

Underdetermined Blind Source Separation from Time-delayed Mixtures Based on Prior Information Exploitation

Liangjun Zhang*, Jie Yang*, Zhiqiang Guo[†] and Yanwei Zhou*

Abstract – Recently, many researches have been done to solve the challenging problem of Blind Source Separation (BSS) problems in the underdetermined cases, and the “Two-step” method is widely used, which estimates the mixing matrix first and then extracts the sources. To estimate the mixing matrix, conventional algorithms such as Single-Source-Points (SSPs) detection only exploits the sparsity of original signals. This paper proposes a new underdetermined mixing matrix estimation method for time-delayed mixtures based on the receiver prior exploitation. The prior information is extracted from the specific structure of the complex-valued mixing matrix, which is used to derive a special criterion to determine the SSPs. Moreover, after selecting the SSPs, Agglomerative Hierarchical Clustering (AHC) is used to automatically cluster, suppress, and estimate all the elements of mixing matrix. Finally, a convex-model based subspace method is applied for signal separation. Simulation results show that the proposed algorithm can estimate the mixing matrix and extract the original source signals with higher accuracy especially in low SNR environments, and does not need the number of sources before hand, which is more reliable in the real non-cooperative environment.

Keywords: Underdetermined blind source separation, Prior information, Automatic clustering, Single-Source-Point, Subspace projection

1. Introduction

The goal of Blind Source Separation (BSS) is to separate source signals only using the mixtures, which has been widely used in the fields of audio, biomedical, digital communications, and array signal processing [1-3]. Recently, many researches focus on the Underdetermined Blind Source Separation (UBSS) problems when the number of observations is less than the sources [4, 5], in which conditions the Independent Component Analysis (ICA) methods are no longer applicable. Instead, “two-step” method is introduced, which estimates the mixing matrix first and then separates the sources [6, 7].

Obviously, mixing matrix estimation is the first and crucial step in the whole process of UBSS, and some classical methods have been developed, such as Sparse Component Analysis (SCA) and statistical-based methods, etc. SCA methods aim to exploit the sparse nature of sources in Time-Frequency (TF) domain and then use the clustering algorithms to estimate the mixing matrix, which have some requirements for sparsity [8, 9]. Differently, statistical-based methods further exploit the mutual statistical independence between sources, and formulate the problem in terms of a decomposition of a tensor (such

as Canonical Decomposition), which has closed form solutions in some cases [10, 11]. For example, X. Luciani *et al.* [12] formulated the Second Characteristic Function (CAF) of the observations, and blindly estimated the underdetermined mixing matrix successfully using tensor decomposition. However, even if the mixing matrix has been estimated, the original signals are still not easily recovered in the underdetermined condition because the model has more unknowns than equations. To solve the ill-conditioned problem, additional assumptions must be set, then sources can be extracted by minimum norm solution using l_p -norm criterion [13], matrix diagonalization [14], and subspace method [15], etc.

Among the existing algorithms, Single-Source-Point (SSP) detection [16-18] is a kind of simple and efficient method for mixing matrix estimation, which achieves good performance in both over and underdetermined cases. Specifically, it sets no conditions on the stationarity, independence or non-Gaussianity of the sources, which made it be widely used. F. Abrard *et al.* [19] first introduced the concept of single-source regions, and developed the Time-Frequency Ratio of Mixtures (TIFROM) algorithm for instantaneous mixtures. Based on this, M. Puigt *et al.* [20] got two extensions, which take both attenuation and delay into account. However, TIFROM-based methods have their limitations. They suppose single-source TF regions can be found, but if only discrete SSPs exist, the performances will turn worse.

Recently, exploitation of prior information and latent component has been proved to be quite powerful for blind

[†] Corresponding Author: The Key Laboratory of Fiber Optic Sensing Technology and Information Processing, Ministry of Education, Wuhan University of Technology, Wuhan, China. (guozhiqiang@whut.edu.cn)

* The Key Laboratory of Fiber Optic Sensing Technology and Information Processing, Ministry of Education, Wuhan University of Technology, Wuhan, China. (jiejyang509@163.com)

Received: March 31, 2014; Accepted: May 26, 2015

separation in some complicated conditions. V. G. Reju [17] found that the directions of real and imaginary parts of Fourier transform coefficients of the same source should stay the same when the sources are linear instantaneous mixed. In general, Reju's method is simple and practicable, but it works only when the elements of mixing matrix are real-valued, which limits their applications. Lihui *et al.* [18] determined another criterion to identify SSPs. After that, binary masking and K-means clustering methods are implemented to get the clustering centers. However, K-means algorithm needs to know the number of sources to ensure the accuracy, which is not able to be guaranteed in non-cooperative environment.

In this paper, a new mixing matrix estimation method for UBSS from delayed mixtures is introduced. Firstly, we exploit prior information carried on by the special structured complex mixing matrix, and then derive the new criterion for SSPs detecting. Then, modified Agglomerative Hierarchical Clustering (AHC) is used for automatic clustering and mixing matrix calculating. Finally, subspace projection method is used to extract the original signals.

The main contribution of this paper is to propose a new algorithm based on SSP detection for complex-valued mixing matrix estimation. Comparing with the existing algorithms, the proposed prior-aided algorithm plays a more rigid role in SSPs detecting, which achieves better performance for mixing matrix estimation, especially in noisy environment.

The remainder of this paper is organized as following. Section 2 provides problem formulation. Section 3 studies the priors and derives the criterion for SSPs detecting. In section 4, AHC is introduced for mixing matrix estimation. Section 5 proposes the separation method based on subspace. Simulation results are given in section 6. Section 7 concludes this paper.

2. Problem Formulation

2.1 Signal mixing model

In this paper, the anechoic mixing model is used taking the time delay of sources into account. Uniform Linear Antenna (ULA) is used here to receive the signals, as shown in Fig. 1. It is a delayed mixture model by considering the time delay τ . ϕ_n is the incident angle of n -th signal. d is the element spacing, which is set to $d = \lambda/2$, where λ is wavelength.

