2D DIRECTION OF ARRIVAL ESTIMATION OF MULTIPLE MOVING SOURCES USING A SPHERICAL MICROPHONE ARRAY

Alastair H. Moore, Christine Evers and Patrick A. Naylor

Imperial College London, Dept. of Electrical and Electronic Engineering

ABSTRACT

Direction of arrival estimation using a spherical microphone array is an important and growing research area. One promising algorithm is the recently proposed Subspace Pseudo-Intensity Vector method. In this contribution the Subspace Pseudo-Intensity Vector method is combined with a state-of-the-art method for robustly estimating the centres of mass in a 2D histogram based on matching pursuits. The performance of the improved Subspace Pseudo-Intensity Vector method is evaluated in the context of localising multiple moving sources where it is shown to outperform competing methods in terms of clutter rate and the number of missed detections whilst remaining comparable in terms of localisation accuracy.

Index Terms— direction of arrival estimation, localisation, tracking, spherical harmonic domain, subspace pseudo-intensity vectors, PIV

1. BACKGROUND

Spherical microphone arrays have received growing attention in recent years thanks to their ability to form direction-independent beam patterns which can be steered anywhere in 2D (azimuth/inclination) space [1–3]. Furthermore, representing a sampled sound field in terms of its spherical harmonic decomposition allows algorithms to be developed which are, to a large extent, independent of the specific microphone positions. This microphone position-independence makes the Spherical Harmonic Domain (SHD) a natural choice for signal processing in the context of robot audition, where it is desirable to integrate a pseudo-spherical microphone array into the head of a humanoid robot [4].

One of the fundamental problems in microphone array signal processing is Direction-of-Arrival (DOA) estimation. For real world applications, and especially in robot audition, it is desirable to localise[1] moving sources. Since localisation algorithms generally assume that the sources are static, one must partition the signals due to moving sources into observation intervals over which the assumption is at least approximately true. Making the observation intervals as short as possible reduces the latency and increases the validity of the static source assumption. However, shorter observation intervals also increase the variance in the estimated DOAs, increase the probability of missed detections (where the source is present but localisation does not return a corresponding DOA) and increase the number of clutter measurements (where a DOA is produced which does not relate to a true source.) The increase in the variance of the estimated DOAs can be reduced by exploiting the time history of the DOAs using a tracker [5]. However, high rates of clutter and missed detections in the DOA estimates generally deteriorate tracking performance. It is therefore desirable to use a DOA estimation algorithm which, given short observation intervals, minimises the missed detections and clutter, whilst maintaining acceptable angular error.

A number of algorithms have been formulated which exploit the SHD [6–15]. The recently proposed Subspace Pseudo-Intensity Vector (SSPIV) method, [14], operates on the spatial covariance matrix. As in [8, 10, 11, 13], frequency smoothing [16] is first used to decorrelate coherent reflections. Also, similar to [8, 10, 13], Singular Value Decomposition (SVD) is used to partition the covariance matrix into signal and noise subspaces. However, uniquely, the SSPIV method operates directly on the signal subspace, whereas previous approaches have used variants of Multiple Signal Classification (MUSIC). Under the assumption that a single source is active in a particular Time-Frequency (TF) region, the signal subspace is one dimensional and proportional to the spherical harmonics evaluated in the DOA. Therefore the Pseudo-Intensity Vector (PIV) method [9] can be applied to the signal subspace to yield a vector which indicates an estimated DOA directly.

In this contribution we combine multiple SSPIVs into a smoothed histogram from which the DOAs can be estimated using the approach of [15]. In contrast to [14] the SSPIV method is not combined with the Direct-Path Dominance (DPD) test proposed in [13]. This results in vastly more vectors being available to contribute to the histogram, at the expense of increased variance in those vectors' directions. The updated approach is compared to a selection of competing algorithms specifically in the context of short observation intervals, as required for localisation of moving sources.

The research leading to these results has received funding from the European Union's Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 609465

[1] We use the terms DOA estimation and localisation interchangeably.

