3D localization of multiple audio sources utilizing 2D DOA histograms

Symeon Delikaris-Manias*, Despoina Pavlidi†‡, Ville Pulkki*, and Athanasios Mouchtaris†‡
*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*—Steered response power (SRP) techniques have been well appreciated for their robustness and accuracy in estimating the direction of arrival (DOA) when a single source is active. However, by increasing the number of sources, the complexity of the resulting power map increases, making it challenging to localize the separate sources. In this work, we propose an efficient 2D histogram processing approach which is applied on the local DOA estimates, provided by SRP, and reveals the DOA of multiple audio sources in an iterative fashion. Driven by the results, we also apply the same methodology to local DOA estimates of a known subspace method and improve its accuracy. The performance of the presented algorithms is validated with numerical simulations and real measurements with a rigid spherical microphone array in different acoustical conditions: for multiple audio sources with different angular separations, various reverberation and signal-to-noise ratio (SNR) values.**

I. Introduction

Direction of arrival (DOA) estimation is required in many applications, such as spatial filtering and enhancement [1], [2], and spatial audio reproduction [3]. In three-dimensional (3D) DOA estimation a wide selection of algorithms is available: subspace [4], intensity-based [5]–[7] and steered-response-power (SRP) [8], each of them with different level of complexity. Selecting the appropriate algorithm depends on the latency and accuracy requirements of the application. Beamforming is an application where inaccurate DOA estimation may lead to reduced audio quality, since the beam might be steered in a noise source instead of the target. Methods such as SRP and multiple signal classification (MUSIC) can provide accurate DOA estimates [4], [8].

Steered-response power methods are based on scanning the sound field with a beamformer. The beamformer is steered in different directions of interest and the output power is then calculated. This signifies the SRP function which is utilized to identify the DOA of active sources. The results of the SRP can be enhanced by applying a phase transform [9]. SRP methods have been proposed for robust real-time applications using a coarse-to-fine search-grid contraction [10] or stochastic search-grid contraction [11].

In our previous work, [7], we have presented a method for DOA estimation based on two-dimensional (2D) histograms processing of selected pseudointensity vector estimates. Motivated by the effectiveness of our post-processing approach, in this contribution we study the problem of multiple audio source localization in the spherical harmonic domain by applying 2D histogram processing on local DOA estimates of

1) SRP methods using four types of axis-symmetric beamformers
2) MUSIC with the direct path dominance test [4],

and we present extensive, comparative results that highlight the characteristics of each beamformer, e.g., noise robustness in case of suppressed sidelobes. The 2D histogram of local DOA estimates experiences lower-complexity compared to the power map for the SRP methods and compared to the pseudospectrum for MUSIC. The histogram is then processed iteratively to obtain the DOAs of multiple sources, achieving accurate results for the SRP methods and an improvement for MUSIC.

The paper is organized as follows. In Section II the background on spherical microphone array processing is presented briefly. Section III describes the application of 2D histograms on the DOA estimates provided by spherical harmonic domain axis-symmetric beamformers and MUSIC. Section IV presents the experimental setup for evaluation and the results using a simulated and a real spherical microphone array in reverberant environments with the presence of multiple speech sources. Section V presents our conclusions.

II. Spherical microphone array processing

In this section, we briefly review the process of obtaining a vector which contains the spherical-harmonic domain signals, $\mathbf{s}$, from a vector of sensor domain signals, $\mathbf{x}$. For a thorough review of spherical microphone array processing the reader is referred to [12]. Let us assume a microphone array that consists of Q microphones positioned at $(\Omega_q, r) = (\theta_q, \phi_q, \mathrm{r})$. We define $\Omega = (\theta, \phi)$, where θ denotes elevation angles and ϕ azimuthal angles with $\theta \in [-\pi/2, \pi/2]$ and $\phi \in [-\pi, \pi]$ and r is the radius. The spherical harmonic signals are then approximated as $s_{lm}(k, \mathrm{r}) \approx \sum_{q=1}^{Q} g_q(\Omega_q) x_q(k, \Omega_q, \mathrm{r}) Y_{lm}^*(\Omega_q)$, where $x_q(k, \Omega_q, \mathrm{r})$ are the separate microphone signals for frequency k, $Y_{lm}^*(\Omega_q)$ are the complex conjugate spherical harmonic functions, and $g_q(\Omega_q)$ is selected so that it provides an accurate approximation of the spherical Fourier transform [13]. The accuracy of this approximation depends on how uniformly the microphones are distributed on the surface of the sphere, the type of the array, the radius r and the frequency k [14]. By omitting the frequency and radial dependency, the equalized spherical harmonic signals can be expressed in matrix form as

