Title	An algorithm for mixing matrix estimation in instantaneous blind source separation
Author(s)	Reju, Vaninirappuputhenpurayil Gopalan; Koh, Soo Ngee; Soon, Ing Yann
Citation	Reju, V. G., Koh, S. N., & Soon, I. Y. (2009). An algorithm for mixing matrix estimation in instantaneous blind source separation. Signal Processing, 89, 1762-1773.
Date	2009
URL	http://hdl.handle.net/10220/7013
Rights	© 2009 Elsevier. This is the author created version of a work that has been peer reviewed and accepted for publication by Signal Processing, Elsevier. It incorporates referee’s comments but changes resulting from the publishing process, such as copyediting, structural formatting, may not be reflected in this document. The published version is available at: [DOI: http://dx.doi.org/10.1016/j.sigpro.2009.03.017].

An algorithm for mixing matrix estimation in instantaneous blind source separation

V.G.Reju, Soo Ngee Koh, Ing Yann Soon

School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore 639798

Abstract

Sparsity of signals in the frequency domain is widely used for blind source separation (BSS) when the number of sources is more than the number of mixtures (underdetermined BSS). In this paper we propose a simple algorithm for detection of points in the Time-Frequency (TF) plane of the instantaneous mixtures where only single source contributions occur. Samples at these points in the TF plane can be used for the mixing matrix estimation. The proposed algorithm identifies the single-source-points (SSPs) by comparing the absolute directions of the real and imaginary parts of the Fourier transform coefficient vectors of the mixed signals. Finally, the SSPs so obtained are clustered using the hierarchical clustering algorithm for the estimation of the mixing matrix. The proposed idea for the SSP identification is simpler than the previously reported algorithms.

Key words: Blind source separation, Sparse component analysis, Underdetermined blind source separation, mixing matrix estimation.

Email addresses: `reju@ntu.edu.sg` (V.G.Reju), `esnkoh@ntu.edu.sg` (Soo Ngee Koh), `eiysoon@ntu.edu.sg` (Ing Yann Soon).
URLs: `http://www.ntu.edu.sg/eee/home/esnkoh` (Soo Ngee Koh),

1 Introduction

Blind source separation is the process of separating the original signals from their mixtures without any knowledge about the mixing process or the signals. Many algorithms have been proposed for both instantaneous and convolutive BSS [1,2]. In the case where the number of sources are less than or equal to the number of mixtures, independent component analysis (ICA) method is the most popularly used. However in practical situations the number of sources may be more than the number of mixed signals, and cases like this are called underdetermined BSS. In this paper we address the problem of estimation of mixing matrix for the separation of sources from their instantaneous underdetermined mixtures. The instantaneous mixing process can be mathematically expressed as:

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) \tag{1}$$

where $\mathbf{x}(t) = [x_1(t), \cdots, x_P(t)]^T$ are the P mixed signals, $\mathbf{A}$ is the mixing matrix of order $P \times Q$ with a_{pq} as its $(p,q)^{\text{th}}$ element, $\mathbf{s}(t) = [s_1(t), \cdots, s_Q(t)]^T$ are the Q sources, t is the time instant and T is the transpose operator.

Generally the algorithms for underdetermined case are suitable for determined and overdetermined cases also. This is also true for the proposed algorithm in this paper. When mixing is underdetermined, the performance of the ICA methods are poor and instead sparse component analysis (SCA) is used. In SCA, sparsity of the signals is utilized to separate the signals from their mixed signals. A signal is said to be sparse in the temporal domain, if the signal

`http://www.ntu.edu.sg/home/eiysoon` (Ing Yann Soon).

amplitude is zero during most of the time period. However, in practice the natural signals like speech are not very sparse in the time domain. In [3], P.Bofill, et al. show that signals like speech are more sparse in the frequency domain than in the time domain and hence if we transform the time domain signal into the frequency domain, the sparsity can be utilized to separate the signals from their mixtures.

The idea of BSS based on TF representation was first reported by Belouchrani and Amin in [4]. The algorithm is for the separation of nonstationary sources in the overdetermined case (number of observations > number of sources) based on joint-diagonalization of a set of Spatial Time Frequency Distributions (STFDs) of the whitened observations at selected TF locations. The algorithm is further extended in [5] to make it suitable for underdetermined cases also, under the assumption that the sources are W-disjoint orthogonal in the TF domain. The idea have been further extended in [6] and [7]. In [8], the algorithm proposed in [4] is extended for the case of stochastic sources and a criterion is proposed for the selection of the points in the TF plane where the spatial matrices should be jointly diagonalized.

By utilizing sparsity in the TF domain, many algorithms have been proposed for blind source separation of underdetermined mixtures [2,3,5–20]. In [19] the fact that, at SSPs, the direction of the modulus of the mixture vectors in the TF domain will be the same as those of the column vectors of the mixing matrix is utilized to develop an algorithm called searching-and-averaging-based method, which relaxes the degree of sparsity needed. Searching-and-averaging-based algorithm for time domain signals is also proposed by the same authors in [20], where for the estimation of the mixing matrix, the algorithm removes the samples which are not in the same or inverse direction of the columns of the mixing matrix.

In [14], it is assumed that the signals are W-disjoint orthogonal in the TF plane, i.e., only one source will occur in each TF window, which is quite restrictive. Later, it is shown that approximate W-disjoint orthogonality is sufficient to separate most speech signals and an algorithm called *Degenerate Unmixing Estimation Technique* (DUET) is proposed in [15]. Aïssa-El-Bey, et al. [6] relax the disjoint orthogonality constraint but assume that at any time the number of active sources in the TF plane is strictly less than the number of mixtures. Hence, when the number of sensors are two, the condition in [6] is exactly the same as W-disjoint orthogonality condition. The algorithm proposed in [11] also assumes that the maximum number of active sources at any instant is less than the number of mixtures. In [16,13], these constrains are again relaxed with the only requirement that each source occurs alone in a *tiny set of adjacent TF windows* while several sources may co-exist everywhere else in the TF plane. This method can therefore be used even when the sources overlap in most of the areas in the TF plane. The algorithm proposed in [16] is based on the complex ratio of the mixtures in the TF domain and it is called the *TIme Frequency Ratio Of Mixtures* (TIFROM) method. In the TF domain if only one source occurs in several adjacent windows, then the complex ratio of the mixtures in those windows will remain constant and it will take different values only if more than one source occur. Hence identifying the area where this ratio remains constant is equivalent to identifying the SSPs. The constant complex ratios of the mixtures at the SSPs are called canceling coefficients and these canceling coefficients can be used for the estimation of the sources from their mixtures. The TIFROM algorithm is further improved in [21].

One of the problems with the TIFROM method is its performance degradation because of the inconsistent estimation of the mixing system. This inconsistency is due to the fact that the TIFROM algorithm uses a series of minimum variances of the ratios of the mixed signals in the TF domain taken over the selected windows for the estimation of the column vectors of the mixing matrix. The absolute values of these variances monotonically increase with the increase in the mean of the corresponding ratios or the corresponding columns of the mixing matrix. Since the TIFROM algorithm looks for the mean corresponding to the minimum variance, in cases where the column matrix and hence the ratios and the corresponding variances are high, the algorithm will end up with a wrong result as it will take the mean of the ratios corresponding to a smaller variance as the column of the mixing matrix. This problem is solved in [17] by normalizing the variances. Even though the normalization of the variances created uniformity, if the TF windows used for estimating one of the column of the mixing matrix is sparser than the TF windows used for estimating another column of the mixing matrix, the variance corresponding to first case will be less than that of the second case [17]. This difference in variance may lead to mixing matrix estimation error, and to solve this problem an algorithm based on k-means clustering is proposed in [17].