2. PROBLEM FORMULATION

In a reverberant environment with N_d simultaneously active sources in the far field, the soundfield at the origin is composed of $N \geq N_d$ planewaves due to the presence of reflected wavefronts. The planewave density of the soundfield is given in the Short Time Fourier Transform (STFT)-SHD as

$$a_{lm}(\nu, \ell) = \sum_{n=1}^{N} \left[Y_l^m(\Psi_n)\right]^* s_n(\nu, \ell) \tag{1}$$

where ν and ℓ denote the discrete frequency and time indices, respectively, $s_n(\nu, \ell)$ is the complex amplitude of the n-th planewave, $\Psi_n = (\theta_n, \phi_n)$ its direction of arrival and Y_l^m is the spherical harmonic of order l and degree m [3]. Considering the $(L+1)^2$ spherical harmonics up to $l \leq L$, (1) is expressed in vector notation as

$$\mathbf{a}_{lm}(\nu, \ell) = \mathbf{Y}(\boldsymbol{\Psi})^H \mathbf{s}(\nu, \ell) \tag{2}$$

where subscript lm on a vector denotes that the elements are spherical harmonic coefficients, $\boldsymbol{\Psi} = [\Psi_1 \ldots \Psi_N]^T$,

$$\mathbf{Y}(\boldsymbol{\Psi}) = \begin{bmatrix} \mathbf{y}(\Psi_1) \\ \vdots \\ \mathbf{y}(\Psi_N) \end{bmatrix}, \tag{3}$$

$\mathbf{y}(\Psi_n) = \left[Y_0^0(\Psi_n)\, Y_1^{-1}(\Psi_n)\, Y_1^0(\Psi_n)\, Y_1^1(\Psi_n) \ldots Y_L^L(\Psi_n)\right]$, $\mathbf{s}(\nu, \ell) = [s_1(\nu, \ell) \ldots s_N(\nu, \ell)]^T$ and $(\cdot)^T$ and $(\cdot)^H$ denote the transpose and conjugate transpose, respectively. Therefore, $\mathbf{a}_{lm}(\nu, \ell)$ and $\mathbf{Y}(\boldsymbol{\Psi})$ have dimensions $(L+1)^2 \times 1$ and $N \times (L+1)^2$, respectively.

To obtain this SHD representation the pressure on the surface of a sphere of radius r is sampled by Q microphones

$$\mathbf{a}_{lm}(\nu, \ell) \approx \mathbf{B}(k_\nu r)^{-1} \mathbf{Y}(\boldsymbol{\Omega})^H \mathbf{W} \mathbf{p}(\nu, \ell, r) \tag{4}$$

where $\mathbf{p}(\nu, \ell, r) = [p_1 \ldots p_Q]^T$ is the pressure at each of the sample points, $\mathbf{W} = \operatorname{diag}\{w_1\, w_2 \ldots w_Q\}$, where $\{w_q\}_1^Q$ are the weights of the sampling scheme, and $\mathbf{Y}(\boldsymbol{\Omega})$ is a $Q \times (L+1)^2$ matrix defined as in (3) but with the spherical harmonics evaluated at $\{\Omega_q\}_1^Q$. The diagonal matrix, $\mathbf{B}(k_\nu r)^{-1}$, is a frequency-dependent compensation for the mode strength of the microphone array, which also depends on whether the microphones are placed on an open or rigid sphere, and k_ν is the wavenumber at the ν-th frequency. The approximation in (4) is valid provided (i) $k_\nu r < L$, (ii) $Q \geq (L+1)^2$ sample positions are approximately equally distributed over the sphere and (iii) the sampling weights are chosen appropriately [17].