$$\mathbf{s} \approx g_q \mathbf{B}^{-1} \mathbf{Y}^H \mathbf{x}, \tag{1}$$

978-0-9928-6265-7/16/$31.00 ©2016 IEEE

where $(^H)$ denotes Hermitian transposition,

$$\mathbf{x} = [x_1, x_2, \ldots, x_Q]^T \in \mathbb{C}^{Q\times 1} \quad (2)$$

are the microphone array input signals,

$$\mathbf{s} = \quad [s_{00}, s_{1-1}, s_{10}, s_{11}, \ldots, s_{LL}]^T \in \mathbb{C}^{(L+1)^2\times 1} \quad (3)$$

are the spherical harmonic signals, $g_q(\Omega_q) = \frac{4\pi}{Q}$, assuming a uniform distribution of microphones on the surface of a sphere,

$$\mathbf{B} = \quad \text{diag}\{[b_0, b_1, b_1, b_1, \ldots, b_L]\} \in \mathbb{C}^{(L+1)^2\times(L+1)^2} \quad (4)$$

is a diagonal matrix containing the equalization weights that depends on the array type, whether it is rigid or open, and is used in (1) to compensate for the effect of the microphone array [13]. $\mathbf{Y} \in \mathbb{C}^{Q\times(L+1)^2}$ is the matrix containing the spherical harmonics up to order L for the Q microphones

$$\mathbf{Y}(\Omega_q) = \begin{bmatrix} Y_{00}(\Omega_1) & Y_{00}(\Omega_2) & \ldots & Y_{00}(\Omega_q) \\ Y_{1-1}(\Omega_1) & Y_{1-1}(\Omega_2) & \ldots & Y_{1-1}(\Omega_q) \\ Y_{10}(\Omega_1) & Y_{10}(\Omega_2) & \ldots & Y_{10}(\Omega_q) \\ Y_{11}(\Omega_1) & Y_{11}(\Omega_2) & \ldots & Y_{11}(\Omega_q) \\ \vdots & \vdots & \vdots & \vdots \\ Y_{LL}(\Omega_1) & Y_{LL}(\Omega_2) & \ldots & Y_{LL}(\Omega_q) \end{bmatrix}^T, \quad (5)$$

where $(^T)$ denotes transposition. The number of microphones to reconstruct L independent spherical harmonics signals is $Q \geq (L+1)^2$ [15].

Axis-symmetric beamforming in the spherical harmonic domain is performed by simply weighting and summing the spherical harmonic signals [12]. The single channel output, y, of the beamformer is

$$y(\Omega_l) = [\mathbf{y}(\Omega_l) \odot \mathbf{s}^T]\mathbf{d}, \quad (6)$$

where $\mathbf{y}(\Omega_l) \in \mathbb{C}^{1\times(L+1)^2}$ is a row of the spherical harmonics matrix (5), $\odot$ denotes the Hadamard product and $\mathbf{d}$ is a vector of weights defined as

$$\mathbf{d} = [d_0, d_1, d_1, d_1, \ldots, d_L, d_L, d_L]^T \in \mathbb{R}^{(L+1)^2\times 1}. \quad (7)$$

Beam steering for rotationally-symmetric beam patterns can be performed by simply adding a set of multipliers in (6) [16] and for arbitrary beam patterns it can be performed by Wigner-D weighting [17] of projection methods [18].

III. Multiple audio source localization based on 2D histograms

The power map of a beamformer can be provided by the output of the SRP function, which is defined as the energy of the beamformer in a grid of directions [8]

$$\Lambda(\Omega_l) = |y(\Omega_l)|^2, \quad (8)$$

where $\Omega_l = (\theta_l, \phi_l)$ consists of all the elevation and azimuthal angles of the search grid. A power map is shown in Fig. 1 (top left) for six simultaneously active audio sources. The peaks in the power map indicate the DOAs of the audio sources. Aiming at enhancing the presence of the DOAs in such 2D representations, we propose to first obtain single DOAs from power maps at each time-frequency (TF) point.

