Signal Processing 100 (2014) 160–168

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

Fast communication

Underdetermined direction of arrival estimation using acoustic vector sensor

Shengkui Zhao [a,*], Tigran Saluev [a], Douglas L. Jones [b]

[a] Advanced Digital Sciences Center, 1 Fusionopolis Way, #08-10 Connexis North Tower, Singapore 138632, Singapore
[b] Coordinated Science Laboratory, University of Illinois at Urbana-Champaign, 1308 West Main Street Urbana, IL 61801, USA

ARTICLE INFO

Article history:
Received 30 July 2013
Received in revised form
29 November 2013
Accepted 23 January 2014
Available online 31 January 2014

Keywords:
Microphone arrays
Vector sensors
Direction of arrival estimation
Covariance matrix

ABSTRACT

This paper presents a new approach for the estimation of two-dimensional (2D) direction-of-arrival (DOA) of more sources than sensors using an Acoustic Vector Sensor (AVS). The approach is developed based on Khatri–Rao (KR) product by exploiting the subspace characteristics of the time variant covariance matrices of the uncorrelated quasi-stationary source signals. An AVS is used to measure both the acoustic pressure and pressure gradients in a complete sound field and the DOAs are determined in both horizontal and vertical planes. The identifiability of the presented KR-AVS approach is studied in both theoretic analysis and computer simulations. Computer simulations demonstrated that 2D DOAs of six speech sources are successfully estimated. Superior root mean square error (RMSE) is obtained using the new KR-AVS array approach compared to the other geometries of the non-uniform linear array, the 2D L-shape array, and the 2D triangular array.

1. Introduction

Two-dimensional (2D) direction of arrival (DOA) estimation with array sensors is essential for source localization in audio surveillance, auditory scene analysis, hearing aids, etc. In these applications, the sources come from not only the horizontal plane but also the vertical plane. In addition, the number of sources can exceed the number of sensors. Using small aperture arrays provides a great convenience in configuration and portability too. Therefore, the DOA estimation in 2D space using small aperture arrays is highly desirable. The conventional linear array approaches [1–4] are only able to estimate the DOAs in the horizontal plane. In addition, they have the front and back ambiguous problem. Therefore, they are less efficient for the situations with sources located at different heights.

The four-element Acoustic Vector Sensor (AVS) first presented in [5] for DOA estimation is an acoustic sensor that is capable of measuring acoustic pressure gradient as well as pressure as in a standard microphone. This combination makes it possible for the sensor to measure the complete sound field.

The AVS has been studied for overdetermined DOA estimation where the number of sources is less than the number of sensors [5–8]. In this case, the 2D DOA can be estimated by employing the subspace approach such as the MUSIC (MUltiple SIgnal Classification) [10,11]. By applying eigenvector decomposition to the local covariance matrix, the source subspace and noise subspace are identified based on their eigenvalues, and then the DOAs are estimated based on searching the steering vectors orthogonal to the noise subspace. However, when the number of sources is equal to or more than the number of senors, the noise subspace cannot be identified using the MUSIC approach on the local covariance matrix. To localize more speech sources, the time–frequency spareness was exploited to find low-rank covariance matrices in [3]. However, it proves difficult to

* Corresponding author.
E-mail addresses: shengkui.zhao@adsc.com.sg (S. Zhao), tigran.saluev@gmail.com (T. Saluev), jones@ifp.illinois.edu (D.L. Jones).

0165-1684/$ - see front matter © 2014 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.sigpro.2014.01.021

identify low rank covariance matrices when ambient noise is high.

In this work, we are interested to study the underdetermined 2D DOA estimation problem where the number of sources is more than the number of sensors using an AVS. We would like to deal with the sources of audio or speech signals which are so-called quasi-stationary signals. It is known that for such quasi-stationary signals the covariance matrix of the array output is locally static over a short period of time and exhibits difference over a long time period. Using the subspace characteristics of the time-variant covariance matrices of the quasi-stationary source signals, the underdetermined DOA estimation problem can be transformed to an overdetermined DOA estimation problem. Therefore, we develop a subspace approach based on Khatri–Rao product [4]. By applying the vectorization to local covariance matrices and stacking the resulting vectors into a virtual matrix, the degree of freedom is increased by square of its original value. Then we apply a detection criterion similar to the MUSIC on the virtual matrix and estimate the azimuth and elevation angles. Our identifiable analysis and simulations show that the 2D DOAs of six speech sources are successfully estimated using an AVS.

2. Problem formulation

We consider a four-element AVS [5] and K wideband sources impinging on the array from far field with azimuth angles of $\theta_k \in (-180^\circ, 180^\circ]$ and elevation angles of $\phi_k \in [-90^\circ, 90^\circ]$, $k=1,\ldots,K(K \geq 4)$. The output signals of the sensors are modeled in time–frequency domain as

$$\mathbf{x}(t,f) = A\mathbf{s}(t,f) + \mathbf{v}(t,f), \quad t = 0,1,2,\ldots \tag{1}$$

where t is the time index and f is the frequency index. Here, $\mathbf{x}(t,f) = [x_1(t,f),\ldots,x_4(t,f)]^T$ is the received signal vector, $\mathbf{s}(t,f) = [s_1(t,f),\ldots,s_K(t,f)]^T$ is the source vector, $\mathbf{v}(t,f) \in \mathbb{C}^4$ represents the spatial noise vector. For convenience we omit the frequency index f during the derivation in the rest of the paper. The matrix $A = [\mathbf{a}(\vec{r}_1),\ldots,\mathbf{a}(\vec{r}_K)] \in \mathbb{R}^{4\times K}$ is the array response matrix, and $\mathbf{a}(\vec{r}_k)$ is the 4×1 AVS array manifold for source k [5]:

$$\mathbf{a}(\vec{r}_k) \triangleq [1, \cos\theta_k \cos\phi_k, \sin\theta_k \cos\phi_k, \sin\phi_k]^T, \tag{2}$$

where the vector $\vec{r} = [\cos\theta_k \cos\phi_k, \sin\theta_k \cos\phi_k, \sin\phi_k]^T$ is the unit source bearing vector where the azimuth and elevation angles are defined as $\theta_k \in [-180^\circ, 180^\circ]$ and $\phi_k \in [-90^\circ, 90^\circ]$. Note that only the bearing angle θ_k is considered for a linear array studied in [4,12]. Next, we will derive the approach for the 2D DOA estimation with both θ_k and ϕ_k.