In practice, the microphone signals, $\mathbf{x}(\nu, \ell)$, are distorted by additive sensor noise, $\mathbf{v}(\nu, \ell)$, such that $\mathbf{x}(\nu, \ell) = \mathbf{p}(\nu, \ell, r) + \mathbf{v}(\nu, \ell)$, where $\mathbf{x}(\nu, \ell)$ and $\mathbf{v}(\nu, \ell)$ have the same dimensions as $\mathbf{p}(\nu, \ell, r)$. Using (2), the noisy planewave decomposition is

$$\tilde{\mathbf{x}}_{lm}(\nu, \ell) = \mathbf{Y}(\boldsymbol{\Psi})^H \mathbf{s}(\nu, \ell) + \tilde{\mathbf{v}}_{lm}(\nu, \ell) \tag{5}$$

where

$$\tilde{\mathbf{v}}_{lm}(\nu, \ell) = \mathbf{B}(k_\nu r)^{-1} \mathbf{Y}(\boldsymbol{\Omega})^H \mathbf{W} \mathbf{v}(\nu, \ell). \tag{6}$$

The aim of the present study is to use $\tilde{\mathbf{x}}_{lm}(\nu, \ell)$ to estimate the DOAs, $\boldsymbol{\Psi}$, of the N_d sources as they move relative to a static spherical microphone array.

3. PROPOSED METHOD

From (2), the covariance matrix of $\mathbf{a}_{lm}(\nu, \ell)$ is [13]

$$\begin{aligned} \mathbf{R}_{\mathbf{a}_{lm}} &= E\left\{\mathbf{a}_{lm}\mathbf{a}_{lm}^H\right\} && (7) \\ &= \mathbf{Y}^H(\boldsymbol{\Psi}) \mathbf{R}_\mathbf{s} \mathbf{Y}(\boldsymbol{\Psi}) && (8) \end{aligned}$$

where $\mathbf{R}_\mathbf{s} = E\left\{\mathbf{s}\mathbf{s}^H\right\}$ is the covariance of the source signals. Considering also the sensor noise, the observed covariance is

$$\begin{aligned} \mathbf{R}_{\tilde{\mathbf{x}}_{lm}} &= E\left\{\tilde{\mathbf{x}}_{lm}\tilde{\mathbf{x}}_{lm}^H\right\} && (9) \\ &= \mathbf{R}_{\mathbf{a}_{lm}} + \mathbf{R}_{\tilde{\mathbf{v}}_{lm}} && (10) \end{aligned}$$

which is approximated in the (ν, ℓ)-th TF-region as [13]

$$\hat{\mathbf{R}}_{\tilde{\mathbf{x}}_{lm}}(\nu, \ell) = \frac{1}{J_\nu J_\ell} \sum_{j_\nu=0}^{J_\nu-1} \sum_{j_\ell=0}^{J_\ell-1} \tilde{\mathbf{x}}_{lm}(\nu + j_\nu, \ell + j_\ell) \times \tilde{\mathbf{x}}_{lm}^H(\nu + j_\nu, \ell + j_\ell). \tag{11}$$

The averaging over time can be seen as approximating the expected value while the averaging over frequency implements frequency smoothing which decorrelates any coherent reflections [8]. Applying SVD to $\hat{\mathbf{R}}_{\tilde{\mathbf{x}}_{lm}}(\nu, \ell)$ leads to

$$\hat{\mathbf{R}}_{\tilde{\mathbf{x}}_{lm}}(\nu, \ell) = \mathbf{U}\boldsymbol{\Sigma}\mathbf{U}^H = [\mathbf{U}_s \mathbf{U}_n] \begin{bmatrix} \boldsymbol{\Sigma}_s & 0 \\ 0 & \boldsymbol{\Sigma}_n \end{bmatrix} \begin{bmatrix} \mathbf{U}_s^H \\ \mathbf{U}_n^H \end{bmatrix} \tag{12}$$

where $\mathbf{U}$ is a unitary matrix, $\boldsymbol{\Sigma}$ is a diagonal matrix containing the singular values of $\mathbf{R}_{\tilde{\mathbf{x}}_{lm}}$ and $\mathbf{U}_s$ and $\mathbf{U}_n$ respectively, represent the conventional partitioning into signal and noise subspaces [18].