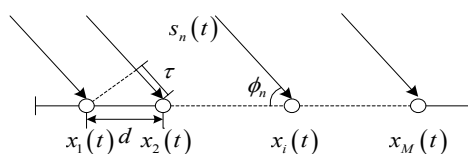


Fig. 1. Uniform Linear Antenna with M elements

The received mixture at i -th element by linear delayed mixing can be written as

$$x_i(t) = \sum_{n=1}^N s_n(t - \tau_{in}) + w_i(t) \quad (i = 1, 2, \dots, M) \quad (1)$$

where

- $x_i(t)$ observed i -th mixture signal,
- $s_n(t)$ n -th source signal,
- $w_i(t)$ additive white Gaussian noise at i -th element,
- τ_{in} time delay for n -th source transmitting to the i -th element, and for ULA antenna, $\tau_{in} = (i - 1)d \cos \phi_n / c$.

For narrow-band signals, (1) can be changed into

$$x_i(t) = \sum_{n=1}^N s_n(t) e^{-j2\pi f_n \tau_{in}} + w_i(t) \quad (i = 1, 2, \dots, M) \quad (2)$$

where f_n is the carrier frequency of each source signal [14]. Particularly, if source signals are real-valued, Hilbert transformation must be applied first to get the analytic signals. Then, rewrite (2) in the form of matrix operation

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) + \mathbf{w}(t) \quad (3)$$

where $\mathbf{x}(t) = [x_1(t), x_2(t), \dots, x_M(t)]^T \in \mathbb{C}^M$,

$\mathbf{s}(t) = [s_1(t), s_2(t), \dots, s_N(t)]^T \in \mathbb{R}^N$ and

$\mathbf{w}(t) = [w_1(t), w_2(t), \dots, w_M(t)]^T \in \mathbb{C}^M$. $\mathbf{A} \in \mathbb{C}^{M \times N}$ is the complex-valued mixing matrix with element $e^{-j2\pi f_n \tau_{in}}$.

2.2 Short-Time Fourier Transform (STFT)

Short-time fourier transform is adopted to exploit the sparsity of signals in TF domain. The STFT of source and mixture signals are defined as

$$\begin{cases} S_n(t, f) = \int_{-\infty}^{+\infty} s_n(\tau) h^*(t - \tau) e^{-j2\pi f \tau} d\tau & (1 \leq n \leq N) \\ X_i(t, f) = \int_{-\infty}^{+\infty} x_i(\tau) h^*(t - \tau) e^{-j2\pi f \tau} d\tau & (1 \leq i \leq M) \end{cases} \quad (4)$$

where $h(t)$ is the window function.

Considering that \mathbf{A} is a constant, and applying STFT to (3) at both sides, the mixing model in TF domain becomes

$$\mathbf{X}(t, f) = \mathbf{A}\mathbf{S}(t, f) + \mathbf{W}(t, f) \quad (5)$$

2.3 Assumptions

The sources are supposed to be non-Gaussian and statistically independent. In the underdetermined case, the number of sources is larger than the number of sensors, i.e.,

$N > M$. To separate all the source signals successfully, the ill-conditioned UBSS model should satisfy the following assumptions.

Assumption 1: For each source, there exist some TF points where the source exists alone. This assumption is necessary only for mixing matrix estimation since we develop the algorithm based on SSP detection.

Assumption 2: Any $M \times M$ submatrix of the mixing matrix \mathbf{A} is of full rank. This assumption helps us in the mixing matrix estimation. To maintain this, sources are assumed to come from different directions in the simulation.

Assumption 3: The number K of sources occur at any TF point is less than the number M of observations, i.e., $K < M$. This assumption is required for original signals extraction using subspace projection method.

3. Proposed SSPs Detection Method

3.1 Prior information exploitation

Prior information can help to improve the performance especial in the complicated environments. Usually, they are derived from the analysis of source signals, which cannot be acquired in the premise of “blind”. Consequently, we exploit the information from the receiver end.

In section 2, time-delayed mixture is introduced. Different from the commonly studied instantaneous mixtures, mixing matrix \mathbf{A} in (3) is a complex-valued with a specific structure, which can be expressed as

$$\mathbf{A} = \begin{bmatrix} 1 & 1 & \dots & 1 \\ e^{-j\pi \cos \phi_1} & e^{-j\pi \cos \phi_2} & \dots & e^{-j\pi \cos \phi_N} \\ \vdots & \vdots & \ddots & \vdots \\ e^{-j\pi(M-1)\cos \phi_1} & e^{-j\pi(M-1)\cos \phi_2} & \dots & e^{-j\pi(M-1)\cos \phi_N} \end{bmatrix} \quad (6)$$

The objective of mixing matrix estimation is to calculate all the elements in (6). The column vector of mixing matrix corresponds to every source signal, which is

$$\mathbf{b}_n(\phi_n) = [1, a_{2,n}, a_{3,n}, \dots, a_{M,n}]^T \\ = [1, e^{-j\pi \cos \phi_n}, \dots, e^{-j\pi(M-1)\cos \phi_n}]^T, (1 \leq n \leq N) \quad (7)$$

where $a_{i,n}$ denotes the element at i -th row and n -th column of \mathbf{A} . Obviously, all the elements at the first row are equal to 1, whereas the others can be written as complex numbers, with modulus equal to 1. Let $a_{i,n} = R_{i,n} + jI_{i,n}$, then the prior information can be achieved as

$$\begin{cases} a_{1,n} = 1 \\ |a_{i,n}| = R_{i,n}^2 + I_{i,n}^2 = 1 \end{cases} \quad (1 \leq i \leq M, 1 \leq n \leq N) \quad (8)$$

Different from the commonly referred “prior information” exploited from the internal characteristics of sources, (8) is derived only using the parameters of receiver antennas, which is feasible in practice.

3.2 Proposed method for SSPs detecting

To estimate the mixing matrix, we detect the SSPs in the TF domain where only one source is active first, and then classify them by clustering method to achieve the estimation of mixing matrix. In this section, a new SSPs detection based on prior information is introduced as follows.

