Multiple DOA Estimation Based on Estimation Consistency and Spherical Harmonic Multiple Signal Classification

Sina Hafezi, Alastair H. Moore and Patrick A. Naylor
Department of Electrical and Electronic Engineering
Imperial College London, UK
{s.hafezi14, alastair.h.moore, p.naylor}@imperial.ac.uk

***Abstract*—A common approach to multiple Direction-of-Arrival (DOA) estimation of speech sources is to identify Time-Frequency (TF) bins with dominant Single Source (SS) and apply DOA estimation such as Multiple Signal Classification (MUSIC) only on those TF bins. In the state-of-the-art Direct Path Dominance (DPD)-MUSIC, the covariance matrix, used as the input to MUSIC, is calculated using only the TF bins over a local TF region where only a SS is dominant. In this work, we propose an alternative approach to MUSIC in which all the SS-dominant TF bins for each speaker across TF domain are globally used to improve the quality of covariance matrix for MUSIC. Our recently proposed Multi-Source Estimation Consistency (MSEC) technique, which exploits the consistency of initial DOA estimates within a time frame based on adaptive clustering, is used to estimate the SS-dominant TF bins for each speaker. The simulation using spherical microphone array shows that our proposed MSEC-MUSIC significantly outperforms the state-of-the-art DPD-MUSIC with less than 6.5° mean estimation error and strong robustness to widely varying source separation for up to 5 sources in the presence of realistic reverberation and sensor noise.**

I. Introduction

Multiple source Direction-of-Arrival (DOA) estimation is considered a fundamental problem in acoustic signal processing due to its wide range of applications including source localization/separation/tracking, speech enhancement, SONAR, SODAR, robot audition and dereverberation. The formulation and implementation of most DOA estimators depends on the scenario and the type of sensor array. In this work we are interested in challenging acoustic scenarios consisting of multiple simultaneously active stationary speech sources in the presence of realistic reverberation and sensor noise. We assume the number of sources is known *a priori* but their minimum angular separation is not. In particular we consider methods targeted at Spherical Microphone Arrays (SMAs) [1], [2] where the signals are processed in the Spherical Harmonic (SH) domain.

There are two categories of approach to Multi-Source (MS) DOA estimation. (1) **MS-based DOA estimation** [3] in which the signal model is formulated based on the assumption of multiple sources. The MUltiple SIgnal Classification (MUSIC) [3] method decomposes the covariance matrix of the observed signal into signal and noise subspaces using eigenvalue decomposition in which the top N_e eigenvalues span the signal space where the N_e is the known number of active sources. Due to narrow-band nature of MUSIC, for high accuracy it is performed in Time-Frequency (TF) domain. In case of multiple speech sources, MUSIC is not always preferred as the number of simultaneously active sources is not consistent due to sparseness of speech. (2) **Single-Source (SS)-based DOA estimation**: a common approach is to assume W-disjoint orthogonality [4] whereby at each TF bin it is assumed that only a single source is active. In practice, many TF bins contain significant contributions from waves arriving from multiple directions, either due to overlapping sources or reflections. Also many bins contain no active sources. A number of methods have been proposed to identify those TF bins, referred to as SS-bins in this work, where only a SS is significantly present [5]–[8].

SS-based approaches consist of three stages: (1) SS-bin detection, (2) SS DOA estimation on selected bins and (3) final DOAs extraction. The state-of-the-art Direct Path Dominance (DPD)-MUSIC [5] uses DPD test as SS-bin detection and SS MUSIC as DOA estimator on the selected bins. The DPD test selects the bins in which direct path of a single source has significant dominance where the metric of dominance is measured as the Singular Value Ratio (SVR) between the two largest singular values of the signal subspace, and significance of dominance is defined by a threshold.

An alternative way to detect SS-bins is our recently proposed Multi-Source Estimation Consistency (MSEC) [9] which was initially proposed as a post-processing technique for accuracy enhancement of any DOA estimator in the TF domain. In MSEC, a computationally fast SS DOA estimator such as Pseudo-intensity vectors (PIVs) [10] is first applied on all TF bins and the most consistent DOA estimates are selected using a weighting strategy. The selected DOA estimates are clustered using K-means clustering to obtain the final DOAs as the centroid of clusters where each cluster is associated to a source. Considering the TF bins of the clustered DOA estimates, we have an estimate of the SS-bins associated with each source.

