DOA ESTIMATION WITH HISTOGRAM ANALYSIS OF SPATIALLY CONSTRAINED ACTIVE INTENSITY VECTORS

Symeon Delikaris-Manias [†], *Despoina Pavlidi* [*‡], *Athanasios Mouchtaris* [*‡], *Ville Pulkki* [†]

[†] Aalto University, Department of Signal Processing and Acoustics, Espoo, FI-00076, Finland
[*] FORTH-ICS, Heraklion, Crete, GR-70013, Greece
[‡] University of Crete, Department of Computer Science, Heraklion, Crete, GR-70013, Greece

ABSTRACT

The active intensity vector (AIV) is a common descriptor of the sound field. In microphone array processing, AIV is commonly approximated with beamforming operations and utilized as a direction of arrival (DOA) estimator. However, in its original form, it provides inaccurate estimates in sound field conditions where coherent sound sources are simultaneously active. In this work we utilize a higher order intensity-based DOA estimator on spatially-constrained regions (SCR) to overcome such limitations. We then apply 1-dimensional (1D) histogram processing on the noisy estimates for multiple DOA estimation. The performance of the estimator is shown with a 7-channel mobile microphone array, in reverberant conditions and under different signal-to-noise ratios.

Index Terms— direction of arrival, higher order active intensity vector, multiple sound sources, microphone arrays

1. INTRODUCTION

Direction of arrival (DOA) estimation is one of the fundamental array processing problems that can be applied in many applications such as spatial sound reproduction [1, 2], acoustic analysis of enclosed spaces [3], or spatial filtering [4, 5]. The selection of a DOA estimator depends on the application's requirements in terms of resolution and computational cost. For example, in sound reproduction such as in [1], the accuracy of the DOA estimator is more forgiving than in spatial filtering where estimation errors will result to spatial noise mixing into the target signal. The most popular approaches for DOA estimation are the steered-response power [6, 7], maximum likelihood, subspace-based [8–10], sensor phase-based [11], and intensity-based [12–14].

In our previous work we proposed to apply histogram processing to the active intensity vector (AIV) estimates, aiming at accurate DOA estimation in the 3D space [13]. The results indicated that very low DOA errors can be achieved. The use of AIV estimates in 2D scenarios requires a minimum of three microphone signals from which the pressure and the particle velocity are approximated. Recent developments in microphone array technology allow the use of much higher number of microphones. Although such arrays can provide beamformers with high directivity, intensity-based DOA estimators utilize the information only up to first order [12]. Hybrid approaches have been proposed in [15–17]. A higher order AIV has been proposed for spatial sound reproduction where multiple higher order active intensity estimates are utilized to estimate a set of parameters which are then used to re-synthesize the sound field for loudspeaker reproduction in the spherical harmonic domain [18, 19].

In this contribution

- we utilize the higher order AIV in a spatially constrained region (SCR) with a prototype compact microphone array with 7 microphones,
- we post-process the instantaneous spatially constrained (SC) AIV estimates with 1D histogram processing to obtain accurate final DOA estimates and avoid noisy estimations,
- we demonstrate the advantage of using SC-AIV in DOA estimation of multiple non-coherent sources inside a SCR when coherent sources are active outside the SCR.

The paper is organized as follows. In Section 2 the AIV background is presented. Section 3 describes the proposed method of estimating the higher-order AIV and post-processing the DOA estimates with 1D histograms. Section 4 presents the experimental setup for evaluation and the results using a real microphone array in reverberant environments with the presence of multiple speech sources. Section 5 presents our conclusions.

2. BACKGROUND

In the current work matrices and vectors are denoted with bold-faced—upper and lower case correspondingly—symbols. The entries of both matrices and vectors are denoted with the

This research has been partly funded by the Aalto ELEC Doctoral School and the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 644283, Project LISTEN.

same non-bold-faced symbols, appended with a subindex. Let us denote with $\mathbf{x}(k,n) \in \mathbb{C}^{Q\times 1}$ the time-frequency (TF) domain signals from a microphone array with Q sensors, where k is the frequency index and n is the time index. The output of a signal-independent beamformer is denoted as $y(k,n) = \mathbf{w}(k)^H\mathbf{x}(k,n)$, where $\mathbf{w}(k) \in \mathbb{C}^{\mathrm{Q}\times 1}$ is a set of complex multipliers that mix the microphone signals to provide the output signal y, and $(\cdot)^H$ denotes Hermitian transposition.