Temporarily ignoring the noises, (5) can be expressed as

$$\mathbf{X}(t, f) = \mathbf{A}\mathbf{S}(t, f) = [\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_N] \begin{bmatrix} S_1(t, f) \\ S_2(t, f) \\ \vdots \\ S_N(t, f) \end{bmatrix} \quad (9) \\ = \mathbf{b}_1 S_1(t, f) + \mathbf{b}_2 S_2(t, f) + \dots + \mathbf{b}_N S_N(t, f)$$

Suppose TF point (t_1, f_1) is a SSP, where only signal $s_n(t)$ is active, then (9) becomes

$$\mathbf{X}(t_1, f_1) = \mathbf{b}_n S_n(t_1, f_1) \quad (10)$$

Apply (7) into (10), and considering the particularity of $a_{i,n}$ equals to 1, we get

$$\begin{cases} X_1(t_1, f_1) = 1 \cdot S_n(t_1, f_1) \\ X_i(t_1, f_1) = a_{i,n} \cdot S_n(t_1, f_1) \end{cases} \quad (2 \leq i \leq M) \quad (11)$$

Obviously, $X_i(t_1, f_1)$, $S_n(t_1, f_1)$ and $a_{i,n} = R_{i,n} + jI_{i,n}$ are all complex numbers. Rewrite (11) in the form of complex multiplication as

$$\begin{aligned} & \text{Re}(X_i(t_1, f_1)) + j \text{Im}(X_i(t_1, f_1)) \\ &= \begin{cases} \text{Re}(S_n(t_1, f_1)) + j \text{Im}(S_n(t_1, f_1)) & (i=1) \\ (R_{i,n} + jI_{i,n}) \cdot [\text{Re}(S_n(t_1, f_1)) + j \text{Im}(S_n(t_1, f_1))] & (2 \leq i \leq M) \end{cases} \end{aligned} \quad (12)$$

where $\text{Re}(X)$ and $\text{Im}(X)$ denote the real and imaginary part of X , respectively. So

$$\begin{cases} \text{Re}(X_1(t_1, f_1)) = \text{Re}(S_n(t_1, f_1)) \\ \text{Im}(X_1(t_1, f_1)) = \text{Im}(S_n(t_1, f_1)) \end{cases} \quad (13)$$

$$\begin{cases} \text{Re}(X_i(t_1, f_1)) = R_{i,n} \cdot \text{Re}(S_n(t_1, f_1)) - I_{i,n} \cdot \text{Im}(S_n(t_1, f_1)) \\ \text{Im}(X_i(t_1, f_1)) = R_{i,n} \cdot \text{Im}(S_n(t_1, f_1)) + I_{i,n} \cdot \text{Re}(S_n(t_1, f_1)) \end{cases} \quad (14)$$

Apply (13) into (14) we can get

$$\begin{cases} \text{Re}(X_i(t_1, f_1)) = R_{i,n} \cdot \text{Re}(X_1(t_1, f_1)) - I_{i,n} \cdot \text{Im}(X_1(t_1, f_1)) \\ \text{Im}(X_i(t_1, f_1)) = R_{i,n} \cdot \text{Im}(X_1(t_1, f_1)) + I_{i,n} \cdot \text{Re}(X_1(t_1, f_1)) \end{cases} \quad (15)$$

Then, considering that $R_{i,n}^2 + I_{i,n}^2 = 1$, (15) can be rewritten as

$$\begin{cases} \text{Re}(X_1(t_1, f_1)) = R_{i,n} \cdot \text{Re}(X_i(t_1, f_1)) + I_{i,n} \cdot \text{Im}(X_i(t_1, f_1)) \\ \text{Im}(X_1(t_1, f_1)) = R_{i,n} \cdot \text{Im}(X_i(t_1, f_1)) - I_{i,n} \cdot \text{Re}(X_i(t_1, f_1)) \end{cases} \quad (16)$$

which can be changed to the form of matrix operation as

$$\begin{bmatrix} \text{Re}(X_1(t_1, f_1)) \\ \text{Im}(X_1(t_1, f_1)) \end{bmatrix} = \begin{bmatrix} \text{Re}(X_i(t_1, f_1)) & \text{Im}(X_i(t_1, f_1)) \\ \text{Im}(X_i(t_1, f_1)) & -\text{Re}(X_i(t_1, f_1)) \end{bmatrix} \begin{bmatrix} R_{i,n} \\ I_{i,n} \end{bmatrix} \quad (17)$$

$$\begin{bmatrix} R_{i,n} \\ I_{i,n} \end{bmatrix} = \begin{bmatrix} \text{Re}(X_i(t_1, f_1)) & \text{Im}(X_i(t_1, f_1)) \\ \text{Im}(X_i(t_1, f_1)) & -\text{Re}(X_i(t_1, f_1)) \end{bmatrix}^{-1} \begin{bmatrix} \text{Re}(X_1(t_1, f_1)) \\ \text{Im}(X_1(t_1, f_1)) \end{bmatrix} \quad (18)$$

Obviously, only $R_{i,n}$ and $I_{i,n}$ are unknown, which can be calculated by (18) directly. Each SSP corresponds to a vector $[R_{i,n}, I_{i,n}]^T$, which satisfies $|a_{i,n}| = |R_{i,n} + jI_{i,n}| = 1$ under ideal conditions. However, because of the noise and computational errors, $|a_{i,n}|$ may not equal to 1 strictly, then the threshold ε (which is very close to 0) is introduced, and the SSPs should satisfy the following inequation as in (19).

$$\left\| \begin{bmatrix} \text{Re}(X_i(t_1, f_1)) & \text{Im}(X_i(t_1, f_1)) \\ \text{Im}(X_i(t_1, f_1)) & -\text{Re}(X_i(t_1, f_1)) \end{bmatrix}^{-1} \begin{bmatrix} \text{Re}(X_1(t_1, f_1)) \\ \text{Im}(X_1(t_1, f_1)) \end{bmatrix} \right\|_2 - 1 < \varepsilon \quad (19)$$

3.3 Feasibility analysis

Remark 1: Multi-Source-Points (MSPs) do not satisfy (19).

Derivation in Section 3.1 proves that criterion (19) holds when a TF point is SSP. However, the actual SSPs detection is a reverse process, that is, a TF point is determined to be a SSP if it satisfies (19). Before doing this, we have to ensure MSPs will not meet the criterion.

