Spatially localized direction of arrival estimation

Symeon Delikaris-Manias, Leo McCormack
Aalto Acoustics Lab, Department of Signal Processing and Acoustics, Aalto University, Espoo, FI-00076, Finland

Despoina Pavlidi, Athanasios Mouchtaris
FORTH-ICS, Heraklion, Crete, Greece, GR-70013,
University of Crete, Department of Computer Science, Heraklion, Crete, Greece, GR-70013.

Summary
Direction-of-arrival (DOA) estimation is a fundamental area of research in the field of array processing, as it applies in a variety of areas, including: sound field analysis, spatial sound reproduction and spatial filtering. There exists a plethora of DOA estimating algorithms, which all vary in terms of accuracy and computational complexity. Among the most popular approaches for DOA estimation are the steered-response power, time-delay based, active intensity based and subspace based methods. In this work, DOA estimation in spatially constrained areas is explored. A spatially constrained version of the active intensity vector is first formulated, which is then utilized for the DOA estimation. This work elaborates on this concept and demonstrates its advantages when compared to conventional approaches.

PACS no. 43.60.Jn, 43.60.Fg

1. Introduction

Direction-of-arrival (DOA) estimation is a signal processing algorithm class with a wide selection of approaches, including: subspace [1, 2], active intensity-based [3, 4, 5, 6], power-spectrum methods [7], each of which yield varying degrees of accuracy and have differing computational requirements. The selection of a DOA algorithm for a given signal processing application should take into account these requirements and also constrain the latency to tolerable limits. While steered-response beamforming and subspace methods can provide accurate DOA estimates, they are only as precise as the degree of separation between the analysis grid points; therefore, they may be too computationally inefficient for real-time applications.

Active intensity-based methods utilize pressure and particle velocity components to analyze the flow of energy in a captured sound field. The pressure and 3-D particle velocity are approximated with one omnidirectional and three dipole microphones, respectively [8]. Due to its tolerable latency, the active intensity vector is an ideal candidate for low-latency DOA estimation and has been previously employed in many time-frequency domain spatial sound processing algorithms [9]. Its performance has been examined in reverberant environments [10], and the formulation has been extended in the spherical harmonic domain in the form of a pseudo-intensity vector [3]. The pseudo-intensity vector has been studied and compared with steered-response power beamformers, where it has been found to be an effective alternative to DOA estimation, due to its low computational complexity [7].

In a previous work, the authors introduced the concept of utilizing this spatially localized active intensity approach for DOA estimation in mobile devices [11]. Whereas for this work, the formulation has been generalized and employs histogram analysis utilizing ambisonic signals. The paper is organized as follows: Section 2 provides the background on estimating the ambisonic signals from the spherical microphone array signals; Section 3 gives an overview of the proposed algorithm; Section 4 evaluates the proposed algorithm by comparing its performance in DOA estimation with the conventional active intensity vector; and Section 5 concludes the paper.

2. Ambisonic encoding

Spherical microphone arrays are a popular means of capturing and analyzing sound fields, as they yield similar performance in all directions when the sensors are placed uniformly or nearly-uniformly on the surface of a spherical baffle or scaffold. A common first procedure is to transform the microphone array signals into an intermediate format, referred to as spher-

ical harmonic signals or ambisonic signals. From a signal processing perspective, this transformation is a beamforming operation, whereby a set of complex weights are applied to the microphone signals. Typically, the aim is to obtain a set of coincident beamformers that correspond to the directional selectivity of a set of orthogonal basis functions; i.e. the spherical harmonics [20]. These ambisonic signals are an intermediate representation of the sound field, which largely abstract away the microphone array specifications from the potential application. They also provide a convenient format for sound-field manipulation. For a detailed overview of these methods, the reader is referred to [12, 13].

For a set of Q microphones, the ambisonic signals may be calculated up to a given order of spherical harmonic expansion L as

$$\mathbf{s}(k,n) = \mathbf{W}(k)\mathbf{x}(k,n), \tag{1}$$

where $\mathbf{x} \in \mathbb{C}^{Q\times 1}$ denotes the microphone signals, $\mathbf{s} \in \mathbb{C}^{(L+1)^2\times 1}$ are the ambisonic signals, $\mathbf{W} \in \mathbb{C}^{(L+1)^2\times Q}$ is a frequency-dependent spatial encoding matrix for frequency bin k and time frame n. For details on the estimation of the transformation matrix, the reader is referred to [13]. The accuracy of this transformation depends on the microphone arrangement, the array construction (i.e., open or a rigid baffle), and the total number of microphones.