In the simplest case of a single planewave, the column vector $\mathbf{U}_s = \left[\hat{a}_{00}\, \hat{a}_{1(-1)}\, \hat{a}_{10}\, \hat{a}_{11} \ldots \hat{a}_{LL}\right]^T$ is proportional to the steering vector for the planewave DOA, $\mathbf{y}(\Psi_n)$. The SSPIV is defined as [9, 14]

$$\tilde{\mathbf{I}}_{\text{ss}} = \frac{4\pi\sqrt{4\pi}}{3} \mathcal{R}\left\{\hat{a}_{00}^* \left[\begin{array}{ccc} \dot{D}_x & \dot{D}_y & \dot{D}_z \end{array}\right]^T\right\} \tag{13}$$

where dependence of all variables on (ν, ℓ) is assumed and $\dot{D}_\varpi = \sum_{m=-1}^{1} Y_1^m(\varphi_\varpi)\hat{a}_{1(m)}$ can be interpreted as a beamformer with dipole directivity pattern aligned with a Cartesian axis, $\varpi \in x, y, z$, such that the look directions of the beamformers are $\varphi_x = (\pi/2, 0)$, $\varphi_y = (\pi/2, \pi/2)$ and $\varphi_z = (0, 0)$. Note that in contrast to the intensity-based vectors in [9,14] which point in the direction of energy flow, i.e. away from the source, the vector in (13) is defined to point towards the DOA. This reversal leads to a more natural discussion of DOAs and vector directions.

Where multiple sources are present, sparsity in the time-frequency domain suggests that most SSPIVs will point towards one of the sources, even if some of them are distorted by an interfering source. Therefore a single histogram formed from the orientations of the vectors over all frequencies and falling within the T frame observation interval $(\ell - T, \ell]$ is expected to have peaks corresponding to the desired source DOAs at time frame ℓ. We estimate the positions of these peaks using the robust method proposed in [15] which first smooths the histogram with a 2D Gaussian kernel before sequentially identifying the largest centres of mass using matched pursuits with 2D Gaussian dictionary elements.

4. EVALUATION

To evaluate the accuracy of the proposed localisation algorithm in the context of moving sources we considered three test cases. In the first, 3 static sources of duration 8 seconds were simulated. By setting the observation interval to be the full signal duration the best case performance was established. In the second case the reduction in accuracy associated with short observation intervals was considered by performing localisation on the same signals but using observation intervals of 250 ms. Finally, the effect of the short term static source assumption was investigated by localising 3 moving sources, again using 250 ms observation intervals.

4.1. Implementation details

Signals were simulated for sources 1.5 m from a 32 channel rigid spherical microphone array with radius 4.2 cm [1] centred at (2.0 m, 2.5 m, 1.5 m) in a 5×6×3 m rectangular room with reverberation time of 0.5 s. Acoustic Impulse Responses (AIRs) from each source position to each microphone were simulated using the image-source method [19] [20]. In the static source cases the sources were placed at (80°, 60°), (100°, 180°) and (80°, 300°). For moving sources the trajectories started at the same positions as in the static case but the azimuth angles were increased linearly with time at a rate of 22.5° per second. Anechoic speech from the TIMIT database was concatenated without gaps to form 8-second segments and convolved with the AIRs. For the moving sources this was achieved using overlap-add processing with the source positions quantised to 5° resolution such that the reverberation was always consistent with the source positions at the time the signal was emitted. The level of each reverberant source signal was normalised according to ITU P56 [21] such that the direct path components of each was equal and Independent and Identically Distributed (i.i.d.) white Gaussian noise was added to each microphone signal to obtain 40 dB SNR with respect to the direct path signals. For each test case the experiment was repeated 30 times using different speech segments and realisations of noise on each repetition. Each 250 ms observation interval overlapped the previous interval by 150 ms, such that DOA were estimated every 100 ms.