When the source signals $s_k(t)$ and noise signals $\mathbf{v}(t)$ are assumed mutually uncorrelated, a local covariance matrix can be defined as

$$\mathbf{R}_m = E\{\mathbf{x}(t)\mathbf{x}^H(t)\} = \mathbf{A}\mathbf{D}_m\mathbf{A}^T + \mathbf{C}, \tag{3}$$

for $\forall t \in [(m-1)L, mL-1]$, where $m=1,2,\ldots$ denotes the frame index, L is the frame size, and $\mathbf{D}_m = \mathrm{Diag}(d_{m1}, d_{m2},\ldots, d_{mK}) \in \mathbb{R}^{K\times K}$ is the source covariance matrix at frame m. These local covariance matrices may be estimated by local averaging. To estimate the DOAs $\theta_1,\ldots,\theta_K$, the conventional MUSIC criterion is based on a single instance of the local covariance matrix $\mathbf{R}_m \in \mathbb{C}^4$. Since the degree of freedom for $\mathbf{R}_m$ is equal to 4, it is insufficient to identify $K \geq 4$ sources.

3. The proposed KR-AVS approach

In this section, we present the approach using the Khatri–Rao product and AVS (KR-AVS) for 2D DOA estimation of $K \geq 4$ sources. The approach will transform the above underdetermined DOA estimation problem into an overdetermined problem.

3.1. The 2D KR-AVS criterion

Let us apply the vectorization computation to the covariance matrix $\mathbf{R}_m$ to obtain

$$\begin{aligned}\mathbf{y}_m &\triangleq \mathrm{vec}(\mathbf{R}_m) = \mathrm{vec}(\mathbf{A}\mathbf{D}_m\mathbf{A}^T) + \mathrm{vec}(\mathbf{C}) \\ &= (\mathbf{A} \odot \mathbf{A})\mathbf{d}_m + \mathrm{vec}(\mathbf{C})\end{aligned} \tag{4}$$

where $\mathrm{vec}(\cdot)$ stands for vectorization computation; i.e., $\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n]$ then $\mathrm{vec}(\mathbf{V}) = [\mathbf{v}_1^T, \mathbf{v}_2^T, \ldots, \mathbf{v}_n^T]^T$, and the symbol $\odot$ stands for the Khatri–Rao (KR) product: $\mathbf{A} \odot \mathbf{A} = [\mathbf{a}(\vec{r}_1) \otimes \mathbf{a}(\vec{r}_1),\ldots,\mathbf{a}(\vec{r}_K) \otimes \mathbf{a}(\vec{r}_K)] \in \mathbb{R}^{16\times K}$, and $\otimes$ denotes the Kronecker product, the vector is $\mathbf{d}_m = [d_{m1},\ldots,d_{mK}]^T$.

Now we can see that the expression of (4) has a similar structure as the signal model in (1). The KR product $(\mathbf{A} \odot \mathbf{A})$ can be considered as the transformed array response matrix, which has virtual array dimension 16 much greater than the physical dimension of 4.

Now consider we have the local covariance matrices $\mathbf{R}_1,\ldots,\mathbf{R}_M$, we stack their vectorization vectors to obtain

$$\mathbf{Y} \triangleq [\mathbf{y}_1,\ldots,\mathbf{y}_M] = (\mathbf{A} \odot \mathbf{A})\mathbf{\Psi}^T + \mathrm{vec}(\mathbf{C})\mathbf{1}_M^T, \tag{5}$$

where the matrix is defined as $\mathbf{\Psi} = [\mathbf{d}_1,\ldots,\mathbf{d}_M]^T \in \mathbb{R}^{M\times K}$ and $\mathbf{1}_M = [1,\ldots,1]^T \in \mathbb{R}^M$. For the quasi-stationary source signals and a large number of frames $M \gg K$, the matrix $[\mathbf{\Psi}\ \mathbf{1}_M] \in \mathbb{R}^{M\times(K+1)}$ can be safely assumed as a full column rank. That is, there exists a collection of K linearly independent columns in $\mathbf{\Psi}$. In a real world, most of the audio and speech signals whose power spectrums are not flat can satisfy this assumption.

Therefore, the noise covariance term $\mathrm{vec}(\mathbf{C})\mathbf{1}_M^T$ in (5) has identical columns and can be eliminated by applying the orthogonal component projection matrix $\mathbf{P}_{\mathbf{1}_M}^\perp = \mathbf{I}_M - (1/M)\mathbf{1}_M\mathbf{1}_M^T$ to (5). Then we have the following decomposition form:

$$\mathbf{Y}\mathbf{P}_{\mathbf{1}_M}^\perp = (\mathbf{A} \odot \mathbf{A})(\mathbf{P}_{\mathbf{1}_M}^\perp \mathbf{\Psi})^T \tag{6}$$

Noted that when a subset $J(J<K)$ of the K sources is stationary, the matrix $[\mathbf{\Psi}\ \mathbf{1}_M] \in \mathbb{R}^{M\times(K+1)}$ cannot be assumed as a full column rank. There exists a collection of $K-J$ linearly independent columns in $\mathbf{\Psi}$. Together with the stationary noise, the J stationary sources are eliminated by the orthogonal component project matrix $\mathbf{P}_{\mathbf{1}_M}^\perp$. Therefore, only the $K-J$ sources are to be considered in the DOA estimation process.

Now consider the subspace of $\mathbf{Y}\mathbf{P}_{\mathbf{1}_M}^\perp$. For ease of exposition of idea, we assume that the decomposition (6) is unique. We will soon provide the conditions under which this assumption is valid. Since we have $\text{rank}(\mathbf{P}_{\mathbf{1}_M}^{\perp}\boldsymbol{\Psi}) = \text{rank}(\boldsymbol{\Psi}) = K$, when the decomposition is unique, we have the subspace

$$\Re(\mathbf{Y}\mathbf{P}_{\mathbf{1}_M}^{\perp}) = \Re(\mathbf{A} \odot \mathbf{A}). \tag{7}$$