The restriction of the TIFROM algorithm, i.e., the requirement of single-source-zone, is further relaxed in [18] where it requires only two adjacent points in the same frequency bin with single source contributions for the estimation of the SSPs. In [18], the fact that at SSPs the mixture vectors in the TF domain will be proportional in magnitude to one of the columns of the mixing matrix is used, i.e., $|\mathbf{X}(t,k)| \simeq \mathbf{a}_j |S_j(t,k)|$. Hence if we plot the scatter diagram using the magnitude of the observed data in the TF domain, it will have a clear orientation towards the directions of the column vectors of the mixing matrix if the sources are sufficiently sparse. In situation where the sources are not sufficiently sparse, the orientation of the scatter diagram will not be very clear. Under such situations, the estimation of the directions of the columns of the mixing vectors will be difficult. Now, at points (t, k) and $(t+1, k)$ in the TF plane of the mixtures, if more than one source component are present, the direction of the mixture vectors $\mathbf{X}(t, k)$ and $\mathbf{X}(t+1, k)$ will be the same only if the amplitudes of all the sources remain the same at both the points (t, k) and $(t+1, k)$, i.e., at two consecutive time frames. Since this condition is very unlikely to happen, the mixture vectors $\mathbf{X}(t, k)$ and $\mathbf{X}(t+1, k)$, $\forall t$, which keeps the directions the same can be considered as SSPs. Utilizing this fact, in [18] the points which satisfy the above condition, i.e.,

$$\angle \left|\mathbf{X}\left(t, k\right)\right| - \angle \left|\mathbf{X}\left(t+1, k\right)\right| < \theta_{\mathrm{th}} \tag{2}$$

where $\angle \mathbf{z}$ represents the angle of the vector $\mathbf{z}$ and θ_{th} is the threshold angle, are selected as the SSPs. Then histogram method is applied to the estimated SSPs for the estimation of the mixing matrix.

In [12], a two stage approach consisting of mixing matrix estimation followed by source estimation is proposed. The restriction of the necessary condition for TIFROM algorithm, i.e., the requirement of some adjacent TF regions where only one of the sources occurs, is relaxed in [12]. The mixing matrix estimation procedure proposed in [12] is an extension of the DUET [15] and TIFROM [16] methods; it is based on the ratios of the TF transforms of the mixtures. From the transformation ratio matrix so obtained, several submatrices each of which has identical columns are detected and these columns will correspond to the points where only one of the sources occurs in the transformation ratio matrix.

It can be seen that the main objective in all these algorithms is the detection of the points in the TF domain where only one source occurs at a time. In this paper we propose a simple algorithm to identify these points and use them for the estimation of the mixing matrix using the hierarchical clustering algorithm which is well known because of its versatility [22]. The proposed algorithm can be used for the mixtures where the sources are overlapped in the TF plane, except for some points. Unlike in [16] and [18], these SSPs need not to be adjacent points in the TF domain and the proposed algorithm is simpler than that in [12], which requires many tuning parameters and a long procedure. Since the algorithms proposed in [16,12] can be directly used for source estimation, either from the identified SSPs [16] or the estimated mixing matrix [16,12], we are not repeating the same and our focus in this paper is on SSP identification and the mixing matrix estimation only.

The paper is structured as follows. The proposed algorithm is derived in Section 2; in Section 3, some experimental results are given and finally conclusions are drawn in Section 4.

2 Proposed method

2.1 Single-source-point identification

The instantaneous mixing model in (1) can be expressed in the TF domain using short time Fourier transform (STFT) as:

$$\begin{aligned} \mathbf{X}(t,k) &= \mathbf{A}\mathbf{S}(t,k) \\ &= \sum_{q=1}^{Q} \mathbf{a}_q S_q(t,k) \end{aligned} \tag{3}$$

where $\mathbf{X}(t,k) = [X_1(t,k), \cdots, X_P(t,k)]^T$ and $\mathbf{S}(t,k) = [S_1(t,k), \cdots, S_Q(t,k)]^T$ are respectively the STFT coefficients of the mixtures and sources in the k^{th} frequency bin at time frame t and $\mathbf{a}_q = [a_{1q}, \cdots, a_{Pq}]^T$ is the q^{th} column of the mixing matrix $\mathbf{A}$. For ease of explanation, assume that there are only two sources, i.e., $Q = 2$, and number of mixtures is P. Now at any point in the TF plane, say (t_1, k_1), if the source component from only one of the sources, say that of s_1, is present, i.e., $S_1(t_1, k_1) \neq 0$ and $S_2(t_1, k_1) = 0$. Then, equation (3) can be written as:

$$\mathbf{X}(t_1, k_1) = \mathbf{a}_1 S_1(t_1, k_1) \tag{4}$$

Equating real and imaginary parts of (4), we will get

$$R\left\{\mathbf{X}(t_1, k_1)\right\} = \mathbf{a}_1 R\left\{S_1(t_1, k_1)\right\} \tag{5}$$

$$I\left\{\mathbf{X}(t_1, k_1)\right\} = \mathbf{a}_1 I\left\{S_1(t_1, k_1)\right\} \tag{6}$$

where $R\{x\}$ and $I\{x\}$ respectively represent the real and imaginary part of x. From (5) and (6) it can be seen that the absolute direction of $R\left\{\mathbf{X}(t_1, k_1)\right\}$ and $I\left\{\mathbf{X}(t_1, k_1)\right\}$ are the same, which is same as that of $\mathbf{a}_1$. Similarly at another point, say (t_2, k_2), if only the contribution from source s_2 is present, i.e., $S_1(t_2, k_2) = 0$ and $S_2(t_2, k_2) \neq 0$, then from (3) we can write

$$R\{\mathbf{X}(t_2, k_2)\} = \mathbf{a}_2 R\{S_2(t_2, k_2)\} \tag{7}$$

$$I\{\mathbf{X}(t_2, k_2)\} = \mathbf{a}_2 I\{S_2(t_2, k_2)\} \tag{8}$$

Hence at (t_2, k_2) the absolute direction of $R\{\mathbf{X}(t_2, k_2)\}$ and $I\{\mathbf{X}(t_2, k_2)\}$ are the same which is same as that of $\mathbf{a}_2$. Now consider another point (t_3, k_3) where the contributions from both the sources are present. Then at (t_3, k_3), the directions of $R\{\mathbf{X}(t_3, k_3)\}$ and $I\{\mathbf{X}(t_3, k_3)\}$ will be

$$R\{\mathbf{X}(t_3, k_3)\} = \mathbf{a}_1 R\{S_1(t_3, k_3)\} + \mathbf{a}_2 R\{S_2(t_3, k_3)\} \tag{9}$$

$$I\{\mathbf{X}(t_3, k_3)\} = \mathbf{a}_1 I\{S_1(t_3, k_3)\} + \mathbf{a}_2 I\{S_2(t_3, k_3)\} \tag{10}$$