The STFT used 8 ms frames with 75% overlap and the SHD representation was achieved according to (4). The maximum spherical harmonic order was $L = 3$ giving a maximum frequency of 3850 Hz to ensure $k_\nu r < L$. The lowest frequency bin was centred at 500 Hz, which avoids excessive noise amplification due to mode strength compensation. $\hat{\mathbf{R}}_{\tilde{\mathbf{x}}_{lm}}(\nu, \ell)$ was obtained with $J_\nu = 2$ and $J_\ell = 13$, which corresponds to averaging over a TF-region of 32 ms and 250 Hz. The SSPIVs were obtained as in (11), (12) and (13). Histograms were formed using 2° resolution in azimuth and inclination, the Gaussian kernel used for smoothing had a standard deviation of 5° and that used for the matching pursuits had a standard deviation of 20°. As in [15], the number of sources was assumed known *a priori* such that the first (largest) 3 peaks were taken as the DOA estimates.

The proposed method was compared to three methods from the literature. The PIV method [9] produces a set of vectors for each observation interval. The same smoothed histogram with matching pursuits method as for the proposed method was used to obtain the estimated DOAs. The Plane-Wave Decomposition-Steered Response Power (PWD-SRP) method used a planewave decomposition beamformer [22] steered in 180×91 directions to produce a power map. The 3 largest peaks were taken as the DOA estimates. The DPD-MUSIC method was implemented as described in [13]. However, for consistency the same STFT, J_ν and J_ℓ parameters were used as for the proposed method. This actually led to an improvement in performance over those used in [13]. The 3 largest peaks of the spatial spectrum were taken as the DOA estimates. Since the power map/spatial spectrum of the PWD-SRP and DPD-MUSIC are not guaranteed to have 3 peaks it is possible for these methods to return fewer than 3 estimated DOAs for a given observation interval.

For each observation interval an estimated DOA was assigned to a source if it was within 30° of the true source direction. The 4 proposed methods are compared in terms of the mean angular error in the assigned DOAs. The mean clutter per interval is the total number of unassigned DOA estimates divided by the number of intervals. Since there were 3 estimated DOAs per interval the clutter rate is between 0 and 3. For each source, the miss rate is the proportion of observation intervals in which none of the estimated DOAs are within 30° of the true DOA. The mean miss rate is the average miss rate across all three sources and is in the range 0-100%.

4.2. Results and discussion

The results of all tests are shown in Table 1. For the static case where the full 8 second signals were used to form a sin-

Algorithm	Mean angular error [°]			Mean clutter per interval		Mean miss rate [%]	
	Static 8s	Static 0.25s	Moving 0.25s	Static 0.25s	Moving 0.25s	Static 0.25s	Moving 0.25s
SSPIV	2.19	6.86	7.84	**0.27**	**0.24**	**9**	**8**
PIV	5.71	8.67	9.28	0.61	0.55	21	18
PWD-SRP	4.91	7.84	8.32	0.92	1.00	31	34
DPD-MUSIC	**0.82**	**3.70**	**5.49**	0.78	0.81	27	28

Table 1. Performance metrics for compared methods.

gle set of DOA estimates, all methods successfully localised all sources. That is, the miss rate and clutter were both 0 for all algorithms. The localisation accuracy was fair in all cases with DPD-MUSIC achieving the lowest error (0.82°). The proposed method was second best (2.19°), followed by PWD-SRP (4.91°) and PIV (5.71°).

Using 250 ms observation intervals the rank order of the algorithms in terms of angular error was unchanged but in all cases there was a substantial increase in the error of about 3-4°. However, the proposed method was better than the competing methods both in terms of the mean clutter and the miss rate. In fact, SSPIV achieved less than half the clutter and miss rate compared to the next best method (PIV).