Suppose $s_{\alpha_p}(t)$, ($p=1, 2, \dots, N_p$) are the N_p sources occur at TF point (t_2, f_2) . That is

$$\mathbf{X}(t_2, f_2) = \sum_{p=1}^{N_p} \mathbf{b}_{\alpha_p} S_{\alpha_p}(t_2, f_2) \quad (20)$$

Then, similar with (12), (21) can be got as

$$\begin{cases} \text{Re}(\mathbf{X}(t_2, f_2)) = \sum_{p=1}^{N_p} [\text{Re}(\mathbf{b}_{\alpha_p}) \cdot \text{Re}(S_{\alpha_p}) - \text{Im}(\mathbf{b}_{\alpha_p}) \cdot \text{Im}(S_{\alpha_p})] \\ \text{Im}(\mathbf{X}(t_2, f_2)) = \sum_{p=1}^{N_p} [\text{Re}(\mathbf{b}_{\alpha_p}) \cdot \text{Im}(S_{\alpha_p}) + \text{Im}(\mathbf{b}_{\alpha_p}) \cdot \text{Re}(S_{\alpha_p})] \end{cases} \quad (21)$$

So that,

$$\begin{cases} \text{Re}(X_1(t_2, f_2)) = \sum_{p=1}^{N_p} \text{Re}(S_{\alpha_p}(t_2, f_2)) \\ \text{Im}(X_1(t_2, f_2)) = \sum_{p=1}^{N_p} \text{Im}(S_{\alpha_p}(t_2, f_2)) \\ \text{Re}(X_i(t_2, f_2)) = \sum_{p=1}^{N_p} \begin{bmatrix} \text{Re}(a_{i,\alpha_p}) \cdot \text{Re}(S_{\alpha_p}) \\ -\text{Im}(a_{i,\alpha_p}) \cdot \text{Im}(S_{\alpha_p}) \end{bmatrix}, \quad 2 \leq i \leq M \\ \text{Im}(X_i(t_2, f_2)) = \sum_{p=1}^{N_p} \begin{bmatrix} \text{Re}(a_{i,\alpha_p}) \cdot \text{Im}(S_{\alpha_p}) \\ +\text{Im}(a_{i,\alpha_p}) \cdot \text{Re}(S_{\alpha_p}) \end{bmatrix}, \quad 2 \leq i \leq M \end{cases} \quad (22)$$

Then, as for $2 \leq i \leq M$, following equation can be got

$$\begin{aligned} & \text{Re}(a_{i,\alpha_q}) \cdot \text{Re}(X_i(t_2, f_2)) + \text{Im}(a_{i,\alpha_q}) \cdot \text{Im}(X_i(t_2, f_2)) \\ &= \text{Re}(S_{\alpha_q}(t_2, f_2)) \\ &+ \sum_{p=1, p \neq q}^{N_p} \left\{ \text{Re}(S_{\alpha_p}(t_2, f_2)) \cdot [\text{Re}(a_{i,\alpha_q}) \cdot \text{Re}(a_{i,\alpha_p}) + \text{Im}(a_{i,\alpha_q}) \cdot \text{Im}(a_{i,\alpha_p})] \right\} \\ &+ \sum_{p=1, p \neq q}^{N_p} \left\{ \text{Im}(S_{\alpha_p}(t_2, f_2)) \cdot [\text{Im}(a_{i,\alpha_q}) \cdot \text{Re}(a_{i,\alpha_p}) - \text{Re}(a_{i,\alpha_q}) \cdot \text{Im}(a_{i,\alpha_p})] \right\} \end{aligned} \quad (23)$$

Suppose (19) also holds at MSPs, equations like (16) can also be got with the help of (22), that is

$$\begin{aligned} & \text{Re}(a_{i,\alpha_q}) \cdot \text{Re}(X_i(t_2, f_2)) + \text{Im}(a_{i,\alpha_q}) \cdot \text{Im}(X_i(t_2, f_2)) \\ &= \text{Re}(X_1(t_2, f_2)) = \sum_{p=1}^{N_p} \text{Re}(S_{\alpha_p}(t_2, f_2)) \end{aligned} \quad (24)$$

It is obvious that (23) and (24) hold at the same time only when the following conditions are satisfied.

$$\begin{cases} \text{Re}(a_{i,\alpha_q}) \cdot \text{Re}(a_{i,\alpha_p}) + \text{Im}(a_{i,\alpha_q}) \cdot \text{Im}(a_{i,\alpha_p}) = 1 \\ \text{Im}(a_{i,\alpha_q}) \cdot \text{Re}(a_{i,\alpha_p}) - \text{Re}(a_{i,\alpha_q}) \cdot \text{Im}(a_{i,\alpha_p}) = 0 \end{cases}, p \neq q \quad (25)$$

Considering that $\text{Re}^2(a_{i,\alpha_q}) + \text{Im}^2(a_{i,\alpha_q}) = 1$ and $\text{Re}^2(a_{i,\alpha_p}) + \text{Im}^2(a_{i,\alpha_p}) = 1$, (26) can be obtained as

$$\begin{cases} \text{Re}(a_{i,\alpha_q}) = \text{Re}(a_{i,\alpha_p}) \\ \text{Im}(a_{i,\alpha_q}) = \text{Im}(a_{i,\alpha_p}) \end{cases}, p \neq q \quad (26)$$

Eq. (26) reveals that there are two different column vectors with same values, that is, $\mathbf{b}_{\alpha_p} = \mathbf{b}_{\alpha_q}$, ($p \neq q$). It is impossible because **Assumption 2** ensures that the mixing matrix proposed in (4) is column full rank, which means the sources come from different directions. Therefore, (26) can not meet, and MSPs can not satisfy criterion (19).

Remark 2: Algorithm can be applied to wideband signals.

For wideband signals, (2) is no longer exist, apply SFTF to (1) directly, so that

$$\begin{aligned} X_i(t, f) &= \int_{-\infty}^{\infty} x_i(t) h^*(t-\tau) e^{-j2\pi f \tau} d\tau \\ &= \int_{-\infty}^{\infty} \left[\sum_{n=1}^N s_n(t-\tau_{in}) \right] h^*(t-\tau) e^{-j2\pi f \tau} d\tau \\ &= \sum_{n=1}^N \int_{-\infty}^{\infty} s_n(t-\tau_{in}) h^*(t-\tau) e^{-j2\pi f \tau} d\tau \end{aligned} \quad (27)$$