This work was supported by the Engineering and Physical Sciences Research Council [grant number EP/M026698/1].

ISBN 978-0-9928626-7-1 © EURASIP 2017

In DPD-MUSIC, the covariance matrix is calculated over a local TF region centred on a bin which indicates the SS-assumption over a local TF region. Unlike DPD, MSEC provides an estimate of the SS-bins assigned to each speaker across the entire TF domain. The idea in this work is to remove the TF-domain regional limitation in DPD-MUSIC by replacing DPD with MSEC as the pre-processing stage to improve the quality of covariance matrix used in the MUSIC algorithm, particularly for the case of multiple simultaneously active speech sources.

This paper is structured as follow: Section II briefly reviews the technical background of MUSIC and DPD-MUSIC. Section III presents the pre-processing stage MSEC and our novel technique MSEC-MUSIC and finally in Section IV we compare our proposed technique with the state-of-the-art.

II. Technical Background

In this section, we briefly define our signal model in the SH domain and review MUSIC and alternative approaches to DPD-MUSIC.

A. Multi-Source Signal Model

In the Short-time Fourier Transform (STFT) domain at frequency k and time frame τ, consider the vector of source signals $\mathbf{S}(\tau,k) = [S_1(\tau,k)\dots S_N(\tau,k)]^T$ and true DOAs $\mathbf{\Omega_u} = [\Omega_{u1}\dots\Omega_{uN}]$ arriving from N plane waves associated with sources in the far-field. The SH transform of the signals gives [11]

$$\mathbf{a_{lm}}(\tau,k) = \mathbf{Y}^H(\mathbf{\Omega_u})\mathbf{S}(\tau,k) + \mathbf{v_{lm}}(\tau,k) \tag{1}$$

and

$$\mathbf{Y}\left(\mathbf{\Omega_u}\right) = \begin{bmatrix} \mathbf{Y_{lm}}^T(\Omega_{u1}) \\ \vdots \\ \mathbf{Y_{lm}}^T(\Omega_{uN}) \end{bmatrix}, \tag{2}$$

where $\mathbf{a_{lm}} = \left[a_{00}, a_{1(-1)}, a_{1(0)}, a_{1(1)}, \dots, a_{LL}\right]^T$ are the eigenbeams of order l and degree m (satisfying $|m| \leq l$) with maximum SH order L, $(.)^H$ indicates Hermitian transpose, $\mathbf{v_{lm}}(\tau,k)$ is a $(L+1)^2 \times 1$ column vector representing the residual due to noise and reverberation and $\mathbf{Y_{lm}} = \left[Y_{00}, Y_{1(-1)}, Y_{1(0)}, Y_{1(1)}, \dots, Y_{LL}\right]^T$ are the spherical harmonic basis functions [11],

$$Y_{lm}\left(\Omega\right) = \sqrt{\frac{(2l+1)}{4\pi}\frac{(l-m)!}{(l+m)!}}P_{lm}\left(\cos\left(\theta\right)\right)e^{im\varphi}, \tag{3}$$

where φ and θ denotes the azimuth and inclination respectively, P_{lm} is the associated Legendre function and $i^2 = -1$. Note that (τ,k) and (Ω) are omitted respectively for $\mathbf{a_{lm}}$ and $\mathbf{Y_{lm}}$ for notational simplicity.

B. Spherical Harmonic MUSIC

In the TF domain, the covariance matrix of eigenbeams is given as

$$\begin{aligned}\mathbf{R}(\tau,k) =& E\left[\mathbf{a_{lm}}(\tau,k)\mathbf{a_{lm}}^H(\tau,k)\right] \\ =& \mathbf{Y}^H(\mathbf{\Omega_u})\mathbf{R_s}(\tau,k)\mathbf{Y}(\mathbf{\Omega_u}) + \mathbf{R_v}(\tau,k),\end{aligned} \tag{4}$$

where $\mathbf{R_s} = E\left[\mathbf{S}\mathbf{S}^H\right]$ and $\mathbf{R_v}$ are respectively the direct paths and residual (noise and reverberation) covariance matrices and $E\left[.\right]$ denotes the expectation.