3. Generalization of the spatially localized active intensity

Active intensity is a vector derived by the scalar component measuring the pressure and the particle velocity vector. These can be approximated by microphones or beamformers with omnidirectional and dipole directivity. The active intensity vector points to the direction of acoustical energy flow. Therefore, it can be utilized as a DOA estimator, by determining the vector pointing in the opposite direction. The instantaneous active intensity vector can be approximated in the time-frequency domain as

$$\mathbf{I} = \Re[p(k,n)^*\mathbf{u}(k,n)], \tag{2}$$

where p is the signal estimating the sound pressure, $\mathbf{u}$ are the signals estimating the particle velocity and $\Re$ denotes the real operator.

3.1. Spatially localized active intensity

Spatially localized intensity has been utilised for parametric sound field reproduction in [17] as a more accurate and comprehensive analysis approach. In this section, the formulation has been generalized for usage with spherical microphone arrays. A real-time implementation of the DOA estimator can be found in [14].

The first step in the calculation of the spatially localized active intensity is the same as the conventional pressure intensity estimation. It is based on four components, an omnidirectional and three dipoles, placed at each axis of the Cartesian coordinate system. However, the difference is that the pressure intensity is now constrained and calculated for different spherical sectors. The spherical sectors can be defined as either almost-uniform or non-uniform. Almost-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. In this work we define these nearly-uniform designs by utilizing the solutions of the sphere packing problem [15].

However, for general spatial audio processing algorithms, these sectors should be defined arbitrarily. For example, in semi-spherical loudspeaker arrangements, in frontal only arrangements or teleconferencing applications, it may be beneficial to perform spatially localized sound-field analysis for only the region of interest; thus, minimizing the effect of interferers from other directions. An example of such arrangement is shown in [11] for a mobile device. In this work, a sector design method based on energy preserving panning functions is proposed. Such a design can accommodate both uniform and non-uniform sector designs. For this work, the vector-base amplitude panning (VBAP) method was utilised [16].

Using a set of points from a sphere packing solution, 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, similar to a spatial sound reproduction system.

For another set of points $\Omega_q = (\theta_q, \phi_q)$, a set of beamformers are defined that follow the directivity of the panning functions, as described in [16]. The directivity of these beamformers $T_s(\Omega_q)$ are then utilized to spatially sharpen the directivity of the conventional pressure and particle velocity components of the active intensity vector.

$$\mathbf{T}(\Omega_q) = T_s(\Omega_q)\mathbf{Y}_{\mathrm{pu}}, \tag{3}$$

where $\mathbf{Y}_{\mathrm{pu}} \in \mathbb{R}^{1\times 4}$ is a matrix containing the directivity of the pressure and particle velocity components of the active intensity and $\mathbf{T}(\Omega_q)$ is the directional pattern of the spatially localized pressure and pressure gradient beamformers. The spatially localized pressure and particle velocity beamformers can be synthesized utilizing least-squares beamforming synthesis where Eq. (3) is used as the target. Therefore, the resulting weights can be estimated as

$$w_s = \mathbf{T}(\Omega_q)\mathbf{Y}_{\mathrm{pu}}^{\dagger}, \tag{4}$$

where † denotes the pseudo-inverse. A spatially localized active intensity-based DOA estimation is then

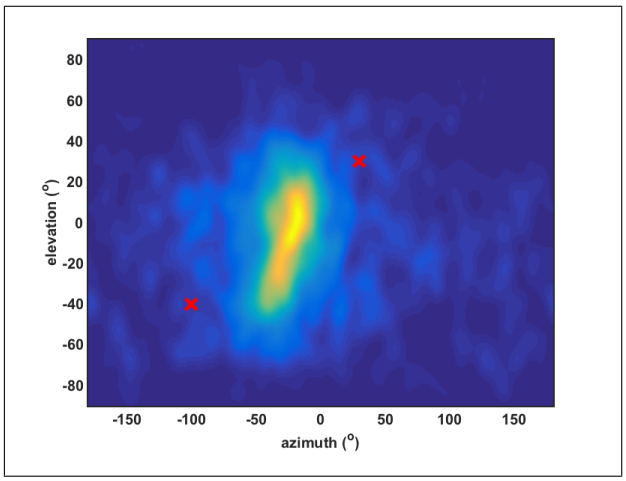

Figure 1. Conventional intensity vector estimates for two coherent sources, one of which is located at $-100°$ azimuth and $-40°$ elevation and the second at 30° azimuth and 30° elevation. The red markers indicate the true DOA of the sources.