Let $t' = t - \tau_{in}$, (27) can be changed into

$$\begin{aligned} X_i(t, f) &= \sum_{n=1}^N \int_{-\infty}^{\infty} s_n(t') h^*(t' - (\tau - \tau_{in})) e^{-j2\pi f(\tau - \tau_{in})} e^{-j2\pi f \tau_{in}} d(\tau - \tau_{in}) \\ &= \sum_{n=1}^N e^{-j2\pi f \tau_{in}} \cdot \int_{-\infty}^{\infty} s_n(t') h^*(t' - (\tau - \tau_{in})) e^{-j2\pi f(\tau - \tau_{in})} d(\tau - \tau_{in}) \\ &= \sum_{n=1}^N e^{-j2\pi f \tau_{in}} \cdot S_n(t, f) \end{aligned} \quad (28)$$

Rewrite (28) in the form of matrix operation as

$$\mathbf{X}(t, f) = \mathbf{A}(f) \mathbf{S}(t, f) + \mathbf{W}(t, f) \quad (29)$$

where $\mathbf{A}_{i,n}(f) = e^{-j2\pi f \tau_{in}}$. Different from (5), the mixing matrix $\mathbf{A}(f)$ is not a constant in the whole TF domain, whereas a function of f instead. However, given a fixed TF point (t_2, f_2) , $\mathbf{A}(f)$ is determined immediately. Then, (29) is equivalent to the standard narrowband model, which means that the proposed SSPs detection method is also available for wideband signals.

Definition 1: Any TF point (t_0, f_0) is determined as a SSP only when criterion (19) is satisfied. And SSPs can be collected as a series of data vector as

$$\begin{aligned} \left[\text{Re}(a_{i,n}), \text{Im}(a_{i,n}) \right]^T &= \left[\begin{array}{cc} \text{Re}(X_i(t_1, f_1)) & \text{Im}(X_i(t_1, f_1)) \\ \text{Im}(X_i(t_1, f_1)) & -\text{Re}(X_i(t_1, f_1)) \end{array} \right]^{-1} \\ \left[\begin{array}{c} \text{Re}(X_1(t_1, f_1)) \\ \text{Im}(X_1(t_1, f_1)) \end{array} \right] & \end{aligned} \quad (30)$$

The only thing to note here is that the proposed SSP detection method works under the condition with at least

two observed signals, i.e., $M \geq 2$, which makes sure (30) can be used to get the data for clustering.

4. Mixing Matrix Estimation Based on AHC

Every detected SSP corresponds to a matrix element $a_{i,n}$ ($2 \leq i \leq M, 1 \leq n \leq N$), and it is a complex of which modulus value is 1. They should distribute on a unit circle when setting X and Y axis correspond to the real and imaginary part respectively, and show a clustering feature to indicate different sources. After applying clustering method, clustering centers can be obtained, which denote different elements of mixing matrix. Once estimate all the matrix elements \hat{a}_{in} ($2 \leq i \leq M, 1 \leq n \leq N$), the mixing matrix can be formulated as

$$\hat{\mathbf{A}} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ \hat{a}_{2,1} & \hat{a}_{2,2} & \cdots & \hat{a}_{2,N} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{a}_{M,1} & \hat{a}_{M,2} & \cdots & \hat{a}_{M,N} \end{bmatrix} \quad (31)$$

K-means clustering method is commonly used for its fast computation and high accuracy especially for large data. However, it needs to know the clustering number and initial cluster centers before hand, which is difficult in non-cooperative environments. It may converge to a local solution if inappropriate initialization is used. Besides, K-means method ascribes every scatter point to one class including the scatter points caused by noises, which undoubtedly increases the estimation error. Fig. 2 shows the distribution of SSPs. C1, C2, and C3 refer to the actual cluster centers. In fact, scatter points usually can not be avoided, which usually locate far from the true centers. They may be caused by noises or calculation errors, which should be discarded.

To overcome the shortage of general clustering methods, a special agglomerative hierarchical clustering is used for automatic clustering. The so-called ‘‘agglomerative’’ refers to the algorithm initialization. It initializes every SSP as a class and merges the two closest clusters at each step, and results in a series of classes. At last, the final clusters are determined after removing the scattered classes caused by noises or computation errors. Here, two parameters are adopted, th_d denotes the minimum Euclidean distance to

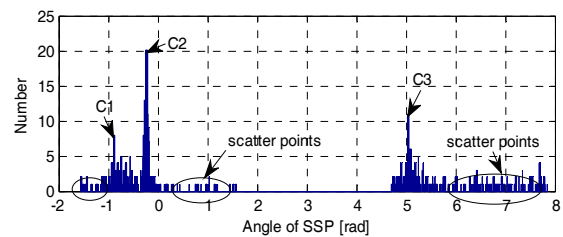


Fig. 2. Schematic plot of SSPs' distribution histogram.

determine when the clustering is completed, and th_N is used to drop discrete error classes. In this way, the number of sources is estimated together with the elements of mixing matrix. To further eliminate the influence of noise, the singular value decomposition method is applied, which takes the singular vectors with maximum singular values as the estimation of mixing matrix elements. The advantage of the AHC method is that in the non-transition merger cases, the noise points have their own clusters and discarded at last, which reduces the interferences for center calculating.

Suppose $U_l = [R_l, I_l]^T (l=1, 2, \dots, L)$ denotes the detected L SSPs. Then, some important parameters are defined as follows. P_l denotes different classes with N_l elements, that is, $P_l = \{U_1, U_2, \dots, U_{N_l}\}$. Define the center of P_l and distance between every two classes as

$$C_l = \frac{1}{N_l} \sum_{k=1}^{N_l} U_k \quad (32)$$

$$D_{l_1 l_2} = \|C_{l_1} - C_{l_2}\|_2 \quad (33)$$

The detailed algorithm for mixing matrix estimation is provided in Table 1.

Table 1. Mixing matrix estimation based on SSP and AHC

Input mixed signals $x_i(t), i=1, 2, \dots, M$.
1). Calculate $\mathbf{X}(t, f)$ using STFT in (4);
2). Detect SSPs using (19), and calculate the dataset $U_l = [R_l, I_l]^T, l=1, 2, \dots, L$ using (30);
3). Initialization for AHC: set th_d and th_N ; generate L classes $P_l = \{U_l\}, l=1, 2, \dots, L$; calculate distances using (32) and (33) between each class as $D_{l_1 l_2}, l_1, l_2 = 1, 2, \dots, L$;
4). while $\{find(D_{l_1 l_2} < th_d) > 0\}$, do: <ul style="list-style-type: none"> (i) Merge two classes with distance less than th_d, L' is the new number of classes; (ii) Calculate new $D_{l_1 l_2}, l_1, l_2 = 1, 2, \dots, L'$ using (32) and (33);
5). Keep classes $P_k (k=1, 2, \dots, N)$ with element number larger than th_N , calculate the final clustering center using (32) as $[Re(\hat{a}_{mn}), Im(\hat{a}_{mn})]^T$;
6). Build matrix $\hat{\mathbf{A}}$ using (31) with all the \hat{a}_{mn} ;
Output estimated mixing matrix $\hat{\mathbf{A}}$.