For any STFT bin (τ,k) with only a single source active, using Singular Value Decomposition (SVD) the covariance matrix of observed eigenbeams is decomposed as

$$\mathbf{R}(\tau,k) = \mathbf{U_s}\mathbf{\Sigma_s}\mathbf{U_s}^H + \mathbf{U_v}\mathbf{\Sigma_v}\mathbf{U_v}^H, \tag{5}$$

where $\mathbf{U_s}$ is the one-dimensional signal subspace matrix, $\mathbf{U_v}$ is noise subspace with $(L+1)^2 - 1$ dimensions and $\mathbf{\Sigma}$ is the rectangular diagonal singular value matrix. Note that (τ,k) are also omitted here for notational simplicity.

Using the estimated noise subspace, the MUSIC spectrum for a single source is given as [3]

$$P_{MUSIC}(\tau,k,\Omega) = \frac{1}{\|\mathbf{U_v}^H(\tau,k)\mathbf{Y_{lm}}^*(\Omega)\|^2}, \tag{6}$$

where $(.)^*$ indicates complex conjugate.

C. DPD-MUSIC for Multiple Sources

In DPD the covariance matrix in (4) is approximated as the average covariance matrix over a local TF neighbourhood [5]

$$\begin{aligned}\mathbf{R}_{DPD}(\tau,k) =& \frac{1}{J_\tau J_k}\sum_{j_\tau=0}^{J_\tau-1}\sum_{j_k=0}^{J_k-1}\mathbf{a_{lm}}(\tau+j_\tau,k+j_k) \\ &\times \mathbf{a_{lm}}^H(\tau+j_\tau,k+j_k),\end{aligned} \tag{7}$$

where J_τ and J_k are the width (number of bins) of averaging window over time and frequency respectively. The TF bins with significant contribution from a direct path are selected as

$$\Upsilon_{\mathrm{DPD}} = \left\{(\tau,k):\ \mathrm{erank}\left(\mathbf{R}_{DPD}(\tau,k)\right) = 1\right\}, \tag{8}$$

where

$$\mathrm{erank}\left(\mathbf{R}_{DPD}(\tau,k)\right) = 1 \text{ if } \eta_{DPD}(\tau,k) > \epsilon \tag{9}$$

is the effective rank, the SVR η_{DPD} is the ratio of the largest and the second largest singular values of $\mathbf{R}_{DPD}$ and ϵ is a threshold. The two alternative approaches [5] to apply MUSIC on the outcome of DPD test are discussed next.

1) Incoherent DPD-MUSIC: In the first approach the MUSIC spectra in (6) are simply summed over the selected TF bins $(\tau,k) \in \Upsilon_{DPDtest}$ so that

$$P_{incoh-MUSIC}(\Omega) = \sum_{(\tau,k)\in\Upsilon_{DPDtest}} P_{MUSIC}(\tau,k,\Omega), \tag{10}$$

where the set of Nhighest peaks in the final spectrum indicates the overall estimated DOAs.

2) Coherent DPD-MUSIC: The second approach performs coherent fusion of the directional information from the selected TF bins. The set of one dimensional signal spaces from the selected TF bins, $\{\mathbf{U_s}(\tau,k)\}_{(\tau,k)\in\Upsilon_{\mathrm{DPD}}}$, are clustered using one-run K-means clustering with random initialization into N clusters with centroids $\{\mathbf{U_s}^n\}_{n=1}^N$ where each centroid signal space is associated with one speaker. The DOA of each

ISBN 978-0-9928626-7-1 © EURASIP 2017

individual speaker is selected as the global peak in the coherent MUSIC spectrum of the speaker n which is given as

$$P^n_{coh-MUSIC}(\Omega) = \frac{1}{\| \left(\mathbf{U_v}^n\right)^H \mathbf{Y_{lm}}^*(\Omega)\|^2} = \frac{1}{\mathbf{Y_{lm}}^T(\Omega)(I - \mathbf{U_s}^n \left(\mathbf{U_s}^n\right)^H)\mathbf{Y_{lm}}^*(\Omega)}. \quad (11)$$

III. Proposed Method

In this section we propose an alternative approach to calculation of covariance matrix for MUSIC. In our technique we aim to enhance the directional information for each source individually by grouping the SS-bins associated with each speaker and calculating a covariance matrix per speaker. First we briefly present our previously proposed MSEC technique that is used to estimate and group the SS-bins for each speaker and then we present our proposed technique MSEC-MUSIC which uses the outcome of MSEC as the input to MUSIC.