The AIV is defined as $\mathbf{I} = 0.5\Re[p^*\mathbf{v}]$, where p denotes sound pressure, $\mathbf{v} = [v_x, v_y]^T \in \mathbb{C}^{2\times 1}$ denotes particle velocity for the 2D case, $\Re$ is the real part operator and $*$ is the conjugate operator. The AIV corresponds to the direction of the sound energy flow, therefore the DOA can be estimated by a vector pointing to the opposite direction. Instead of measuring the pressure and particle velocity in sound reproduction and DOA estimation methods, the sound intensity is approximated by measuring the pressure and particle velocity components with an omnidirectional, s_{p}, and dipole microphones, $s_{\mathrm{x}}, s_{\mathrm{y}}$ respectively [1]. When using a microphone array, these signals can be synthesized with signal-independent beamforming designs.

3. DOA ESTIMATION WITH SPATIALLY-CONSTRAINED ACTIVE INTENSITY VECTORS

Recently, a higher-order AIV was introduced for spatial sound reproduction, where the active intensity is estimated for multiple spatially-constrained areas that sum up to an omnidirectional pattern [18]. In this work we investigate the use of the higher order AIV as a DOA estimator in the SCR of interest when a coherent source occurs outside, referred to as SC-AIV. The SCR is assumed to be known since it can be user-defined or indicated by the application or the deployed device (e.g., the front area of a mobile or tablet device). We apply 1D histogram post-processing to retrieve accurate DOA estimates of multiple sources. A demonstrative scenario is illustrated in Fig. 1, where multiple talkers are active within a SCR (grey region), with a coherent source outside this area. This is a typical scenario where a mobile device is used for recording a sound scene. Using the AIV in such a scenario will provide inaccurate estimates that modulate between the two coherent sources. In constrast SC-AIV copes with the presence of coherent sources as it will be demonstrated in Section 4.

3.1. Higher order active intensity vector

The higher order AIV is defined as

$$\mathbf{I}_{\mathrm{HO}}(k,n) = \frac{1}{2}\Re\left\{ s_{\mathrm{pHO}}(k,n)^* \begin{bmatrix} s_{\mathrm{xHO}}(k,n) \\ s_{\mathrm{yHO}}(k,n) \end{bmatrix} \right\}, \quad (1)$$

where $s_{\mathrm{pHO}}, s_{x_{\mathrm{HO}}}, s_{y_{\mathrm{HO}}}$ are signals that approximate the spatially constrained pressure and particle velocity for the x- and y-axis respectively. The directional patterns $T_{\mathrm{pHO}}(\phi)$, $T_{\mathrm{xHO}}(\phi)$ and $T_{\mathrm{yHO}}(\phi)$ of the spatially constrained pressure and particle velocity components $s_{\mathrm{pHO}}, s_{\mathrm{xHO}}, s_{\mathrm{yHO}}$, with $\phi \in [-180, 180)$, are

$$T_{\mathrm{pHO}}(\phi) = c(\phi), \quad (2)$$
$$T_{\mathrm{xHO}}(\phi) = c(\phi)\cos(\phi), \quad (3)$$
$$T_{\mathrm{yHO}}(\phi) = c(\phi)\sin(\phi), \quad (4)$$

where $c(\phi)$ is a spatial windowing function that focuses on the direction of interest.

The instantaneous DOA is then estimated as $\theta(k,n) = \angle[-\mathbf{I}_{\mathrm{HO}}(k,n)]$, where $\angle$ gives the angle of a vector. The advantage of such spatial windowing is that DOA estimates within the spatial window are not affected by sources outside the window while it remains as computationally efficient as the first order intensity estimator. The design of the spatially constrained pressure and particle velocity components is based on signal-independent beamforming techniques via l_2 minimization in the space domain. By setting each function of (2,3,4) as a target pattern $\mathbf{t} \in \mathbb{R}^{N\times 1}$ defined at N points, we consider to minimize the squared error between the actual and target pattern at directions ϕ. The regularized least squares solution is given by

$$\mathbf{w}(k) = \left[\mathbf{V}^{\mathrm{H}}(k)\mathbf{V}(k) + \lambda\mathbf{I}_{\mathrm{Q}}\right]^{-1}\mathbf{V}^{\mathrm{H}}(k)\mathbf{t}, \quad (5)$$

where $\mathbf{V} \in \mathbb{C}^{N\times Q}$ is the matrix of steering vectors, $\mathbf{I}_{\mathrm{Q}} \in \mathbb{R}^{Q\times Q}$ is the identity matrix, and λ is a regularization parameter [20].

