Audio Engineering Society

Convention Paper

Presented at the 144th Convention
2018 May 23 – 26, Milan, Italy

Real-time conversion of sensor array signals into spherical harmonic signals with applications to spatially localised sub-band sound-field analysis

Leo McCormack[1], Symeon Delikaris-Manias[1], Angelo Farina[2], Daniel Pinardi[2], and Ville Pulkki[1]

[1]*Aalto Acoustics Lab, Aalto University, Espoo, Finland*
[2]*University of Parma, Industrial Engineering Dept., Parma, Italy*

Correspondence should be addressed to Leo McCormack (leo.mccormack@aalto.fi)

ABSTRACT

This paper presents two real-time audio plug-ins for processing sensor array signals for sound-field visualisation. The first plug-in utilises spherical or cylindrical sensor array specifications to provide analytical spatial filters, which encode the array signals into spherical harmonic signals. The second plug-in utilises these intermediate signals to estimate the direction-of-arrival of sound sources, based on a spatially localised pressure-intensity (SLPI) approach. The challenge with the traditional pressure-intensity (PI) sound-field analysis, is that it performs poorly when presented with multiple sound sources with similar spectral content. Test results indicate that the proposed SLPI approach is capable of identifying sound source directions with reduced error in various environments, when compared to the PI method.

1 Introduction

Capturing a sound-field by utilising an array of sensors offers many advantages over using a single sensor. These advantages largely stem from the fact that certain spatial attributes of the sound-field are encoded into the array signals, where the spatial resolution of this encoding is predominately influenced by the number of sensors and their orientation. Essentially, whereas a single sensor can be used to establish whether a sound source is present in a sound scene, a sensor array can be utilised to generate spatially selective filters (referred to as beamformers) or extract meaningful spatial parameters, such as the direction-of-arrival (DoA) of sound sources. These beamformers and spatial parameters can be utilised in parametric reproduction methods for sound-field auralisation purposes, or utilised by sound-field visualisation systems. This paper is primarily focused on the latter function; however, parametric reproduction methods are also discussed as they have had influence on the algorithms presented later in the paper.

While is it possible to derive spatial parameters and generate beamformers using the sensor array signals directly, this approach generally results in the algorithms being tailored for a specific array; as the array specifications become an integral part of the algorithms. Therefore, a more flexible approach is to first transform the sensor array signals into an array-independent domain; whereby the sound-field is decomposed into orthonormal basis functions about the unit sphere, by means of a spherical harmonic transform (SHT) [1]. Essentially, by spatially encoding the sensor array sig-

nals into spherical harmonic signals (also known as Ambisonic or B-Format signals), the array specifications are largely abstracted away from the algorithms that utilise them.

There are two main approaches for performing a sensor array SHT. In the case of spherical and cylindrical arrays, analytical solutions are available [2, 3], which describe how plane waves interact with the sensor array. These formulae take into account certain aspects of the array, such as the radius and whether it has an open or rigid baffle construction. However, currently there exists no generalised real-time software for spatially encoding sensor array signals into spherical harmonic signals, using this theoretical approach. For atypical sensor array shapes, for which no analytical solution can be easily derived, the more generalised measurement-based filters approach is more appropriate. This approach requires the measurement of the sensor array impulse responses (IRs) from many directions, in order to design the encoding filters [4, 5]. These measurement-based filters can be applied in real-time using, for example, the X-volver audio plug-in [6]. Note that the spatial resolution of this transform is largely dependent on the number of sensors, as this determines the highest possible order of spherical harmonic expansion [1].

Today, there exists a plethora of algorithms and methods that utilise spherical harmonic signals. For instance, many spatial audio systems rely on beamforming as a means to reproduce or visualise a captured sound-field. Regarding the latter function, one approach is to generate a beamformer and steer it towards multiple directions that sample a particular area of interest; an approach commonly referred to as *scanning beamforming*. Popular beamformers include the Plane-wave Decomposition (PWD) algorithm, which (in its most basic form) is simply derived as a weighted sum of the spherical harmonic signals; where the weights are the spherical harmonics for a direction on the sphere [7]. On the other hand, adaptive beamformers, such as the Minimum-variance Distotion-less Response (MVDR) algorithm, can yield increased spatial selectivity; however, they generally require increased computation and are susceptible to coherent interferers [8]. Spatial post-filters, such as the Cross-Pattern Coherence (CroPaC) algorithm [9, 10], may also be utilised to improve the performance of beamformers, especially in reverberant or noisy environments.

Once an area has been subjected to a scanning beamformer, the DoA of the most prevalent sound sources in the target area can be ascertained by determining which directions yielded the highest beamformer signal energy. Naturally, this approach is precise to the degree of separation between the scanning grid directions; therefore, the computational requirements can be significant when high precision is required. If the relative energy of a scanning beamformer is represented with a colour gradient, the result can be described as a *power-map*, which is a convenient format for representing the relative acoustical energy for multiple directions. Systems which overlay this information on top of a corresponding video stream in real-time, are referred to as *acoustic cameras*; for which a spherical harmonic domain implementation can be found in [11].