5. Source Signals Extraction

After estimating the mixing matrix, source separation algorithm is then used to extract the original signals. However, source signals can not be directly acquired

since we can not calculate the inverse matrix of mixing matrix in the underdetermined case. The UBSS model is ill-conditioned, which has an infinite number of possible solutions exactly. However, after exploiting the sparsity of sources and under some assumptions, it can be solved successfully. Aissa-El-Bey *et al.* [15] first introduced the subspace projection method for signal extraction supposing the number of signals appeared at any TF point was less than the number of sensors.

Suppose the number of active source at TF point (t_0, f_0) is K , and the signals are $s_{\alpha_1}(t), \dots, s_{\alpha_K}(t)$, respectively, $K < M$ and $\{\alpha_1, \dots, \alpha_K\} \subset \{1, 2, \dots, N\}$. Define the actual mixing matrix at (t_0, f_0) as $\mathbf{A}_K = [\mathbf{a}_{\alpha_1}, \dots, \mathbf{a}_{\alpha_K}]$, which corresponds to the source signal matrix $\tilde{\mathbf{S}}(t) = [s_{\alpha_1}(t), \dots, s_{\alpha_K}(t)]^T$. Then, temporarily ignoring the additive noise item, the mixing model after STFT in (5) at point (t_0, f_0) can be written as $\mathbf{X}(t_0, f_0) = \mathbf{A}_K \tilde{\mathbf{S}}(t_0, f_0)$.

The orthogonal projection matrix \mathbf{Q} for $s_{\alpha_1}(t), \dots, s_{\alpha_K}(t)$ projecting to noise space is calculated by

$$\mathbf{Q} = \mathbf{I} - \mathbf{A}_K (\mathbf{A}_K^H \mathbf{A}_K)^{-1} \mathbf{A}_K^H \quad (34)$$

$$\begin{cases} \mathbf{Q} \mathbf{a}_i = 0, & i \in \{\alpha_1, \dots, \alpha_K\} \\ \mathbf{Q} \mathbf{a}_i \neq 0, & i \in \{1, 2, \dots, N\}, i \notin \{\alpha_1, \dots, \alpha_K\} \end{cases} \quad (35)$$

In real environment, because of the noises and calculation errors, $\mathbf{Q} \mathbf{a}_i (i \in \{\alpha_1, \dots, \alpha_K\})$ are not strictly equal to 0, whereas very close to 0 instead. Considering that \mathbf{A} has been estimated, the column vectors of \mathbf{A}_K can be detected by minimizing

$$\{\beta_1, \dots, \beta_K\} = \arg \min_{\alpha_1, \dots, \alpha_K} \left\{ \|\mathbf{Q} \mathbf{X}(t_0, f_0)\|_{\mathbf{A}_K} \right\} \quad (36)$$

Then the real mixing matrix at TF point (t_0, f_0) can be formed as $\mathbf{A}_K = [\mathbf{a}_{\beta_1}, \dots, \mathbf{a}_{\beta_K}]$.

The STFT values of K sources at (t_0, f_0) are estimated by pseudo-inverse calculation

$$\tilde{\mathbf{S}}(t_0, f_0) = \mathbf{A}_K^{-1} \mathbf{X}(t_0, f_0) \quad (37)$$

At last, transform $\tilde{\mathbf{S}}(t_0, f_0)$ back to time domain using inverse STFT.

In general, the subspace-based method works well in most cases. However, the authors [21] have proved that it may “over-estimated”, which means the estimated mixing matrix at TF point (t_0, f_0) is larger than the actual ones. To solve the problem, a modified subspace method based on convex model is proposed by replace (36) with new objective function

$$\{\beta_1, \dots, \beta_r\} = \arg \min_{\alpha_1, \dots, \alpha_r} \left\{ \|\mathbf{Q}_r \mathbf{X}(t_0, f_0)\| + \gamma \cdot rank(\mathbf{A}_r) \mid \mathbf{A}_r \right\} \quad (38)$$

where γ is the balance factor, which is set to 0.4 in the simulation.

6. Simulation and Results

Estimation error E_A and average Signal to Interference Ratio (SIR) are adopted to measure the performance of mixing matrix estimation and signal extraction, respectively, as in (39).

$$\begin{cases} E_A = 10 \lg \left(\frac{1}{N} \|\mathbf{A} - \hat{\mathbf{A}}\|_F \right) & (dB) \\ \overline{SIR} = \frac{1}{N} \sum_{n=1}^N \left(10 \lg \frac{E[s_n(t)^2]}{E[s_n(t) - \hat{s}_n(t)]^2} \right) & (dB) \end{cases} \quad (39)$$

where N is the number of sources, \mathbf{A} and $\hat{\mathbf{A}}$ are the actual and estimation of mixing matrix. $\|\cdot\|_F$ calculates the Frobenius norm. In the following simulations, speech signals are used. The time window length of STFT is 32, overlapping with 0.75, FFT size is 256, and threshold for SSPs detecting is $\varepsilon = 0.001$. Set th_d and th_N to 0.3 and 90, respectively.