A. MSEC

Our recently proposed technique MSEC [9] is used as a post-processing stage after DOA estimation for all TF bins. The initial DOA estimates (one per TF bin) can be obtained by any SS DOA estimator such as PIV [10], as used in this work, Augmented Intensity Vectors (AIVs) [12], [13], Steered Response Power-based methods [14] or SS MUSIC. The initial DOA estimates are weighted based on their consistency within a time interval and the ones with the strongest weights are selected as the most consistent DOAs using the assumption of stationary sources. MSEC is based on MS assumption in a time frame where the number of active incoherent sources is unknown. It uses the properties of distribution of DOA estimates as the metric of consistency. Adaptive clustering is used to efficiently group the DOA estimates since the number of active sources in time frame is unknown.

1) Adaptive Clustering: In order to have strong concentrations of DOA estimates for the purpose of robust clustering, adaptive clustering at frame τ is performed on data set $\Psi_{MSEC}(\tau)$ including all DOA estimates from frame $\tau - T$ to τ, which is defined as

$$\Psi_{MSEC}(\tau) = \{\hat{\mathbf{u}}(t,k) : \forall k,\, t \in \{\tau, \tau-1, \ldots, \tau-T\}\}, \quad (12)$$

where $\hat{\mathbf{u}}(\tau, k)$ is the initial estimated DOA unit vector.

K-means for $K = \{1, \ldots, K_{max}\}$ is performed on data set $\Psi_{MSEC}(\tau)$ with random initializations. Using Akaike Information Criterion (AIC) [15] which trades-off distortion against the model complexity, the best number of clusters, K_c, is estimated as [16]

$$K_c(\tau) = \arg\min_K \left[\mathrm{RSS}(K(\tau)) + 2QK(\tau)\right], \quad (13)$$

where $\mathrm{RSS}(.)$ is the residual sum of squared (sum of squared distance of each member to its cluster centroid) and Q denotes the number of dimensions of centroid which leads to QK parameters for K clusters. This results in the optimum number of clusters $K_c(\tau)$, the clusters $\{S_i(\tau)\}_{i=1}^{K_C(\tau)}$ and the centroids unit vector $\{\hat{\mathbf{c}}_i(\tau)\}_{i=1}^{K_C(\tau)}$ where i is the cluster index.

2) Weighting: The DOA estimate at each TF bin is assigned a cluster weight and a member weight. The cluster weight representing the normalized measure of concentration in its associated cluster is

$$\psi_{MSEC}(\tau, k) = 1 - \frac{D_i(\tau)}{\pi},\ \hat{\mathbf{u}}(\tau, k) \in S_i(\tau), \quad (14)$$

and the member weight representing how close it is to its associated centroid is

$$\lambda_{MSEC}(\tau, k) = 1 - \frac{\angle(\hat{\mathbf{u}}(\tau, k), \hat{\mathbf{c}}_i(\tau))}{\pi}, \hat{\mathbf{u}}(\tau, k) \in S_i(\tau), \quad (15)$$

where $\angle(.)$ denotes the angle between two vectors.
The MSEC weight in the TF domain is then formed as

$$w_{MSEC}(\tau, k) = \psi_{MSEC}(\tau, k)\lambda_{MSEC}(\tau, k). \quad (16)$$

Having performed MSEC, the DOA estimates from the bins associated with the top $M\%$ strongest MSEC weights over all TF bins are selected.