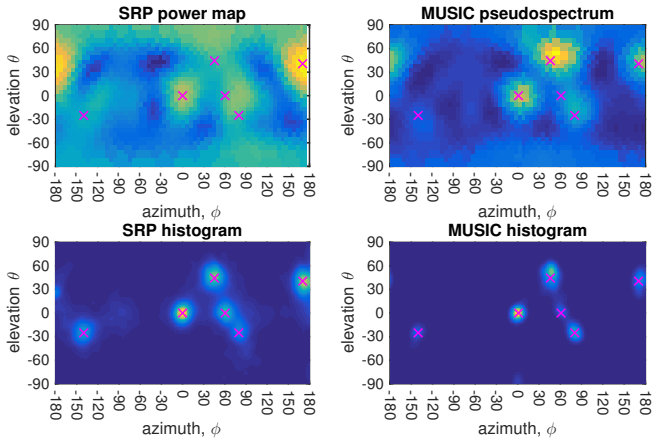

Fig. 1. SRP power map (top left), SRP 2D-histogram (bottom left), MUSIC-DPD pseudospectrum (top right) and MUSIC-DPD 2D-histogram (bottom right) snapshots of a scenario with 6 simultaneously active sources at a simulated environment of RT_{60}=0.3 sec. The pink markers denote the actual positions of the audio sources

Adopting the W-disjoint orthogonality assumption [19], we could then identify the highest peak of the power map as the DOA of the specific TF point which we define as a local DOA. By collecting all DOAs from the TF points of interest, we can form 2D histograms which we can process and infer the azimuthal and elevation angles of multiple sources, assuming their number to be known a priori. In Section III-A we present the four types of beamformers which are utilized for the estimation of the power maps. In Section III-B we describe the formulation and processing of 2D histograms. In Section III-C we describe how 2D histograms can be used to improve the accuracy of the MUSIC algorithm with the direct-path-dominance test (MUSIC-DPD).

A. Rotationally-symmetric beamformers

The types of axis-symmetric beamformers utilized in the present work for DOA estimation are:

- A *regular* beamformer, $\mathbf{d}_{\text{r}}$, with unity gains [12]

$$d_{\text{r}}(l) = 1, \ \forall l = [0, \ldots, L]. \quad (9)$$

- A *minimum-sidelobe* beamformer, $\mathbf{d}_{\text{ms}}$, [20]. It smooths the sidelobes of the regular beamformer and provides complete sidelobe suppression with the cost of a wider main lobe. The weights are given by

$$d_{\text{ms}}(l) = g_0 \frac{\Gamma(L+1)\Gamma(L+2)}{\Gamma(L+1+l)\Gamma(L+2+l)}, \quad (10)$$

  where Γ is the gamma function, and $g_0 = \sqrt{\frac{(2L+1)}{(L+1)^2}}$.
- A *maximum-energy* beamformer, $\mathbf{d}_{\text{maxE}}$, that maximizes the energy concentration towards the look direction [21], [22]

$$d_{\text{maxE}}(l) = C P_l(\text{E}), \quad (11)$$

  where $P_l(\text{E})$ is the l^{th} Legendre polynomial, E the largest root of P_{L+1} and C a normalization constant [23].
- A *Dolph-Chebyshev* beamformer

$$\mathbf{d}_{\text{dc}} = \frac{2\pi}{R}\mathbf{PACTx_0}, \quad (12)$$

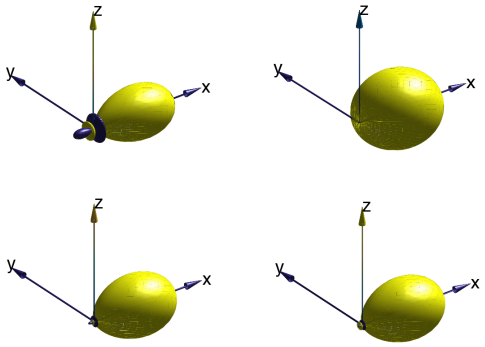

Fig. 2. Directivity patterns of the axis-symmetric beamformers: regular (top left), minimum sidelobes (top right), maximum energy (bottom left) and Dolph-Chebyshev (bottom right).

where $\mathbf{P}, \mathbf{A}, \mathbf{C}, \mathbf{T} \in \mathbb{R}^{(L+1)\times(L+1)}$, $\mathbf{x_0} \in \mathbb{R}^{(L+1)\times 1}$ and R are defined as in [12] in (6.66-6.71). The elements of the vector $\mathbf{d}_{\mathrm{dc}}$ are re-arranged so that they match (7), where the same weight is applied to the spherical harmonic signals of the same order.