Regarding sound-field reproduction methods for auralisation purposes, perhaps the most prominent approach is the Ambisonics method [12], which is a linear mapping of the spherical harmonic signals to the loudspeaker channels using appropriate gain factors. There are several perceptually motivated approaches to ambisonic decoding [13, 14, 15] and the most basic ambisonic decoder is simply a matrix of static beamforming steering vectors. However, the perceptual performance of these methods is largely dependent on the spatial resolution of the input format, which can result in poor localisation accuracy and colouration [16], when utilising first-order and lower-order spherical harmonic signals.

Therefore, perceptually-motivated parametrically enhanced alternatives, such as: Directional Audio Coding (DirAC) [17], enhancement of linear decoders with excessively narrow beamformers [18], and High Angular Resolution plane-wave Expansion (HARPEX) [19] have been introduced. These approaches yield improved perceived spatial accuracy, at the cost of increased computational and implementation complexity [20]. The enhancement of linear decoders using excessively directive beamformers [18] is an approach whereby the directivity of linear decoders is sharpened. This is performed by imposing parameters derived from a set of beamformers that match the spatial selectivity of a panning function or head-related transfer function [21]. Whereas, in the case of DirAC, the perceptual improvements lay with the extraction of the DoA and diffuseness parameters per time-frequency index from the pressure-intensity (PI) vector, which DirAC utilises to synthesise individual direct and diffuse audio

streams. The HARPEX algorithm follows a similar approach to DirAC and also visualises the DoA estimates in the software implementation, but it does not utilise a diffuseness estimate for the parametric reproduction.

Ascertaining the DoA from the flow of acoustical energy, described by the PI vector, is regarded as a computationally efficient method when compared to power-map based alternatives [20]. However, in scenarios where there are multiple sound sources over-lapping in both time and frequency, the flow of energy will be located in a direction between the sources. Therefore, DirAC has recently been expanded to utilise higher-order spherical harmonic signals [22, 23] to impose spatially selective sectors on the first-order signals, and subsequently extract a separate DoA estimate from each spatially localised pressure-intensity (SLPI) vector. Provided that sound sources are located in their own individual sector, this approach has been shown to be more robust for reproduction purposes; resulting in higher perceived spatial accuracy.

While the recent formulations of DirAC have utilised the SLPI approach to improve sound-field parametric reproduction for auralisation purposes, there is currently no real-time system that utilises this approach for computationally efficient sound-field visualisation. Furthermore, current spatial encoders that are deployed in the market place today, are confined to operate only on specific microphone arrays and offer limited flexibility for tuning the filter parameters. This is especially problematic for situations where the spatial encoding filters are required to be algorithm dependent; for example, ambisonic reproduction systems are very sensitive to sensor noise amplification, while certain post-filters can mitigate this noise [9]. Therefore, the main contributions for this paper can be summarised as:

- The development of a real-time audio plug-in that utilises theoretical filters to spatially encode sensor array signals into spherical harmonic signals, up to 7th order of spherical harmonic expansion.
- The design and development of a real-time DoA estimator based on a generalised SLPI algorithm.

This paper is organised as follows: Section 2 provides background regarding encoding sensor array signals into spherical harmonic signals; Section 3 describes the traditional PI algorithm for DoA estimation; Section 4 details the proposed SLPI algorithm; Section 5 then presents the real-time implementations for both theoretical spatial encoding of sensor array signals and the proposed SLPI-based DoA estimator; Section 6 evaluates the algorithms utilised in both plug-ins; and Section 7 concludes the paper.

2 Spatial encoding

Before expanding upon the proposed SLPI-based DoA estimator, the sensor array signals must first be transformed into the spherical harmonic domain.

2.1 Theoretical encoding

Since spatial encoding is frequency-dependent, the first step is to transform the sensor array signals $\mathbf{x}(t)$ into the time-frequency domain $\hat{\mathbf{x}}(t,f)$ by means of either a short-time Fourier transform, or perfect reconstruction filterbank; were t and f refer to the down-sampled time and frequency indices, respectively.

Now consider a spherical or cylindrical sensor array, denoted with Q sensors 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 spherical harmonic transform can then be utilised to decompose the array signals, $\hat{\mathbf{x}} \in \mathbb{C}^{Q\times 1}$, into a set of spherical harmonic signals for each frequency band. The accuracy of this decomposition depends on the sensor distribution on the surface of the array, the type of the array and the radius [1]. The total number of sensors defines the highest order of spherical harmonic signals L that can be estimated. Please note that the frequency and time indices are omitted for the brevity of notation.

