphone Speech Enhancement," in *Microphone Arrays*, pp. 111–132 (Springer, 2001).

[4] O. Thiergart, M. Taseska, and E. A. Habets, "An Informed Parametric Spatial Filter Based on Instantaneous Direction-of-Arrival Estimates," *IEEE/ACM Trans. Audio, Speech and Lang. Proc. (TASLP)*, vol. 22, no. 12, pp. 2182–2196 (2014).

[5] D. P. Jarrett, E. A. Habets, and P. A. Naylor, *Theory and Applications of Spherical Microphone Array Processing* (Springer, 2017).

[6] F. J. Fahy and V. Salmon, "Sound Intensity," *J. Acoust. Soc. Amer.*, vol. 88, no. 4, pp. 2044–2045 (1990).

[7] L. McCormack, S. Delikaris-Manias, A. Farina, D. Pinardi, and V. Pulkki, "Real-Time Conversion of Sensor Array Signals Into Spherical Harmonic Signals with Applications to Spatially Localized Sub-band Sound-Field Analysis," presented at the *144th Convention of the Audio Engineering Society* (2018 May), convention paper 9939.

[8] S. Delikaris-Manias, D. Pavlidi, A. Mouchtaris, and V. Pulkki, "DOA Estimation with Histogram Analysis of Spatially Constrained Active Intensity Vectors," *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 526–530 (2017).

[9] S. Delikaris-Manias, L. McCormack, D. Pavlidi, and A. Mouchtaris, "Spatially Localized Direction of Arrival Estimation," *European Congress and Exposition on Noise Control Engineering (Euronoise)*, pp. 2549–2554 (2018).

[10] L. McCormack, A. Politis, and V. Pulkki, "Sharpening of Angular Spectra Based on a Directional Re-Assignment Approach for Ambisonic Sound-Field Visualization," *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 576–580 (2019).

[11] H. L. Van Trees, *Optimum Array Processing: Part IV of Detection, Estimation, and Modulation Theory* (John Wiley & Sons, 2004).

[12] M. A. Gerzon, "Periphony: With-Height Sound Reproduction," *J. Audio Eng. Soc.*, vol. 21, no. 1, pp. 2–10 (1973).

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

[14] H. Teutsch, *Modal Array Signal Processing: Principles and Applications of Acoustic Wavefield Decomposition*, vol. 348 (Springer, 2007).

[15] S. Moreau, J. Daniel, and S. Bertet, "3D Sound Field Recording with Higher Order Ambisonics—Objective Measurements and Validation of Spherical Microphone," presented at the *120th Convention of the Audio Engineering Society* (2006 May), convention paper 6857.

[16] C. T. Jin, N. Epain, and A. Parthy, "Design, Optimization and Evaluation of a Dual-Radius Spherical Microphone Array," *IEEE/ACM Trans. Audio, Speech and Lang. Proc. (TASLP)*, vol. 22, no. 1, pp. 193–204 (2014).

[17] A. Politis and H. Gamper, "Comparing Modeled and Measurement-Based Spherical Harmonic Encoding Filters for Spherical Microphone Arrays," presented at the *2017 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)* (2017).

[18] A. Farina, S. Campanini, L. Chiesi, A. Amendola, and L. Ebri, "Spatial Sound Recording With Dense Microphone Arrays," presented at the *AES 55th International Conference: Spatial Audio* (2014 Aug.), conference paper P-10.

[19] C. Schörkhuber and R. Höldrich, "Ambisonic Microphone Encoding With Covariance Constraint," *Proceedings of the International Conference on Spatial Audio* (2017).

[20] D. P. Jarrett, E. A. Habets, and P. A. Naylor, "3D Source Localization in the Spherical Harmonic Domain Using a Pseudointensity Vector," *2010 18th European Signal Processing Conference*, pp. 442–446 (2010).

[21] D. L. Alon and B. Rafaely, "Spatial Decomposition by Spherical Array Processing," in *Parametric Time-Frequency Domain Spatial Audio*, pp. 25–48 (Wiley Online Library, 2017).

[22] M. D. Zoltowski, "On the Performance Analysis of the MVDR Beamformer in the Presence of Correlated Interference," *IEEE Trans. Acoust., Speech, and Signal Proc.*, vol. 36, no. 6, pp. 945–947 (1988).

[23] D. P. Jarrett and E. A. Habets, "On the Noise Reduction Performance of a Spherical Harmonic Domain Tradeoff Beamformer," *IEEE Signal Proc. Lett.*, vol. 19, no. 11, pp. 773–776 (2012).