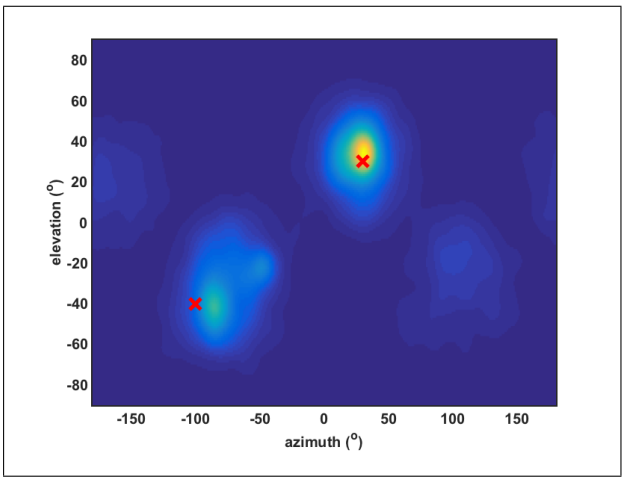

Figure 2. Spatially constrained intensity vector estimates for two coherent sources, one of which is located at $-100°$ azimuth and $-40°$ elevation and the second at 30° azimuth and 30° elevation.The red markers indicate the true DOA of the sources.

formulated for each sector by applying the weights from Eq. (4) to the active intensity for each sector.

3.2. Histogram processing

The DOA estimates from each sector are then analyzed utilizing histograms. Once all of the DOAs have been gathered from all sectors from a block of consecutive time frames, a 2D histogram is formed. This constant size block slides one frame for each iteration. The 2D histograms are processed in order to extract the final DOA estimates by smoothing the 2D histogram, utilizing a circularly symmetric Gaussian window of zero mean and standard deviation σ_a and by iteratively locating its most significant peaks with another Gaussian window of zero mean and standard deviation σ_c, assuming the number of simultaneously active sources to be known, as described in [4]. Exemplary histogram results are shown in Figures 1 and 2, where two coherent sources are simultaneously active at azimuth-elevation pairs of $(-100°,-40°)$ and $(30°,30°)$ for the active intensity and the spatially constrained active intensity, respectively.

4. Results

The algorithm is evaluated in a simulated reverberant environment. A rigid spherical microphone array is utilized with radius of 42 cm comprising 32 microphones placed at the center of the faces of a truncated icosahedron. A room impulse response simulator was used based on the image source method by Allen and Berkley, [19], to simulate a reverberant room of $\mathrm{RT}_{60} = 0.3\,\mathrm{s}$ with approximate dimensions of $5.5 \times 6 \times 3\,\mathrm{m}^3$. The simulator is capable of providing impulse responses for arbitrary spherical microphone arrays [19]. The spherical array was placed in the center of the room and the sound sources were placed 1 m away from the center of the array. The sampling frequency was 48 kHz and the time frame and FFT size was 2048 samples in length.

The 2D histograms were formed utilizing DOA estimates from the current and previous time frames corresponding to one second data. The histograms were updated at each next time frame. The windows used for the histogram processing had standard deviation of $\sigma_a = 5°$ and $\sigma_c = 20°$. The speed of sound was c $= 343\,\mathrm{m/s}$ and the frequency range utilized was 500-3800 Hz, such that only the non-spatially-aliased frequencies were utilized for the processing.