The spherical harmonic signals can be estimated as

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

where

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

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

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

where $\mathbf{Y}_e \in \mathbb{R}^{Q\times(L+1)^2}$ is the spherical harmonic encoding matrix which estimates the pressure on the surface

of the sphere or cylinder; and $\mathbf{W}_l \in \mathbb{C}^{(L+1)^2 \times (L+1)^2}$ is an equalisation matrix that eliminates the effect of the sphere or cylinder.

The frequency-independent spherical harmonic encoding matrix can be calculated for either uniform or non-uniform arrangements as

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

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

$$\mathbf{Y}(\Omega_q) = \begin{bmatrix} Y_{00}(\Omega_1) & Y_{-11}(\Omega_1) & \dots & Y_{LL}(\Omega_1) \\ Y_{00}(\Omega_2) & Y_{-11}(\Omega_2) & \dots & Y_{LL}(\Omega_2) \\ Y_{00}(\Omega_3) & Y_{-11}(\Omega_3) & \dots & Y_{LL}(\Omega_3) \\ \vdots & \vdots & \vdots & \vdots \\ Y_{00}(\Omega_Q) & Y_{-11}(\Omega_Q) & \dots & Y_{LL}(\Omega_Q) \end{bmatrix}, \tag{5}$$

where Y_{lm} are the individual real-valued spherical harmonics of order $l \geq 0$, degree $m \in [-l, l]$ and sensor direction Ω.

The equalisation matrix can be defined as

$$\mathbf{W}_l = \begin{bmatrix} w_0 & & & & & \\ & w_1 & & & & \\ & & w_1 & & & \\ & & & w_1 & & \\ & & & & \ddots & \\ & & & & & w_L \end{bmatrix}, \tag{6}$$

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

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

where the modal coefficients take into account whether the array construction is open or rigid and the directivity of the sensors (cardioid, dipole 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 sensors; in order to cater for cases in which the sensors protrude from the surface of the array. For more details on calculating these modal coefficients, the reader is directed to [2, 3].

However, the problem with directly inverting the modal coefficients in (7), is that the internal noise of the sensors can be amplified significantly, especially at low frequencies and higher orders. Therefore, a regularised inversion is more preferable. One popular approach is the Tikhonov method [4]

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

where λ is a regularisation parameter that influences the sensor noise amplification. Another popular approach is the soft limiting regularisation, formulated as [24]

$$w_l = \frac{2\lambda}{\pi}\frac{|b_l|}{b_l}\arctan\left(\frac{\pi}{2\lambda|b_l|}\right). \tag{9}$$

Both approaches allow 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 calculating the equalisation matrix $\mathbf{W}_l$, the reader is referred to [25, 26, 7].

2.2 Arbitrary encoding via impulse response measurements

While the theoretical encoding approach is applicable for spherical arrays with uniform or nearly-uniform sensors arrangement and cylindrical arrays with sensors distributed on one plane. Analytical solutions for arbitrary array shapes and unconventional arrangements is not always possible. Therefore, IR measurements in free-field environments for several directions on the surface of the array must be obtained, in order to derive a suitably accurate spatial encoding. A regularised least-square solution [4] or direct spherical harmonic domain approach, like in [5], can be utilised for these cases.

3 Pressure-Intensity DOA estimation

The DoA of sound sources can be estimated from the active-intensity vector, which is derived from the zeroth and first-order spherical harmonic signals. This approach is less computationally demanding when compared to the scanning beamforming approach and has been utilised, for example, in DirAC for parametric reproduction purposes [17].

The active-intensity vector can be estimated as [27]

$$\mathbf{i}_{\mathrm{a}} = \Re[p^*\mathbf{u}], \tag{10}$$

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

$$\mathbf{u} \simeq -\frac{1}{\rho_0 c\sqrt{2}} \begin{bmatrix} \mathbf{s}_{1-1} \\ \mathbf{s}_{10} \\ \mathbf{s}_{11} \end{bmatrix}, \tag{11}$$

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

The DoA estimate for the most prominent sound source is then

$$\mathrm{DoA}(\theta_s, \phi_s) = \angle \, \mathrm{E}\left[\frac{\mathbf{i}_\mathrm{a}}{||\mathbf{i}_\mathrm{a}||} \right], \tag{12}$$

where $\angle$ is the angle of a three-dimensional vector.

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 played simultaneously; because in these scenarios, the DoA estimate oscillates between the different sound source directions.

4 Spatially localised pressure-intensity

Although DoA estimation utilising the active-intensity vector is computationally efficient, it becomes inaccurate when multiple sources and/or reflections are present in the same time-frequency tile. However, higher-order signals can be utilised to segregate the sound-field into individual spatially selective sectors, in order to estimate multiple active-intensity vectors. This approach has been shown to improve parametric sound-field reproduction techniques in [22, 23]. Furthermore, SLPI estimation has also been shown to improve DoA estimation for mobile device applications [28].

