Narrowband multi-source Direction-of-Arrival estimation in the spherical harmonic domain

Sina Hafezi,[1, a] Alastair H. Moore,[1, b] and Patrick A. Naylor[1, c]

Department of Electrical and Electronic Engineering, Imperial College London, SW7 2AZ, UK

(ΩDated: 14 March 2021)

A conventional approach to wideband Multi-Source (MS) Direction-of-Arrival (DOA) estimation is to perform Single Source (SS) DOA estimation in Time-Frequency (TF) bins for which a SS assumption is valid. Such methods use W-Disjoint Orthogonality (WDO) assumption, due to the speech sparseness. As the number of sources increases, the chance of violation of WDO assumption increases. As shown, in the challenging scenarios with multiple simultaneously active sources over a short period of time masking each other, it is possible for a strongly masked source (due to inconsistency of activity or quietness) to be rarely dominant in a TF bin. SS-based DOA estimators fail in detection or accurate localization of masked sources in such scenarios. Two analytical approaches are proposed for narrowband DOA estimation based on MS assumption in a bin in the spherical harmonic domain. In the first approach EigenValue Decomposition (EVD) is used to decompose a MS scenario into multiple SS scenarios on each of which a SS-based analytical DOA estimation is performed. The second approach analytically estimates two DOAs per bin assuming the presence of two active sources per bin. The evaluation validates the improvement to double accuracy and robustness to sensor noise compared to the baseline methods.

[a] s.hafezi14@imperial.ac.uk

[b] alastair.h.moore@imperial.ac.uk

[c] p.naylor@imperial.ac.uk

I. INTRODUCTION

Multi-Source (MS) Direction-of-Arrival (DOA) estimation is a fundamental task for acoustic source separation, localization, tracking, spatial filtering, environment mapping, dereverberation and speech enhancement where multiple sources are simultaneously active. As such it can be used in a wide range of applications including hearing aids, robot audition, meeting diarization and teleconferencing.

This work is based on Spherical Microphone Arrays (SMAs) which have become popular due to their ability to analyse the sound field with equal resolution in all directions. Using a SMA, the sound field can be represented in the Spherical Harmonic Domain (SHD), which enables the problem formulation to be independent of the specific geometry, size and type of the SMA, e.g. rigid/open. The studies on SMAs can be categorized into two groups of array design and processing. Numerous works[1,2,5,7,9,10,12] have provided studies of various designs for SMAs and analysis of their advantages or limitations. There have also been extensive studies on soundfield modelling[3,4,15], beamforning[8,11,13,14] and localization[6,16,17] for SMAs, those of which related to the context of this paper will be discussed later. The first works on DOA estimation using SMA can be found in[18] and the first works on signal processing for SMAs are presented in[2,19,20].

The degradation factors for the performance of DOA estimators are typically reverberation, sensor/environmental noises and sources' activity, strength and movement. The interested case of study in this work is the challenging scenarios in which the sources are stationary and simultaneously active with different average strength where one source is

mostly and significantly masked by others. This often occurs in scenarios where the number of sources exceeds two and one or more sources are at further distance than others.

Several existing approaches to wideband MS DOA estimation for speech sources exploit W-Disjoint Orthogonality (WDO)[21], assuming sparseness of speech in the Time-Frequency (TF) domain, in combination with subspace decomposition to decompose the noisy observation covariance matrix into signal and noise subspaces[22]. Such methods often find the TF bins where the Single Source (SS) assumption is valid to some extent using various metrics[6,23–27], and apply a SS-based narrowband DOA estimator at the identified SS bins. Such approaches benefit from high accuracy and reliability of the SS-based narrowband DOA estimators due to the simplicity of the SS problem.

Narrowband DOA estimation methods either provide an analytical closed-form solution and or are open-form, requiring a cost function to be evaluated over many possible directions and the optimal direction(s) selected. In analytical-based methods, the DOA is directly calculated from the signals whereas the steering-based methods combine the signals with the steering vector of the array for all the possible directions to find the direction with the most reliable outcome. Pseudointensity vectors (PIVs)[17] uses up to the first-order spatial information of the captured sound field to directly estimate the DOA. A noise-robust variation of PIV, SSPIV, is proposed in[28] using subspace decomposition in which the first eigenvector of the sample covariance matrix is used as the input to the PIV formulation. Although analytical approaches such as PIV are computationally fast by avoiding many possible directions and estimating the DOA directly instead, they suffer from limitation to low spatial

information which results in erroneous DOA estimate in case of narrowband MS scenario as shown later in this work.

On the other hand, Steered Response Power (SRP)[29] and MUltiple SIgnal Classification (MUSIC)[6,30] are two of the well-known steering-based methods. In SRP a beamformer such as Plane-Wave Decomposition (PWD)[11,14,31] is applied to the raw signals. The recent works[36,37] use model-based Bayesian analysis with beamforming to determine the number of sources and their DOAs. They generate and evaluate number of models, each defined by a specific number of sources and DOAs, to find the best fit to the data. The noise-robust MUSIC uses subspace decomposition and minimizes the output of the beamforming on the noise subspace (in fact maximises the inverse of the beamforming output). The noise subspace in MUSIC depends on the rank of the covariance matrix. For SS assumption, the first eigenvector is taken as the signal subspace and the remaining eigenvectors as the noise subspace. Steering-based approaches benefit from their formulation's extendibility to high order spatial information. However, they can be computationally expensive when used to obtain a a narrowband DOA estimate per TF bin.

SS-based narrowband DOA estimators only result in accurate DOA if the WDO assumption is valid in that bin. Where multiple sources are simultaneously active in a bin, SS-based DOA estimator accuracy is reduced due to violation of this assumption. In the most optimistic cases where subspace decomposition is used, the SS-based methods estimate the DOA of the dominant source and, when taken as a set of DOA estimates, fail in detecting a quiet source if the source rarely has the chance of domination in a bin as shown later in this work.

A solution to this problem is the use of MS narrowband DOA estimation. A well-known example of it, is MSMUSIC[22,30] where the noise subspace is defined by the eigenvectors excluding the first N such vectors which represent the signal subspace, assuming the presence of N simultaneously active sources in the bin. The DOAs per bin are obtained as the top N peaks in the MUSIC spectrum. This approach suffers from high computational cost of steering and multi-peak detection per TF bin. The single peak detection in the SS-based approaches is relatively fast as it aims for the global maximum in a wrapped space whereas multi-peak detection requires multiple 2D-neighbouring check for peak detection, which is significantly more computationally expensive than the single peak detection.

The motivation behind this work is to study and propose a relatively low-computational and analytical narrowband MS DOA estimation algorithm which uses low-order spatial information to directly estimate single, two or more DOAs per TF bin. Two alternative methods for an analytical narrowband MS DOA estimation are proposed and named as Multi-Source Pseudointensity Vectors (MSPIVs) and Dual-Intensity Vectors (DIVs). The first method, MSPIV, is based on de-mixing a mixture of N sources into a multiple SS scenario on each of which PIV is performed. The second method DIV is based on a two-source assumption and analytically derives up to two DOAs from the mixture signals. A subspace version of the second method is also proposed in order to improve its robustness to noise and scenarios with more than two sources.

This paper is structured as follows. Section II demonstrates the signal model in the SHD and briefly reviews the conventional SS- and MS-based methods for the problem. Section III presents the proposed methods MSPIV and DIV. It also presents a narrowband

and a wideband analysis for simulated examples. Section IV evaluates the performance of the proposed methods against the conventional methods reviewed in Section II under simulated scenarios. Finally Section V provides an analysis of the computational complexity of each method.

II. TECHNICAL BACKGROUND

The point location in the spherical coordinates is represented as $(r, \Omega) = (r, \theta, \varphi)$ with range r, inclination θ, and azimuth φ. Let τ, k and κ denote the time frame, frequency index and wavenumber respectively.