Then by applying singular value decomposition (SVD) on matrix $\mathbf{Y}\mathbf{P}_{\mathbf{1}_M}^{\perp}$, we can write

$$\mathbf{Y}\mathbf{P}_{\mathbf{1}_M}^{\perp} = [\mathbf{U}_s \mathbf{U}_n] \begin{bmatrix} \boldsymbol{\Sigma}_s & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{V}_s^H \\ \mathbf{V}_n^H \end{bmatrix}, \tag{8}$$

where $\mathbf{U}_s \in \mathcal{R}^{16\times K}$ and $\mathbf{V}_s \in \mathcal{R}^{M\times K}$ are the left and right singular matrices associated with the nonzero singular values, respectively, $\mathbf{U}_n \in \mathcal{R}^{16\times(16-K)}$ and $\mathbf{V}_n \in \mathcal{R}^{M\times(16-K)}$ are the left and right singular matrices associated with the zero singular values, respectively. $\boldsymbol{\Sigma}_s$ is a diagonal matrix whose diagonals contain the nonzero singular values. Using the SVD result we have that $\Re(\mathbf{A} \odot \mathbf{A}) = \mathbf{U}_s$ and $\Re(\mathbf{A} \odot \mathbf{A})$ is orthogonal to $\mathbf{U}_n$. Therefore, we have that

$$\mathbf{U}_n^T[\mathbf{A} \odot \mathbf{A}] = \mathbf{U}_n^T[\mathbf{a}(\vec{r}_1)\otimes\mathbf{a}(\vec{r}_1), \ldots, \mathbf{a}(\vec{r}_K)\otimes\mathbf{a}(\vec{r}_K)] = \mathbf{0}. \tag{9}$$

The 2D KR-AVS criterion similar to the MUSIC algorithm is derived as

$$J(\theta,\varphi) = \frac{1}{\|U_n^H(\mathbf{a}(\vec{r}) \otimes \mathbf{a}(\vec{r}))\|^2}. \tag{10}$$

The directional spectra $J(\theta,\varphi)$ from all valid frequency bins are summed to yield a final directional spectrum. The bearing vectors $[\theta,\phi]$ that give the spectrum peaks on the final directional spectrum correspond to the source DOAs.

3.2. Conditions for unique identification

Before we can apply the KR-AVS criterion (10) for the DOA estimation, it is crucial to determine conditions under which the DOAs of sources are uniquely identifiable. It is similarly to say under what conditions the decomposition (6) is unique.

We say that decomposition (6) is unique if $\hat{\boldsymbol{\Psi}} = \mathbf{P}_{\mathbf{1}_M}^{\perp}\boldsymbol{\Psi}$ and $\mathcal{A} = \mathbf{A} \odot \mathbf{A}$ are unique up to permutation of $\vec{r}_i$. Let the element of $\mathbf{A}$ at the jth row and kth column be denoted as $\mathbf{A}(j,k)$ $(j=1,\ldots,4,\ k=1,\ldots,K)$.

Then if we consider $\hat{\mathbf{Y}} = \mathbf{Y}\mathbf{P}_{\mathbf{1}_M}^{\perp}$ as 3-dimensional tensor of size $M\times4\times4$, we can write $\hat{\mathbf{Y}}(m,j_1,j_2)$ $(m=1,\ldots,M,$ $j_1=1,\ldots,4,\ j_2=1,\ldots,4)$ into the following canonical decomposition (CANDECOMP/PARAFAC) as

$$\hat{\mathbf{Y}}(m,j_1,j_2) = \sum_{k=1}^{K} \hat{\boldsymbol{\Psi}}(m,k)\cdot\mathbf{A}(j_1,k)\cdot\mathbf{A}(j_2,k). \tag{11}$$

So the decomposition (6) is further decomposed to the form (11) with factors $\hat{\boldsymbol{\Psi}}, \mathbf{A}, \mathbf{A}$. And this means that now we can study uniqueness of decomposition (11) instead of decomposition (6).

Now we are going to apply the following theorem [9].

Theorem 1. *Consider a real 3-dimensional array* $\mathbf{A}(i_1,i_2,i_3)$ *of size* $n_1\times n_2\times n_3$ *and its canonical decomposition*

$$\mathbf{A}(i_1,i_2,i_3) = \sum_{k=1}^{r} \mathbf{A}_1(i_1,k)\cdot\mathbf{A}_2(i_2,k)\cdot\mathbf{A}_3(i_3,k). \tag{12}$$

If

$$2r+2 \le \text{krank}(\mathbf{A}_1)+\text{krank}(\mathbf{A}_2)+\text{krank}(\mathbf{A}_3),$$

where $\text{krank}(\mathbf{A})$ *is the maximum number k such that any k columns of matrix A are linearly independent, then the decomposition (12) is unique up to permutation and rescaling of factors.*

By Kruskal theorem we can get sufficient condition for uniqueness of decomposition (11) up to permutation and rescaling (rescaling is not important for us if we assume $(\mathbf{a}(\vec{r}))_1 \equiv 1$) in (2) in terms of Kruskal ranks of $\hat{\boldsymbol{\Psi}}$ and $\mathbf{A}$: if

$$2K+2 \le \text{krank}(\hat{\boldsymbol{\Psi}})+2\text{krank}(\mathbf{A}), \tag{13}$$

then the decomposition (11) is essentially unique.

By using the result $\text{krank}(\hat{\boldsymbol{\Psi}}) = \text{rank}(\hat{\boldsymbol{\Psi}}) = K$, the condition (13) is then simply rewritten as

$$K \le 2\text{krank}(\mathbf{A}) - 2. \tag{14}$$

Eq. (14) implies that the number of sources K that satisfies the unique decomposition (11) is essentially up-bounded by the $\text{krank}(\mathbf{A})$. To investigate $\text{krank}(\mathbf{A})$, we provide the following lemma:

Lemma 2. *For the AVS array in (2) and a set of* $K\ge4$ *distinguishable sources (i.e.,* $\vec{r}_i \ne \vec{r}_j$ *for all* $i\ne j$*), the following is true:*

$$\text{krank}(\mathbf{A}) = \begin{cases} 3 & \textit{if 4 or more sources lie on the same plane} \\ 4 & \textit{otherwise} \end{cases}$$

Proof. Consider any four sources with unit source bearing vectors $\vec{r}_k \in \mathbb{R}^3$, $k=1,2,3,4$. The array response matrix is given as $\mathbf{A} = [\mathbf{a}(\vec{r}_1), \mathbf{a}(\vec{r}_2), \mathbf{a}(\vec{r}_3), \mathbf{a}(\vec{r}_4)]$, where $\mathbf{a}(\vec{r}_k) = [1\ \vec{r}_k^T]^T$. To study the $\text{krank}(\mathbf{A})$, let us investigate the linearly independent rows of the matrix $\mathbf{A}$. Let the row vector $\vec{v}_i$ denote the ith row of $\mathbf{A}, i=1,2,3,4$, then we have $\vec{v}_1 = [1\ 1\ 1\ 1]$, $\vec{v}_2 = [\cos\theta_1 \cos\phi_1, \ldots, \cos\theta_4 \cos\phi_4]$, $\vec{v}_3 = [\sin\theta_1 \cos\phi_1, \ldots, \sin\theta_4 \cos\phi_4]$, and $\vec{v}_4 = [\sin\phi_1, \ldots, \sin\phi_4]$. If we assume that the vectors $\vec{v}_i$, $i=1,2,3,4$ are linearly independent, we have that the matrix $\mathbf{A}$ is a full rank and $\text{rank}(\mathbf{A}) = 4$. It implies that $\text{krank}(\mathbf{A}) = 4$. If we assume that the vectors $\vec{v}_i$, $i=1,2,3,4$ are linearly dependent, then there exist scalars $c_i, i=1,2,3,4$ not all zeros such that