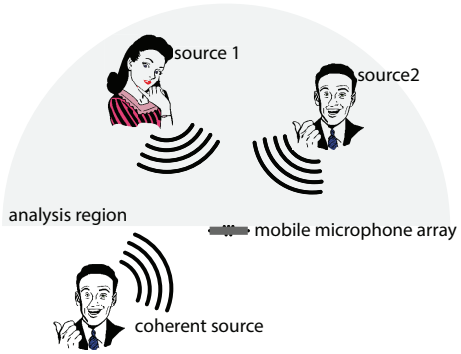

Fig. 1. Recording scenario

3.2. 1D histogram processing

We collect instantaneous DOA estimates, $\theta(k,n)$, (Section 3.1) from B consecutive time frames and post process them by forming 1D histograms for final multiple sources DOA estimation retrieval. The block of B time frames slides one frame each time. The 1D histogram is further smoothed with a Gaussian window $\mathbf{h}_{\mathrm{A}}(\phi)$ of zero mean and standard deviation (std) equal to σ_{A}, leading to

$$\mathbf{y}_{\mathrm{s}}(\phi) = \sum_i \mathbf{y}(i)\mathbf{h}_{\mathrm{A}}(\phi - i), \quad (6)$$

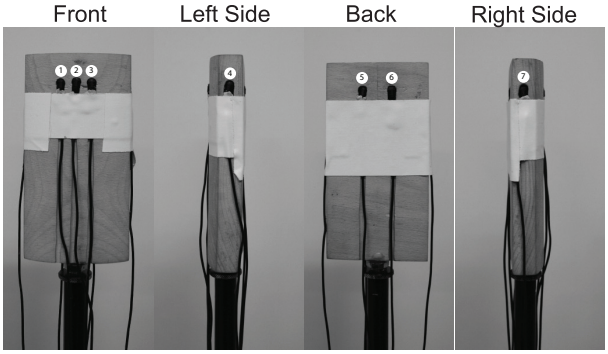

Fig. 2. Photos of the microphone array prototype.

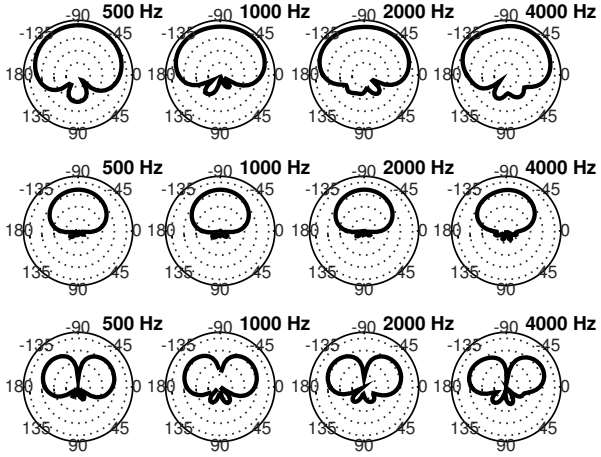

Fig. 3. Synthesized spatially constrained intensity pressure (top) and spatially constrained particle velocity (middle and bottom). Each dotted circle indicates a drop of 10 dB.

where $\mathbf{h}(\phi) = \frac{1}{2\pi\sigma^2} e^{-\frac{1}{2}\frac{\phi^2}{\sigma^2}}$ is the Gaussian window, $\mathbf{y}(\phi)$ is the original 1D histogram and $\mathbf{y}_{\mathrm{s}}(\phi)$ is the smoothed one. We then iteratively detect the highest peak of the smoothed histogram $\mathbf{y}_s^g(\phi)$, identify its index as the DOA of a source, $\phi_g = \arg\max_{\phi} \mathbf{y}_s^g(\phi)$ and remove its contribution from the histogram, $\delta_g = \mathbf{y}_s(\phi) \odot \mathbf{h}_{\mathrm{C}}(\phi - \phi_g)$ by applying a second Gaussian window $\mathbf{h}_{\mathrm{C}}(\phi)$ of zero mean and std equal to σ_{C} until we reach the number G of sources, which is assumed to be known. Thus the smoothed histogram at each next iteration would be $\mathbf{y}_s^{g+1}(\phi) = \mathbf{y}_s^g(\phi) - \delta_g$. The described manipulation of the 1D histograms follows the principles in [13, 15]. However, one could explore the use of other smoothing windows, e.g., a Blackman one [21] or the possibility of use of a more formal Bayesian framework.