A. Signal Model in the SHD

Consider a SMA with radius r_a and Q microphones, each with angle Ω_q. Let $p_q(\tau, k)$ denotes the signal captured by microphone q in the Short-Time Fourier Transform (STFT) domain. Using Spherical Harmonic (SH) Transform, also known as Spherical Fourier Transform, the signal in the SHD, referred to as eigenbeam of order l and degree m (satisfying $|m| \leq l$) is given by[32]

$$a_{lm}(\tau, k) \approx \frac{1}{b_l(\kappa r_a)} \sum_{q=1}^{Q} w_{q,lm} p_q(\tau, k), \tag{1}$$

where b_l is the range dependence mode strength of the SMA. In the case of a uniform sensor distribution, the weights $w_{q,lm}$ are

$$w_{q,lm} = \frac{4\pi}{Q} Y_{lm}^*(\Omega_q), \tag{2}$$

where $(.)^*$ denotes the complex conjugate. The SH basis functions $Y_{lm}(\Omega)$ are given by[32]

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

where P_{lm} is the associated Legendre function and $i^2 = -1$. To avoid spatial aliasing due to discrete sampling of the sound field the lower bound on Q is[7]

$$Q \geq (L+1)^2, \tag{4}$$

where L is the maximum SH order considered. Further information regarding the analysis of spatial aliasing errors and the selection of an appropriate spatial sampling scheme can be found in[33].

In the case of multiple N_s plane waves $\{S_n\}_{n=1}^{N_s}$ with DOAs $\{\Omega_n = (\theta_n, \varphi_n)\}_{n=1}^{N_s}$, the plane-wave representation of the eigenbeams is[12,33]

$$a_{lm}(\tau, k) = (\sum_{n=1}^{N_s} S_n(\tau, k) Y_{lm}^*(\Omega_n)) + \nu_{lm}, \tag{5}$$

where ν_{lm} is the sensor noise in the SHD.

B. Subsapce Decomposition

Subspace-based methods use eigenbeams covariance matrix,

$$\mathbf{R} = E\{\mathbf{a}_{lm}\mathbf{a}_{lm}^H\}, \tag{6}$$

where $E\{.\}$ and $(.)^H$ denote the expectation and conjugate transpose, respectively, and $\mathbf{a_{lm}} = \left[a_{00},\, a_{1(-1)},\, a_{1(0)},\, a_{1(1)}, \ldots,\, a_{LL}\right]^T$ is the $(L+1)^2$-length column vector of eigenbeams. Note that (τ, k) are omitted here for notational simplicity.

In order to obtain the covariance matrix $\mathbf{R}$, it is approximated as the average covariance matrix over a local TF region[6,28]

$$\mathbf{R_a}(\tau,k) = \frac{1}{J_\tau J_k}\sum_{j_\tau=0}^{J_\tau-1}\sum_{j_k=0}^{J_k-1}\mathbf{a}(\tau-j_\tau,k+j_k) \times \mathbf{a}^H(\tau-j_\tau,k+j_k), \tag{7}$$

where J_τ and J_k are the widths (number of bins) of the averaging windows over time and frequency respectively.

The EigenValue Decomposition (EVD) of the covariance matrix gives

$$\mathbf{R_a} = \mathbf{U}\mathbf{\Sigma}\mathbf{U}^H, \tag{8}$$

where $\mathbf{\Sigma}$ is the diagonal matrix containing the decreasingly sorted eigenvalues and $\mathbf{U}$ is the square matrix containing the eigenvectors as its columns.

Assuming the presence of N uncorrelated active sources, where $N < (L+1)^2$, let $\mathbf{U_s}^{(N)}$, representing the signal subspace, be the first N columns of $\mathbf{U}$ and $\mathbf{U_v}^{(N)}$, representing the noise subspace, be the remaining

$$\mathbf{U_s}^{(N)} = [\mathbf{U}_1 \dots \mathbf{U}_N], \tag{9}$$

$$\mathbf{U_v}^{(N)} = \left[\mathbf{U}_{N+1} \dots \mathbf{U}_{(L+1)^2}\right]. \tag{10}$$

C. MSMUSIC

MSMUSIC uses spatial steering vector and the noise subspace obtained from (10). The estimated DOAs of the N sources are the largest N peaks in the MUSIC spectrum given

by[22]

$$P_{MUSIC}(\Omega) = \frac{1}{\sum_{n=1}^{N} \|(\mathbf{U_v}^{(N)})^H \mathbf{Y_{lm}}^*(\Omega)\|^2}, \tag{11}$$

where $\mathbf{Y_{lm}} = \left[Y_{00}, Y_{1(-1)}, Y_{1(0)}, Y_{1(1)}, \ldots, Y_{LL}\right]^T$ is the steering vector for look direction Ω. Note that (Ω) are omitted for notational simplicity. Note that it requires the sources to be sufficiently separated in angular domain.

D. PWD-SRP

A baseline DOA estimator used for comparison is the SRP approach[29] implemented in the SHD using a PWD beamformer[31,38]. The narrow-band power response of SRP steered to look direction Ω is given as[11]

$$P_{PWD-SRP}(\Omega) = \mid \mathbf{a}_{lm}^T \mathbf{Y_{lm}}(\Omega) \mid^2 . \tag{12}$$

E. PIVs

Unlike steering-based methods such as MUSIC, PIV directly estimates a single DOA whose Cartesian coordinates are resulted from steering the first-order beams (dipoles) into the negative of the coordinate. It is calculated as[17]

$$\mathbf{I_{piv}}(\tau, k) = \frac{1}{2}\Re\left\{ a_{00}(\tau, k)^* \begin{bmatrix} D_{-x}(\tau, k, \mathbf{a_{lm}}) \\ D_{-y}(\tau, k, \mathbf{a_{lm}}) \\ D_{-z}(\tau, k, \mathbf{a_{lm}}) \end{bmatrix} \right\}, \tag{13}$$

where

$$D_\nu(\tau, k, \mathbf{a_{lm}}) = \sum_{m=-1}^{1} Y_{1m}(\phi_\nu) a_{1m}(\tau, k), \ \nu \in \{-x, -y, -z\} \tag{14}$$

are dipoles steered in the negative direction of Cartesian axes, given by $\phi_{-x} = (\pi/2, \pi)$, $\phi_{-y} = (\pi/2, -\pi/2)$ and $\phi_{-z} = (\pi, 0)$.

The PIV DOA unit vector $\mathbf{u_{piv}}$ is the negative of $\mathbf{I_{piv}}$ direction given by

$$\mathbf{u_{piv}}(\tau, k) = -\frac{\mathbf{I_{piv}}(\tau, k)}{\|\mathbf{I_{piv}}(\tau, k)\|}, \tag{15}$$

where $\|.\|$ indicates ℓ_2-norm.

F. SSPIVs

PIV is based on single source assumption and fails when there are more than one source active in a bin as shown later. To enhance the accuracy of PIV, in[28] the authors proposed SSPIV, which estimates the DOA of the most dominant source in a TF bin using subspace decomposition and PIV formulation in (13) and (14). In SSPIV, the raw eigenbeams are replaced with the signal subspace $\mathbf{U_s}^{(1)} = \mathbf{U}_1$, which is the first column of $\mathbf{U}$ in (8) giving

$$\mathbf{I_{sspiv}}(\tau, k) = \frac{1}{2}\Re\left\{ a_{00}(\tau, k)^* \begin{bmatrix} D_{-x}(\tau, k, \mathbf{U}_1) \\ D_{-y}(\tau, k, \mathbf{U}_1) \\ D_{-z}(\tau, k, \mathbf{U}_1) \end{bmatrix} \right\}. \tag{16}$$

Note that in the original SSPIV paper, Singular Value Decomposition (SVD) is used instead of EVD, which are equivalent as $\mathbf{R_a}$ is symmetric.

III. PROPOSED METHODS

In this section two methods for analytical narrowband DOA estimation based on MS assumption are proposed. The first method, MSPIV, is based on de-mixing a mixture

of N sources into multiple SS scenarios on each of which PIV is performed. The second method, DIV, is based on a two-source assumption and analytically derives the DOAs from the mixture signals. A subspace version of the second method is also proposed in order to improve its robustness to noise and scenarios with more than two sources.