[24] S. Delikaris-Manias and V. Pulkki, "Cross Pattern Coherence Algorithm for Spatial Filtering Applications Utilizing Microphone Arrays," *IEEE Trans. Audio, Speech, and Lang. Proc.*, vol. 21, no. 11, pp. 2356–2367 (2013).

[25] S. Delikaris-Manias, J. Vilkamo, and V. Pulkki, "Signal-Dependent Spatial Filtering Based on Weighted-Orthogonal Beamformers in the Spherical Harmonic Domain," *IEEE/ACM Trans. Audio, Speech and Lang. Proc. (TASLP)*, vol. 24, no. 9, pp. 1507–1519 (2016).

[26] L. McCormack, S. Delikaris-Manias, and V. Pulkki, "Parametric Acoustic Camera for Real-Time Sound Capture, Analysis and Tracking," *Proceedings of the 20th International Conference on Digital Audio Effects (DAFx-17)*, pp. 412–419 (2017).

[27] K. Han and A. Nehorai, "Improved Source Number Detection and Direction Estimation With Nested Arrays and ULAs Using Jackknifing," *IEEE Trans. Signal Proc.*, vol. 61, no. 23, pp. 6118–6128 (2013).

[28] K. M. Wong, Q.-T. Zhang, J. P. Reilly, and P. Yip, "On Information Theoretic Criteria for Determining the Number of Signals in High Resolution Array Processing," *IEEE Trans. Acoust., Speech, and Signal Proc.*, vol. 38, no. 11, pp. 1959–1971 (1990).

[29] R. Goossens and H. Rogier, "Unitary Spherical ESPRIT: 2-D Angle Estimation With Spherical Arrays for Scalar Fields," *IET Signal Proc.*, vol. 3, no. 3, pp. 221–231 (2009).

[30] S. Tervo and A. Politis, "Direction of Arrival Estimation of Reflections From Room Impulse Responses Using a Spherical Microphone Array," *IEEE/ACM Trans. Audio, Speech and Lang. Proc. (TASLP)*, vol. 23, no. 10, pp. 1539–1551 (2015).

[31] L. Bianchi, M. Verdi, F. Antonacci, A. Sarti, and S. Tubaro, "High Resolution Imaging of Acoustic Reflections With Spherical Microphone Arrays," presented at the *2015 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)* (2015).

[32] B. Gunel, H. Hacihabiboglu, and A. M. Kondoz, "Acoustic Source Separation of Convolutive Mixtures Based on Intensity Vector Statistics," *IEEE Trans. Audio, Speech and Lang. Proc.*, vol. 16, no. 4, p. 748 (2008).

[33] O. Thiergart, R. Schultz-Amling, G. Del Galdo, D. Mahne, and F. Kuech, "Localization of Sound Sources in Reverberant Environments Based on Directional Audio Coding Parameters," presented at the *127th Convention of the Audio Engineering Society* (2009 Oct.), convention paper 7853.

[34] S. Tervo, "Direction Estimation Based on Sound Intensity Vectors," *17th European Signal Processing Conference, 2009*, pp. 700–704 (2009).

[35] D. Levin, E. A. Habets, and S. Gannot, "On the Angular Error of Intensity Vector Based Direction of Arrival Estimation in Reverberant Sound Fields," *J. Acoust. Soc. Amer.*, vol. 128, no. 4, pp. 1800–1811 (2010).

[36] V. Pulkki, "Spatial Sound Reproduction with Directional Audio Coding," *J. Audio Eng. Soc.*, vol. 55, pp. 503–516 (2007 Jun.).

[37] V. Pulkki, A. Politis, M.-V. Laitinen, J. Vilkamo, and J. Ahonen, "First-Order Directional Audio Coding (DirAC)," in *Parametric Time-Frequency Domain Spatial Audio*, pp. 89–138 (John Wiley & Sons, 2007).

[38] A. Politis, J. Vilkamo, and V. Pulkki, "Sector-Based Parametric Sound Field Reproduction in the Spherical Harmonic Domain," *IEEE J. Selec. Topics Signal Proc.*, vol. 9, no. 5, pp. 852–866 (2015).

[39] A. Politis, L. McCormack, and V. Pulkki, "Enhancement of Ambisonic Binaural Reproduction Using Directional Audio Coding with Optimal Adaptive Mixing," presented at the *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics* (2017).

[40] A. Politis and V. Pulkki, "Acoustic Intensity, Energy-Density and Diffuseness Estimation in a Directionally-Constrained Region," *arXiv:1609.03409* (2016).