$$c_1\vec{v}_1 + c_2\vec{v}_2 + c_3\vec{v}_3 + c_4\vec{v}_4 = 0. \tag{15}$$

Let us consider a 3-dimensional Cartesian coordinate system and the Cartesian coordinates $(x_i, y_i, z_i) = (\cos\theta_i \cos\phi_i, \sin\theta_i \cos\phi_i, \sin\phi_i)$, $i=1,2,3,4$. Using (15), we have

$$c_2x_i + c_3y_i + c_4z_i + c_1 = 0, \quad i=1,2,3,4 \tag{16}$$

Eq. (16) is a plane equation with the four Cartesian coordinates (x_i, y_i, z_i) located on the plane. It implies that the four sources lie on the sample plane. Similarly, if we assume that the four sources lie on the same plane, it can be easily showed that the vectors $\vec{v}_i, i=1,2,3,4$ are linear dependent. In this case, we eliminate the first row of $\mathbf{A}$ and obtain $\mathbf{A}' = [\vec{r}_1, \vec{r}_2, \vec{r}_3, \vec{r}_4]$. Since the four sources are distinguishable, we have $\text{rank}(\mathbf{A}') = 3$. It implies that $\text{krank}(\mathbf{A}) = 3$. Therefore, we conclude that Lemma 2 is true. □

Using Lemma 2, it follows from the condition (14) that $K \le 6$ for any four sources not lying on the same plane and $K \le 4$ for four or more sources lying on the same plane. Since the rank of the matrix $\mathbf{A} \odot \mathbf{A}$ determines the identifiability too, it is necessary to further investigate the full column rank of the matrix $\mathbf{A} \odot \mathbf{A}$. We will use the following lemma (see Property 2 in [4]):

Lemma 3. *For two matrices* $\mathbf{A} \in \mathbb{C}^{n \times k}$ *and* $\mathbf{B} \in \mathbb{C}^{m \times k}$, *with* $\mathrm{krank}(\mathbf{A}) \ge 1$ *and* $\mathrm{krank}(\mathbf{B}) \ge 1$, *it holds that*

$$\mathrm{krank}(\mathbf{A} \odot \mathbf{B}) \ge \min\{k, \mathrm{krank}(\mathbf{A}) + \mathrm{krank}(\mathbf{B}) - 1\}. \qquad (17)$$

When there are 4 or more sources lying on the same plane, we have that $\mathrm{krank}(\mathbf{A}) = 3$ and $\mathrm{rank}(\mathbf{A} \odot \mathbf{A}) \ge \min\{K, 2\mathrm{krank}(\mathbf{A}) - 1\}$. It implies that the matrix $\mathbf{A} \odot \mathbf{A} \in \mathbb{R}^{16 \times K}$ is of full column rank for $K \le 5$. Considering the condition $K \le 4$ given by (14), we can conclude that for the AVS array in (2), the DOAs of K sources are uniquely identifiable if $K \le 5$ and there are no more than 4 sources lying on the same plane. In case that all the sources lying on the same plane, the DOAs of K sources are uniquely identifiable if $K \le 4$.

When there are no 4 or more sources lying on the same plane, we have that $\mathrm{rank}(\mathbf{A} \odot \mathbf{A}) \ge \mathrm{krank}(\mathbf{A} \odot \mathbf{A}) \ge \min\{K, 2\mathrm{krank}(\mathbf{A}) - 1\}$, and it follows $\mathrm{rank}(\mathbf{A} \odot \mathbf{A}) \ge K$, for $K \le 6$. Therefore, the matrix $\mathbf{A} \odot \mathbf{A} \in \mathbb{R}^{16 \times K}$ is of full column rank for $K \le 6$. We can conclude that for the AVS array in (2), the DOAs of K sources are uniquely identifiable if $K \le 6$.

Note that the degree of freedom can be further increased by forming an array of AVSs [14]. However, the identifiability analysis of the KR-AVS criterion using an array of AVSs becomes very challenging as a proof of the full column rank of $\mathbf{A} \odot \mathbf{A}$ is not straightforward. The study of an array of AVSs is beyond the scope of this paper.

4. Experiment results

In this section, the performance of the proposed KR-AVS approach is evaluated through computer simulations and compared with the other array geometries with a similar number of sensors. The simulations are carried out with real wideband speech signals employed as the sources. The six speech source signals that were used in our simulation were plotted in Fig. 1. All the speech signals are downloaded from BBC News. The speech sources consist of different speakers. We cut the signals to 2.3 s long and re-sampled to a sampling rate of 22 kHz. An additive Gaussian noise is added to the source signals with various signal to noise ratio (SNR) to be defined as

$$\mathrm{SNR} = \frac{\frac{1}{T}\sum_{t=0}^{T-1} E\{\|\mathbf{A}\mathbf{s}(t)\|^2\}}{E\{\|\mathbf{v}(t)\|^2\}}, \qquad (18)$$

where T is the total number of samples. The wideband speech signals are transformed into the frequency-domain by the short-time Fourier transform (STFT) with frame length $L=512$ or around 25 ms. In speech processing, it is generally assumed that the speech signals are stationary within 25 ms. There are total $M=98$ frames. The frequency band to be processed is [800 Hz, 6000 Hz]. The output