Form (9) and (10) it can be seen that the absolute direction of $R\{\mathbf{X}(t_3, k_3)\}$ will be the same as that of $I\{\mathbf{X}(t_3, k_3)\}$ only if

$$\frac{R\{S_1(t_3, k_3)\}}{I\{S_1(t_3, k_3)\}} = \frac{R\{S_2(t_3, k_3)\}}{I\{S_2(t_3, k_3)\}} \tag{11}$$

However, in practice, the probability to satisfy the above condition is very low. This fact is experimentally shown in Fig.1, where the mean of the percentage of the points in the TF plane which are below the *absolute value of difference between the ratios*, i.e., $\left|\frac{R\{S_1(t,k)\}}{I\{S_1(t,k)\}} - \frac{R\{S_2(t,k)\}}{I\{S_2(t,k)\}}\right|$, calculated for 15 pairs of randomly selected speech utterances of length 10 s each is shown. For example, from Fig.1, there is only 0.3% of the total multi-source-points (MSPs) (i.e., the point in the TF plane of the mixture where more than one source occur) in the TF plane with difference between the ratios of less than 0.01, i.e., $\left|\frac{R\{S_1(t,k)\}}{I\{S_1(t,k)\}} - \frac{R\{S_2(t,k)\}}{I\{S_2(t,k)\}}\right| < 0.01$. It can also be seen from Fig.1 that the probability to satisfy the condition in (11) is almost zero. Hence we can say that,

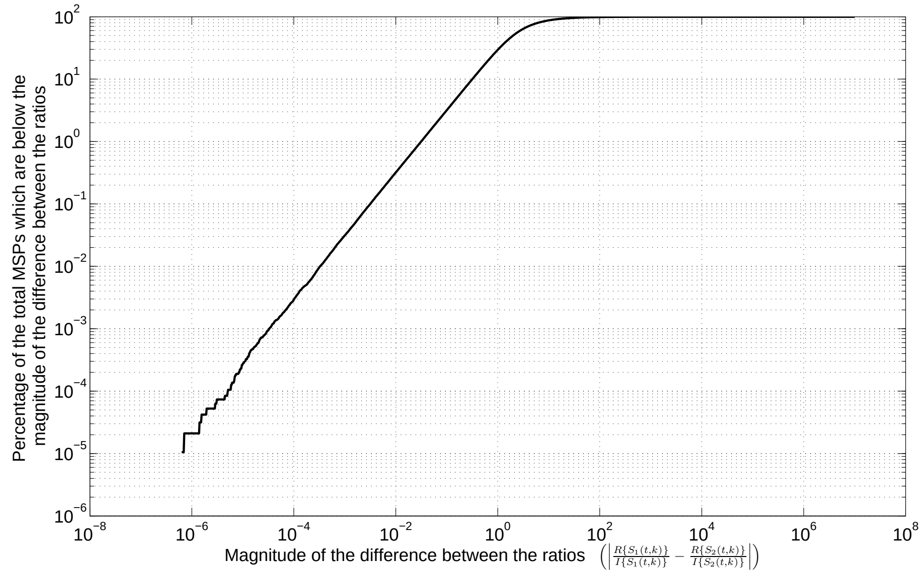


Fig. 1. Percentage of samples which are below the magnitude of the difference between the ratios of the real and imaginary parts of the DFT coefficient of the signals.

in practice, a point, (t, k), in the TF plane of the mixture will be a SSP if the absolute direction of $R\{\mathbf{X}(t,k)\}$ is the same as that of $I\{\mathbf{X}(t,k)\}$; otherwise, it will be a MSP.

For a general case of P mixtures and Q sources, at a MSP (t, k), the real and imaginary parts of $\mathbf{X}(t,k)$ can be written as:

$$R\{\mathbf{X}(t,k)\} = \sum_{q=1}^{Q} \mathbf{a}_q R\{S_q(t,k)\} \tag{12}$$

$$I\{\mathbf{X}(t,k)\} = \sum_{q=1}^{Q} \mathbf{a}_q I\{S_q(t,k)\} \tag{13}$$

Now, the angle between (12) and (13) is given by;

$$\theta = \cos^{-1}\left(\frac{R\left\{\mathbf{X}\left(t,k\right)\right\}^{T} I\left\{\mathbf{X}\left(t,k\right)\right\}}{\sqrt{R\left\{\mathbf{X}\left(t,k\right)\right\}^{T} R\left\{\mathbf{X}\left(t,k\right)\right\}}\sqrt{I\left\{\mathbf{X}\left(t,k\right)\right\}^{T} I\left\{\mathbf{X}\left(t,k\right)\right\}}}\right)$$
$$= \cos^{-1}\left(\frac{\sum_{p=1}^{P}\left(\left(\sum_{q=1}^{Q} a_{pq} R\left\{S_q\left(t,k\right)\right\}\right)\left(\sum_{q=1}^{Q} a_{pq} I\left\{S_q\left(t,k\right)\right\}\right)\right)}{\sqrt{\sum_{p=1}^{P}\left(\sum_{q=1}^{Q} a_{pq} R\left\{S_q\left(t,k\right)\right\}\right)^2 \sum_{p=1}^{P}\left(\sum_{q=1}^{Q} a_{pq} I\left\{S_q\left(t,k\right)\right\}\right)^2}}\right) \tag{14}$$

In the above equation, θ will become 0^o or 180^o if

$$\frac{R\left\{S_1\left(t,k\right)\right\}}{I\left\{S_1\left(t,k\right)\right\}} = \cdots = \frac{R\left\{S_q\left(t,k\right)\right\}}{I\left\{S_q\left(t,k\right)\right\}} = \cdots = \frac{R\left\{S_Q\left(t,k\right)\right\}}{I\left\{S_Q\left(t,k\right)\right\}} \tag{15}$$

Hence, for the absolute directions of $R\left\{\mathbf{X}(t,k)\right\}$ and $I\left\{\mathbf{X}(t,k)\right\}$ to be the same, at any point (t,k) in the TF plane, either the point must be a SSP or the ratios between the real and imaginary parts of the Fourier transform coefficients of all the signals at that points must be the same. However, as shown previously, the probability for the second case is extremely low and this probability will decrease as the number of sources increases. Hence we can conclude that SSPs in the TF plane are the points where the absolute direction of $R\left\{\mathbf{X}(t,k)\right\}$ is the same as that of $I\left\{\mathbf{X}(t,k)\right\}$.

The probability of getting SSPs where the amplitudes of all the source contributions except one are exactly equal to zero is very low in a practical situation. Hence, we relax the condition for SSP as the point in the TF plane where the component of one of the sources is significantly higher than that of the remaining sources. As a result, the point in the TF plane where the difference between the absolute directions of $R\left\{\mathbf{X}(t,k)\right\}$ and $I\left\{\mathbf{X}(t,k)\right\}$ is less than $\Delta\theta$ is taken as SSP, i.e., SSPs are the points in the TF plane where the following condition is satisfied:

Table 1
Algorithm for the detection of the single-source-points

Step 1:	Convert $\mathbf{x}$ in the time domain to the TF domain to get $\mathbf{X}$.
Step 2:	Check the condition in (16).
Step 3:	If the condition in (16) is satisfied, then $\mathbf{X}(t,k)$ is a sample at the SSP, which we keep for mixing matrix estimation; otherwise, discard the point.
Step 4:	Repeat Steps 2 to 3 for all the points in the TF plane or until sufficient number of SSPs are obtained.

$$\left|\frac{R\left\{\mathbf{X}\left(t,k\right)\right\}^{T} I\left\{\mathbf{X}\left(t,k\right)\right\}}{\left\|R\left\{\mathbf{X}\left(t,k\right)\right\}\right\| \left\|I\left\{\mathbf{X}\left(t,k\right)\right\}\right\|}\right| > \cos\left(\Delta\theta\right) \tag{16}$$

where $|\cdot|$ represent the absolute value and $||\mathbf{y}|| = \sqrt{\mathbf{y}^T\mathbf{y}}$. Samples at these SSPs are used for the clustering algorithm in Section 2.2. The algorithm to locate the SSPs in the TF plane is summarized in Table.1.