For moving sources there was a consistent increase in the angular error of about 0.5-1.5° which is likely due to the lag introduced by the observation interval. However, the clutter and miss rates were consistent with those in the static case. Figure 1 shows the azimuth component estimated by each algorithm as a function of time for a representative trial. In general the correspondence between the ground truth trajectories and the estimated DOAs is clear. Whilst the variance of the estimates obtained with the proposed method are visibly bigger than for DPD-MUSIC, the lower clutter rates and absence of missed detections suggest that the proposed method may be better suited to tracking moving sources.

5. CONCLUSIONS

The SSPIV method has been presented in the context of moving source DOA estimation and compared to state-of-the-art algorithms in a simulated environment with moderate reverberation time (0.5 s). In addition to localisation accuracy, tracking moving sources also requires low rates of clutter and missed detections. Of the methods compared, only DPD-MUSIC achieved lower localisation error for (5.49° vs 7.84°). On the other hand, the proposed method had substantially lower rates of clutter (0.24 vs 0.81) and misses (8% vs 28%) compared to DPD-MUSIC suggesting that SSPIV is well suited for tracking multiple moving sources.

REFERENCES

[1] J. Meyer and G. Elko, "A highly scalable spherical microphone array based on an orthonormal decomposition of the soundfield," in *Proc. IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, May 2002, vol. 2, pp. 1781–1784.

[2] B. Rafaely, "Analysis and design of spherical microphone arrays," *IEEE Trans. Speech Audio Process.*, vol. 13, no. 1, pp. 135–143, Jan. 2005.

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

[4] V. Tourbabin and B. Rafaely, "Theoretical framework for the optimization of microphone array configuration for humanoid robot audition," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 22, no. 12, pp. 1803–1814, Dec. 2014.

[5] C. Evers, J. Sheaffer, A. H. Moore, B. Rafaely, and P. A. Naylor, "Bearing-only acoustic tracking of moving speakers for robot audition," in *Proc. IEEE Intl. Conf. Digital Signal Processing (DSP)*, Singapore, July 2015.

[6] B. Rafaely, "Plane-wave decomposition of the pressure on a sphere by spherical convolution," *J. Acoust. Soc. Am.*, vol. 116, no. 4, pp. 2149–2157, Oct. 2004.

[7] H. Teutsch and W. Kellermann, "EB-ESPRIT: 2D localization of multiple wideband acoustic sources using eigen-beams," in *Proc. IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, Mar. 2005, vol. 3, pp. iii/89–iii/92.

[8] D. Khaykin and B. Rafaely, "Coherent signals direction-of-arrival estimation using a spherical microphone array: Frequency smoothing approach," in *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, New Paltz, NY, USA, Oct. 2009, pp. 221–224.

[9] 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.

[10] B. Rafaely, Y. Peled, M. Agmon, D. Khaykin, and E. Fisher, "Spherical microphone array beamforming," in *Speech Processing in Modern Communication: Challenges and Perspectives*, I. Cohen, J. Benesty, and S. Gannot, Eds., chapter 11. Springer, Jan. 2010.