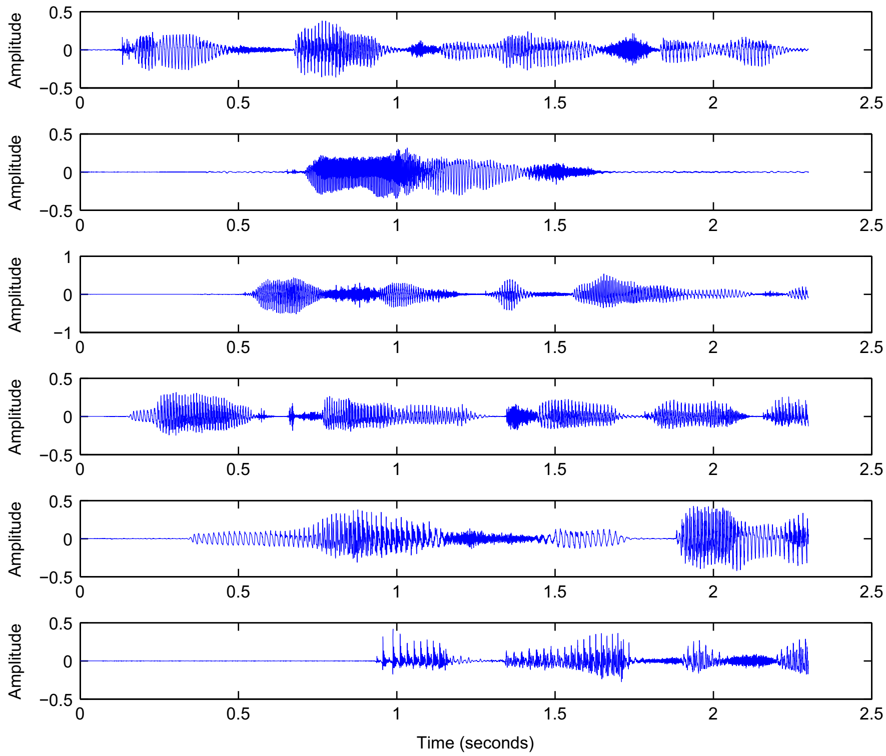

Fig. 1. The six speech source waveforms used for the 2D DOA estimation. The sources are randomly distributed to the six directions.

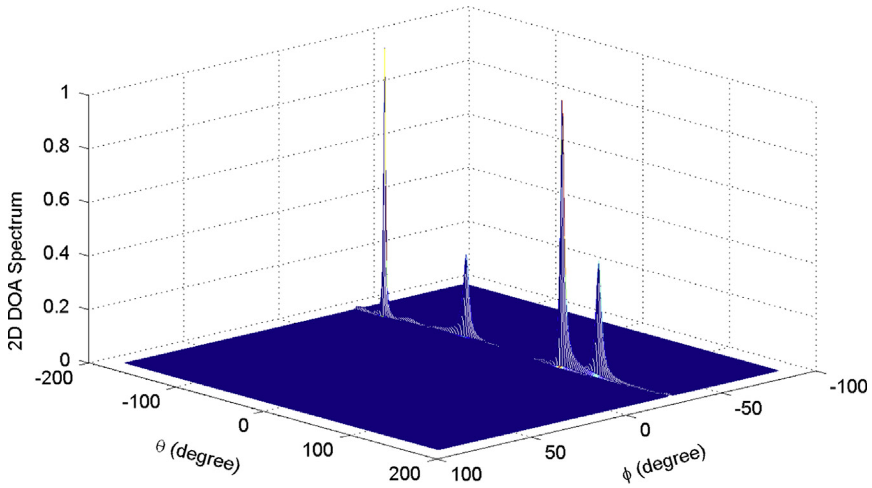

Fig. 2. 2D DOA spectrum of the KR-AVS approach using four-element AVS with four wideband real-speech sources on the same plane (with the same elevation angles -30°).

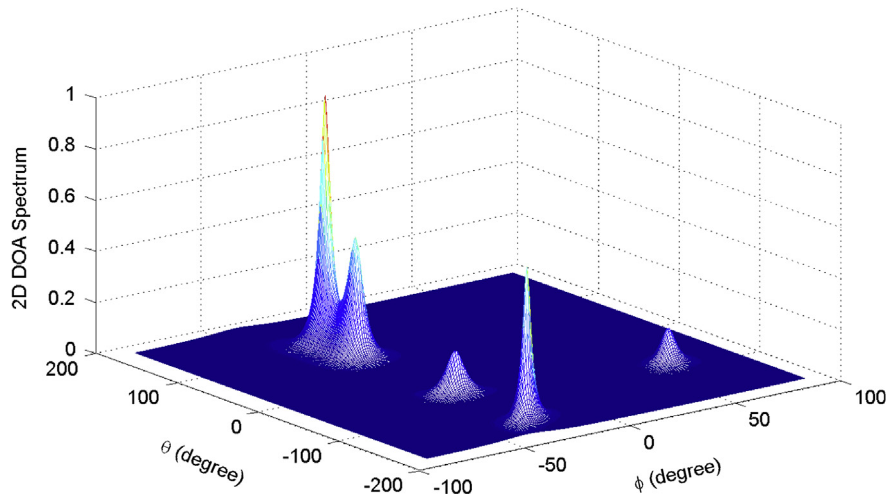

Fig. 3. 2D DOA spectrum of the KR-AVS approach using four-element AVS with five wideband real-speech sources (four of the sources lie on the same plane).

signals of the sensors were obtained using the model given in (1). The DOA spectrum for each frequency bin is computed based on (10), then we summed the DOA spectra over all the valid frequency bins. The angle indices corresponding to peaks of the combined DOA spectrum are identified as the source directions.

4.1. Identification verification of the KR-AVS approach

We verify the identification of the proposed KR-AVS approach in three different source and position settings. We set the SNR as 10 dB in this evaluation and plotted the normalized DOA spectrum for a better visibility. Fig. 2 shows the DOA spectrum of the KR-AVS approach for the four speech sources lying on the same plane (here we set the sources to be the same elevation angles) and the true DOAs are defined as $\{(\theta_1, \phi_1), \ldots, (\theta_4, \phi_4)\} = \{(-150^{\circ}, -30^{\circ}), (-60^{\circ}, -30^{\circ}), (60^{\circ}, -30^{\circ}), (100^{\circ}, -30^{\circ})\}$. It is seen that there are four peaks corresponding to the true angles in a line at $\phi = -30^{\circ}$. By mapping the peaks to the angle indices, we can obtain the estimated DOAs of the four speech sources. Fig. 3 shows the DOA spectrum of the KR-AVS approach for the five speech sources with four sources lying on the same plane where the true DOAs are defined as $\{(\theta_1, \phi_1), \ldots, (\theta_5, \phi_5)\} = \{(-150^{\circ}, -30^{\circ}), (-60^{\circ}, -30^{\circ}), (60^{\circ}, -30^{\circ}), (100^{\circ}, -30^{\circ}), (-120^{\circ}, 60^{\circ})\}$. It is seen that there are five peaks corresponding to the true angles. The DOAs of the five sources are successfully identified. Fig. 4 shows the DOA spectrum of the KR-AVS approach for the six speech sources where no four sources lie on the same plane, and the true DOAs are defined as $\{(\theta_1, \phi_1), \ldots, (\theta_6, \phi_6)\} = \{(120^{\circ}, -30^{\circ}), (10^{\circ}, 10^{\circ}), (110^{\circ}, 30^{\circ}), (-140^{\circ}, 10^{\circ}), (-90^{\circ}, 60^{\circ}), (-60^{\circ}, -60^{\circ})\}$. It is seen that all the sources are successfully identified and both the