The directivity patterns of the aforementioned beamformers are shown in Fig. 2. The input parameters of the Dolph-Chebyshev beamformer were: $R = 10^{\lambda/20}$ with $\lambda = 30\,\mathrm{dB}$.

B. 2D Histogram processing on SRP map

By estimating the beamformers' power map and locating its highest peak index at each TF point, we obtain a local DOA estimate. We form a 2D histogram from a set of local DOA estimates in all TF points of interest in a block of N consecutive time frames. This constant size block slides one frame each time in order to allow a batch mode DOA estimation of multiple sources. We process the 2D histogram in order to extract the final DOA estimates. Our proposed processing has been previously applied on intensity vector estimates in [7]. In this work we apply it to beamformer's local DOA estimates. In order to enhance the presence of DOA estimates of a source, we first smooth the 2D histogram by applying a circularly symmetric window. We have employed a Gaussian window $\mathbf{w}_{\mathrm{A}}(\theta, \phi)$ of zero mean and standard deviation (std) equal to σ_{A}, leading to

$$\mathbf{h}_{\mathrm{s}}(\theta, \phi) = \sum_i \sum_j \mathbf{h}(i, j)\mathbf{w}_{\mathrm{A}}(\theta - i, \phi - j), \qquad (13)$$

where $\mathbf{w}(\theta, \phi) = \frac{1}{2\pi\sigma^2} e^{-\frac{1}{2}\frac{\theta^2+\phi^2}{\sigma^2}}$ is the Gaussian window, $\mathbf{h}(\theta, \phi)$ is the original 2D histogram and $\mathbf{h}_{\mathrm{s}}(\theta, \phi)$ is the smoothed one. We show examples of 2D histograms in Fig. 1 (bottom) where one could easily detect the sources as well as their contribution to the 2D histogram when compared to the power map and pseudospectrum (top). In an iterative fashion we then detect the highest peak of the smoothed histogram $\mathbf{h}_s^g(\theta, \phi)$, identify its index as the DOA of a source, $(\theta_g, \phi_g) = \arg\max_{\theta,\phi} \mathbf{h}_s^g(\theta, \phi)$ and remove its contribution from the histogram, $\delta_g = \mathbf{h}_s(\theta, \phi) \odot \mathbf{w}_{\mathrm{C}}(\theta - \theta_g, \phi - \phi_g)$ by applying a second Gaussian window $\mathbf{w}_{\mathrm{C}}(\theta, \phi)$ of zero mean and std equal to σ_{C} until we reach the known number G of sources. Thus the smoothed histogram at each next iteration would be $\mathbf{h}_s^{g+1}(\theta, \phi) = \mathbf{h}_s^g(\theta, \phi) - \delta_g$.

C. 2D Histogram processing on MUSIC

One of the most well known methods for multiple source localization is the MUSIC algorithm, which has been recently formulated in the spherical harmonic domain [4]. The authors proposed to estimate the narrowband MUSIC pseudospectrum only in TF points that are identified as dominated by a single source by the direct-path dominance test. In the aforementioned implementation of MUSIC-DPD algorithm, the selected incoherent narrowband pseudospectra are averaged to provide one pseudospectrum, the local peaks of which reveal the DOAs of the active sources. Our work on 2D histograms of local DOAs motivated us to modify the MUSIC-DPD algorithm aiming at improving its accuracy in localizing multiple sources. We propose to estimate a local DOA as the index of the highest peak of each narrowband pseudospectrum at TF points approved by the DPD test. All the local DOAs for a block of N consecutive time frames are then provided as input to our 2D histogram processing algorithm as described in the preceding section III-B. In Fig. 1 we show an example of the MUSIC pseudospectrum (top right) and the proposed MUSIC 2D histogram (bottom right).