The sound sources were speech signals of duration equal to two seconds each. The signals had equal power and the SNR at each microphone was estimated as the ratio of the power of each source signal to the power of the noise signal. The performance of the proposed algorithm is demonstrated by the mean estimation error (MEE), which measures the angular distance between a unit vector pointing at the true DOA and a unit vector pointing at the estimated DOA, over all sound sources and frames of the source signals. The error is defined as in [4]. We also compare our proposed approach with our previously proposed work which is based on the active intensity vector (IV) [4]. The results are shown in Figure 3. The top row shows the MEE estimation results for one, two, three and four sources for the IV and the spatially localized active intensity vector (SLDOA) for different SNRs of 15, 20, 25 and 30 dB taking into account estimates where the error was lower than 15°. Directly below are the success scores (SS) as percentages of frames of DOA estimates that exhibit error not higher than 15° for cases with incoherent (I) and cases with a pair of coherent (C) sources, respectively. For the IV es-

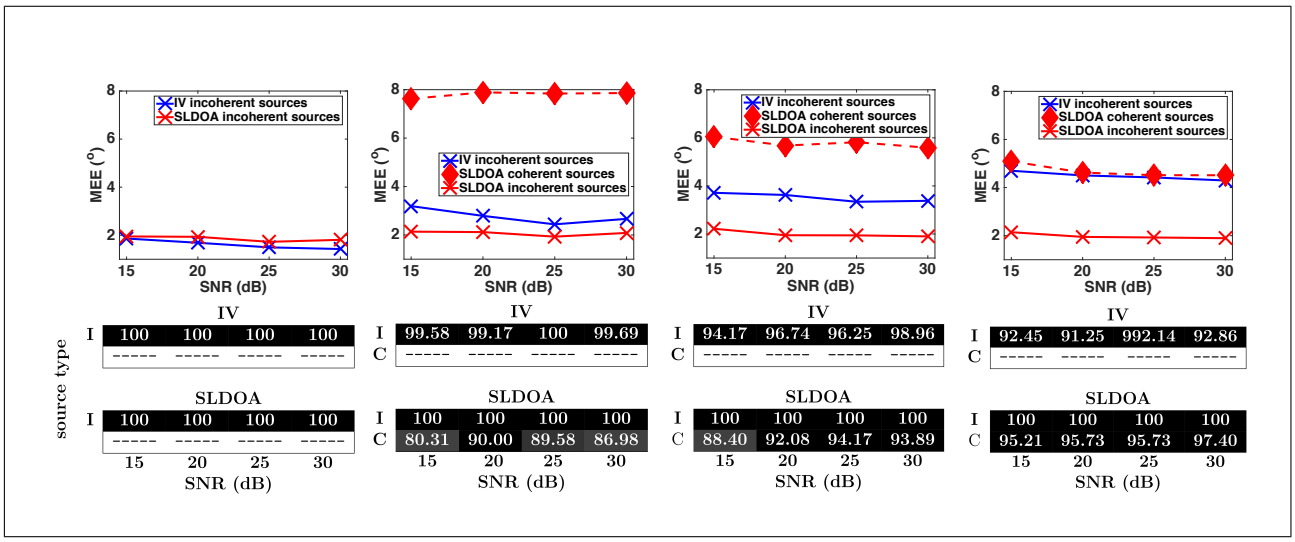

Figure 3. Mean absolute estimation error results for the active intensity vector(IV) and the spatially localized active intensity vector (SLDOA). Top row shows the error for different SNRs and different number of sources: one, two, three and four (from left to right). Below are the success rates in DOA estimation for coherent and incoherent sources.

timator, only the case of incoherent sources is presented, since first order IV fails at providing sensible DOA estimates for coherent sources. An example of such case is also shown in Figure 1. For the case of incoherent sources, both IV and SLDOA perform very well with very high success scores for all SNR conditions. The SLDOA performs also very well in cases where a pair of coherent sources were simultaneously active, again with high success rates. The IV estimator exhibits high SS for incoherent sources, but is heavily influenced by the presence of coherent sources. Thus the SS results are not applicable for this scenario. On the other hand, the proposed SLDOA estimator achieves high SS for both incoherent and coherent sound sources scenarios, for all tested SNR conditions.

5. Conclusions

This paper has proposed a DOA estimation algorithm based on a generalization of the spatially constrained active intensity vector, which is then followed by histogram analysis for the visualization. By constraining the active intensity estimation in sectors and by post processing the estimates with histogram analysis, more accurate directional information can be derived when compared to the traditional active-intensity based DOA estimation. Most importantly the DOA estimation in one sector is not affected by estimates of the other sectors. This becomes especially relevant when coherent sounds arrive in the same time-frequency tile, which is a typical case where the assumptions of a traditional active intensity vector-based DOA estimator are violated. Examples of such cases include competing speakers and sound sources in high reverberant environments, where the sounds and the corresponding reflections may arrive at the same time.