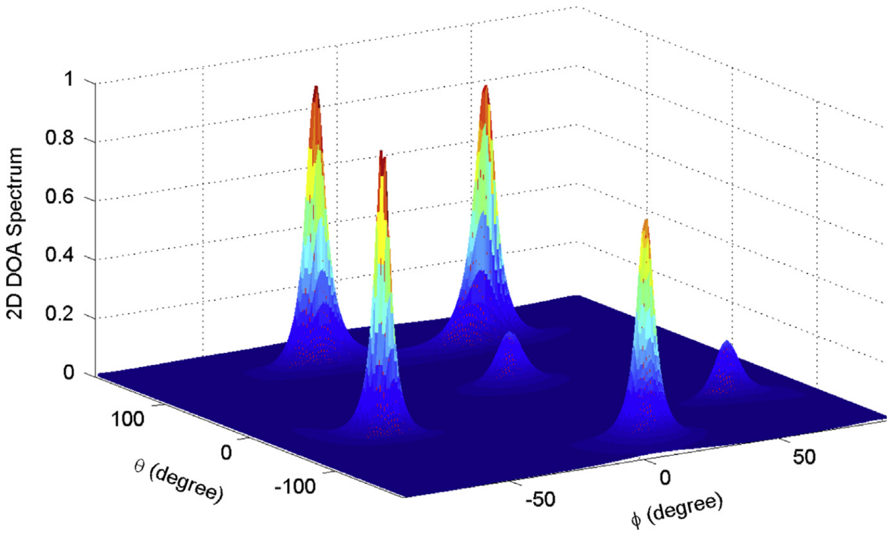

Fig. 4. 2D DOA spectrum of the KR-AVS approach using four-element AVS with six wideband real-speech sources.

azimuth and elevation angles are well estimated. The above results verified our theoretic analysis of the proposed KR-AVS approach in Section 3.

4.2. Comparison of various array geometries

We now compare the AVS array with the non-uniform linear array, the 2D L-shape array [13], and the 2D triangular array. To the author's knowledge, only the identification ability of the linear array [4] and the L-shape array [13] geometries has been studied for DOA estimation using the Khatri–Rao criterion. In this paper, we first time demonstrate the non-uniform linear array and the triangular array geometries for underdetermined DOA estimation using the Khatri–Rao criterion. The setup of each array is given as follows and the microphone 1 is chosen as the reference sensor for all the arrays. The setup of the four-element non-uniform linear array is illustrated in Fig. 5. The inter-sensor spacing for the non-uniform linear array is d, $2d$ and $3d$. The array manifold of the non-uniform linear array is given as

$$\mathbf{a}(\theta_k) = \left\{1, e^{-j2\pi fd\ \sin(\theta_k)/c}, e^{-j2\pi f3d\ \sin(\theta_k)/c}, e^{-j2\pi f6d\ \sin(\theta_k)/c}\right\}, \tag{19}$$

where c denotes the sound speed. The setup of the L-shape array uses five sensors (the minimum number of sensors required to identify 6 sources stated in [13]) and it is illustrated in Fig. 6. The inter-sensor spacing between neighboring sensors is d. The array manifold of the L-shape array is given as

$$\mathbf{a}(\theta_k, \phi_k) = \{1, e^{-j2\pi fd\ \cos(\theta_k)\sin(\phi_k)/c}, e^{-j2\pi f2d\ \cos(\theta_k)\sin(\phi_k)/c}, e^{-j2\pi fd\ \sin(\theta_k)\sin(\phi_k)/c}, e^{-j2\pi f2d\ \sin(\theta_k)\sin(\phi_k)/c}\}. \tag{20}$$

The setup of the triangular array uses four sensors and it is illustrated in Fig. 7. The inter-sensor spacing between neighboring sensors is d. The array manifold of the triangular array is given as

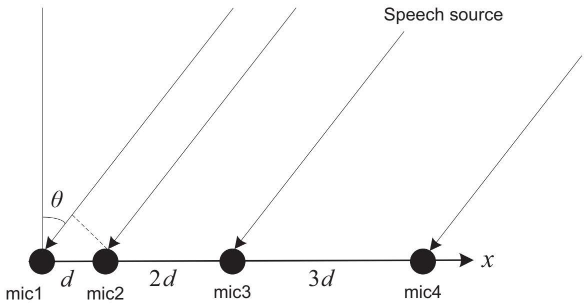

Fig. 5. Non-uniform linear array geometry with four sensors for one-dimensional DOA estimation.

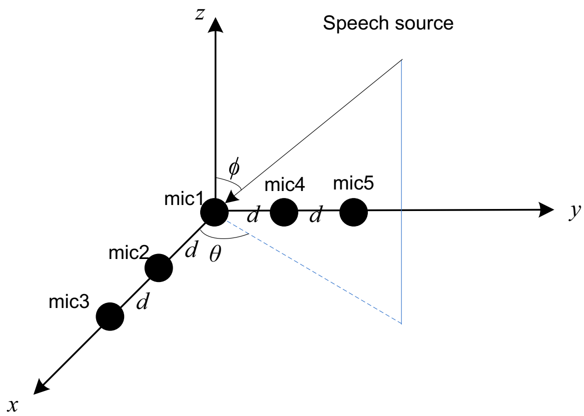

Fig. 6. L-shape array geometry with five sensors for two-dimensional DOA estimation.

$$\mathbf{a}(\theta_k, \phi_k) = \{1, e^{-j2\pi fd\ \cos(\theta_k)\sin(\phi_k)/c}, e^{-j2\pi fd\ \sin(\theta_k)\sin(\phi_k)/c}, e^{-j2\pi fd\ \cos(\phi_k)/c}\}. \tag{21}$$

For all the above arrays, we used the sound speed $c=346$ m/s and set the value as $d=2.87$ cm for avoiding spatial aliasing. It is noted that the AVS array uses one omni-directional sensor and three bi-directional sensors.

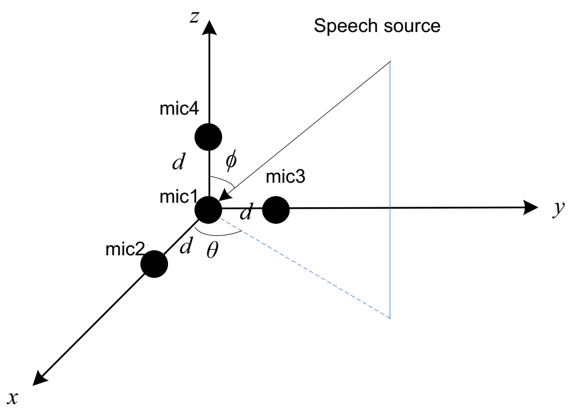

Fig. 7. Triangular array geometry with four sensors for two-dimensional DOA estimation.