[41] R. Baumgartner, H. Pomberger, and M. Frank, "Practical Implementation of Radial Filters for Ambisonic Recordings," presented at the *Proc. Int. Conf. Spatial Audio (ICSA)* (2011).

[42] B. Bernschütz, C. Pörschmann, S. Spors, and S. Weinzierl, "Soft-Limiting der modalen Amplitudenverstärkung bei sphärischen Mikrofonarrays im Plane Wave Decomposition Verfahren,"*Proceedings of the 37th Deutsche Jahrestagung für Akustik (DAGA)*, pp. 661–662 (2011).

[43] S. Lösler and F. Zotter, "Comprehensive Radial Filter Design for Practical Higher-Order Ambisonic Recording," *Fortschritte der Akustik (DAGA)*, pp. 452–455 (2015).

[44] F. Zotter, "A Linear-Phase Filter-Bank Approach to Process Rigid Spherical Microphone Array Recordings," presented at the *Proc. IcETRAN* (2018).

[45] H. Hacihabiboglu, "Theoretical Analysis of Open Spherical Microphone Arrays for Acoustic Intensity Measurements," *IEEE/ACM Trans. Audio, Speech and Lang. Proc. (TASLP)*, vol. 22, no. 2, pp. 465–476 (2014).

[46] A. Politis, *Microphone Array Processing for Parametric Spatial Audio Techniques*, *Ph.D. thesis*, Aalto University, Finland (2016).

[47] F. Zotter and M. Frank, "All-Round Ambisonic Panning and Decoding," *J. Audio Eng. Soc.*, vol. 60, pp. 807–820 (2012 Oct.).

[48] V. Pulkki, "Virtual Sound Source Positioning Using Vector Base Amplitude Panning," *J. Audio Eng. Soc.*, vol. 45, pp. 456–466 (1997Jun.).

[49] A. Farina and L. Tronchin, "3D Sound Characterization in Theaters Employing Microphone Arrays," *Acta acustica united with Acustica*, vol. 99, no. 1, pp. 118–125 (2013).

[50] D. Pavlidi, S. Delikaris-Manias, V. Pulkki, and A. Mouchtaris, "3D DOA Estimation of Multiple Sound Sources Based on Spatially Constrained Beamforming Driven by Intensity Vectors," *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 96–100 (2016).

[51] P. Flandrin, F. Auger, and E. Chassande-Mottin, "Time-Frequency Reassignment: From Principles to Algorithms," in *Applications in Time-Frequency Signal Processing*, p. 102 (CRC Press, 2003).

[52] S. Delikaris-Manias, D. Pavlidi, V. Pulkki, and A. Mouchtaris, "3D Localization of Multiple Audio Sources Utilizing 2D DOA Histograms," *2016 24th European Signal Processing Conference (EUSIPCO)*, pp. 1473–1477 (2016).

[53] A. Wabnitz, N. Epain, C. Jin, and A. Van Schaik, "Room Acoustics Simulation for Multichannel Microphone Arrays," *Proceedings of the International Symposium on Room Acoustics*, pp. 1–6 (2010).

THE AUTHORS


Leo McCormack


Symeon Delikaris-Manias


Archontis Politis


Despoina Pavlidi


Daniel Pinardi


Angelo Farina


Ville Pulkki

Leo McCormack received his M.Sc. degree in computer communications and information sciences, majoring in acoustics and audio technology, at Aalto University, Finland, and his B.Sc. in music technology and audio systems at the University of Huddersfield, UK. He was a student intern at Fraunhofer IIS, Erlangen, Germany, in 2013–2014 and received an AES Convention "best peer-reviewed paper" award in 2018. He is currently a doctoral candidate for the Department of Signal Processing and Acoustics at Aalto University, Finland, researching into parametric spatial audio technologies. His research interests include signal processing for sound-field manipulation and reproduction, microphone and hydrophone array processing, and sound source localization. Leo McCormack is also a lead developer for the open-source SPARTA and COMPASS audio plug-in libraries and is a member of the Audio Engineering Society and IEEE Signal Processing Society.

•

Dr. Symeon Delikaris-Manias received his D.Sc. in electrical engineering with distinction at Aalto Acoustics Lab at Aalto University, Finland, his M.Sc. degree in sound and vibration from the Institute of Sound and Vibration Research (ISVR), UK, and his B.Sc. degree in mathematics from the University of Crete, Greece. He received awards from the Audio Engineering Society, The Foundation of Aalto Science and Technology, the Nokia Foundation, and the International College of the European University of Brittany. He is an editor and author of the first parametric spatial audio book from Wiley and IEEE Press. Between 2008–2010, he was employed by GSacoustics in London working on acoustic modeling and auralization of concert halls. In 2010–2011 he was at the Center for Virtual Reality, Brest, France, developing and evaluating multichannel recording and reproduction techniques for virtual reality applications. His main research interests are signal-dependent spatial audio processing, array signal processing, and underwater spatial audio. Delikaris-Manias is a member of IEEE Signal Processing Society. In his free time he enjoys scuba diving.