In this paper, a new generalised formulation of the SLPI approach is presented. First, the pressure-intensity is calculated for different spherical sectors defined as either nearly-uniform or non-uniform. Uniform arrangements are based on positioning a number of equidistant points on the surface of a sphere and then defining regions around these points with almost equal surface area. For nearly-uniform sector designs, there exist a variety of different techniques for obtaining the points on a sphere; these include: t-designs, geodesic methods, electron equilibrium, sphere packing and sphere covering. For spatial sound reproduction applications, the division is generally performed by using t-designs [22]. However, for general spatial audio processing algorithms, these sectors should be defined arbitrarily. For example, in semi-spherical loudspeaker arrangements, frontal only arrangements or teleconferencing applications, it is beneficial to perform spatially localised sound-field analysis for only the region of interest; thus minimising the effect of interferers from other directions.

The proposed generalised design is based on panning functions and can accommodate the design of both uniform or non-uniform sectors. In this paper, the vector-base amplitude panning (VBAP) method is employed [29]; however, the principles can also be extended to utilise other panning methods. The sector design is proposed in the spherical harmonic domain. First, the analysis sectors are defined at the following points: $\Omega_s = (\theta_s, \phi_s)$, for $s = 1, \ldots, S$ where S is the total number of sectors. These points are then used as the directions for calculating the panning functions. For a set of points Ω_q, beamformers are defined that follow the directivity of the panning functions, as described in [29]. These beamformers $T_s(\Omega_q)$ are then utilised to spatially sharpen the traditional pressure, s_{00}, and pressure-gradient, $\mathbf{u}$, beamformer patterns as follows

$$\mathbf{T}_\mathrm{SLDoA}(\Omega_q) = T_s(\Omega_q)\mathbf{Y}_\mathrm{pu}, \tag{13}$$

where $\mathbf{Y}_\mathrm{pu} \in \mathbb{R}^{1\times 4}$ is a matrix containing spherical harmonics for zeroth and first-order; and $\mathbf{T}_\mathrm{SLDoA}(\Omega_q)$ is the directional pattern of the spatially localisation pressure and pressure gradient beamformers. The spatially localised pressure and pressure-gradient beamformers can be synthesised in the spherical harmonic domain, utilising a least squares optimisation, where Eq. (13) is used as the target. The resulting weights can be estimated as

$$w_s = \mathbf{T}_\mathrm{SLDoA}(\Omega_q)\mathbf{Y}_\mathrm{pu}^\dagger. \tag{14}$$

A spatially localised active intensity-based DoA estimation may then be formulated for each sector or only for the sectors of interest as in Eq. 10. This provides two advantages over the traditional active intensity-based DoA estimation:

- higher-order spherical harmonics signals can now be utilised for more accurate active intensity DoA estimation.

- estimates from one sector do not affect the accuracy of the estimation for other sectors. This has been a common disadvantage of the traditional active-intensity-based DoA estimation for coherent sources or reflections arriving simultaneously.

5 Real-time implementations

Real-time software implementations were developed for both the spatial encoding of spherical/cylindrical sensor array signals into spherical harmonic signals and the proposed SLPI-based DoA estimation algorithm[1]. These software implementations were realised as Virtual Studio Technology (VST) plug-ins, using the JUCE framework.

The time-frequency transform chosen for both plug-ins was the alias-free STFT[2], which optimises window lengths to mitigate temporal aliasing artefacts. A hop size of 128 samples was selected, yielding 128 uniformly distributed frequency bands, of which the lowest three bands were further subdivided using additional filtering, such that the centre frequencies more closely resemble the Bark scale; an approach similar to filter-banks used in perceptually-motivated audio codecs.

5.1 Spatial encoder

The spatial encoder, named Array2SH, generates theoretical filters up to 7th order, as described in Section 2, for several different array types and sensor directivity patterns. Additionally, the plug-in offers control over the speed of sound of the medium, in order to support both microphone and hydrophone arrays. It also accommodates different spherical harmonic conventions and normalisation schemes, including support for the Ambix convention, so that it may also be utilised for Ambisonics reproduction purposes.

The spatial encoding equalisation curves for each order are depicted on the user-interface to provide visual feed-back to the user, as shown in Fig. 1. Presets are included for several tetrahedral arrays, such as the Sennheiser Ambeo, and higher-order arrays, such as the MH Acoustics Eigenmike32 and the more recent Zylia array.