6.1 Performance of mixing matrix estimation

In this part, ULA with 2 sensors is adopted for simplicity. Three speech signals are taken as sources, and the DOAs are $\pi/8$, $\pi/4$ and $5\pi/8$, respectively, SNR is 25dB, the actual mixing matrix can be calculated using (6) as

$$\mathbf{A} = \begin{bmatrix} 1 & 1 & 1 \\ -0.9715+0.2369j & -0.6057+0.7957j & 0.3603-0.9328j \end{bmatrix} \quad (40)$$

Fig. 3 shows the scatter plot of detected SSPs. Theoretically, all the column vectors of mixing matrix should be different when setting different DOAs, and then clustering methods can be used to distinguish. In Fig. 3(a), all the SSPs locate on a unit circle, and obvious clustering characteristic is shown according to the SSPs' density of distribution. More points stay around the actual centers. Fig. 3(b) shows the SSPs' distribution after AHC, with most discrete points removed. Clearly, the proposed method can cluster the SSPs into three groups automatically, and the estimated centers almost have the same position as the actual centers. The estimated mixing matrix is $\hat{\mathbf{A}}$ as in (41), with E_A equals to -19.5298dB.

$$\hat{\mathbf{A}} = \begin{bmatrix} 1 & 1 & 1 \\ -0.9604+0.2520j & -0.5862+0.7811j & 0.3492-0.9261j \end{bmatrix} \quad (41)$$

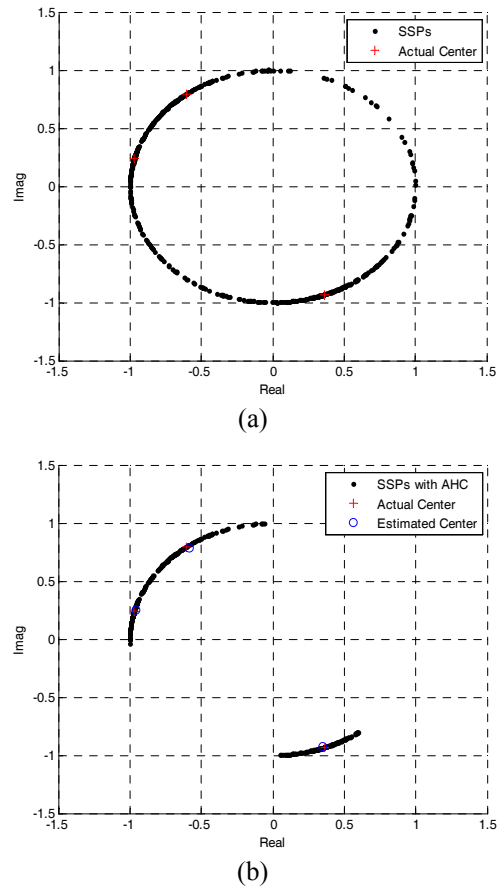


Fig. 3. Performance of AHC for SSPs: (a) SSPs before clustering; (b) SSPs after AHC. Blue and red points represent the estimated and actual centers, respectively.

Fig. 4 shows the performance of mixing matrix estimation with different SNRs, in comparison with the TIFROM [20] and the algorithms proposed by Reju [16] and Lihui [18]. When the SNR increases, the estimation errors of all the algorithms reduce. Lihui's method seems a little different, it achieves the best performance with large SNRs, but the degradation is obvious in low SNRs conditions. It is reasonable, because the scatter points caused by noises greatly reduced in high SNR conditions, which is more convenient for clustering. On the contrary, lower SNR results in more error scatter points, and K-means method ascribes every scatter point to one class, which undoubtedly increases the estimation error. The algorithm proposed in this paper has the most stable and reliable performance, with error changing less than 5dB when SNR decreases from 40dB to 15dB. It outperforms the ones developed by Reju and TIFROM, with about 6dB error reduction when SNR is lower than 30dB. Particularly, the proposed method plays a stricter rule in detecting the SSPs by exploiting prior information of mixing matrix, which makes it more effective in low SNR environments. The advantage is more obvious when the number of sources increases, and the optimal SNR range expands.

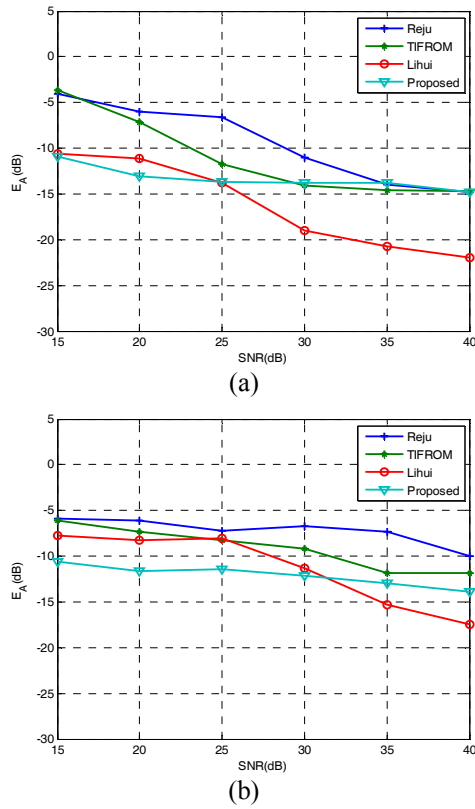


Fig. 4. Performances with different SNRs, $M=2$: (a) $N=3$; (b) $N=4$.

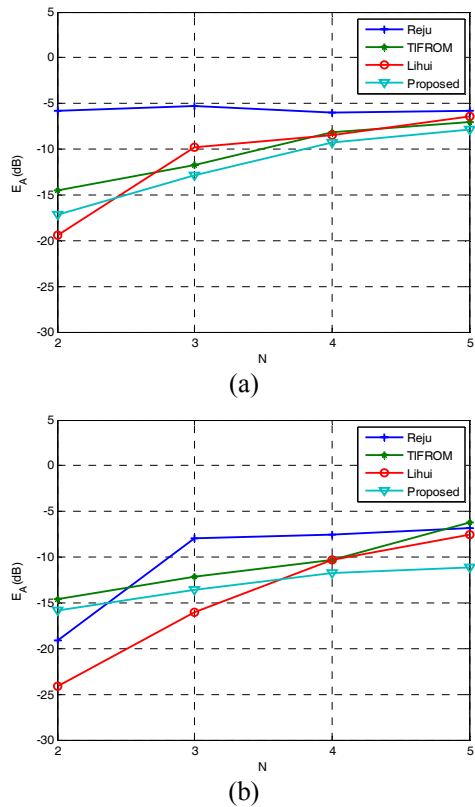


Fig. 5. Performances with different number of sources, $M=2$. (a) SNR=15dB; (b) SNR=25dB.