B. MSEC-MUSIC

Having obtained the selected DOA estimates using MSEC, for the purpose of robust clustering, the potential outlier DOAs are removed if the average cardinality over a spatial window of 1×1 degree (azimuth $\times$ inclination) centred on DOA estimate is below a threshold γ. Applying K-means with $K = N$ on DOA estimates after outlier removal, we obtain the clusters $\{S_n\}_{n=1}^N$ and the centroids unit vectors $\{\hat{\mathbf{c}}_n\}_{n=1}^N$. The MSEC covariance matrix for source n is formed using the SS-bins across all TF domain that are assigned to source n

$$\mathbf{R}^n_{MSEC} = \frac{1}{|S_n|} \sum_{(\tau,k)\in S_n} \mathbf{a_{lm}}(\tau, k)\mathbf{a_{lm}}^H(\tau, k), \quad (17)$$

where $|S_n|$ indicates the number of members in cluster S_n. Using SVD on $\mathbf{R}^n_{MSEC}$ as in (5), the MSEC-MUSIC spectrum for each source is given as

$$P^n_{MSEC-MUSIC}(\Omega) = \frac{1}{\| \left(\mathbf{U_v}^n\right)^H \mathbf{Y_{lm}}^*(\Omega)\|^2}, \quad (18)$$

in which the global peak indicates the estimated DOA for that source.

IV. Evaluations

An evaluation of methods were conducted using simulation for a $5 \times 6 \times 4\,\mathrm{m}$ shoebox room with $T_{60} = 0.4\,\mathrm{s}$ [17]. The Spherical Microphone arrays Impulse Response Generator (SMIRgen) [18] based on Allen & Berkley's image method [19] was used for 32-element rigid SMA with radius of $42\,\mathrm{cm}$ placed at $(2.52, 4.48, 1.45)\,\mathrm{m}$. N sources were randomly placed with azimuth interval of $\triangle\phi$, at distance of $1\,\mathrm{m}$ from the centre of SMA on the same horizontal plane as SMA. Per each pair of N and $\triangle\phi$, we used 100 random trials where in each trial the first azimuth was randomly selected from a uniform circular distribution around the SMA. The source signals consist of different anechoic speech signals randomly selected for each trial from the APLAWD database [20]. The active level of each speech source according to ITU-T P.56

ISBN 978-0-9928626-7-1 © EURASIP 2017

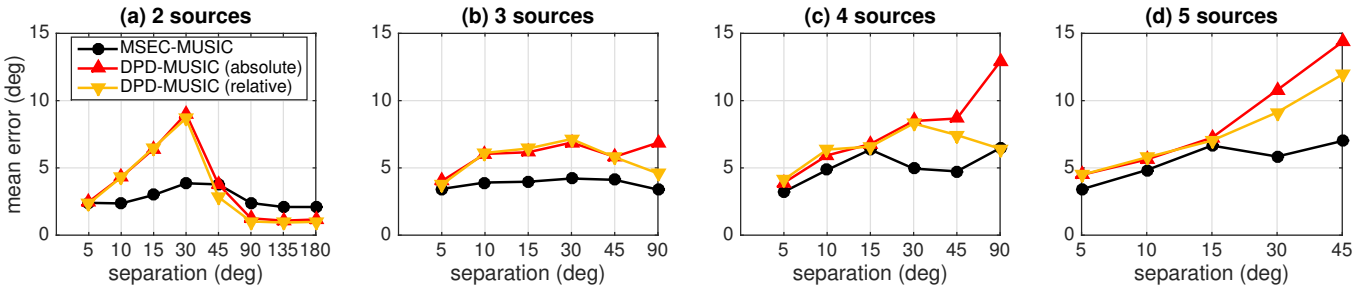

Fig. 1: Mean error of MSEC-MUSIC, relative and absolute DPD-MUSIC for varying N and $\triangle\phi$.

[21], as measured at omnidirectional eigenbeam, is set to be equal across all trials. Spatio-temporally white Gaussian noise is added to the microphone signals to produce a signal to incoherent noise ratio (iSNR) of 25 dB for each source. A sampling frequency of 8 kHz was used with frame length of 4 ms and 50% overlapping of time frames.

MSEC was performed on initial DOA estimates obtained by PIV [10] DOA estimator. We empirically chose $K_{max} = 4$, $T = 4$ frames, $\gamma = 0.3$ for the average cardinality threshold in outlier removal. DPD test had $J_\tau = 4$ and $J_k = 5$ as the size of its averaging window in the TF domain in (7). We empirically chose the threshold in (9) as $\epsilon = 6$ which also matches the recommended value in the original paper [5]. The MUSIC spectrum in (11) and (18) was calculated with 1° resolution across azimuth and inclination (360×181). Incoherent DPD-MUSIC was excluded from our evaluation since studies in [13] show that incoherent DPD-MUSIC fails in case of low angular separation of sources as the two peaks associated with two adjacent sources can be merged into one peak over summation of MUSIC spectra which causes the second highest peak to be detected far from the sources.