[1]The VST plug-ins are available for download on the companion web-page: `http://research.spa.aalto.fi/publications/papers/aes18-array2sh-sldoa/`

[2]`https://github.com/jvilkamo/afSTFT`

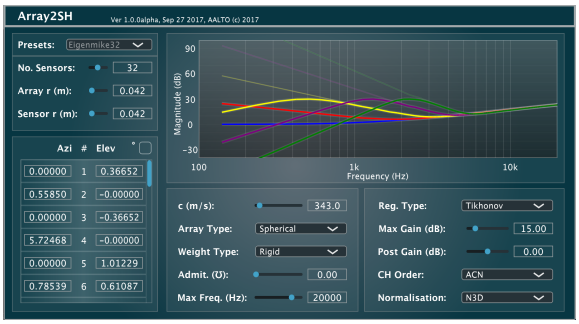


Fig. 1: The GUI for the Array2SH VST plug-in.

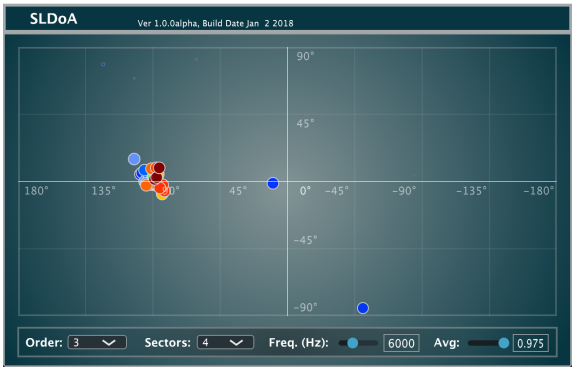


Fig. 2: The GUI for the SLDoA VST plug-in.

5.2 DoA estimator based on spatially localised pressure-intensity

The second plug-in is an SLPI-based DoA estimator, named SLDoA, which is the real-time implementation of the algorithms described in Section 4; and operates on spherical harmonic signals up to 7th order. The plug-in displays the individual DoA estimates, up to a maximum frequency and averaged over a user specified amount of time, as opposed to a power-map. The estimates for each sector are scaled by their relative sector energies and depicted by a circle icon, on an equirectangular representation of the sphere. Circles coloured in red indicate estimates for higher frequencies, while blue denotes lower frequencies. For sectors with higher energies their DoA icons appear larger and more opaque on the user-interface, while icons associated with lower energy estimates appear smaller and more transparent.

An example of the SLDoA plug-in in the process of analysing a real recoding, is shown in Fig. 2. The

recoding consisted of white noise located at $[-90, 0]$ relative to the listening position. The sound-field was first captured by an Eigenmike32 and the array signals were encoded into third-order spherical harmonic signals using the Array2SH plug-in, before being fed into SLDoA; where four sectors were utilised to mitigate reflections and interferers.

6 Evaluation

The performance of both the Array2SH plug-in and the SLPI algorithm used by the SLDoA plug-in were evaluated. Several simulations were performed using the MCRoomSim MatLab library [30]; which is a ray-based multi-channel acoustical modeller for rectangular rooms, with a wide range of user customisation options.

The first test involved impulse responses of three microphone arrays, in order to obtain an estimation of the array signals using the simulator; these included: one first-order array, Sennheiser *Ambeo*, and two fourth-order arrays, an *Eigenmike32* and a 32 sensor cylindrical array designed at the University of Parma, named *Cylindrico*. The *MediumRoom* preset was selected for the simulation, which configures the room dimensions, absorption coefficients and scattering coefficients to resemble a typical 10x8x3m office room, with broadband reverberation time (RT60) of approximately 0.3 s. The receiver position was placed directly in the centre of the room. Two omnidirectional sound sources, one male speech and one female speech interferer, were placed one metre away from the receiver, with directions $[-90, 45]$ and $[-30, -30]$, respectively. For the Ambeo and Eigenmike, both theoretical filters (from Array2SH) and measurement-based filters (using the least-squares solution in [4]) were utilised to spatially encode the simulated array signals into spherical harmonic signals. While for the Cylindico array, only measurement-based filters were utilised, as it has sensors located on several circular rings on its surface, which Array2SH does not take into account for cylindrical arrays.

The spherical harmonic signals were then passed through the first-order PI and the proposed SLPI algorithms, where the latter utilised fourth-order signals and 8 uniformly distributed sectors. The mean estimation error (MEE) between the true source direction and the SLPI estimated direction was derived as [31]

$$\text{MEE} = \frac{1}{N_\text{F}} \sum_f \cos^{-1}(\mathbf{v}_\text{f}^T \hat{\mathbf{v}}_\text{f}) \qquad (15)$$

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

In Fig. 3, the MEE for the male speaker was averaged over a 200 ms moving time window and frequency bands $\in [1, 5]$ kHz; as all three arrays gave high spatial coherence [4] for these frequencies. It can be seen that the MEE fluctuates between 30 and 95 degrees when using the PI algorithm, as it is unable to distinguish between the two sound sources. Whereas, the SLPI yields a much reduced MEE over time, as the sound sources are located in two different sectors. Additionally, the difference in performance between the Array2SH (denoted with *_theory*) and measurement-based (denoted with *_measurement*) filters can be seen to be minimal. Therefore, the algorithms in both audio plug-ins are validated for this scenario.