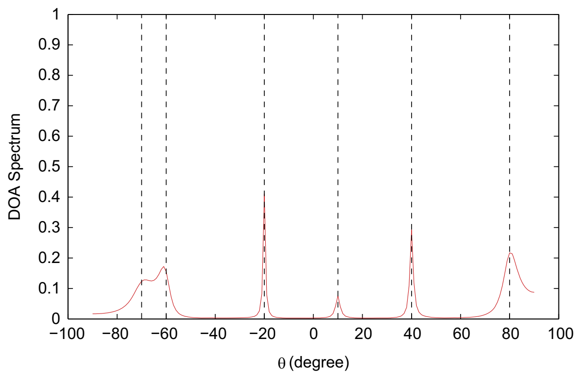

Fig. 8. One-dimensional DOA spectrum of the KR-linear approach using four-element non-uniform linear array with six wideband real-speech sources.

There is no spacing requirement for the AVS array [5]. Compared to the above array geometries, the AVS array aperture is much smaller. Consider that the non-uniform linear array and the L-shape array have ambiguity between front and back directions. In this simulation, we set the six speech sources to the front directions as follows: $\{(\theta_1,\phi_1), \ldots, (\theta_6,\phi_6)\}=\{(80^\circ,-30^\circ), (10^\circ,10^\circ), (40^\circ,40^\circ), (-60^\circ, 10^\circ), (-70^\circ, -60^\circ), (-20^\circ,80^\circ)\}$. The SNR was taken as 10 dB. The DOA spectrum results of all the compared arrays using the KR-AVS criterion are shown in Figs. 8–11. Here normalized DOA spectrum was taken for a better visibility. It is seen that all the arrays produce peaks on the directions of six sources. For the non-uniform linear array, only the azimuth angles are identified between $[-90^\circ, 90^\circ]$. Here $\theta=0^\circ$ is the front direction of the broadside for the linear array. It can be observed that the non-uniform linear array introduces bigger DOA estimation errors when the sources are too close as seen for the azimuth angles of -70° and -60°. In addition, the resolution of detection decreases when the source moves towards the endfire as seen from the more flat peak at $\theta=80^\circ$. Both the L-shape array and the triangular array produce side peaks along the main peaks. For the two sources that are at $\{(\theta_2,\phi_2),(\theta_4,\phi_4)\}=\{(10^\circ,10^\circ),(-60^\circ, 10^\circ)\}$ on the same plane, the resolution of DOA estimation for both the L-shape array and the triangular array decreases. By comparison, the AVS array only produces the main peaks on the directions of sources. There is also no side peaks on the DOA spectrum from the AVS array. Therefore, this simulated evaluation demonstrated that the AVS array seems more preferred for underdetermined 2D DOA estimation.

To evaluate the capabilities of the DOA estimation using different array geometries, we compare the root mean square error (RMSE) for various SNR conditions. The RMSE is defined as

$$\mathrm{RMSE}=\sqrt{E\left\{\frac{1}{K}\sum_{k=1}^{K}\left(|\theta_k-\hat{\theta}_k|^2+|\phi_k-\hat{\phi}_k|^2\right)\right\}}, \tag{22}$$

where the parameters $\hat{\theta}_k$ and $\hat{\phi}_k$ are the estimates of θ_k and ϕ_k. Here we matched the estimated angles with the actual

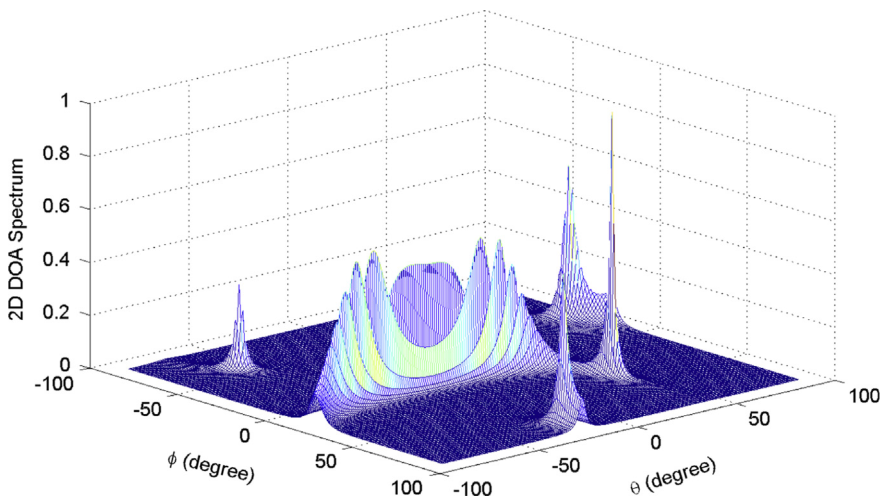

Fig. 9. Two-dimensional DOA spectrum of the KR-L-shape approach using five-element L-shape array with six wideband real-speech sources.

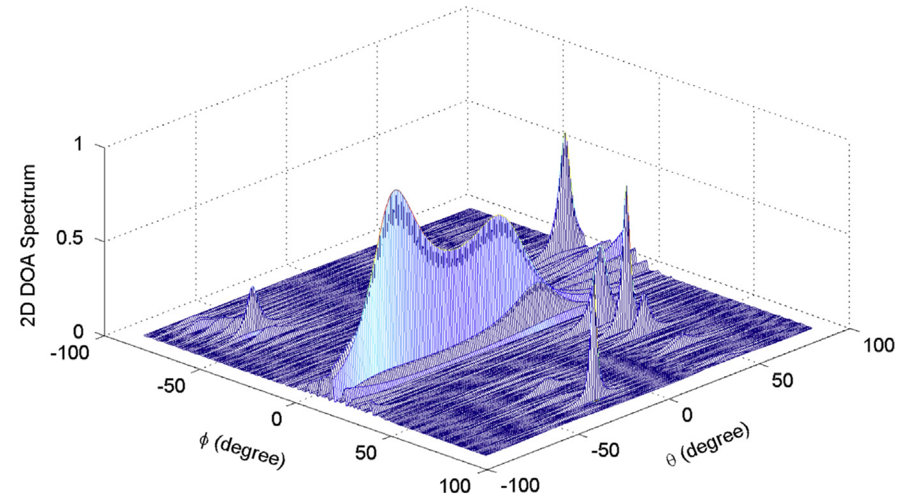

Fig. 10. Two-dimensional DOA spectrum of the KR-triangular approach using four-element triangular array with six wideband real-speech sources.