4. EVALUATION

Seven microphones (DPA 4060) are fitted in a wooden rectangular object, similar to the dimensions of a mobile device, of size $5.5 \times 2 \times 11$ cm, shown in Fig. 2. The array steering vectors were obtained in an anechoic environment. Measurements in a reverberant environment were performed by placing the microphone array and the rotator in a room of $\mathrm{RT}_{60} = 0.3$ sec and a loudspeaker at 2 m distance. The recording scenarios were generated by convolving the recorded reverberant impulse responses with a dry signal and adding white noise with different signal-to-noise ratios (SNR). The frequency range for DOA estimation was set in $[500, 3500]$ Hz, while the std values in the histogram processing were $\sigma_{\mathrm{A}} = 10^{\circ}$ and $\sigma_{\mathrm{C}} = 40^{\circ}$. The spatially constrained pressure and particle velocity beampatterns are calculated by setting

$$c(\phi) = \begin{cases} 1 & \phi \in [-180, 0] \\ 0 & \phi \in (0, 180), \end{cases} \quad (7)$$

as shown in the grayed region in Fig. 1. The synthesized beampatterns are shown in Fig. 3.

We investigate the performance of the proposed algorithm and the employed microphone array in two different sets of real conditions, i.e., when only incoherent sources are active and when there is one pair of coherent sources. The employed sources were speech signals. The coherent pair consists of the same speech signal positioned in and out of the SCR of interest. For both sets we show results using the first-order AIV estimator and the SC-AIV one (1). The advantage of the SC-AIV estimator is that we obtain DOA estimates only for those sources that are in the analysis region and are not affected by a source outside. Estimating the number of sources is a separate research problem, thus in this work the number of active sources for each estimator is assumed to be known.

We demonstrate the aforementioned scenarios along with the performance of each estimator in Figs. 4 and 5. For these results the SNR was equal to 20 dB. We observe that when the involved sources are incoherent both the AIV and the SC-AIV estimators exhibit accurate DOA estimation performance (Fig. 4). On the other hand, for coherent sources, the AIV estimator fails to accurately estimate the DOAs as indicated by the estimates in between the true DOAs of the involved sources (Fig. 5 (a) and (b)), while the SC-AIV shows robust performance, providing accurate DOA estimates for the sources in its field of view. The results in Fig. 5 (a) and (c) involve two coherent sources, one in the analysis area and one outside, while the results in Fig. 5 (b) and (d) follow a scenario similar to the one demonstrated in Fig. 1, where the source outside the SCR is coherent with one of the sources in the analysis area.

The AIV and SC-AIV estimators are evaluated by utilizing the mean absolute estimation error (MAEE), as in [21], for three different SNR conditions, when two sources are simultaneously active, both for the coherent and incoherent case in Fig. 6. The sources were positioned in 10 random direction pairs around the array, assuring that one source is always outside the analysis area and the other is inside. The MAEE involves estimates that exhibit error not higher than 15°. For

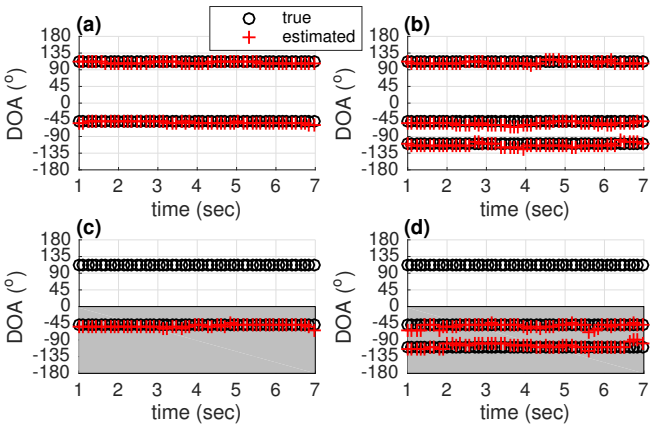

Fig. 4. DOA estimation result for two and three simultaneously active incoherent sources with (a)-(b) the AIV estimator, and (c)-(d) the SC-AIV estimator. The gray region in the plots indicates the analysis area.