2.2 Mixing matrix estimation

After identifying the SSPs in the TF plane, the next stage is the estimation of the mixing matrix. Here we are using the hierarchical clustering technique [22,23] for the estimation of mixing matrix. We are not claiming that this is the best algorithm to cluster the samples as other algorithms can also be used [7]. Readers are also referred to [23] and the references therein for a detailed review on clustering algorithms. The main contribution of this paper is the efficient algorithm proposed in Section.2.1 for the detection of SSPs. The real and imaginary parts of $\mathbf{X}\left(t,k\right)$ at the SSPs in the TF plane are stacked into an array, $\tilde{\mathbf{X}}$, and this array is used as the input for clustering. It can be seen that either the real or imaginary parts of the sample vectors at the SSPs are sufficient for clustering as the absolute direction of $R\left\{\mathbf{X}(t,k)\right\}$ and $I\left\{\mathbf{X}(t,k)\right\}$

are the same, except for a difference of maximum $\Delta\theta$. See Section 3 for more explanation.

For hierarchical clustering, we used $1-|\cos(\theta)|$ as the distance measure, where $\cos(\theta) = \tilde{\mathbf{X}}_m^T \tilde{\mathbf{X}}_n / \left(\left\|\tilde{\mathbf{X}}_m\right\| \left\|\tilde{\mathbf{X}}_n\right\|\right)$ is the cosine of the angle between m^{th} and n^{th} sample vectors (column vectors) $\tilde{\mathbf{X}}_m$ and $\tilde{\mathbf{X}}_n$ respectively in $\tilde{\mathbf{X}}$. This clustering is illustrated with a simple example in Fig.2, where the scatter diagram of the data and its dendrogram are shown. To get a clear idea about the clustering algorithm used, Matlab code (only up to the hierarchical tree generation) is also provided in Table.2. In hierarchical clustering, data are partitioned into different clusters by cutting the dendrogram at suitable distance, as shown in Fig.2. If the data contains outliers, the selection of the distance and hence the selection of the number of clusters is important. For example in Fig.2, if we divide the dendrogram into two clusters, we will have one point (point 15), which is the outlier, in one cluster and the remaining will be in the second cluster. In this particular case we must form three clusters and discard the cluster with the least number of samples so that the outlier will be removed. Automatic selection of the number of clusters without any knowledge about the data is difficult [1]. Hence, here we assume that out of the valid clusters (if there are Q sources, there must be Q valid clusters), the cluster with the minimum number of samples will contain at least 5% of the average number of samples in the remaining valid clusters. We also assume that the maximum number of outliers is less than 5% of the total number of samples in the valid clusters. Hence in the algorithm for cutting the dendrogram to form clusters, we will first cut the dendrogram at a suitable height to form Q clusters and if

[1] It may be noted that there are some advanced techniques for the automatic estimation of the number of sources, e.g.,[7]

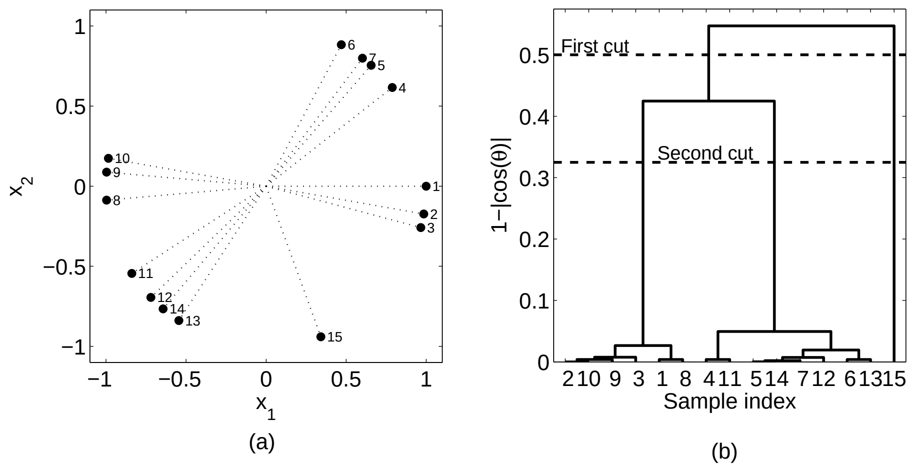


Fig. 2. Hierarchical clustering (a) scatter diagram (b) dendrogram

Table 2
Matlab code for clustering algorithm

```
Y = pdist(Xtilde,'cosine');
Y = 1-abs(1-Y);
Z = linkage(Y,'average');
T = cluster(Z,'maxclust',C);
```

the clusters do not satisfy the above conditions, the dendrogram will be cut at another height to form $Q+1$ clusters. This process is repeated until the above condition is satisfied or when the maximum number of clusters is equal to two times the number of sources. In our experiments the total number of clusters never exceeded $2Q$.

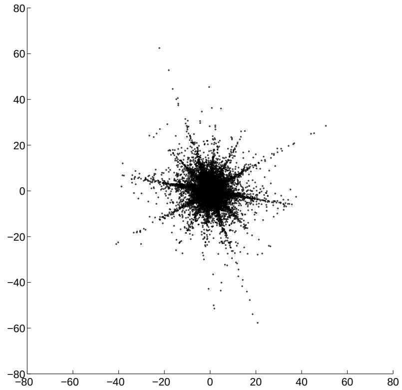
Since $\tilde{\mathbf{X}}$ contains only the samples at SSPs, the scatter plot will have a clear orientation towards the directions of the column vectors in the mixing matrix, as shown in Fig.3, and hence the points in $\tilde{\mathbf{X}}$ will cluster into Q groups. After clustering, the column vectors of the mixing matrix are determined by calculating the centroid of each cluster. The points lying in the left hand side of the vertical axis in the scatter diagram (for two mixture case) are mapped to the right hand side (by changing their sign) before calculating the centroid; otherwise, very small value or zero will result.

The mixing matrix estimation error can be further reduced by removing the points which are away from the mean direction of the cluster. Here we remove all the points which are away from the mean direction of the cluster by $\epsilon\sigma_{\phi_q}$, where ϵ is a constant and σ_{ϕ_q} is the standard deviation of the directions of the samples in the q^{th} cluster. In other words, i^{th} sample in the q^{th} cluster is removed if $\left|\phi_q(i) - \mu_{\phi_q}\right| > \epsilon\sigma_{\phi_q}$, where $\phi_q(i)$ is the absolute direction of the i^{th} sample in the q^{th} cluster and μ_{ϕ_q} is the mean of the absolute direction of the samples in the q^{th} cluster. This is illustrated in Fig.3.