A. MSPIVs

The idea for the first proposed method is to extend the concept of SSPIV to multi-source in a TF bin rather than the dominant source only. Assuming N active uncorrelated sources in a bin, the MSPIV for each source $n \in \{1...N\}$ is

$$\mathbf{I}^{(n)}_{\mathbf{mspiv}}(\tau,k) = \frac{1}{2}\Re\left\{ a_{00}(\tau,k)^* \begin{bmatrix} D_{-x}(\tau,k,\mathbf{U}_n) \\ D_{-y}(\tau,k,\mathbf{U}_n) \\ D_{-z}(\tau,k,\mathbf{U}_n) \end{bmatrix} \right\}, \tag{17}$$

where $\mathbf{U}_n$ is the n^{th} column of the matrix $\mathbf{U}$. Note that $\mathbf{I}^{(1)}_{\mathbf{mspiv}}$ is equivalent to $\mathbf{I}_{\mathbf{sspiv}}$.

B. DIVs

So far all variations of PIV either take the input assuming as a single source scenario (PIV) or decompose the input into multiple single source scenarios (SSPIV and MSPIV). In DIV the input is assumed to be the mixture of two sources and then both DOAs are analytically estimated.

For the formulation in this section, (τ,k) are omitted for notational simplicity. In a TF bin, consider two simultaneously active sources with plane-wave signals S_1 and S_2, assuming

$S_1 \neq -S_2$ (no opposite phase with equal amplitude), and their DOAs (θ_1, φ_1) and (θ_2, φ_2) respectively. Writing (5) in a noise-free situation for up to $L = 1$ gives

$$\bar{a}_{0(0)} = S_1 + S_2, \tag{18}$$

$$\bar{a}_{1(-1)} = S_1 \sin(\theta_1) e^{i\varphi_1} + S_2 \sin(\theta_2) e^{i\varphi_2}, \tag{19}$$

$$\bar{a}_{1(0)} = S_1 \cos(\theta_1) + S_2 \cos(\theta_2), \tag{20}$$

$$\bar{a}_{1(+1)} = S_1 \sin(\theta_1) e^{-i\varphi_1} + S_2 \sin(\theta_2) e^{-i\varphi_2}, \tag{21}$$

where

$$\bar{a}_{lm} = a_{lm} / \sqrt{\frac{(2l+1)}{4\pi} \frac{(l-m)!}{(l+m)!}} \tag{22}$$

are the eigenbeams compensated by the square root component in (3) for notational simplicity.

Using (18) and (20)

$$S_2 = \bar{a}_{0(0)} - S_1, \tag{23}$$

$$S_1 = \frac{\bar{a}_{1(0)} - \bar{a}_{0(0)} \cos(\theta_2)}{\cos(\theta_1) - \cos(\theta_2)}. \tag{24}$$

Summing and subtracting (19) with (21) respectively gives

$$B^+ = S_1 \sin(\theta_1) \cos(\varphi_1) + S_2 \sin(\theta_2) \cos(\varphi_2), \tag{25}$$

$$B^- = S_1 \sin(\theta_1) \sin(\varphi_1) + S_2 \sin(\theta_2) \sin(\varphi_2), \tag{26}$$

where $B^+ = (\bar{a}_{1(-1)} + \bar{a}_{1(+1)})/2$ and $B^- = (\bar{a}_{1(-1)} - \bar{a}_{1(+1)})/2i$.

Using (23) and (24), substituting S_1 and S_2 both as a function of θ_1 and θ_2 into (25) and (26) results in two complex equations with four real unknowns (θ and φ for sources 1 and

2). Splitting the real and imaginary parts of the two complex equations (25) and (26) gives four real equations which can be simplified into

$$\begin{bmatrix} E_\theta & F_\theta \\ H_\theta & J_\theta \end{bmatrix} \begin{bmatrix} \cos(\varphi_1) \\ \cos(\varphi_2) \end{bmatrix} = \begin{bmatrix} D_\theta \\ G_\theta \end{bmatrix}, \tag{27}$$

$$\begin{bmatrix} E_\theta & F_\theta \\ H_\theta & J_\theta \end{bmatrix} \begin{bmatrix} \sin(\varphi_1) \\ \sin(\varphi_2) \end{bmatrix} = \begin{bmatrix} K_\theta \\ L_\theta \end{bmatrix}, \tag{28}$$

where

$$\begin{cases} E_\theta = & \sin(\theta_1)\left(\Re\left\{\bar{a}_{1(0)}\right\} - \Re\left\{\bar{a}_{0(0)}\right\}\cos(\theta_2)\right), \\ F_\theta = & \sin(\theta_2)\left(\Re\left\{\bar{a}_{0(0)}\right\}\cos(\theta_1) - \Re\left\{\bar{a}_{1(0)}\right\}\right), \\ H_\theta = & \sin(\theta_1)\left(\Im\left\{\bar{a}_{1(0)}\right\} - \Im\left\{\bar{a}_{0(0)}\right\}\cos(\theta_2)\right), \\ J_\theta = & \sin(\theta_2)\left(\Im\left\{\bar{a}_{0(0)}\right\}\cos(\theta_1) - \Im\left\{\bar{a}_{1(0)}\right\}\right), \\ D_\theta = & \Re\left\{B^+\right\}\left(\cos(\theta_1) - \cos(\theta_2)\right), \\ G_\theta = & \Im\left\{B^+\right\}\left(\cos(\theta_1) - \cos(\theta_2)\right), \\ K_\theta = & \Re\left\{B^-\right\}\left(\cos(\theta_1) - \cos(\theta_2)\right), \\ L_\theta = & \Im\left\{B^-\right\}\left(\cos(\theta_1) - \cos(\theta_2)\right). \end{cases} \tag{29}$$

Using linear matrix algebra

$$\begin{cases} \cos(\varphi_1) = & \frac{JD-FG}{EJ-FH} \\ \sin(\varphi_1) = & \frac{JK-FL}{EJ-FH} \end{cases}, \tag{30}$$

$$\begin{cases} \cos(\varphi_2) = & \frac{EG-HD}{EJ-FH} \\ \sin(\varphi_2) = & \frac{EL-HK}{EJ-FH} \end{cases}. \tag{31}$$

Note that index $(.)_\theta$ is omitted from now on for notational simplicity. Using the Pythagorean identity $(\sin^2(\varphi) + \cos^2(\varphi) = 1)$ for (30) and (31) gives a system of two equations

$$\begin{cases} (JD - FG)^2 + (JK - FL)^2 & = (EJ - FH)^2, \\ (EG - HD)^2 + (EL - HK)^2 & = (EJ - FH)^2, \end{cases} \tag{32}$$

where unknowns are θ_1 and θ_2.

Substituting E, F, H, J, D, G, K, L from (29) into (32) and then solving them for θ_1 results in a quadratic equation

$$C_2 \cos^2(\theta_1) - C_1 \cos(\theta_1) + C_0 = 0, \tag{33}$$

where

$$\begin{cases} C_2 = & \vartheta_R A_I^2 + \vartheta_I A_R^2 - 2\chi A_I A_R + (a_R A_I - a_I A_R)^2, \\ C_1 = & 2\left(\vartheta_R A_I a_I + \vartheta_I A_R a_R - \chi\left(a_R A_I + a_I A_R\right)\right), \\ C_0 = & \vartheta_R a_I^2 + \vartheta_I a_R^2 - 2\chi a_I a_R - (a_R A_I - a_I A_R)^2, \end{cases} \tag{34}$$

and

$$\begin{cases} \vartheta_R = & \left(\Re\left\{B^+\right\}\right)^2 + \left(\Re\left\{B^-\right\}\right)^2, \\ \vartheta_I = & \left(\Im\left\{B^+\right\}\right)^2 + \left(\Im\left\{B^-\right\}\right)^2, \\ \chi = & \Re\left\{B^+\right\}\Im\left\{B^+\right\} + \Re\left\{B^-\right\}\Im\left\{B^-\right\}, \end{cases} \tag{35}$$

where $a_R = \Re\left(\bar{a}_{1(0)}\right)$, $a_I = \Im\left(\bar{a}_{1(0)}\right)$, $A_R = \Re\left(\bar{a}_{0(0)}\right)$ and $A_I = \Im\left(\bar{a}_{0(0)}\right)$.