Having established that the theoretical filters in Array2SH perform as intended, the second round of testing investigated the performance of the SLPI algorithm further; utilising only the Eigenmike32 and theoretical encoding. Different numbers of sound sources (between 1 and 8), spherical harmonic orders (between 1 and 4), and simulated rooms were utilised for the MEE calculations. For the following results, the MEE was averaged using a weighted-sum over a 2 s window, for each DoA estimate; the weights for which, were derived from the normalised sector energies. Fig. 4(a) depicts the results of a MediumRoom simulation, where uncorrelated white-noise sound sources (1 m from the receiver position) were introduced into each sector one-by-one, up to the number of sectors utilised for each processing order $S = 2L$. In Fig. 4(b), the same approach was applied using the 49x19x18 m *Concerthall* preset; with broadband RT60 of 2.4 s. Figs. 4(c-d), then employed the same approach as Figs. 4(a-b), except 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 noise, clapping, and finally an acoustic guitar. The sound files were all normalised to have the same peak level.

For a single source, it can be seen in Figs. 4(a-d), that the SLPI algorithm yields a reduced MEE for all tested scenarios. This is essentially because the sector-based approach provides de-reverberation, as the reflections and diffuse sound in the remaining sectors have reduced influence on the DoA estimation. As the number of

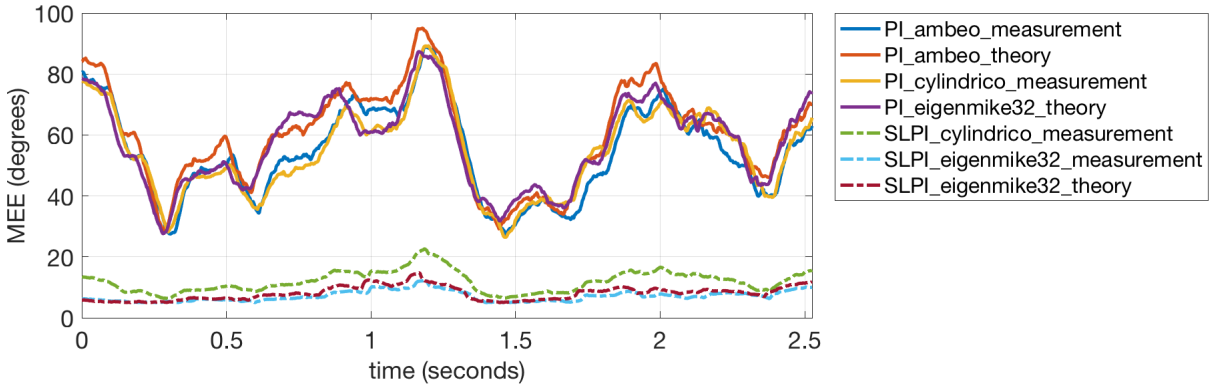
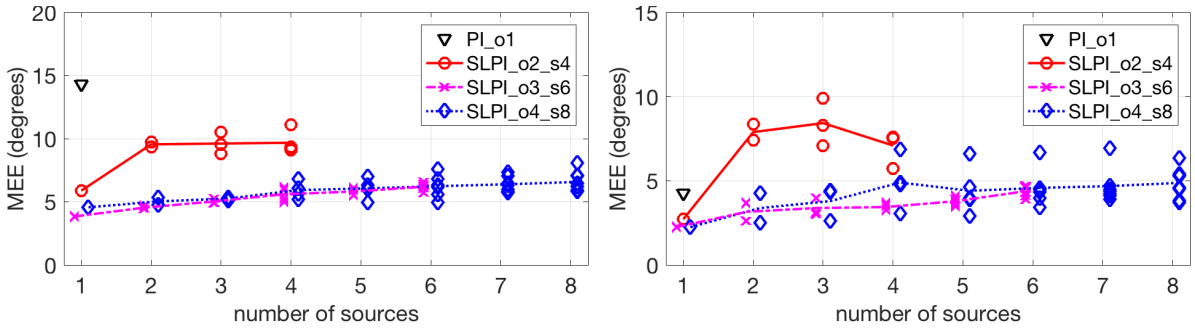
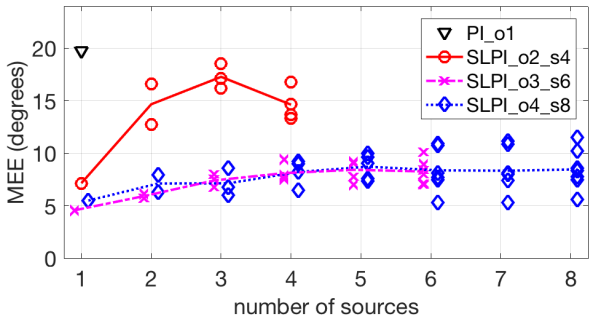