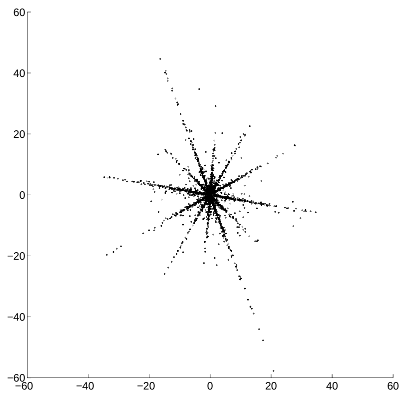
3 Experimental Results

In all the experiments in this paper, except for the cluster diagram and in Section.3.1, average performance of 100 randomly selected combinations, from a set of 11 speech utterances, which are not sparse in the time domain, are used. The other experimental conditions are: sampling frequency 16 kHz, STFT size 1024, Hanning window as the weighting function and $\epsilon = 0.5$.

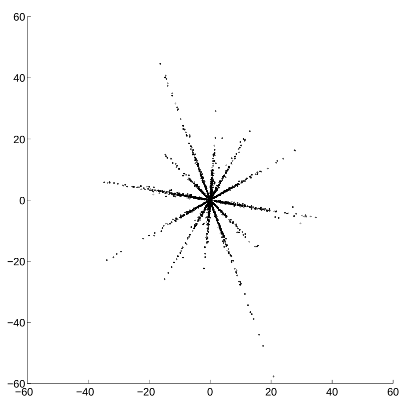
To show that the proposed algorithm is effective in identifying the SSPs and hence in estimating the mixing matrix, six speech utterances are mixed using the mixing matrix $\mathbf{A} = \begin{bmatrix} 0.0872 & 0.3420 & 0.7071 & 0.9848 & 0.8660 & 0.5000 \\ 0.9962 & -0.9397 & -0.7071 & -0.1736 & 0.5000 & 0.8660 \end{bmatrix}$ The scatter diagram in Fig.3 clearly shows the effectiveness of the proposed method for selecting the SSPs, which are in the direction of the column vectors of the mixing matrix, and rejecting the other points. The mixing matrix estimation



(a)



(b)



(c)

Fig. 3. Scatter diagram of the mixtures taking samples from 40 frequency bins; $P = 2$; $Q = 6$; and $\Delta\theta = 0.8^o$ (a) all the DFT coefficients (b) samples at SSPs obtained by comparing the directions of $R\{\mathbf{X}(t,k)\}$ with that of $I\{\mathbf{X}(t,k)\}$ (c) samples at SSPs obtained after elimination of the outliers.

error obtained is:

$$\mathbf{A}-\hat{\mathbf{A}}=\begin{bmatrix} -0.0020 & 0.0049 & 0.0032 & -0.0005 & 0.0007 & 0.0056 \\ 0.0002 & 0.0018 & 0.0032 & -0.0029 & -0.0012 & -0.0032 \end{bmatrix}$$

where $\hat{\mathbf{A}}$ is the estimated mixing matrix, which corresponds to -47.61dB normalized mean square error (NMSE). The NMSE in dB is defined as:

$$\mathrm{NMSE}=10\log_{10}\left(\frac{\sum\limits_{p,q}\left(\hat{a}_{pq}-a_{pq}\right)^{2}}{\sum\limits_{p,q}\left(a_{pq}\right)^{2}}\right) \tag{17}$$

where $\hat{a}_{pq}$ is the $(p,q)^{\mathrm{th}}$ element of the estimated matrix $\hat{\mathbf{A}}$. Since the number of samples to be used for clustering and estimation of the mixing matrix is significantly reduced, the computational time and memory requirement for the clustering algorithm are also reduced. For hierarchical clustering the computational complexity is $O(N^2)$, where N is the number of samples to be clustered [23]. Generally, in the TF domain, the number of samples having very small values dominates and these samples can be removed without much impact on the mixing matrix estimation error. In all our experiments, except where it is mentioned, samples in the TF domain having magnitude below 0.25 (i.e., $||R\left\{\mathbf{X}\left(t,k\right)\right\}|| < 0.25$) are removed.

The advantage of elimination of the outliers from the samples at SSPs estimated by comparing the absolute direction of $R\{\mathbf{X}(t,k)\}$ and $I\{\mathbf{X}(t,k)\}$ is illustrated in Fig.4, where the mixing matrix is $\mathbf{A}$ whose q^{th} column vector is $[\cos\left(\theta_q\right),\ \sin\left(\theta_q\right)]^T$, with $\theta_q=\left(\frac{-\pi}{2.4}+\frac{(q-1)\pi}{6}\right)$ and $q=1,2,\cdots,6$. In Fig.4, the mixing matrix estimation error when the initial samples at SSPs obtained by comparing the absolute direction of $R\{\mathbf{X}(t,k)\}$ and $I\{\mathbf{X}(t,k)\}$ is shown by dotted lines and that obtained after eliminating the outliers as explained in

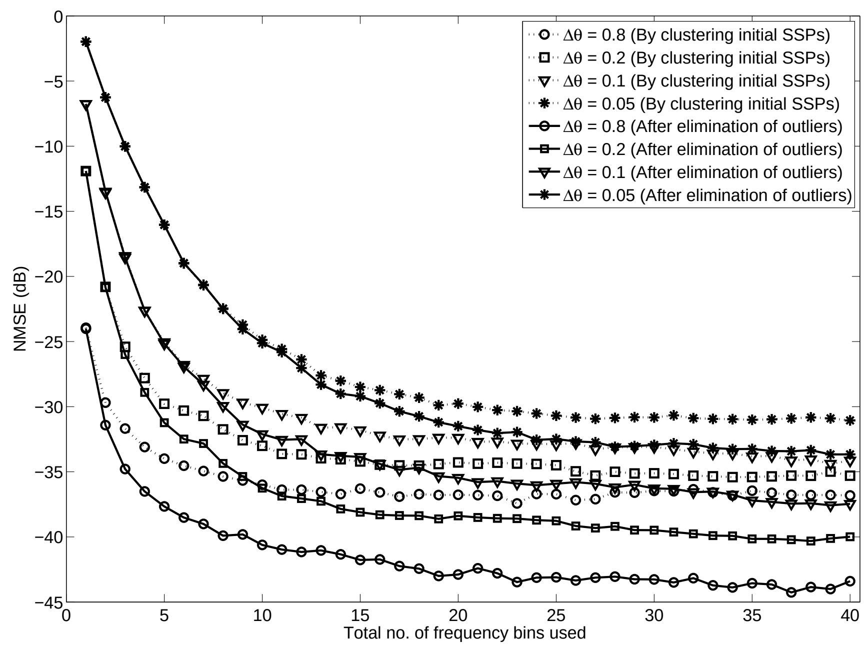


Fig. 4. NMSE on mixing matrix estimation before (dotted lines) and after (solid lines) elimination of the outliers from the initial estimated samples at SSPs for various values of $\Delta\theta$; $P = 2$ and $Q = 6$.