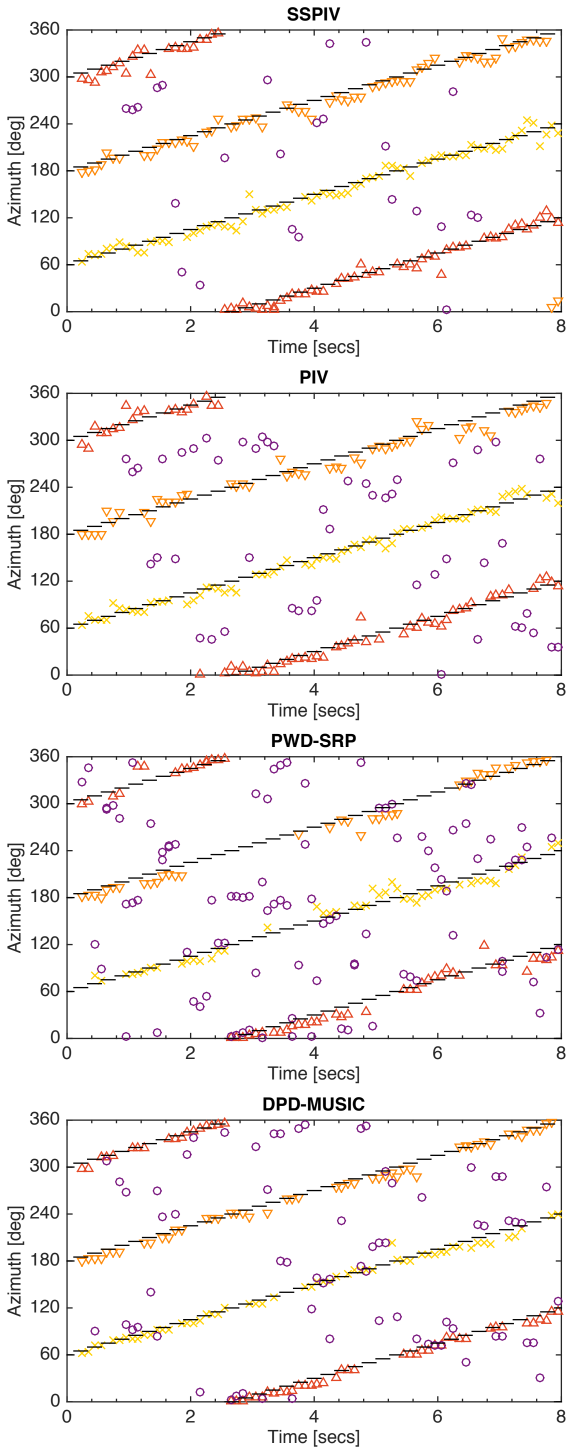

Fig. 1. Azimuth component of DOAs as a function of time estimated using the SSPIV, PIV, PWD-SRP and DPD-MUSIC algorithms. Source assignments used for calculating performance metrics are indicated by symbol shape (× source 1, ▽ source 2, △ source 3). Broken lines show ground truth trajectories and ○ indicates unassigned (clutter) estimates.

[11] H. Sun, E. Mabande, K. Kowalczyk, and W. Kellermann, "Localization of distinct reflections in rooms using spherical microphone array eigenbeam processing," *J. Acoust. Soc. Am.*, vol. 131, pp. 2828–2840, 2012.

[12] C. Evers, A. H. Moore, and P. A. Naylor, "Multiple source localisation in the spherical harmonic domain," in *Proc. Intl. Workshop on Acoustic Signal Enhancement (IWAENC)*, Nice, France, July 2014.

[13] 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 Trans. Audio, Speech, Lang. Process.*, vol. 22, no. 10, pp. 1494–1505, Oct. 2014.

[14] A. H. Moore, C. Evers, P. A. Naylor, D. L. Alon, and B. Rafaely, "Direction of arrival estimation using pseudo-intensity vectors with direct-path dominance test," in *Proc. European Signal Processing Conf. (EUSIPCO)*, 2015.

[15] 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)*, 2015.

[16] H. Wang and M. Kaveh, "Coherent signal-subspace processing for the detection and estimation of angles of arrival of multiple wide-band sources," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 33, no. 4, pp. 823–831, Aug. 1985.

[17] B. Rafaely, B. Weiss, and E. Bachmat, "Spatial aliasing in spherical microphone arrays," *IEEE Trans. Signal Process.*, vol. 55, no. 3, pp. 1003–1010, Mar. 2007.

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

[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] D. P. Jarrett, E. A. P. Habets, M. R. P. Thomas, and P. A. Naylor, "Rigid sphere room impulse response simulation: algorithm and applications," *J. Acoust. Soc. Am.*, vol. 132, no. 3, pp. 1462–1472, Sept. 2012.

[21] "Objective measurement of active speech level," Mar. 1993.

[22] B. Rafaely, "Phase-mode versus delay-and-sum spherical microphone array processing," *IEEE Signal Process. Lett.*, vol. 12, no. 10, pp. 713–716, Oct. 2005.