The two solutions for (33) give the inclination of sources

$$\begin{cases} \theta_1 = & \arccos(\frac{C_1+\sqrt{\triangle}}{2C_2}), \\ \theta_2 = & \arccos(\frac{C_1-\sqrt{\triangle}}{2C_2}), \end{cases} \tag{36}$$

where

$$\triangle = C_1^2 - 4C_2C_0. \tag{37}$$

Note that since inclination is within $[0, \pi]$ there is no sign ambiguity for θ_1 and θ_2.

Having calculated inclinations θ_1 and θ_2, then E, F, H, J, D, G, K, L are calculated using (29). The tangent of azimuths are calculated using (30) and (31)

$$\begin{cases} \tan\left(\varphi_1\right) = & \frac{JK-FL}{JD-FG}, \\ \tan\left(\varphi_2\right) = & \frac{EL-HK}{EG-HD}. \end{cases} \tag{38}$$

Since the azimuths are within $[0, 2\pi]$, the sign of cosine of azimuths from (30) and (31) are used to solve the sign ambiguity of φ_1 and φ_2 giving

$$\begin{cases} \varphi_1 = & \arctan(\frac{JK-FL}{JD-FG}) + (1 - \mathrm{sgn}(\frac{JD-FG}{EJ-FH}))\frac{\pi}{2}, \\ \varphi_2 = & \arctan(\frac{EL-HK}{EG-HD}) + (1 - \mathrm{sgn}(\frac{EG-HD}{EJ-FH}))\frac{\pi}{2}, \end{cases} \tag{39}$$

where sgn(.) denotes the sign operator.

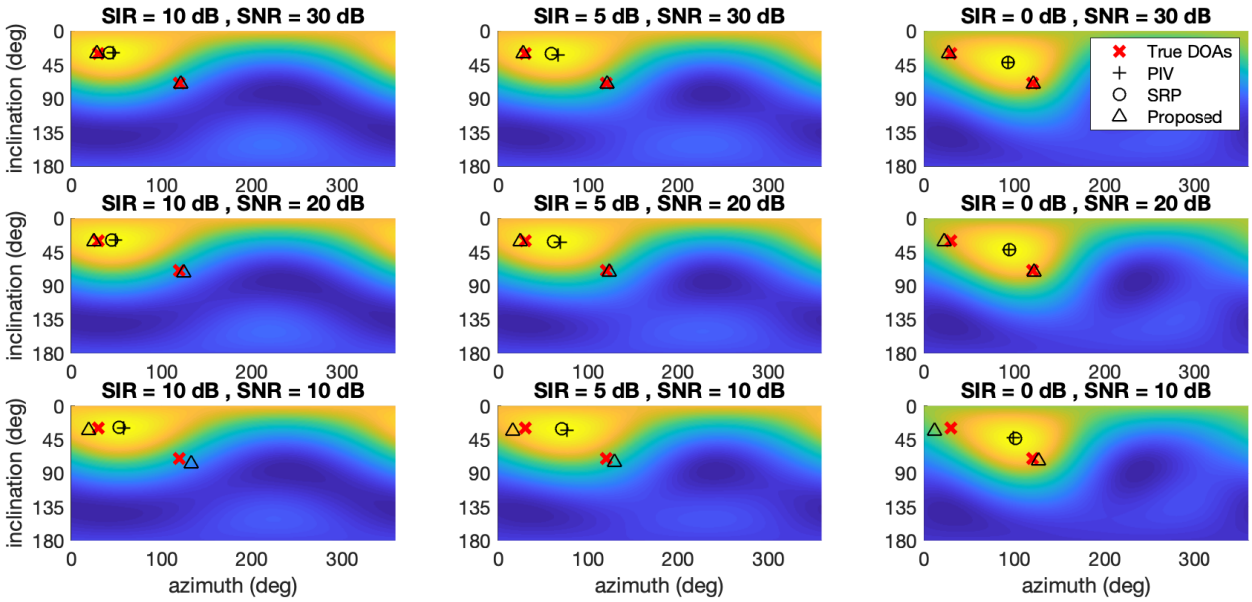


FIG. 1. The SRP spectrums for two sources with varying SIR=$\{10, 5, 0\}$ dB (columns) and SNR=$\{30, 20, 10\}$ dB (rows). The true DOAs are marked as red cross ($\times$) and the estimates by PIV, SRP and proposed method (DIV) are marked as black plus ($+$), circle ($\circ$) and triangle ($\triangle$) respectively.

Note that such analytical solution may have no or only one valid answer if $\triangle \leq 0$ in (37) or $|\cos(\theta_1)| > 1$ in (33).

C. Special case of equal inclinations

In case of equal inclinations for both sources, $\theta_1 = \theta_2$, (24) is undefined and therefore a different solution is required.

Let $\theta_1 = \theta_2 = \theta$, then (20) can be rewritten as

$$\bar{a}_{1(0)} = \cos(\theta)\,(S_1 + S_2). \tag{40}$$

209 Using (18) and (40)

$$\theta = \theta_1 = \theta_2 = \arccos\left(\Re(\frac{\bar{a}_{1(0)}}{\bar{a}_{0(0)}})\right). \tag{41}$$

Having obtained θ, (25) and (26) can be rewritten as

$$\bar{B}^{+} = S_1 \cos(\varphi_1) + S_2 \cos(\varphi_2), \tag{42}$$

$$\bar{B}^{-} = S_1 \sin(\varphi_1) + S_2 \sin(\varphi_2), \tag{43}$$

210 where $\bar{B}^{\pm} = B^{\pm}/\sin(\theta)$.

From (42), (43) and (18)

$$S_1 = \frac{\bar{a}_{0(0)} \cos(\varphi_2) - \bar{B}^{+}}{\cos(\varphi_2) - \cos(\varphi_1)}, \tag{44}$$

$$S_2 = \frac{\bar{B}^{+} - \bar{a}_{0(0)} \cos(\varphi_1)}{\cos(\varphi_2) - \cos(\varphi_1)}. \tag{45}$$

211 Substituting S_1 and S_2, from (44) and (45), into (43) and then simplifying it results in

$$\bar{m} X_x + \bar{n} Y_x + \bar{q} = Z_x, \tag{46}$$

212 where the known coefficients are

$$\begin{cases} \bar{m} = & \frac{-(\bar{a}_{1(-1)}+\bar{a}_{1(+1)})\sqrt{\left(\bar{a}_{0(0)}\right)^2-\left(\bar{a}_{1(0)}\right)^2}}{\bar{a}_{1(-1)}\bar{a}_{1(+1)}-\left(\bar{a}_{0(0)}\right)^2+\left(\bar{a}_{1(0)}\right)^2}, \\ \bar{n} = & \frac{\bar{a}_{1(-1)}\bar{a}_{1(+1)}+\left(\bar{a}_{0(0)}\right)^2-\left(\bar{a}_{1(0)}\right)^2}{\bar{a}_{1(-1)}\bar{a}_{1(+1)}-\left(\bar{a}_{0(0)}\right)^2+\left(\bar{a}_{1(0)}\right)^2}, \\ \bar{q} = & \frac{(\bar{a}_{1(-1)}-\bar{a}_{1(+1)})^2}{2(\bar{a}_{1(-1)}\bar{a}_{1(+1)}-\left(\bar{a}_{0(0)}\right)^2+\left(\bar{a}_{1(0)}\right)^2)}, \end{cases} \tag{47}$$