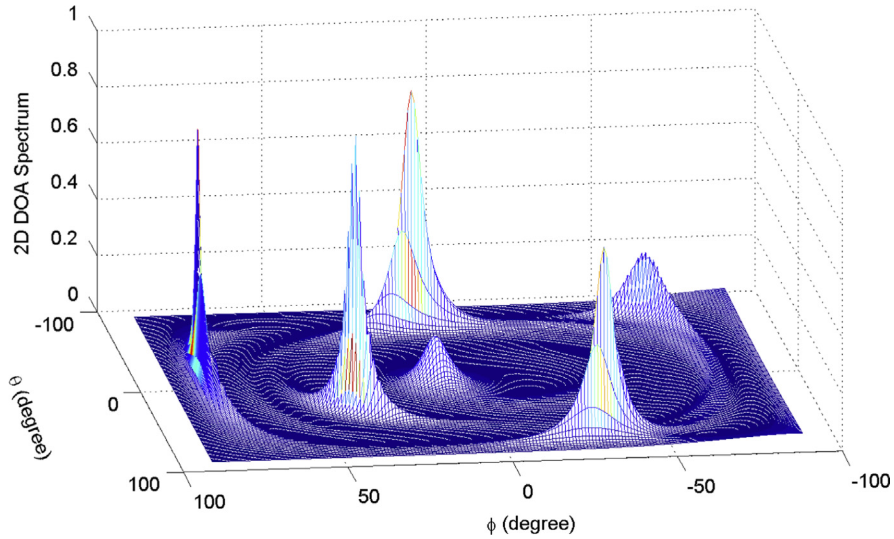

Fig. 11. Two-dimensional DOA spectrum of the KR-AVS approach using four-element AVS array with six wideband real-speech sources.

source angles for the minimum RMSE for each trial. The final RMSE values were taken from the ensemble average of 100 trials. For each trial, we randomly located six speech sources at different directions. All sources were positioned in the front directions and no sources were allowed to be on the same plane. Fig. 12 shows the comparison of the RMSE for various array geometries using the KR-AVS criterion given in (10). We see that for all the array geometries, the RMSE values decrease when the SNR values increase. For all the SNR conditions, the proposed KR-AVS approach achieves better RMSE than the other array geometries.

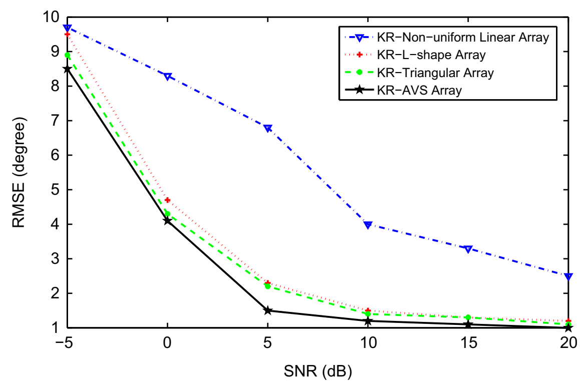

Fig. 12. Comparison of RMSE for the DOA estimation with various array geometries using the KR-AVS criterion versus SNR.

5. Conclusions

We have addressed the underdetermined 2D DOA estimation of the quasi-stationary signals using the KR product and an AVS array. Our analysis for unique identification indicates that the proposed approach is able to identify up to six distinguishable sources in the full 2D space. The experiment results have elaborated the validation of the analysis. The performance evaluation shows that the RMSE of the proposed AVS array approach is superior over the geometries of the non-uniform linear array, the L-shape array, and the triangular array. As the AVS has much smaller aperture size and covers both the full horizontal and vertical planes, the proposed KR-AVS approach is more preferred for practical applications.

References

[1] B. Porat, B. Friedlander, Direction finding algorithms based on high-order statistics, IEEE Trans. Signal Process. 39 (September (9)) (1991) 2016–2023.
[2] P. Chevalier, A. Ferreol, On the virtual array concept for the fourth-order direction finding problem, IEEE Trans. Signal Process. 47 (September (9)) (1999) 2592–2595.
[3] S. Mohan, M.E. Lockwood, M.L. Kramer, D.L. Jones, Localization of multiple acoustic sources with small arrays using a coherence test, J. Acoust. Soc. Am. 123 (April (4)) (2008) 2136–2147.
[4] W.-K. Ma, T.-H. Hsieh, C.-Y. Chi, Doa estimation of quasi-stationary signals with less sensors than sources and unknown spatial covariance: a Khatri–Rao subspace approach, IEEE Trans. Signal Process. 58 (April (4)) (2010) 2168–2180.
[5] A. Nehorai, E. Paldi, Acoustic vector-sensor array processing, IEEE Trans. Signal Process. 42 (September (9)) (1994) 2481–2491.
[6] M. Hawkes, A. Nehorai, Effects of sensor placement in the presence of a reflecting boundary, IEEE J. Ocean. Eng. 24 (1999) 33–44.
[7] P. Tichavsky, K.T. Wong, M.D. Zoltowski, Nearfield/far-field azimuth and elevation angle estimation using a single vector hydrophone, IEEE Trans. Signal Process. 5 (2001) 543–551.
[8] B. Hochwald, A. Behorai, Identifiability in array processing models with vector-sensor applications, IEEE Trans. Signal Process. 44 (January) (1996) 83–95.
[9] J.B. Kruskal, Three-way arrays: rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics, Linear Algebra Appl. 18 (2) (1977) 95–138.
[10] R.O. Schmidt, Multiple emitter location and signal parameter estimation, IEEE Trans. Antennas Propag. 34 (3) (1986) 276–280.
[11] S. Zhao, S. Ahmed, Y. Liang, K. Rupnow, D. Chen, D.L. Jones, A real-time 3D sound localization system with miniature microphone array for virtual reality, in: Proceedings of the 7th IEEE Conference on Industrial Electronics and Applications (ICIEA '12), Singapore, July 2012, pp. 1853–1857.
[12] D. Feng, M. Bao, Z. Ye, L. Guan, X. Li, A novel wideband DOA estimator based on Khatri–Rao subspace approach, Signal Process. 91 (2011) 2415–2419.
[13] P. Palanisamy, C. Kishore, 2-D DOA estimation of quasi-stationary signals based on Khatri–Rao subspace approach, in: Proceedings of the IEEE International Conference on Recent Trends in Information Technology (ICRTIT-2011), Chennai, India, June 3–5, 2011.
[14] J.P. Kitchens, Acoustic vector-sensor array processing (Ph.D. thesis), Massachusetts Institute of Technology, 2010.