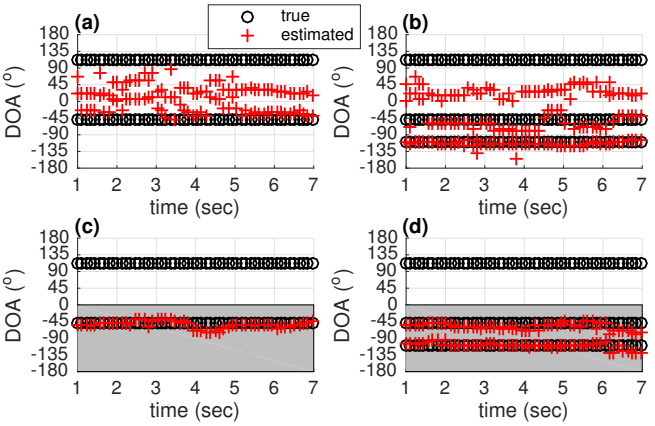

Fig. 5. DOA estimation result with coherent sources when two and three source are active with (a)-(b) the AIV estimator, and (c)-(d) the SC-AIV. The gray region in the plots indicates the analysis area.

the AIV estimator only the case of incoherent sources is presented, since first order AIV fails at providing sensible DOA estimates for coherent sources (see also Fig.7). For SC-AIV the MAEE refers to the target source, i.e., the source in the analysis area. In Fig. 7 we provide the success scores (SS) of the AIV (left) and the SC-AIV (right) estimators, i.e., the percentage of times that the evaluated DOA is in the range of $\pm 15^{\circ}$ from the true DOA. We observe that when the active sources are incoherent both estimators achieve accurate DOA estimation for all different SNR conditions. Moreover, SC-AIV achieves robust DOA estimation of the target source when a coherent source is simultaneously active. The AIV estimator exhibits high SS for incoherent sources (D), but is severely affected by the presence of a coherent source (C), as also demonstrated in Fig. 5. Thus the SS results are not applicable for this scenario. On the other hand the proposed SC-AIV estimator achieves high SS for both different and coherent sources scenarios. Lower SNR conditions lead to reduced SS, due to the noise boost caused by the beamforming operations, but the performance of the estimator remains in a functional range of values.

5. CONCLUSION

We presented a method to obtain accurate DOA estimates in sound field scenarios with multiple sound sources. We relied on the estimation of a higher order, spatially constrained, active intensity vector and 1D histogram post-processing. The method was evaluated using a real microphone array, mounted on a rigid, mobile-like device, in a reverberant environment with different signal-to-noise ratio conditions. The proposed method achieves to deliver accurate DOA estimates even in scenarios with coherent sources, while its performance is comparable with the first order active intensity for simultaneously active, incoherent sources.

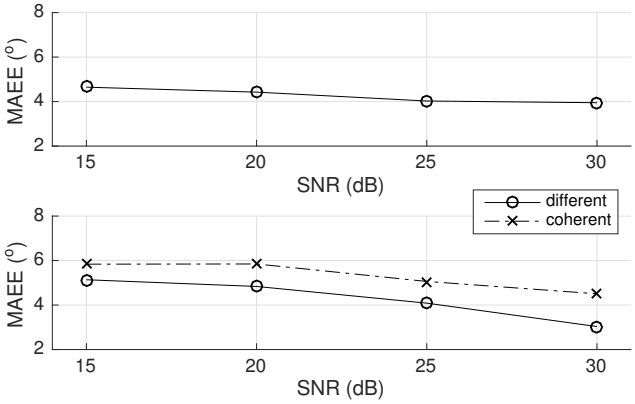

Fig. 6. Mean absolute estimation error using AIV for different sources (top) and SC-AIV for different and coherent sources (bottom).

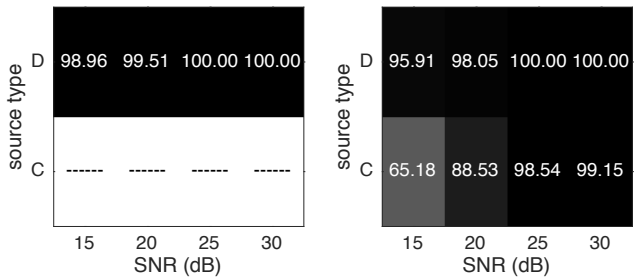

Fig. 7. Success scores in DOA estimation using the AIV (left) and the SC-AIV (right) for incoherent (D) and coherent (C) sources.

6. REFERENCES

[1] V. Pulkki, "Spatial sound reproduction with directional audio coding," *Journal of the Audio Engineering Society*, vol. 55, no. 6, pp. 503–516, 2007.