and the unknowns are

$$X_x = \cos(\varphi_1) + \cos(\varphi_2), \tag{48}$$

$$Y_x = 1 + \cos(\varphi_1)\cos(\varphi_2), \tag{49}$$

$$Z_x = \sin(\varphi_1)\sin(\varphi_2). \tag{50}$$

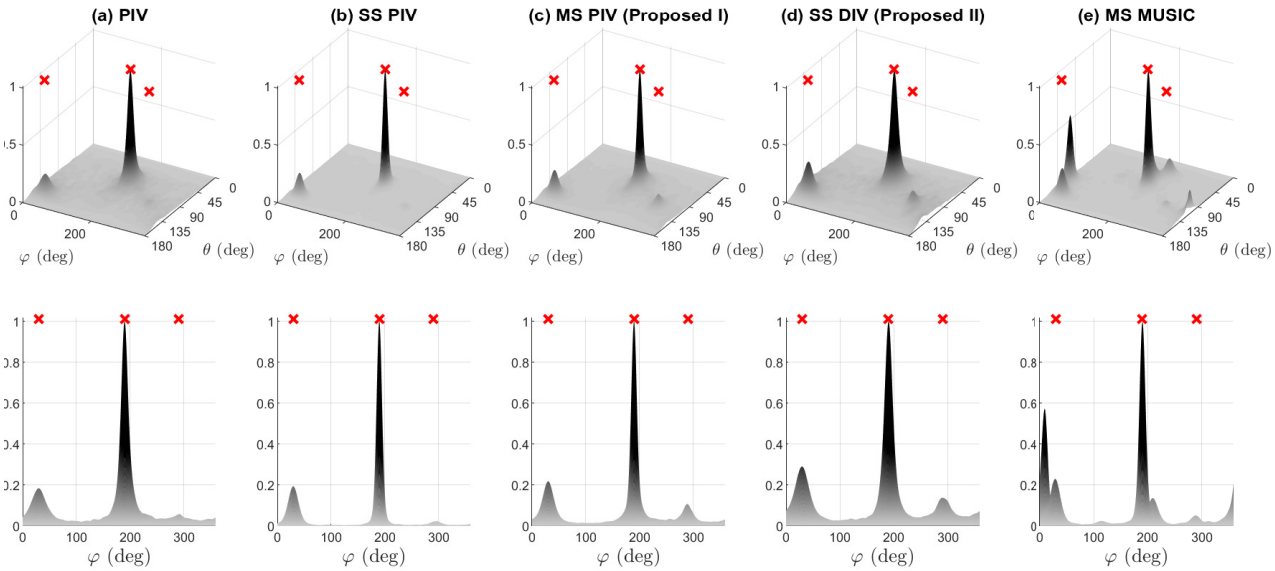


FIG. 2. Corner view (top row) and the side view (bottom row) of the normalized smoothed 2D histograms of DOAs estimated by (a) PIV, (b) SSPIV, (c) MSPIV, (d) SSDIV and (e) MSMUSIC. The true DOAs are marked by red cross (×)

Splitting the real and imaginary parts of (46) result into two real equations which, using linear matrix algebra, give

$$\begin{bmatrix} X_x \\ Y_x \end{bmatrix} = \begin{bmatrix} \bar{m}_R & \bar{n}_R \\ \bar{m}_I & \bar{n}_I \end{bmatrix}^{-1} \begin{bmatrix} Z_x - \bar{q}_R \\ -\bar{q}_I \end{bmatrix}, \tag{51}$$

where $(.)_R = \Re(.)$ and $(.)_I = \Im(.)$.

On the other hand Pythagorean identity ($\sin^2(\varphi) + \cos^2(\varphi) = 1$) for (48), (49) and (50) gives

$$Y_x^2 - X_x^2 = Z_x^2. \tag{52}$$

Substituting X_x and Y_x as a function of Z_x from (51) into (52) results in a quadratic equation

$$B_2(Z_x - \bar{q}_R)^2 + B_1(Z_x - \bar{q}_R) + B_0 = 0, \tag{53}$$

where

$$\begin{cases} B_2 = & (\bar{m}_I^2 - \bar{n}_I^2) - \bar{d}^2, \\ B_1 = & \bar{q}_I(\bar{m}_R\bar{m}_I - \bar{n}_R\bar{n}_I) - 2\bar{q}_R\bar{d}^2, \\ B_0 = & \bar{q}_I^2(\bar{m}_R^2 - \bar{n}_R^2) - \bar{q}_R^2\bar{d}^2, \end{cases} \tag{54}$$

and $\bar{d} = \bar{m}_R\bar{n}_I - \bar{m}_I\bar{n}_R$. Then Z_x is obtained as

$$Z_x = \sin(\varphi_1)\sin(\varphi_2) = \frac{-B_1 \pm \sqrt{\triangle_Z}}{2B_2} + \bar{q}_R, \tag{55}$$

where $\triangle_Z = (B_1)^2 - 4B_2B_0$. Having obtained Z_x, then X_x and Y_x are obtained using (51). Using (48) and (49), the cosine of the azimuths are

$$\begin{cases} \cos(\varphi_1) = & (X_x + \sqrt{\triangle_\varphi})/2, \\ \cos(\varphi_2) = & (X_x - \sqrt{\triangle_\varphi})/2, \end{cases} \tag{56}$$

where $\triangle_\varphi = (X_x)^2 - 4(Y_x - 1)$. In order to overcome the problem of sign ambiguity of φ (where $\varphi \in (-\pi, \pi]$ or the ambiguity between φ and $2\pi - \varphi$ where $\varphi \in [0, 2\pi)$) having its cosine, the following procedure can be done. Substituting the cosine of the azimuths in (56) into (44) and (45) gives S_1 and S_2 which along with (43) and (50) can be used to obtain the correct sign for φ_1 and φ_2.

D. Narrowband analysis

This section provides an analysis of the derived DIV formulation under its initial assumption, which is two active sources in the far-field giving two impinging plane waves in the same TF bin in an anechoic environment. This analysis is based on numerical simulation of the input signals in the SHD based on (5). The main evaluation using SMAs and with more sources in a more challenging scenario is presented in Section IV.

In (5), consider two simultaneously active uncorrelated sources ($N_s = 2$) in a TF bin with DOAs $(\varphi_1, \theta_1) = (30, 30)^\circ$ and $(\varphi_2, \theta_2) = (120, 70)^\circ$ and no reverberation. Let the Signal-to-Interference Ratio (SIR) denote the ratio of their amplitude ($|S_1|/|S_2|$) in decibels. Note that in case of more than two sources, SIR is the ratio of the amplitude of a single source and sum of all other sources. The phases of S_1 and S_2 are selected randomly from a uniform distribution. Using (5), numerically simulated eigenbeams a_{lm} are obtained for varying SIRs and SNRs. Assuming the DOAs are unknown and only the eigenbeams are available, the aim is to analyse the accuracy of DOAs estimated by DIV, PIV and PWD-SRP.

Figure 1 illustrates the SRP spectrums and the DOAs estimated by PIV using (13), PWD-SRP using (12) and the proposed method (DIV) for such scenario with varying SIR=$\{10, 5, 0\}$ dB and additive white Gaussian sensor noise giving SNR=$\{30, 20, 10\}$ dB. Table I presents the associated mean errors for PIV and DIV. The DOA estimation error ε between the estimated DOA unit vector $\mathbf{u}_e$ and the closest true DOA unit vector $\mathbf{u}_o$ was computed in degrees as

$$\varepsilon = \cos^{-1}\left(\mathbf{u}_o^T \mathbf{u}_e\right). \tag{57}$$

PIV , DIV	SIR = 10 dB	SIR = 5 dB	SIR = 0 dB
SNR = 30 dB	8.2° , 1.4°	18.5° , 1.0°	35.9° , 1.0°
SNR = 20 dB	9.4° , 4.0°	19.7° , 3.2°	35.1° , 3.1°
SNR = 10 dB	13.6° , 10.3°	23.7° , 8.8°	32.1° , 8.7°

TABLE I. Mean error of narrowband PIV and DIV, in black and red respectively, for varying SNR and SR as shown in Figure 1.