Fig. 3: The MEE over time while utilising the PI and SLPI algorithms with the *ambeo*, *cylindrico* and *eigenmike32* arrays. A male speaker was placed at $[-90, 45]$ and one female speaker was placed at $[-30, -30]$ as an interferer. The proposed SLPI algorithm was configured to utilise fourth-order spherical harmonic signals and 8 uniformly distributed sectors. Also shown is a comparison between measurement-based spatial encoding, denoted with *_measurement*, and theoretical spatial encoding using the Array2SH plug-in, denoted with *_theory*.

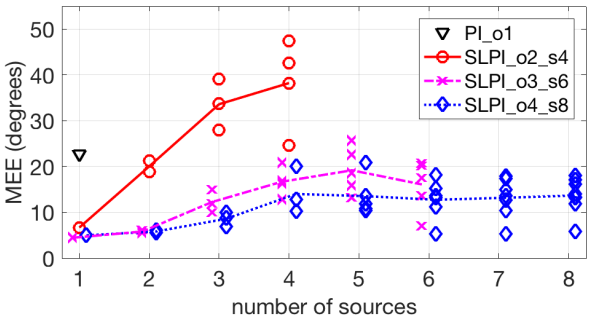


(a) MediumRoom preset, with white-noise sources.

(b) ConcertHall preset, with white-noise sources.



(c) MediumRoom preset, with sound samples as sources.



(d) ConcertHall preset, with sound samples as sources.

Fig. 4: The MEE values when utilising the SLPI approach with different orders and number of sectors. The markers denote the individual DoA estimates per source; whereas the lines denote the mean of the DoA estimates for each number of sources (where applicable).

sources increases, it can be seen that the general trend is an increase in the MEE. While on the other hand, the higher the spherical harmonic order and number of sectors, the lower the MEE becomes. There are some outliers, such as the test cases that utilise 4 sources and second-order signals in Figs. 4(b,c), which appear to perform better with 3 sources; which can possibly be explained by rogue early reflections from other sources. Furthermore, when comparing the white-noise sources to the audio sound sources, the pauses in the mono audio appear to have resulted in DoA estimates of diffuse sound and/or reflections from other sources, which are then included in the weighted-average.

7 Summary

This paper has presented real-time audio plug-ins for sound-field visualisation and conversion of spherical and cylindrical sensor array signals into spherical harmonic signals. The former utilises a proposed generalised algorithm based on SLPI, in order to visualise multiple DoA estimates per time and frequency index. Whereas the latter, employs theoretical filters in order to perform the spatial encoding and can be applied to either spherical or cylindrical arrays; offering control over both the array specifications and the degree of sensor noise amplification.

After testing the plug-ins using acoustical simulations of dry and reverberant environments, it was found that the proposed SLPI method is capable of identifying the DoA of multiple sound sources with reduced error, when compared to the traditional PI approach. It was also discovered that the theoretical spatial encoding filters yielded similar error values to the more laborious measurement-based approach.

References

[1] Rafaely, B., *Fundamentals of spherical array processing*, volume 8, Springer, 2015.

[2] Williams, E. G., *Fourier acoustics: sound radiation and nearfield acoustical holography*, Academic press, 1999.

[3] Teutsch, H., *Modal array signal processing: principles and applications of acoustic wavefield decomposition*, volume 348, Springer, 2007.

[4] Moreau, S., Daniel, J., and Bertet, S., "3D sound field recording with higher order ambisonics–Objective measurements and validation of a 4th order spherical microphone," in *120th Convention of the AES*, pp. 20–23, 2006.

[5] Jin, C. T., Epain, N., and Parthy, A., "Design, optimization and evaluation of a dual-radius spherical microphone array," *IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP)*, 22(1), pp. 193–204, 2014.

[6] Farina, A., "X-volver VST plug-in, for matrix convolution of audio signals," 2013.

[7] Alon, D. L. and Rafaely, B., "Spatial decomposition by spherical array processing," in *Parametric time-frequency domain spatial audio*, pp. 25–48, Wiley Online Library, 2017.

[8] Zoltowski, M. D., "On the performance analysis of the MVDR beamformer in the presence of correlated interference," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 36(6), pp. 945–947, 1988.

[9] Delikaris-Manias, S. and Pulkki, V., "Cross pattern coherence algorithm for spatial filtering applications utilizing microphone arrays," *IEEE Transactions on Audio, Speech, and Language Processing*, 21(11), pp. 2356–2367, 2013.

[10] Delikaris-Manias, S., Vilkamo, J., and Pulkki, V., "Signal-dependent spatial filtering based on weighted-orthogonal beamformers in the spherical harmonic domain," *IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP)*, 24(9), pp. 1507–1519, 2016.