IV. Evaluation

The evaluation of the 2D histogram-based multiple DOAs estimation algorithms, presented in this work, is based on numerical simulation and real measurements in reverberant environments. For the numerical simulations a baffled spherical microphone array is simulated with radius $r = 4.2\,\mathrm{cm}$, with 32 microphone capsules arranged on the faces of a truncated icosahedron. The array's geometry in the simulations matched the geometry of the measurements array. The numerical simulations utilized a room impulse response (RIR) generator for spherical microphone arrays based on the image-source method [24], [25]. A room of $5.6 \times 6.3 \times 2.7\ \mathrm{m}^3$ was simulated with the same dimensions as the room where the real measurements took place. 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 equal to 48 kHz and the time frame and FFT size was 2048 samples. We applied 50% overlapping in time. The beamformers' order was $L = 3$. For the specific array configuration and assuming an SNR of 45 dB the low frequency limits for processing are 11, 115 and 480 Hz for orders of 1, 2 and 3 respectively [12]. The scanning area for the beamformers and the MUSIC pseudospectra comprises a set of 1002 points on a sphere. The distribution of the points is defined from a geodesic sphere constructed from an icosahedron with an iterative process [26]. The windows used at the histograms processing had std equal to $\sigma_{\mathrm{A}} = 5°$ and $\sigma_{\mathrm{C}} = 20°$. The speed of sound was $c = 343\,\mathrm{m/s}$ while the frequency range used was 500-3800 Hz to avoid aliasing phenomena [4]. We have utilized speech files of duration approximately 7 seconds, with any silent periods removed. The

2D histograms and the MUSIC pseudospectra have resulted from 1 second, i.e., $N = 46$ frames, of data.

As a metric for the evaluation of the algorithms we use the mean estimation error (MEE) defined as

$$\text{MEE} = \frac{1}{N_{\text{F}} G} \sum_{n} \sum_{g} \cos^{-1}\left(\mathbf{v}_{ng}^{T} \hat{\mathbf{v}}_{ng}\right), \qquad (14)$$

where $\cos^{-1}\left(\mathbf{v}_{ng}^{T} \hat{\mathbf{v}}_{ng}\right)$ expresses the angular distance between the true DOA of the g^{th} active source in the n^{th} frame and the estimated one. N_{F} is the total number of frames after subtracting $N-1$ frames of the initialization period and G is the number of active sources, assumed to be known. We note that the DOA estimation results for the four axis-symmetric beamformers (see also Section III-A) that we present hereafter are solely from the proposed 2D DOA histograms processing. Due to the energy spread of the beamformers, obtaining multiple sources DOA estimates directly from the power maps had a significant error and was considered very inaccurate [6].

A. DOA results with simulated room impulse responses

In our first set of simulations, Fig. 3, we plot the MEE versus the number of active sources in a simulated reverberant environment of RT_{60}=0.6 sec for SNR=$\{0, 10, 20\}$ dB for the four different types of beamformers of Section III. We notice that all four beamformers provide very good results for medium and high SNR conditions even when the number of active sources increases. However, their performance degrades as the SNR decreases with the minimum-sidelobe beamformer exhibiting the best performance.

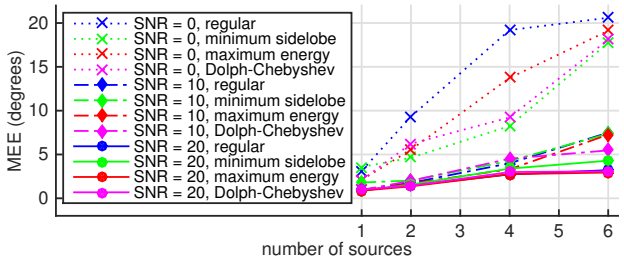

Fig. 3. MEE versus number of sources for $\text{RT}_{60} = 0.6$ sec and various SNR conditions for four types of axis-symmetric beamformers.

In Fig. 4 we explore the performance of the MUSIC algorithm when the DOA results from the averaged 2D pseudospectra, denoted as "MUSIC-SH DPD-incoh", and when it results from the 2D histograms, denoted as "MUSIC-SH DPD-2Dhist" (see also Section III-C). It is clear that for high and moderate SNR values the proposed 2D histogram based processing exhibits an advantage versus the averaged pseudospectra approach. In low SNR conditions both approaches fail to provide a reasonable MEE especially when the number of active sources is increased. We show the performance of all presented methodologies for various angular separation values between two sources in a reverberant environment of RT_{60}=0.4 sec and SNR=20 dB in Fig. 5. All four beamformers as well as the MUSIC-SH DPD-2Dhist show very low MEE for all angular separations. As expected, when the sources get