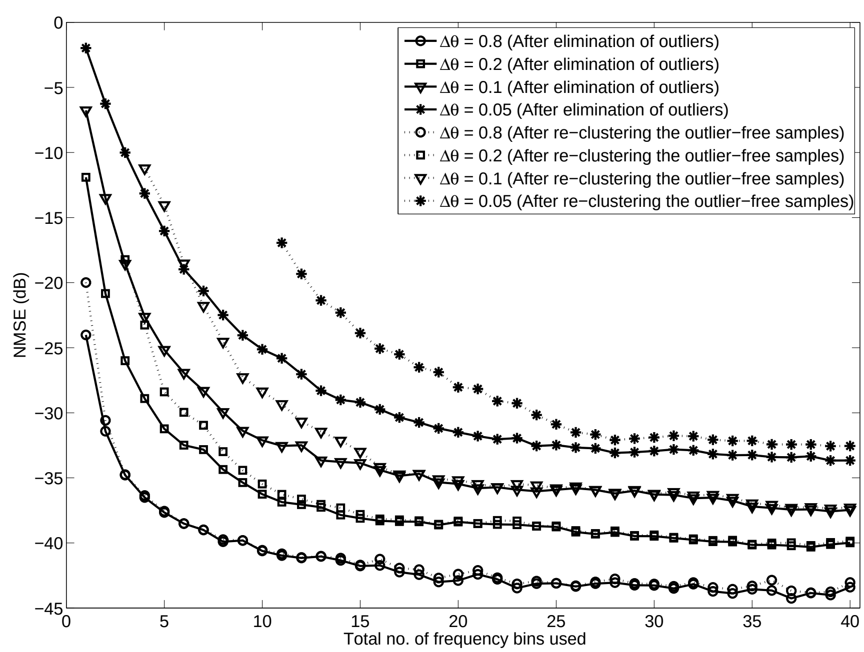


Fig. 5. NMSE on mixing matrix estimation before and after re-clustering the outlier-free samples for various values of $\Delta\theta$; $P = 2$ and $Q = 6$.

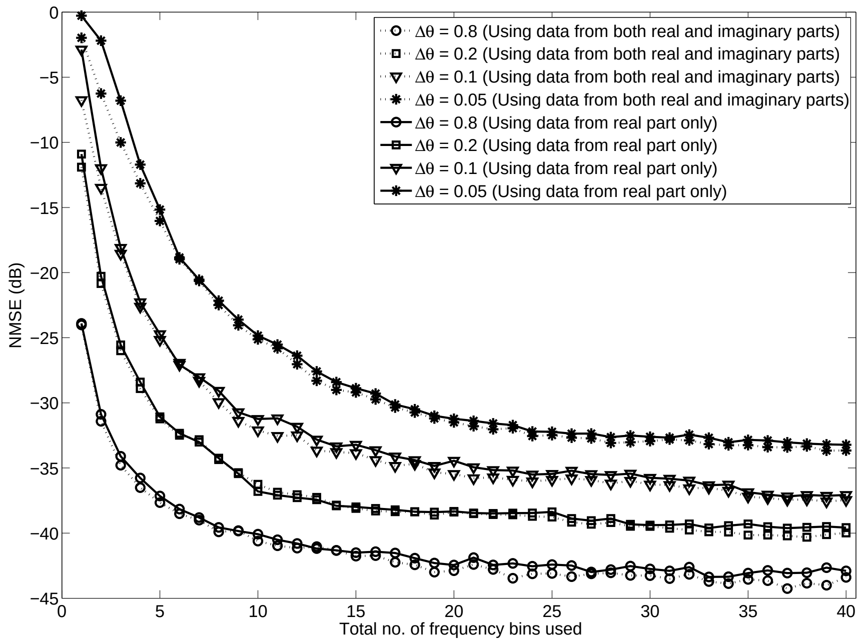


Fig. 6. Comparison of mixing matrix estimation error when samples at SSPs from $R\{\mathbf{X}(t,k)\}$ alone is used with that when samples at SSPs from both $R\{\mathbf{X}(t,k)\}$ and $I\{\mathbf{X}(t,k)\}$ are used, for various values of $\Delta\theta$; $P = 2$ and $Q = 6$.

section.2.2 and recalculating the centroid is shown with solid lines.

In Fig.5, the mixing matrix estimation error obtained by recalculating the centroid of the clusters after eliminating the outliers is compared with that obtained by re-clustering the outlier free samples. It can be seen from the figure that there is no advantage in re-clustering samples after eliminating the outliers. Since the SSPs are identified by comparing the absolute direction of $R\{\mathbf{X}(t,k)\}$ with that of $I\{\mathbf{X}(t,k)\}$, at SSPs the maximum difference in direction between the two vectors will be only $\Delta\theta$. Hence there will not be much difference in performance even if we use $R\{\mathbf{X}(t,k)\}$ or $I\{\mathbf{X}(t,k)\}$ alone instead of both. This is illustrated in Fig.6 where the variation of mixing matrix estimation error for different values of $\Delta\theta$, when $R\left\{\mathbf{X}(t,k)\right\}$ alone (solid lines) and $R\left\{\mathbf{X}(t,k)\right\}$ together with $I\left\{\mathbf{X}(t,k)\right\}$ (dotted lines) are used as the data for clustering, as a function of total number of frequency bins

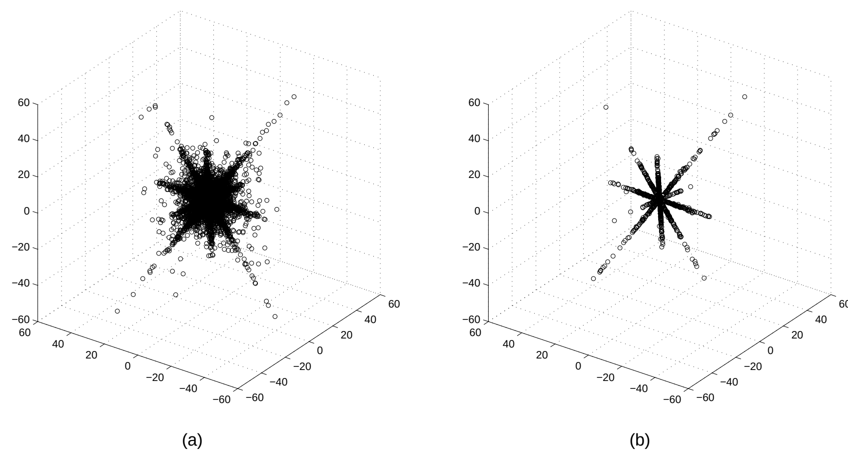


Fig. 7. Scatter diagram of the mixtures taking samples from 40 frequency bins; $P = 3$; $Q = 6$; and $\Delta\theta = 0.8^o$ (a) all the DFT coefficients (b) samples at SSPs after elimination of the outliers.

taken is shown.

In all the experiments in this paper the frequency bins corresponding to one mixture, $\mathbf{x}_1$, were sorted in the descending order of their variance and the order of the frequency bins of other mixtures were modified according to that of $\mathbf{x}_1$ before starting the SSP detection. This is because most of the energy will be concentrated in nearly 10% of the frequency bins [10] and by sorting, the unnecessary computation in the frequency bins where the energy is low can be avoided. From Figs.4, 5, 6 and 8, it is clear that with a properly selected $\Delta\theta$, only 2 to 4% of the frequency bins are sufficient to obtain an accurate estimate of the mixing matrix. When the number of sources are fewer, the first few bins will be sufficient to obtain an accurate estimate of the mixing matrix because the number of SSPs will increase as the number of sources decreases [6].