It can be seen that the decrease of SIR (through the columns per each row), due to the increase in the violation of WDO assumption, results in more error for PIV and SRP estimates whereas the proposed method shows strong robustness to SIR due to two-source assumption, which in this case is correct number of sources. Across all cases, the proposed method also maintains the relatively high accuracy of 1° to 8° mean error (with respect to the closest true DOA) compared to PIV whose mean error dramatically varies between 8° to 35°.

E. Sub-Space DIV (SSDIV): Noise-robustness improvement

In order to improve DIV's robustness to noise the mixture of signal subspaces $\mathbf{U}_1$ and $\mathbf{U}_2$, obtained using EVD in (8), are formed as

$$\mathbf{U}_{1,2} = \mathbf{U}_1 + \mathbf{U}_2, \tag{58}$$

which represents the de-noised eigenbeams mixture of the two most dominant sources. Instead of $\{a_{lm}\}_{l=0}^{1}$, the first four elements of the vector $\mathbf{U}_{1,2}$ are treated as the inputs to DIV formulation to give SSDIV.

F. Wideband analysis

In this section an analysis of the methods is provided for a simulated case of speech talkers and a 32-element rigid SMA with radius of 4.2 cm (corresponding to the em32 Eigenmike®) where the impulse responses were simulated using Spherical Microphone arrays Impulse Response Generator (SMIRgen)[34]. Any setup parameter that is not stated in this section is similar to the ones used in Section IV.

Consider three simultaneously active male talkers for a duration of 2 s with DOAs $(\varphi_1, \theta_1) = (190, 70)^\circ$, $(\varphi_2, \theta_2) = (290, 110)^\circ$ and $(\varphi_3, \theta_3) = (30, 150)^\circ$ all 1 m away from the array in an anechoic environment. Random white Gaussian noise with SNR= 10 dB is added to the sensors. The sources 1 and 3 have an equal average strength level while source 2 has SIR=-26 dB (10% of the strength of sources 1 and 3).

Figure 3 shows $|a_{00}|$ for each source in isolation before the mixing (a,b and c) and for the mixture of sources (d) in the STFT domain. Figure 3(e) displays the ID of the dominant source in the STFT domain where dominance is defined as the maximum contribution of $|a_{00}|$ in the mixture. It can be seen that source 2 is rarely dominant compared to the other sources due to its low level strength and overlap with others. Hence, compared to other sources, very few bins with accurate DOA estimates belonging to source 2 are expected from PIV and SSPIV.

Figure 2 illustrates the smoothed 2D histograms of DOAs estimated by PIV, SSPIV, MSPIV, SSDIV and MSMUSIC. Both MSPIV using (17) and MSMUSIC using (11) were applied using $N = 3$, where each of three DOA estimates per TF bin is individually considered in the overall spatial histogram of DOAs for final wideband DOA estimation. Note that PIV and SSPIV estimate a single DOA per TF bin whereas MSPIV estimates three DOAs per bin. SSDIV and MSMUSIC respectively estimate up to two and three DOAs per bin. As expected, for PIV and SSPIV there are very few bins where source 2 is dominant as shown in Fig. 3(e) and so the associated peak is very weak. On the other hand, both proposed methods estimate enough DOAs belonging to source 2, which forms a visibly distinct peak in the histogram. This is due to successful estimation of DOAs belonging to source 2 even in TF bins where it is not dominant. MSMUSIC shows slightly better localization of source 2 compared to PIV and SSPIV as it is based on multi-source assumption but is outperformed by our two proposed methods. It can also be seen that MSMUSIC results in extra peaks at the opposite directions of true DOAs. Considering MUSIC as equivalent to beamforming since it provides spatial selectivity, this phenomenon is the result of estimating a peak at the main lobe and a second peak at the rear lobe of the beam pattern in the MUSIC spectrum in the TF bins where only a single source is present or significantly dominant.

IV. EVALUATIONS

The performance of the proposed (MSPIV and SSDIV) and comparative (PIV, SSPIV and MSMUSIC) methods is evaluated using recorded anechoic speech signals convolved with direct-path impulse response with varying additive white Gaussian sensor noise for

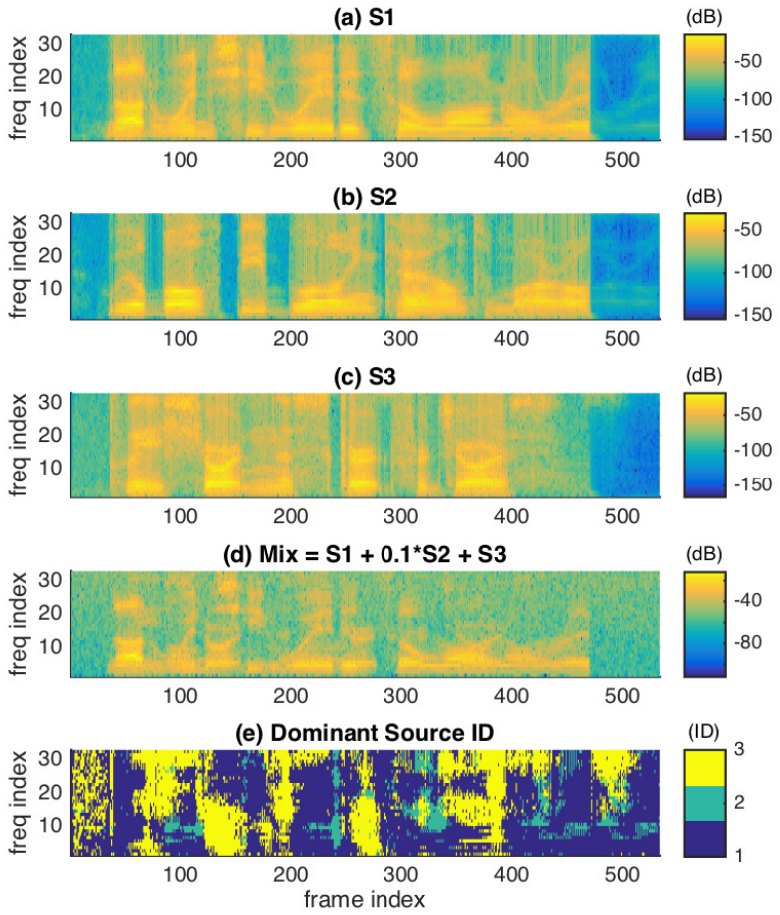


FIG. 3. Magnitude of the omnidirectional (zeroth order) eigenbeam for (a) S_1, (b) S_2, (c) S_3 and (d) the mixture of three sources. Part (e) shows the ID of the dominant source where dominance is defined as the maximum magnitude contribution in the mixture.

a 32-element rigid SMA with radius of 4.2 cm (corresponding to the em32 Eigenmike®) where the impulse responses were simulated using Spherical Microphone arrays Impulse Response Generator (SMIRgen)[34]. The noise at each microphone is uncorrelated. Four anechoic speech sources were employed in total (3 males and 1 female) each with duration of 2 seconds selected from the APLAWD database[35]. The number of sources is assumed to

be known for all compared methods. Some segments (with overall duration of 1 second) of source 2 were removed to make inconsistency of activity as illustrated in Figure 4(b). The sources were mixed with mixing coefficient of 1 for sources 1, 3 and 4 and 0.2 (SIR=-24 dB) for source 2 so that source 2 is strongly masked by others. The sources were all 1.5 m away from the array. In each trial, four random source positions were drawn from a uniform distribution around the sphere. Only combinations of positions where the angular separation of each two sources was $\geq 70^\circ$ and no two sources had equal azimuths or inclinations were retained. Exact equality is considered for exclusion of equal azimuths or inclinations. 100 trials were used for each SNR=$\{10, 20, 30, \text{Inf}\}$ dB.

A sampling frequency of 8 kHz was used with 50% overlapping time-frames of 8 ms duration. The sample covariance matrix in (7) had $J_\tau = 6$ and $J_k = 4$ as the size (number of bins) of the averaging windows over time and frequency respectively. This gives 32 ms and 500 Hz window-size in the TF domain based on our time and frequency resolution. MSMUSIC and MSPIV assumes $N = N_s = 4$ active sources per bin. Hence MSMUSIC and MSPIV estimate four DOAs whereas PIV and SSPIV estimate one DOA and SSDIV estimates up to two DOAs per bin. Only the eigenbeams for up to SH order $L = 1$ (first four eigenbeams) are used for each method in order to exclude the impact of high-order ($l > 1$) harmonics in the performance.