In order to avoid any ambiguity due to data association uncertainty in our results, best case data association was used to obtain the mean estimation error between true DOAs and estimated DOAs.

The original DPD test is based on absolute selection due to comparison of SVR with a fixed threshold. This results in reduction of selected TF bins as the number of sources increases for DPD unlike MSEC which is based on relative selection of top $M\%$ best DOA estimates. For the purpose of fairness in the selection process in our evaluation, we also include an alternative DPD-MUSIC based on relative selection which selects the TF bins with the top $M\%$ SVR, η_{DPD}.

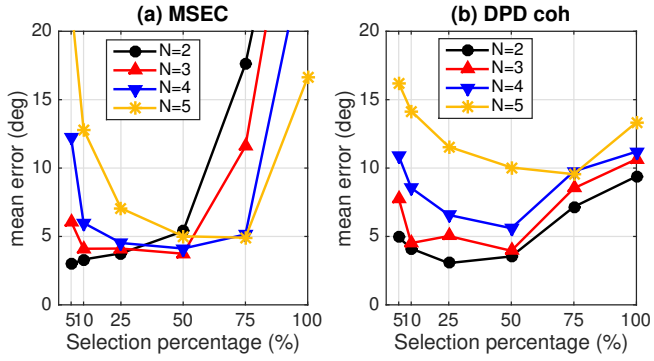

Fig. 2: Mean estimation error as a function of M for MSEC- and relative DPD-MUSIC for varying N and $\triangle\phi = 45^\circ$.

A. *Effect of selection percentage*

In this section, we evaluate the effect of selection percentage for MSEC and DPD-MUSIC with relative selection in order to find the optimum M for both methods.

Figure 2 shows the mean estimation error as a function of selection percentage M for $N = \{2, 3, 4, 5\}$ and $\triangle\phi = 45^\circ$. As we can see in Fig. 2, low ($\leq 10\%$) and high ($> 50\%$) values of M respectively cause underestimation and overestimation of number of bins which both result in high estimation error. As expected, the optimum M increases with increasing N. We can also observe that the average performance of MSEC, compared to DPD, is more dependant on the M since the value of M directly affects the quality of the covariance matrix in (17) which is the input to SVD of MUSIC unlike DPD in which the covariance matrix calculation in (7) is independent of M. According to these findings, the value $M = 25\%$ is selected for both MSEC and relative-DPD.

B. *Overall Evaluation*

In this section, we evaluate the best performing approaches of MSEC-MUSIC ($M = 25\%$), absolute ($\epsilon = 6$) and relative ($M = 25\%$) coherent DPD-MUSIC for $N = \{2, 3, 4, 5\}$ and widely varying $\triangle\phi$.

As can be seen in Fig. 1, in all cases of N for all methods the performance of DOA estimation improves as $\triangle\phi$ decreases below 30°. Since the spatial resolution of Y_{lm} depends on the maximum SH order L, below a certain $\triangle\phi$ multiple sources active in a TF bin are considered as a single source spatially between the true sources and therefore that bin is selected as a SS-bin. In such cases, the lower the angular separation of sources is, the lower the estimation error will be. For separation of $\triangle\phi \geq 30^\circ$, both DPD-MUSIC methods in our experiments lose accuracy and robustness to N and $\triangle\phi$ unlike MSEC-MUSIC which shows relatively strong robustness. As expected, relative DPD shows higher robustness to N as it uses dynamic selection process unlike absolute DPD with static selection. Overall MSEC-MUSIC, due to

ISBN 978-0-9928626-7-1 © EURASIP 2017

global consideration of SS-bins, shows stronger robustness to source separation and number of sources as it varies from 2.4° to 6.5° compared to DPD-MUSIC which is based on local consideration of SS-bin and changes from 2.2° to 15° mean estimation error.