•

Archontis Politis obtained his M.Eng. degree in civil engineering from Aristotle University, Thessaloniki, Greece, and his M.Sc. degree in sound & vibration studies from the Institute of Sound and Vibration Research (ISVR), Southampton, UK, in 2006 and 2008, respectively. From 2008 to 2010 he worked as a graduate acoustic consultant in Arup Acoustics, UK, and as a researcher in a joint collaboration between Arup Acoustics and the Glasgow School of Arts on architectural auralization using spatial sound techniques. In 2016 he obtained a Doctor of Science degree on parametric spatial sound recording and reproduction from Aalto University, Finland. He also completed an internship at the Audio and Acoustics Research Group of Microsoft Research during the summer of 2015. He received an AES Convention best student paper award in 2013. He has held workshops on parametric spatial audio processing, and he has co-authored a book on the topic, published in 2017. He is currently a post-doctoral researcher at Tampere University, Finland. His research interests include spatial audio technologies, virtual acoustics, array signal processing, and acoustic scene analysis.

•

Despoina Pavlidi received the diploma degree in electrical and computer engineering in 2009 from the National Technical University of Athens (NTUA), Greece, and the M.Sc. and Ph.D. degrees in computer science in 2012 and 2018, respectively, from the Computer Science Department of the University of Crete, Greece. Since 2010 she is affiliated with the Institute of Computer Science at the Foundation for Research and Technology-Hellas (FORTH-ICS) and the Computer Science Department, University of Crete (UOC-CSD) as a research assistant. Since March 2019 she is also affiliated with the institute of Applied and Computational Mathematics at the Foundation for Research and Technology-Hellas (FORTH-IACM) as a postdoctoral Stavros Niarchos Foundation fellow. Her research interests include underwater acoustics, hyrdophone arrays, cetaceans passive localization, audio signal processing, microphone arrays, 2D and 3D

sound source localization, spherical harmonic domain analysis, networked music performance (NMP) systems, and audio coding.

•

Angelo Farina got his Master Degree in civil engineering in December 1982 at the University of Bologna, Italy, with a thesis on the acoustics and vibrations inside a tractor cab. In 1987 he got a PhD in applied physics at the University of Bologna with a thesis on experimental assessment of concert hall acoustics. He was a full-time researcher since 1st November 1986 at the University of Bologna and since 1st March 1992 at the University of Parma, where he became Associate Professor on 1st November 1998 and Full Professor of Environmental Applied Physics since 1st May 2005, where he has the chair of Applied Acoustics (M.D. Engineering) and Environmental Applied Physics (M.D. Architecture). During his academic career Angelo worked in several fields of applied acoustics, including noise and vibration, concert hall acoustics, simulation software, and advanced measurement systems. In the last 10 years Angelo focused mostly on applications involving massive microphone and loudspeaker arrays. In 2008 Angelo Farina was awarded with the AES fellowship for his pioneering work on electroacoustic measurements based on exponential sine sweeps. Angelo is author of more than 250 scientific papers and three widely employed software packages (Ramsete, Aurora Plugins, DISIA).

•

Daniel Pinardi got a Master Degree in mechanical engineering at the University of Parma, Italy, in July 2016. During his studies met Angelo Farina, who became his supervisor for an experimental thesis on automotive loudspeakers. Now he is continuing his studies on acoustics as a Ph.D. student and research fellow, working on a project about the application of active cancellation of random noise on automotive vehicles. Daniel is also working on a new project making use of microphone arrays for monitoring environmental noise both in air and underwater.

•

Ville Pulkki a professor in the Department of Signal Processing and Acoustics in Aalto University, Helsinki, Finland. He has been working in the field of spatial audio for over 20 years. He developed the vector-base amplitude panning (VBAP) method in his Ph.D. (2001) and directional audio coding after the Ph.D. with his research group. He also has contributions in perception of spatial sound, laser-based measurement of room responses, and binaural auditory models. He has received the Samuel Warner memorial medal from SMPTE and Silver Medal from AES. He enjoys being with his family, building his summer house, and performing in musical ensembles.