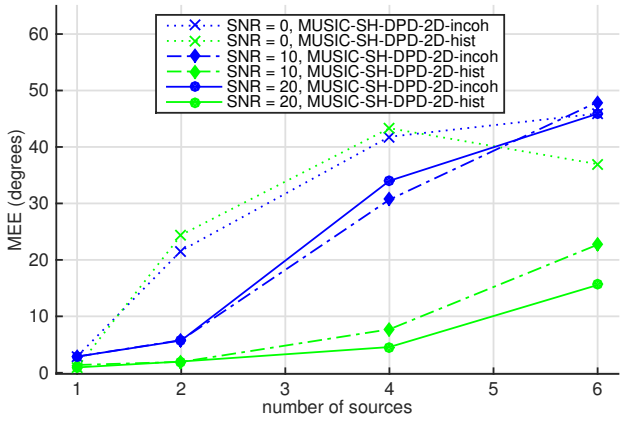

Fig. 4. MEE versus number of sources for $\text{RT}_{60} = 0.6$ sec and various SNR conditions for two approaches of the MUSIC-DPD algorithm

closer the MEE is higher but still in a very reasonable range of values except for the MUSIC-SH DPD-incoh which exhibits the highest error.

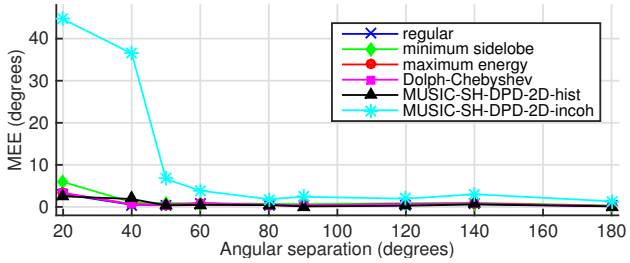

Fig. 5. MEE vs angular separation for $\text{RT}_{60} = 0.4$ sec and SNR=20 dB.

B. DOA results with measured room impulse responses

The real measurements were performed by recording RIRs with the EigenMike [27] in a listening room with approximately the same dimensions as in the numerical simulations. The reverberation time in the recording room was approximately equal to RT_{60}=0.3 sec. We show our first set of results in Fig. 6 at the left plot, while at the right we plot a simulated counterpart. The SNR for both environments was at 45 dB. The performance of all beamformers is very good - the MEE is below three degrees in all cases - following similar tendency between the simulated and real results. In our second set of results with real RIRs we demonstrate the performance of the MUSIC-SH DPD-incoh and MUSIC-SH DPD-2Dhist approaches in Fig. 7 at the left plot with a simulated counterpart at the right side of the figure, also at SNR=45 dB. The MUSIC-SH DPD-2Dhist approach outperforms the MUSIC-SH DPD-incoh for all tested number of sources.

V. Conclusion

In this work we proposed a 2D histogram processing approach which is applied to local DOA estimates for enhanced DOA estimation of multiple audio sources. We have presented our 2D histograms processing methodology under four axis-symmetric beamforming schemes and by enhancing the DOA estimation of a subspace algorithm, both defined in the

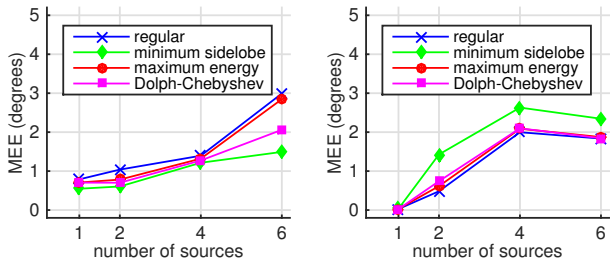

Fig. 6. MEE versus number of sources for $RT_{60} = 0.3$ sec for (a) real and (b) simulated measurements.

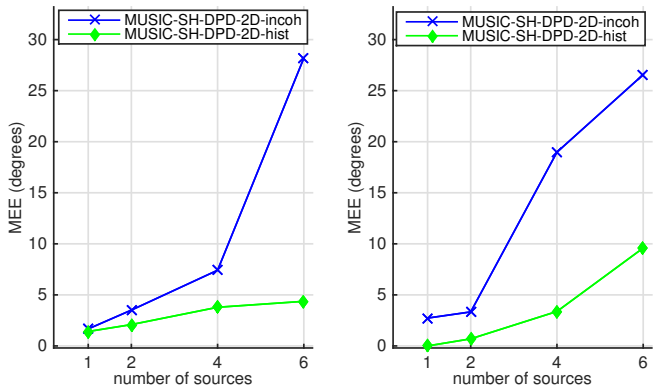

Fig. 7. MEE versus number of sources for $RT_{60} = 0.3$ sec for (a) real and (b) simulated measurements.

spherical harmonic domain. The evaluation was conducted in simulated reverberation and different SNR conditions and with real RIRs for various number of sources, revealing accurate DOA estimation in all examined scenarios.