V. Conclusions

A DOA estimation method has been proposed for multiple active sources. The method exploits a variant of multi-source clustering of speaker-dominant time frequency bins to make a fundamental change to the computation of the spatial covariance matrix used in the MUSIC algorithm. The effectiveness of this approach has been tested for multiple simultaneously active speech sources in a simulated acoustic environment with $0.4\,\mathrm{s}$ reverberation time, and using a spherical microphone array. The simulation shows that our technique MSEC-MUSIC significantly outperforms the state-of-the-art DPD-MUSIC with less than 6.5° mean estimation error, 4° and 2.5° robustness to number of sources and source separation respectively for up to 5 sources with widely varying source separations in the presence of realistic reverberation and sensor noise. As a conclusion, our work indicates that estimation of a global covariance matrix per speaker, compared to clustering of local signal spaces derived from local covariance matrices, leads to a more accurate global signal space per speaker.

References

[1] D. P. Jarrett, E. A. Habets, and P. A. Naylor, *Theory and Applications of Spherical Microphone Array Processing*, Springer Topics in Signal Processing. Springer, Berlin Heidelberg, 2016.

[2] B. Rafaely, *Fundamentals of Spherical Array Processing*, Springer Topics in Signal Processing. Springer, Berlin Heidelberg, 2015.

[3] R. O. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antennas Propag.*, vol. 34, no. 3, pp. 276–280, 1986.

[4] 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*, 2016.

[5] O. Nadiri and B. Rafaely, "Localization of multiple speakers under high reverberation using a spherical microphone array and the direct-path dominance test," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 22, no. 10, pp. 1494–1505, Oct. 2014.

[6] 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 *Proc. European Signal Processing Conf. (EUSIPCO)*, Bucharest, Romania, August 2012.

[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 *Proc. European Signal Processing Conf. (EUSIPCO)*, Nice, France, September 2015.

[8] S. Hafezi, A. H. Moore, and P. A. Naylor, "Multiple source localization using estimation consistency in the time-frequency domain," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing (ICASSP)*, New Orleans, LA, USA, March 2017.

[9] S. Hafezi, A. H. Moore, and P. A. Naylor, "Multi-source estimation consistency for improved multiple direction-of-arrival estimation," in *Joint Workshop on Hands-free Speech Communication and Microphone Arrays (HSCMA)*, San Francisco, CA, USA, March 2017.

[10] D. P. Jarrett, E. A. P. Habets, and P. A. Naylor, "3D source localization in the spherical harmonic domain using a pseudointensity vector," in *Proc. European Signal Processing Conf. (EUSIPCO)*, Aalborg, Denmark, Aug. 2010, pp. 442–446.

[11] E. G. Williams, *Fourier Acoustics: Sound Radiation and Nearfield Acoustical Holography*, Academic Press, London, first edition, 1999.

[12] S. Hafezi, A. H. Moore, and P. A. Naylor, "3D acoustic source localization in the spherical harmonic domain based on optimized grid search," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing (ICASSP)*, Shanghai, China, March 2016.

[13] S. Hafezi, A. H. Moore, and P. A. Naylor, "Multiple source localization in the spherical harmonic domain using augmented intensity vectors based on grid search," in *Proc. European Signal Processing Conf. (EUSIPCO)*, Budapest, Hungary, September 2016.

[14] B. D. van Veen and K. M. Buckley, "Beamforming: A versatile approach to spatial filtering," *IEEE Acoustics, Speech and Signal Magazine*, vol. 5, no. 2, pp. 4–24, Apr. 1988.

[15] H. Akaike, "A new look at the statistical model identification," *IEEE Trans. Autom. Control*, vol. AC-19, no. 6, pp. 716–723, Dec. 1974.

[16] C. D. Manning, P. Raghavan, and H. Schutze, *Introduction to Information Retrieval*, Cambridge University Press, Cambridge, UK, 2008.

[17] P. A. Naylor and N. D. Gaubitch, Eds., *Speech Dereverberation*, Springer, 2010.

[18] D. P. Jarrett, "Spherical Microphone array Impulse Response (SMIR) generator," http://www.ee.ic.ac.uk/sap/smirgen/.

[19] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *J. Acoust. Soc. Am.*, vol. 65, no. 4, pp. 943–950, Apr. 1979.

[20] G. Lindsey, A. Breen, and S. Nevard, "SPAR's archivable actual-word databases," Technical report, University College London, June 1987.

[21] ITU-T, "Objective measurement of active speech level," Dec. 2011.

ISBN 978-0-9928626-7-1