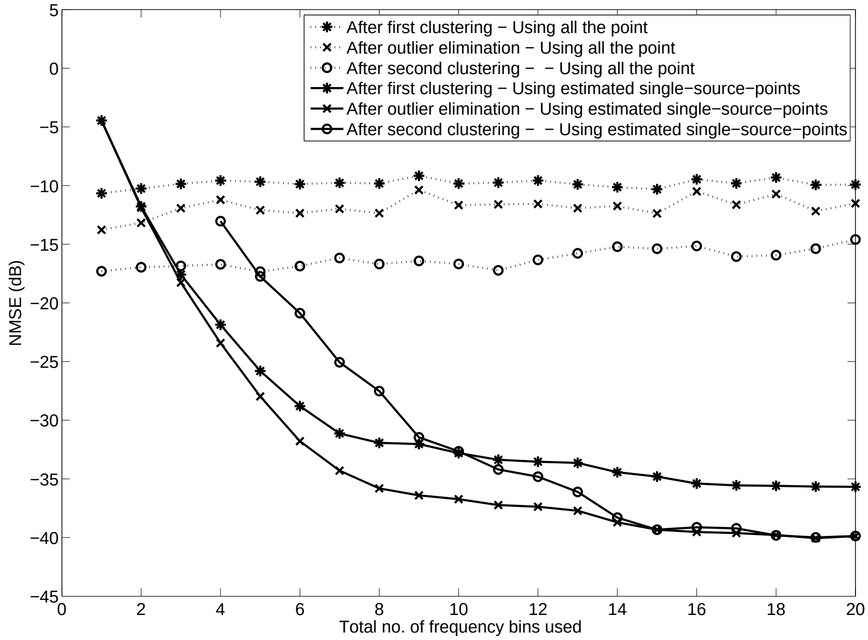


Fig. 8. Comparison of NMSE on estimation of the mixing matrix using all the DFT coefficients in the TF plane with that using the estimated SSPs; $P = 3$; $Q = 6$; and $\Delta\theta = 0.8^o$.

For the case when $P = 3$, $Q = 6$ and a randomly selected mixing matrix

$$\mathbf{A} = \begin{bmatrix} 0.6330 & 0.7650 & 0.0612 & -0.7455 & -0.1988 & -0.6284 \\ 0.5179 & -0.2892 & -0.8156 & 0.3364 & -0.8156 & -0.5201 \\ 0.5754 & 0.6843 & 0.5621 & 0.4994 & -0.7589 & -0.5804 \end{bmatrix}$$

is illustrated in Fig.7 and the performance is shown in Fig.8. In Fig.8 the error in mixing matrix estimation obtained when all the samples in $R\{\mathbf{X}\}$ are used is also shown, where the same procedure described in Section.2.2 is used for the mixing matrix estimation.

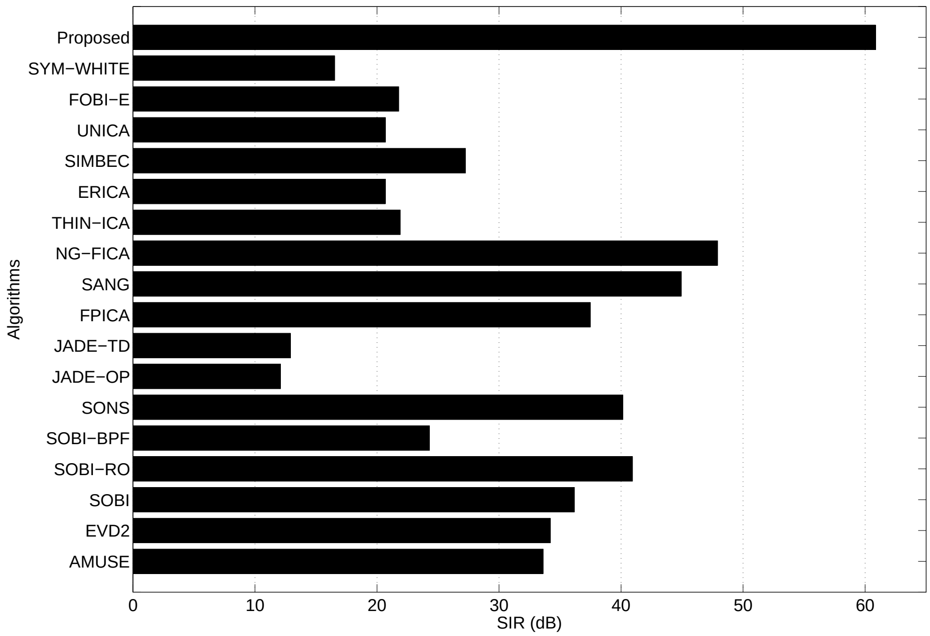


Fig. 9. Comparison of the proposed algorithm with classical algorithms for determined case, $P = Q = 2$

3.1 Comparison with other algorithms

3.1.1 Determined case

Here we compare the performance of the proposed algorithm with several classical algorithms, using the *ICALAB Ver 3* toolbox available at [24]. The references for the algorithms are available in *Help* included in the toolbox. In this experiment the separation performance of each of the algorithm is obtained for five pairs of speech utterances each of length 10 s, and for each pair, the performance is obtained for 100 randomly selected 2×2 mixing matrices. The mean Signal to Interference Ratios (SIR) in dB, so obtained, for the different algorithms are shown in Fig.9. From the figure, it can be seen that the proposed algorithm outperforms the other classical algorithms.

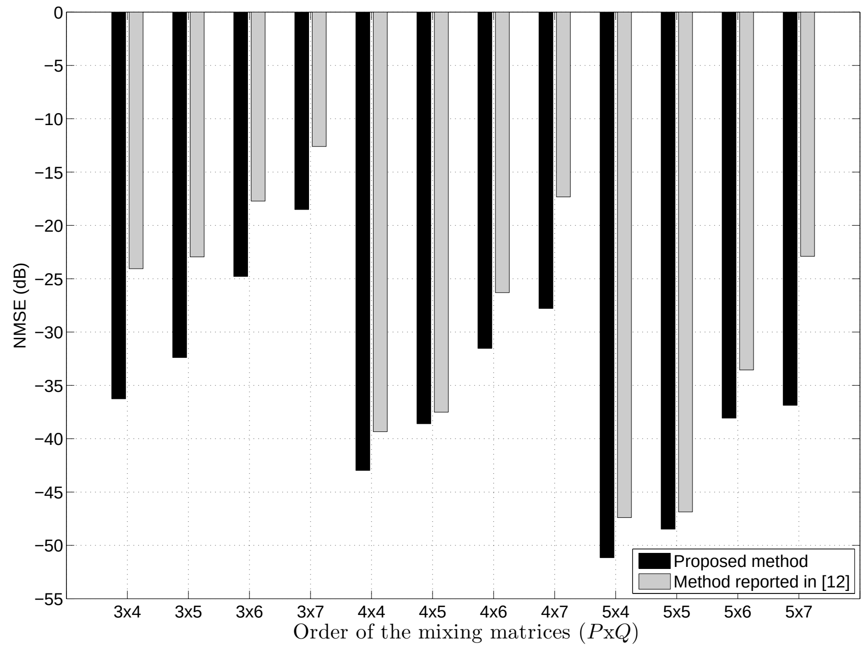


Fig. 10. Comparison of the proposed algorithm with that proposed in [12]

3.1.2 Underdetermined case

In this experiment we compare the proposed algorithm with one of the recently reported algorithms. The algorithm presented in [12], is an extension of the DUET and TIFROM algorithms. Unlike the DUET method, for [12] the spectra of the sources can overlap in the TF domain, i.e, the W-disjoint orthogonality condition need not be met. Furthermore, unlike the TIFROM algorithm, the 'single source region' is also not needed. This is true for the proposed algorithm also. Hence we compare the proposed algorithm with that reported in [12]. Here we conducted 12 experiments (3-sensor, 4-sensor and 5-sensor cases each for 4 to 7 sources) as shown in Fig.10. Each experiment is repeated with 100 different randomly generated mixing matrices and the mean NMSEs so obtained are shown in Fig.10. From the figure, it can be seen that the proposed algorithm outperforms that in [12] in all the cases.