ACKNOWLEDGMENT

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 and the Aalto ELEC Doctoral School.

REFERENCES

[1] 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.

[2] S. Delikaris-Manias and V. Pulkki, "Cross pattern coherence algorithm for spatial filtering applications utilizing microphone arrays," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 21, no. 11, pp. 2356–2367, Nov 2013.

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

[4] 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.

[5] 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.

[6] C. Evers, A. H. Moore, and P. A. Naylor, "Multiple source localisation in the spherical harmonic domain," in *14th International Workshop on Acoustic Signal Enhancement (IWAENC)*, Sept 2014, pp. 258–262.

[7] 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 *Proceedings of the European Signal Processing Conference (EUSIPCO)*, 2015, pp. 1556–1560.

[8] 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.

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

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

[11] H. Do, H. F. Silverman, and Y. Yu, "A real-time SRP-PHAT source location implementation using stochastic region contraction (src) on a large-aperture microphone array," in *Acoustics, Speech and Signal Processing, 2007. ICASSP 2007. IEEE International Conference on*, vol. 1. IEEE, 2007, pp. I–121.

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

[13] H. Teutsch, *Modal array signal processing: principles and applications of acoustic wavefield decomposition*. Springer Science & Business Media, 2007, vol. 348.

[14] B. Rafaely, "Analysis and design of spherical microphone arrays," *IEEE Transactions on Speech and Audio Processing*, vol. 13, no. 1, pp. 135–143, 2005.

[15] T. D. Abhayapala, "Generalized framework for spherical microphone arrays: Spatial and frequency decomposition," in *IEEE International Conference on Acoustics, Speech and Signal Processing, (ICASSP)*, 2008, pp. 5268–5271.

[16] J. Meyer and G. Elko, "A highly scalable spherical microphone array based on an orthonormal decomposition of the soundfield," in *Acoustics, Speech, and Signal Processing (ICASSP), 2002 IEEE International Conference on*, vol. 2. IEEE, 2002, pp. II–1781.

[17] B. Rafaely and M. Kleider, "Spherical microphone array beam steering using wigner-d weighting," *Signal Processing Letters, IEEE*, vol. 15, pp. 417–420, 2008.

[18] J. Atkins, "Robust beamforming and steering of arbitrary beam patterns using spherical arrays," in *Applications of Signal Processing to Audio and Acoustics (WASPAA), 2011 IEEE Workshop on*. IEEE, 2011, pp. 237–240.

[19] O. Yilmaz and S. Rickard, "Blind separation of speech mixtures via time-frequency masking," *IEEE Transactions on Signal Processing*, vol. 52, no. 7, pp. 1830–1847, 2004.

[20] J. Daniel, "Représentation de champs acoustiques, applicationa la reproduction eta la transmission de scenes sonores complexes dans un contexte multimédia," Ph.D. dissertation, Ph. D. thesis, University of Paris 6, Paris, France, 2000.

[21] F. Zotter, H. Pomberger, and M. Noisternig, "Energy-preserving ambisonic decoding," *Acta Acustica united with Acustica*, vol. 98, no. 1, pp. 37–47, 2012.

[22] S. Bertet, J. Daniel, and S. Moreau, "3d sound field recording with higher order ambisonics-objective measurements and validation of spherical microphone," in *Audio Engineering Society Convention 120*. Audio Engineering Society, 2006.

[23] J. Daniel, J.-B. Rault, and J.-D. Polack, "Ambisonics encoding of other audio formats for multiple listening conditions," in *Audio Engineering Society Convention 105*. Audio Engineering Society, 1998.

[24] 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.

[25] 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.

[26] F. Hollerweger, *Periphonic sound spatialization in multi-user virtual environments*. Citeseer, 2006.

[27] mh acoustics, "EM32 eigenmike microphone array release notes (v17. 0)," mh acoustics, Summit, NJ, USA, Tech. Rep., 2013.