Acknowledgement

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

References

[1] Bo Wang, Yanping Zhao, and Juanjuan Liu. Mixed-order MUSIC algorithm for localization of far-field and near-field sources. IEEE Signal Processing Letters, 20(4): 31–314, April 2013.

[2] Or Nadiri and Boaz Rafaely. Localization of multiple speakers under high reverberation using a spherical microphone array and the direct-path dominance test. Audio, Speech, and Language Processing, IEEE/ACM Trans- actions on, 22(10):1494–1505, 2014.

[3] Daniel P Jarrett, Emanuël AP Habets, and Patrick A Naylor: 3D source localization in the spherical harmonic domain using a pseudo-intensity vector. 18th European Signal Processing Conference (EUSIPCO), 2010.

[4] Despoina Pavlidi, Symeon Delikaris-Manias, Ville Pulkki, and Athanasios Mouchtaris: 3D localization of multiple sound sources with intensity vector estimates in single source zones. 23rd European Signal Processing Conference (EUSIPCO), 2015, 1556–1560.

[5] Sakari Tervo: Direction estimation based on sound intensity vectors. 7th European Signal Processing Conference (EUSIPCO),2009, 700–704.

[6] Huseyin Hacihabiboglu: Theoretical analysis of open spherical microphone arrays for acoustic intensity measurements. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 22(2):465–476, 2014.

[7] Symeon Delikaris-Manias, Despoina Pavlidi, Ville Pulkki, and Athanasios Mouchtaris: 3D localization of multiple audio sources utilizing 2D DOA histograms. 24th European Signal Processing Conference (EUSIPCO), 2016, 1473–1477.

[8] Juha Merimaa and Ville Pulkki: Spatial impulse response rendering i: Analysis and synthesis. Journal of the Audio Engineering Society, 53(12):1115–1127, 2005.

[9] Ville Pulkki, Symeon Delikaris-Manias, and Archontis Politis: Parametric Time-frequency Domain Spatial Audio. John Wiley & Sons, 2017.

[10] Dovid Levin, EmanuÃńl a P Habets, and Sharon Gannot. On the angular error of intensity vector based direction of arrival estimation in reverberant sound fields. The Journal of the Acoustical Society of America, 128(4):1800–11, October 2010.

[11] Symeon Delikaris-Manias, Despoina Pavlidi, Athanasios Mouchtaris, and Ville Pulkki: DOA estimation with histogram analysis of spatially constrained active intensity vectors. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2017, 526–530.

[12] Boaz Rafaely: Fundamentals of Spherical Array Processing. volume 8. Springer, 2015.

[13] David Lou Alon, and Boaz Rafaely: Spatial Decomposition by Spherical Array Processing. Parametric Time-Frequency Domain Spatial Audio (2017): 25.

[14] Leo McCormack, Symeon Delikaris-Manias, Angelo Farina, Daniel Pinardi and Ville Pulkki: Real-time conversion of sensor array signals into spherical harmonic signals with applications to spatially localised sub-band sound-field analysis. AES Convention 144, 2018.

[15] John Horton Conway, and Neil James Alexander Sloane: Sphere packings, lattices and groups. Vol. 290. Springer Science & Business Media, 2013.

[16] Ville Pulkki: Virtual sound source positioning using vector base amplitude panning. Journal of the Audio Engineering Society, 45(6):456–466, 1997.

[17] Archontis Politis, Juha Vilkamo, and Ville Pulkki: Sector-based parametric sound field reproduction in the spherical harmonic domain. IEEE Journal of Selected Topics in Signal Processing 9, no. 5 (2015): 852–866.

[18] D. P. Jarrett, E. A. P. Habets, M. R. P. Thomas, and P. A. Naylor: Rigid sphere room impulse response simulation: Algorithm and applications. The Journal of the Acoustical Society of America, vol. 132, no. 3, pp. 1462–1472, 2012.

[19] J. B. Allen and D. A. Berkley: Image method for efficiently simulating smallroom acoustics. The Journal of the Acoustical Society of America, vol. 65, no. 4, pp. 943–950, 1979.

[20] Earl G. Williams: Fourier acoustics: sound radiation and nearfield acoustical holography. Academic press, 1999.