Figure 4 illustrates the STFT representation of each source in isolation, (a) to (d), the mixture (e) as well as the ID of the dominant source in each TF bin (f). It can be seen from Fig 4(f) that source 2 is rarely dominant which makes this situation a challenging scenario for SS-based DOA estimators.

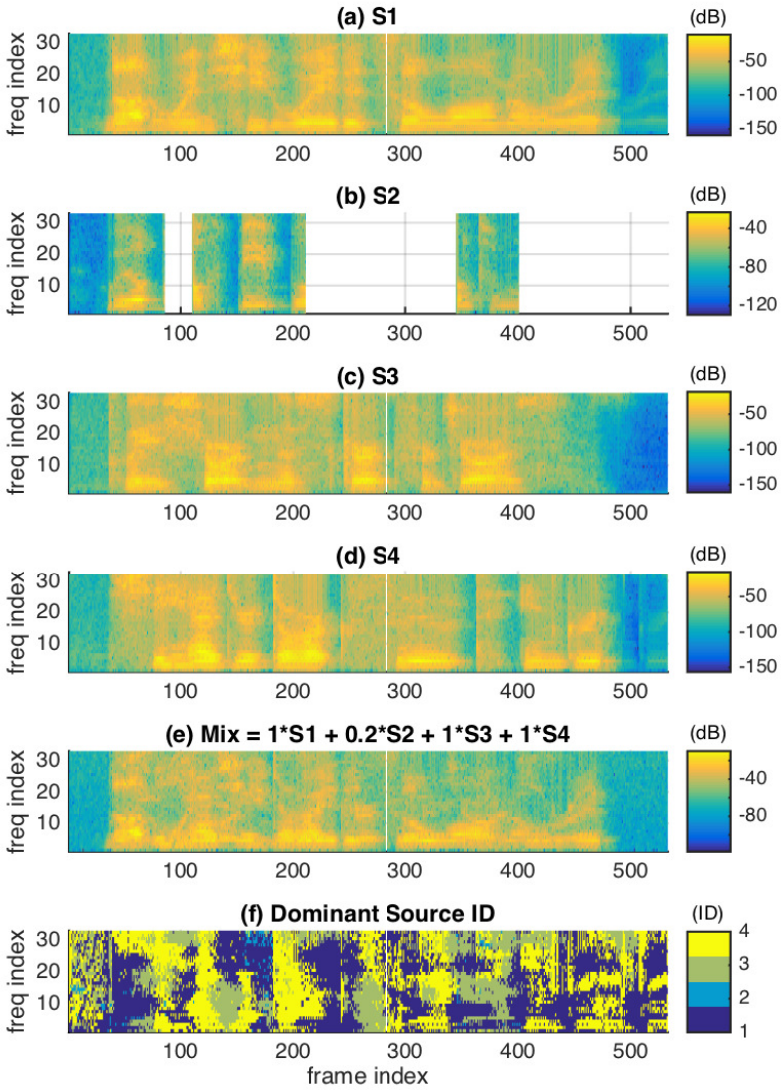


FIG. 4. Magnitude of the omnidirectional (zeroth order) eigenbeam for (a) S_1, (b) S_2, (c) S_3, (d) S_4 and (e) the mixture of four sources. Part (f) shows the ID of the dominant source where dominance is defined as the maximum magnitude contribution in the mixture.

In each trial, the largest $N_s = 4$ peaks in the smoothed histogram were selected as the estimated DOAs. 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.

Figure 5 presents the mean error of each method for SNR=$\{10, 20, 30, \text{Inf}\}$ dB. MSMUSIC fails with high error due to the presence of erroneous peaks of rear lobe as explained before. This is caused as a fixed number of peaks are selected per bin for each method. Other techniques do not suffer from rear lobe problem since they are not steering-based. Note that DIV (and SSDIV) estimate two identical DOAs in case of a SS scenario, which makes it a robust algorithm for a SS scenario although it is a MS DOA estimator. The proposed MSPIV and SSDIV both significantly outperform the baseline techniques PIV and SSPIV with almost double accuracy due to the utilisation of narrowband MS assumption. MSPIV shows 1° to 4° better accuracy than SSDIV since it assumes more accurate number of sources per bin. Note that although some MSPIV estimates can be erroneous in TF bins with < 4 active sources, the irregular directions of the errors avoid the formation of a prominent peak in the histogram. The same phenomenon happens for SSDIV in TF bins with < 2 active sources. As an example, using dataset with SNR=20 dB, analysis of SSDIV estimates in the TF bins where only source 1 had SIR of > 10 dB (assumed condition of single source activity in a TF bin) showed standard deviation of 5.2° for the estimates closest to source 1 whereas the other estimates in the same TF bins had standard deviation of over 30°.

Table II presents the errors (averaged among 100 trials) per each source for dataset with sensor noise of SNR=10 dB. As expected, source 2, which is strongly masked due to lower strength and shorter duration, has the highest estimation error in all methods. However, the

Error	PIV	SSPIV	MSMUSIC	MSPIV	SSDIV
Source 1	6.2°	9.7°	5.4°	7.6°	6.8°
Source 2	36.8°	26.9°	72.7°	10.1°	18.4°
Source 3	11.5°	10.3°	13.7°	5.5°	10.5°
Source 4	7.9°	7.7°	14.2°	9.2°	9.9°

TABLE II. Mean error (averaged among 100 trials) per each source for SNR=10 dB.

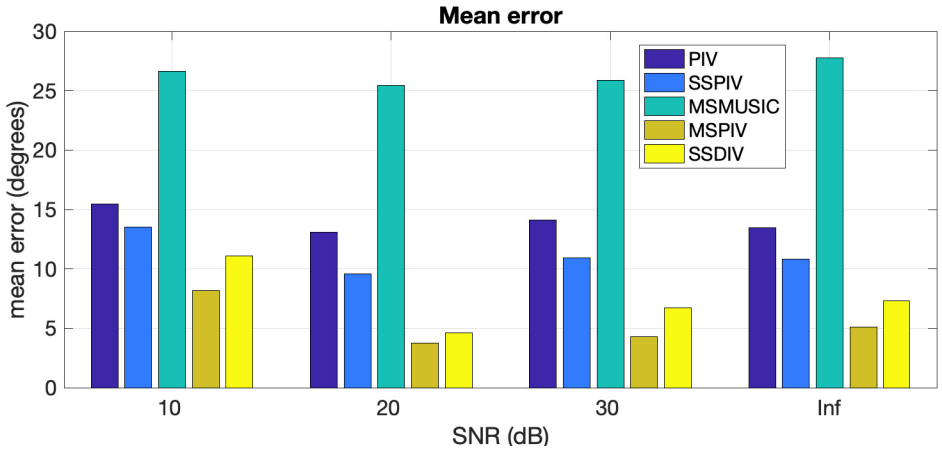


FIG. 5. The mean error of PIV, SSPIV, MSMUSIC, MSPIV and SSDIV for varying SNR.

two proposed methods MSPIV and SSDIV show significant superiority in accuracy compared to other methods due to their multi-source assumption per TF bin.

V. COMPUTATIONAL COMPLEXITY

This section evaluates the computational complexity of each technique. Table III presents the approximate number of real multiplications and MATLAB™ average running time (eval-

	PIV	SSPIV	MSMUSIC	MSPIV	SSDIV
Approx. # real ×	50	50	10^6	$50N$	100
Mean run time (s)	0.203	1.024	329.738	1.682	1.241

TABLE III. Computationality of each method

uated for comparison purposes all on the same computer, MacBook Pro with 2.6 GHz Intel Core i5 processor and 8 GB 1600 MHz DDR3 memory) for each method. Note that the number of real multiplications were calculated per bin for narrowband DOA estimator only excluding the covariance matrix calculation, EVD operation and peak picking.