Both the algorithms are implemented in the STFT domain and the number of frequency bins used for both the algorithms are the same. To decide the number of frequency bins to be used, for each experiment, the number of frequency bins is increased until the proposed algorithm detects a minimum of 1000 SSPs.

For the proposed algorithm, the magnitude of the real part of the mixture vectors at any point (t, k) which are less than 5% of the maximum magnitude of all the vectors in the TF plane, i.e., the points with $\|R\{\mathbf{X}(t,k)\}\| < 0.05 \max(\|R\{\mathbf{X}\}\|)$, are discarded.

In the algorithm used for comparison, to cluster the column vectors in $\mathbf{E}$ (please refer [12]) according to their directions, the hierarchical clustering algorithm proposed in this paper is used. The other parameters used are the same as that used in [12], i.e., $M0 = 400$, $J1 = 100$ and $J2 = 100$ (please refer [12] for more details). In cases where the algorithm fails to identify sufficient number of sub-matrices, $M0$, $J1$ and $J2$ are divided by two to obtain their new values and the experiment is repeated using the new values.

4 Conclusion

In this paper we have derived a simple and effective algorithm for single-source-point identification in the TF plane of the mixture signals for the estimation of mixing matrix in underdetermined blind source separation. The algorithm can be used for the mixtures where the spectra of the sources overlap and the single-source-points occur only at a small number of locations. The proposed algorithm does not have any restriction on the number of sources and mixtures and the single-source-points need not be in adjacent locations in the TF plane. Since only the samples at single-source-points are used for the clustering algorithm for the estimation of the mixing matrix, the estimation error, computation time and memory requirement are reduced as compared to using all the samples in the TF plane.

References

[1] A. Hyvärinen, J. Karhunen, E. Oja, Independent Component Analysis, John Wiley & Sons Ltd, New York, 2001.

[2] A. Cichocki, S. Amari, Adaptive Blind Signal and Image Processing Learning Algorithms and Applications, John Wiley & Sons Ltd, New York, 2002.

[3] P. Bofill, M. Zibulevsky, Underdetermined blind source separation using sparse representation, Signal Processing 81 (11) (2001) 2353–2362.

[4] A. Belouchrani, M. G. Amin, Blind source separation based on time-frequency signal representations, IEEE Transactions on Signal Processing 46 (11) (1998) 2888–2897.

[5] L.-T. Nguyen, A. Belouchrani, K. Abed-Meraim, B. Boashash, Separating more sources than sensors using time-frequency distributions, in: Proceedings of the International Symposium on Signal Processing and its Applications, 2001, pp. 583–586.

[6] A. Aïssa-El-Bey, N. Linh-Trung, K. Abed-Meraim, A. Belouchrani, Y. Grenier, Underdetermined blind separation of nondisjoint sources in the time-frequency domain, IEEE Transactions on Signal Processing 55 (3) (2007) 897–907.

[7] Y. Luo, W. Wang, J. A. Chambers, S. Lambotharan, I. Proudler, Exploitation of source nonstationarity in underdetermined blind source separation with

advanced clustering techniques, IEEE Transactions on Signal Processing 54 (6) (2006) 2198–2212.

[8] C. Févotte, C. Doncarli, Two contributions to blind source separation using Time-Frequency distributions, IEEE Signal Processing Letters 11 (3) (2004) 386–389.

[9] N. Mitianoudis, T. Stathaki, Batch and online underdetermined source separation using laplacian mixture models, IEEE Transactions on Audio, Speech and Language Processing 15 (6) (2007) 1818–1832.

[10] R. Saab, O. Yilmaz, M. J. McKeown, R. Abugharbieh, Underdetermined anechoic blind source separation via l^q-Basis-Pursuit with $q < 1$, IEEE Transactions on Signal Processing 55 (8) (2007) 4004–4017.

[11] P. Georgiev, F. Theis, A. Cichocki, Sparse component analysis and blind source separation of underdetermined mixtures, IEEE Transactions on Neural Networks 16 (4) (2005) 992–996.

[12] Y. Li, S. Amari, A. Cichocki, D. W. C. Ho, S. Xie, Underdetermined blind source separation based on sparse representation, IEEE Transactions on Signal Processing 54 (2) (2006) 423–437.

[13] Y. Deville, M. Puigt, Temporal and time-frequency correlation-based blind source separation methods. Part I: Determined and underdetermined linear instantaneous mixtures, Signal Processing 87 (3) (2007) 374–407.

[14] A. Jourjine, S. Rickard, O. Yilmaz, Blind separation of disjoint orthogonal signals: Demixing N sources from 2 mixtures, in: Proceedings of the ICASSP, 2000, pp. 2986–2988.

[15] O. Yilmaz, S. Rickard, Blind separation of speech mixtures via time-frequency masking, IEEE Transactions on Signal Processing 52 (7) (2004) 1830–1847.

[16] F. Abrard, Y. Deville, A time-frequency blind signal separation method applicable to underdetermined mixtures of dependent sources, Signal Processing 85 (7) (2005) 1389–1403.

[17] D. Smith, J. Lukasiak, I. S. Burnett, A two channel, block-adaptive audio separation technique based upon time-frequency information, in: Proceedings of the 12th European Signal Processing Conference, pp. 393–396.

[18] P. Bofill, Identifying single source data for mixing matrix estimation in instantaneous blind source separation, in: ICANN (1), 2008, pp. 759–767.

[19] M. Xiao, S. Xie, Y. Fu, A novel approach for underdetermined blind sources separation in frequency domain, Advances in Neural Networks ISNN 2005 2005 (3497) (2005) 484–489.

[20] X. Ming, X. ShengLi, F. YuLi, Searching-and-averaging method of underdetermined blind speech signal separation in time domain, Science in China Series F: Information Sciences 50 (5) (2007) 771–782.

[21] Y. Deville, M. Puigt, B. Albouy, Time-frequency blind signal separation : extended methods, performance evaluation for speech sources, in: Proceedings of the International Joint Conference on Neural Networks, 2004, pp. 255–260.

[22] A. K. Jain, M. N. Murty, P. J. Flynn, Data clustering: A review, ACM Computing Surveys 31 (3) (1999) 264–323.

[23] R. Xu, Donald II Wunsch, Survey of clustering algorithms, IEEE Transactions on Neural Networks 16 (3) (2005) 645–678.

[24] A. Cichocki, S. Amari, K. Siwek, T. Tanaka, A. H. P. et al, ICALAB Toolboxes, http://www.bsp.brain.riken.jp/ICALAB.