[2] B. Gunel, H. Hacihabiboglu, and A. M. Kondoz, "Acoustic source separation of convolutive mixtures based on intensity vector statistics," *IEEE transactions on audio, speech, and language processing*, vol. 16, no. 4, pp. 748–756, 2008.

[3] G. Del Galdo, M. Taseska, O. Thiergart, J. Ahonen, and V. Pulkki, "The diffuse sound field in energetic analysis," *The Journal of the Acoustical Society of America*, vol. 131, no. 3, pp. 2141–2151, 2012.

[4] S. Delikaris-Manias, J. Vilkamo, and V. Pulkki, "Signal-dependent spatial filtering based on weighted-orthogonal beamformers in the spherical harmonic domain," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 24, no. 9, 2016.

[5] O. Thiergart, M. Taseska, and E. A. P. Habets, "An informed parametric spatial filter based on instantaneous direction-of-arrival estimates," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 22, no. 12, pp. 2182–2196, 2014.

[6] J. H DiBiase, H. F Silverman, and M. S. Brandstein, "Robust localization in reverberant rooms," in *Microphone Arrays*, pp. 157–180. Springer, 2001.

[7] M. Cobos, A. Marti, and J. J. Lopez, "A modified SRP-PHAT functional for robust real-time sound source localization with scalable spatial sampling," *IEEE Signal Processing Letters*, vol. 18, no. 1, pp. 71–74, 2011.

[8] O. Nadiri and B. Rafaely, "Localization of multiple speakers under high reverberation using a spherical microphone array and the direct-path dominance test," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 22, no. 10, pp. 1494–1505, 2014.

[9] R. Roy and T. Kailath, "ESPRIT-estimation of signal parameters via rotational invariance techniques," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 37, no. 7, pp. 984–995, 1989.

[10] M. Costa, A. Richter, and V. Koivunen, "Doa and polarization estimation for arbitrary array configurations," *IEEE Transactions on Signal Processing*, vol. 60, no. 5, pp. 2330–2343, 2012.

[11] O. Thiergart and W. Huang, "A low complexity weighted least squares narrowband doa estimator for arbitrary array geometries," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2016, pp. 340–344.

[12] D. P. Jarrett, E. A. P. Habets, and P. A. Naylor, "3D source localization in the spherical harmonic domain using a pseudointensity vector," in *Proceedings of the European Signal Processing Conference (EUSIPCO)*, 2010, pp. 442–446.

[13] D. Pavlidi, S. Delikaris-Manias, V. Pulkki, and A. Mouchtaris, "3D localization of multiple sound sources with intensity vector estimates in single source zones," in *Signal Processing Conference (EUSIPCO), 2015 23rd European*. IEEE, 2015, pp. 1556–1560.

[14] S. Tervo, "Direction estimation based on sound intensity vectors," in *Proceedings of the European Signal Processing Conference (EUSIPCO)*, 2009, pp. 700–704.

[15] 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," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, March 2016, pp. 96–100.

[16] S. Hafezi, A. H. Moore, and P. A. Naylor, "3D acoustic source localization in the spherical harmonic domain based on optimized grid search," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, March 2016, pp. 415–419.

[17] A. H. Moore, C. Evers, and P. A. Naylor, "Direction of arrival estimation in the spherical harmonic domain using subspace pseudointensity vectors," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 25, no. 1, pp. 178–192, Jan 2017.

[18] A. Politis, J. Vilkamo, and V. Pulkki, "Sector-based parametric sound field reproduction in the spherical harmonic domain," *IEEE Journal of Selected Topics in Signal Processing*, vol. 9, no. 5, pp. 852–866, 2015.

[19] A. Politis and V. Pulkki, "Acoustic intensity, energy-density and diffuseness estimation in a directionally-constrained region," in *arXiv:1609.03409*, 13/9/2016.

[20] O. Kirkeby, P. A. Nelson, H. Hamada, and F. Orduna-Bustamante, "Fast deconvolution of multichannel systems using regularization," *IEEE Transactions on Speech and Audio Processing*, vol. 6, no. 2, pp. 189–194, 1998.

[21] A. Griffin, D. Pavlidi, M. Puigt, and A. Mouchtaris, "Real-time multiple speaker DOA estimation in a circular microphone array based on matching pursuit," in *Proceedings of the 20th European Signal Processing Conference (EUSIPCO)*, Aug 2012, pp. 2303–2307.