[11] McCormack, L., Delikaris-Manias, S., and Pulkki, V., "Parametric acoustic camera for real-time sound capture, analysis and tracking," in *Proceedings of the 20th International Conference on Digital Audio Effects (DAFx-17)*, pp. 412–419, 2017.

[12] Gerzon, M. A., "Periphony: With-height sound reproduction," *Journal of the Audio Engineering Society*, 21(1), pp. 2–10, 1973.

[13] Zotter, F. and Frank, M., "All-round ambisonic panning and decoding," *Journal of the audio engineering society*, 60(10), pp. 807–820, 2012.

[14] Epain, N., Jin, C., and Zotter, F., "Ambisonic decoding with constant angular spread," *Acta Acustica united with Acustica*, 100(5), pp. 928–936, 2014.

[15] Zotter, F., Pomberger, H., and Noisternig, M., "Energy-preserving ambisonic decoding," *Acta Acustica united with Acustica*, 98(1), pp. 37–47, 2012.

[16] Santala, O., Vertanen, H., Pekonen, J., Oksanen, J., and Pulkki, V., "Effect of listening room on audio quality in Ambisonics reproduction," in *Audio Engineering Society Convention 126*, Audio Engineering Society, 2009.

[17] Pulkki, V., "Spatial sound reproduction with directional audio coding," *Journal of the Audio Engineering Society*, 55(6), pp. 503–516, 2007.

[18] Delikaris-Manias, S. and Vilkamo, J., "Adaptive Mixing of Excessively Directive and Robust Beamformers for Reproduction of Spatial Sound," in *Parametric time-frequency domain spatial audio*, pp. 201–214, Wiley Online Library, 2017.

[19] Barrett, N. and Berge, S., "A new method for B-format to binaural transcoding," in *Audio Engineering Society Conference: 40th International Conference: Spatial Audio: Sense the Sound of Space*, Audio Engineering Society, 2010.

[20] Pulkki, V., Delikaris-Manias, S., and Politis, A., *Parametric Time-Frequency Domain Spatial Audio*, John Wiley & Sons, 2017.

[21] Vilkamo, J. and Delikaris-Manias, S., "Perceptual reproduction of spatial sound using loudspeaker-signal-domain parametrization," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 23(10), pp. 1660–1669, 2015.

[22] Politis, A., Vilkamo, J., and Pulkki, V., "Sector-based parametric sound field reproduction in the spherical harmonic domain," *IEEE Journal of Selected Topics in Signal Processing*, 9(5), pp. 852–866, 2015.

[23] Politis, A., McCormack, L., and Pulkki, V., "Enhancement of ambisonic binaural reproduction using directional audio coding with optimal adaptive mixing," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics,*, 2017.

[24] Bernschütz, B., Pörschmann, C., Spors, S., Weinzierl, S., and der Verstärkung, B., "Soft-limiting der modalen amplitudenverstärkung bei sphärischen mikrofonarrays im plane wave decomposition verfahren," *Proceedings of the 37. Deutsche Jahrestagung für Akustik (DAGA 2011)*, pp. 661–662, 2011.

[25] Alon, D. L., Sheaffer, J., and Rafaely, B., "Robust plane-wave decomposition of spherical microphone array recordings for binaural sound reproduction," *The Journal of the Acoustical Society of America*, 138(3), pp. 1925–1926, 2015.

[26] Lösler, S. and Zotter, F., "Comprehensive radial filter design for practical higher-order Ambisonic recording," *Fortschritte der Akustik, DAGA*, pp. 452–455, 2015.

[27] Fahy, F. J. and Salmon, V., "Sound intensity," *The Journal of the Acoustical Society of America*, 88(4), pp. 2044–2045, 1990.

[28] Delikaris-Manias, S., Pavlidi, D., Mouchtaris, A., and Pulkki, V., "DOA estimation with histogram analysis of spatially constrained active intensity vectors," in *Acoustics, Speech and Signal Processing (ICASSP), 2017 IEEE International Conference on*, pp. 526–530, IEEE, 2017.

[29] Pulkki, V., "Virtual sound source positioning using vector base amplitude panning," *Journal of the Audio Engineering Society*, 45(6), pp. 456–466, 1997.

[30] Wabnitz, A., Epain, N., Jin, C., and Van Schaik, A., "Room acoustics simulation for multichannel microphone arrays," in *Proceedings of the International Symposium on Room Acoustics*, pp. 1–6, 2010.

[31] Delikaris-Manias, S., Pavlidi, D., Pulkki, V., and Mouchtaris, A., "3D localization of multiple audio sources utilizing 2D DOA histograms," in *Signal Processing Conference (EUSIPCO), 2016 24th European*, pp. 1473–1477, IEEE, 2016.