The PIV and MSMUSIC have the lowest and the highest computational complexity respectively as expected. SSDIV has approximately double number of real multiplications than PIV and SSPIV while MSPIV has N times more multiplications than PIV as it performs PIV N times where N is the assumed number of active sources per bin. MSMUSIC has extremely high computational cost due to steering to all directions per bin (180×360 directions) excluding even the high computational cost of multi-peak picking per bin.

VI. CONCLUSION

Two narrowband analytical DOA estimators based on Multi-Source (MS) assumption have been proposed. Using EVD, MS-PIV decomposes a MS scenario into multiple single source scenarios and Sub-Space Dual-Intensity Vectors (SSDIVs) decomposes the MS scenario into mixture of two most dominant sources. The first approach performs Pseu-

dointensity vectors (PIVs) on each time-frequency bin based on single source assumption whereas SSDIVs directly estimates two DOAs based on a two-source assumpotion per bin. An evaluation was performed for a challenging scenario of one source being strongly masked by three other simultaneously active sources in an anechoic environment with noisy sensors. The results show an improvement of almost double accuracy in average at the cost of double computational complexity compared to the conventional single-source DOA estimators. They also significantly outperform the steering-based and MS-based DOA estimator with much less computational complexity.

REFERENCES

[1]I. Balmages, and B. Rafaely, "Open-sphere designs for spherical microphone arrays,", in *IEEE Trans. Audio, Speech, Lang. Process.*, **15**(2), 727–732 (2007).

[2]B. N. Gover, J. G. Ryan, and M. R. Stinson, "Microphone array measurement system for analysis of directional and spatial variations of sound fields,,", JASA Journal Acoust. Soc. Am. **112**, 1980–1991 (2002).

[3]E. Fisher and B. Rafaely, "The nearfield spherical microphone array,", in *Proc. IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP)* (2008), pp. 5272–5275.

[4]E. Fisher and B. Rafaely, "Near-field spherical microphone array processing with radial filtering,,", in *IEEE Trans. Audio, Speech, Lang. Process.*, **19**(2), 256–265 (2011).

[5]Z. Li and R. Duraiswami, "Flexible and optimal design of spherical microphone arrays for beamforming,,", in *IEEE Trans. Audio, Speech, Lang. Process.*, **15**(2), 702–714 (2007).

[6]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. **22**(10), 1494–1505 (2014).

[7]B. Rafaely, "Analysis and design of spherical microphone arrays,", IEEE Trans. Audio, Speech, Lang. Process. **13**(1), 135–143 (2005).

[8]B. Rafaely, "Phase-mode versus delay-and-sum spherical microphone array processing,", IEEE Signal Process. Lett. **12**(10), 713–716 (2005).

[9]B. Rafaely, B. Weiss, and E. Bachmat, "Spatial aliasing in spherical microphone arrays,,", in *IEEE Trans. Signal Process.*, **55**(3), 1003–1010 (2007).

[10]B. Rafaely, "The spherical-shell microphone array,", IEEE Trans. Audio, Speech, Lang. Process. **16**(4), 740–747 (2008).

[11]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. Springer, **11**, (2010).

[12]B. Rafaely, *Fundamentals of Spherical Array Processing*, Eds. Springer, (2015).

[13]H. C. Schau, and A. Z. Robinson, "Passive source localization employing intersecting spherical surfaces from time-of-arrival differences,", IEEE Trans. Audio, Speech, Lang. Process. **35**(8), 1223–1225 (1987).

[14]S. Yan, H. Sun, U. P. Svensson, X. Ma, and J. M. Hovem, "Optimal modal beamforming for spherical microphone arrays,", IEEE Trans. Audio, Speech, Lang. Process. **19**(2), 361–371 (2011).

[15]D. N. Zotkin, R. Duraiswami, and N. A. Gumerov, "Plane-wave decomposition of acoustical scenes via spherical and cylindrical microphone arrays,", IEEE Trans. Audio, Speech, Lang. Process. **18**(1), 2–16 (2010).

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

[17]D. P. Jarrett, E. A. 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 (2010), pp. 442–446.

[18]B. N. Gover, J. G. Ryan, and M. R. Stinson, "Measurements of directional properties of reverberant sound fields in rooms using a spherical microphone array,,", JASA Journal Acoust. Soc. Am. **116**, 2138–2148 (2004).

[19]J. Meyer,and T. Agnello, "Spherical microphone array for spatial sound recording,", in *Proc. Audio Eng. Soc. (AES) Convention*, New York, NY, USA (2003), pp. 1–9.

[20]T. D. Abhayapala ,and D. B. Ward, "Theory and design of high order sound field microphones using spherical microphone array,", in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing (ICASSP)*, Orlando, FL, USA (2002), pp. 1949–1952.

[21]S. Rickard,and Z. Yilmaz, "On the approximate W-disjoint orthogonality of speech,", in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing (ICASSP)*, (2002), pp. 529–532.

[22]R. O. Schmidt, "Multiple emitter location and signal parameter estimation,", IEEE Trans. Antennas Propag. **34**(3), 276–280 (1986).

[23]S. Mohan, M. E. Lockwood, M. L. Kramer, and D. L. Jones, "Localization of multiple acoustic sources with small arrays using a coherence test,,", JASA Journal Acoust. Soc. Am. **123**, 2136–2147 (2008).

[24]D. Pavlidi, M. Puigt, A. Griffin, and A. Mouchtaris, "Real-time multiple sound source localizaation using a circular microphone array based on single-source confidence measures,", in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing (ICASSP)*, Kyoto, Japan (2012), pp. 2625–2628.

[25]S. Hafezi, A. H. Moore, and P. A. Naylor, "Multiple DOA estimation based on estimation consistency and spherical harmonic multiple signal classification,", in *Proc. European*

Signal Processing Conf. (EUSIPCO), Kos, Greece (2017), pp. 1280–1284.

[26]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 (2017).

[27]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 (2017).

[28]A. H. Moore, C. Evers, and P. A. Naylor, "Direction of arrival estimation in the spherical harmonic domain using subspace pseudo-intensity vectors,", IEEE Trans. Audio, Speech, Lang. Process. (2016).

[29]B. D. van Veen, and K. M. Buckley, "Beamforming: A versatile approach to spatial filtering,", IEEE Acoustics, Speech and Signal Magazine, **5**(2), 4–24 (1988).

[30]D. Khaykin, and B. Rafaely, "Acoustic analysis by spherical microphone array processing of room impulse responses,,", JASA Journal Acoust. Soc. Am. **132**(1), 261–270 (2012).

[31]B. Rafaely, "Plane-wave decomposition of the pressure on a sphere by spherical convolution,,", JASA Journal Acoust. Soc. Am. **116**(4), 2149–2157 (2004).

[32]E. G. Williams, *Fourier Acoustics: Sound Radiation and Nearfield Acoustical Holography*, 1st Ed. London: Academic Press, (1999).

[33]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).

[34]D. P. Jarrett, E. A. P. Habets, M. R. P. Thomas, and P. A. Naylor, "Simulating room impulse responses for spherical microphone arrays,", in *Proc. IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, Prague, Czech Republic (2011), pp. 129–132.

[35]G. Lindsey, A. Breen, and S. Nevard, "SPAR's archivable actual-word databases,", Technical Report, University College London (1987).

[36]N. Xiang, and C. Landschoot, "Bayesian Inference for Acoustic Direction of Arrival Analysis Using Spherical Harmonics,", Entropy **21**(6), 579 (2019).

[37]C. Landschoot, and N. Xiang, "Model-based Bayesian direction of arrival analysis for sound sources using a spherical microphone array,,", JASA Journal Acoust. Soc. Am. **146**(6), 4936–4946 (2019).

[38]M. Park, and B. Rafaely, "Sound-field analysis by plane-wave decomposition using spherical microphone array,,", JASA Journal Acoust. Soc. Am. **118**, 3094–3103 (2005).