Fig. 5 describes the performance with different number of sources. The increase of source number results in the estimation error expansion. Lihui's method has the least estimation error when $N=2$, but when N increases, its performance drops together with the others. The proposed method has better performance when the source number continues to increase, and gets the lowest E_A at last. In

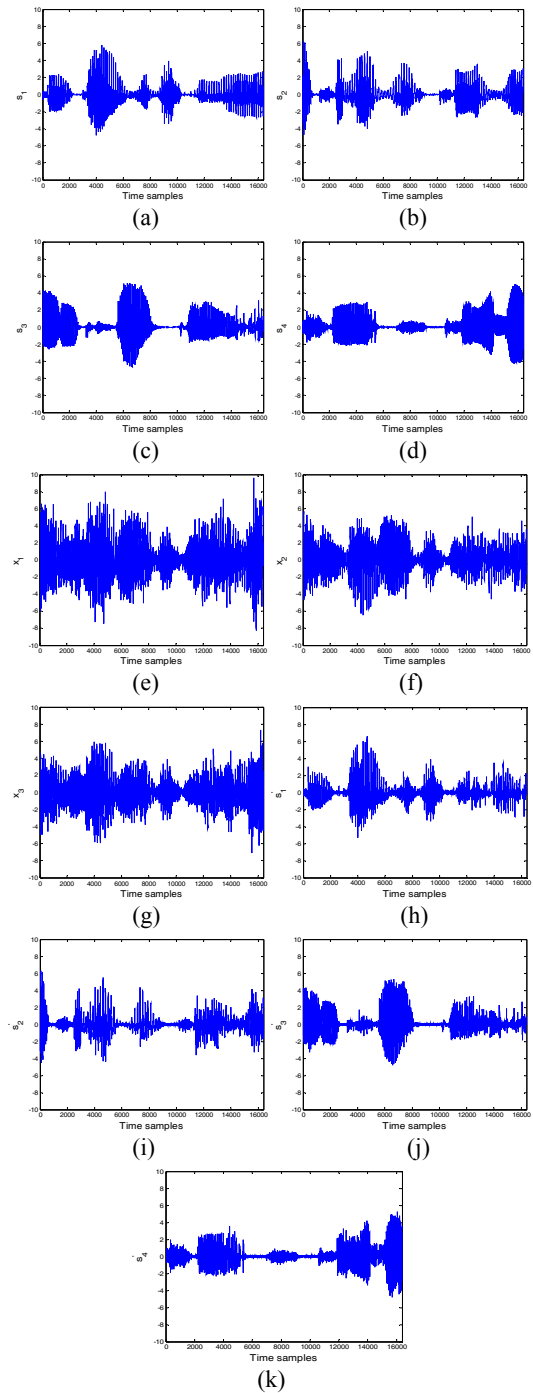


Fig. 6. Separation results with speech signals, SNR=25dB. (a)-(d) are three original signals; (e)-(g) are two observed mixtures; (h)-(k) represent the estimated source signals.

general, our method has significant advantages in low SNRs and more source environment, which make it more reliable in real applications.

6.2 Performance of source signal extraction

In this part, the number N of source is four and the number M of sensor is three. The DOAs of the signals are $\pi/11$, $3\pi/11$, $5\pi/11$ and $7\pi/11$, respectively. Fig. 6 shows that all the signals have been reconstructed successfully. The top four plots show the original signals; the middle three plots represent the three mixtures and the bottom four plots describe the estimation of sources by the proposed algorithm and subspace method. The recovered waves almost have the same shapes as the original ones.

Fig. 7 shows the SIR comparison of different methods. It is clear that the proposed method achieves highest SIR when SNR is less than 30dB. It is reasonable because our method estimate the mixing matrix more accurately in the low SNR environments, which is consistent with the result expressed in Fig. 4(b). In particular, the performance tends towards stability when the SNR is higher than 30dB, which almost have the same performance as Lihui's method.

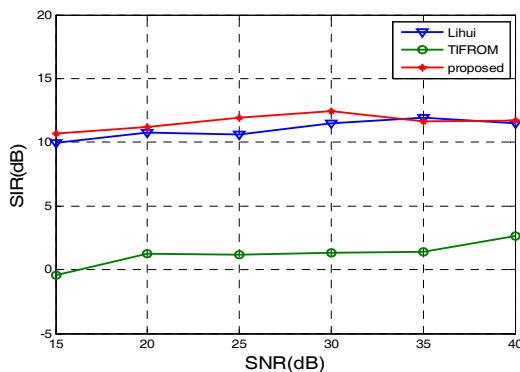


Fig. 7. SIR comparison with different SNRs.

7. Conclusion

This paper focuses on the underdetermined blind source separation from linear time-delayed mixtures, and a new complex-valued mixing matrix estimation method based on prior information exploitation is developed. The proposed method relaxes the sparsity of signals as long as there are SSPs in TF domain, no matter continuous or discrete, and does not need the number of sources before hand, which makes it more reliable in the real non-cooperative environment. Simulation results indicate that the proposed method can estimate the mixing matrix and extract the original sources with higher accuracy, and has significant advantages especially in low SNR and more sources conditions. In the future work, our attention should be focused on exploiting more effective methods for underdetermined blind source separation in the non-

disjoint cases. Another trend for blind source separation is to extract more prior information about true nature, morphology or structure of latent variables or sources, to get better performance.

Acknowledgements

The authors are grateful for the support from the National Science Foundation of China (51479159) and the Soft Science Project of China's Ministry of Transport (2013-322-811-470).

References

- [1] C. Jutten and J. Herault, "Blind Separation of Sources, Part I: An Adaptive Algorithm based on Neuro-mimetic Architecture," *Signal Processing*, vol. 24, no. 1, pp. 1-20, Jul. 1991.
- [2] A. Hyvarinen, J. Karhunen and E. Oja, *Independent Component Analysis*; New York: John Wiley&Sons, 2001, p.407-440.
- [3] P. Comon and C. Jutten, *Handbook of Blind Source Separation: Independent Component Analysis and Applications*; Academic Press of Elsevier, 2010.
- [4] P. Bofill and M. Zibulevsky, "Blind Separation of More Sources than Mixtures using Sparsity of Their Short-Time Fourier Transform," in *Proceedings of ICA 2000*, pp. 87-92, Jun. 2000.
- [5] P. Bofill and M. Zibulevsky, "Underdetermined Blind Source Separation using Sparse Representations," *Signal Processing*, vol. 81, no. 11, pp. 2353-2362, Nov. 2001.
- [6] P. Comon, "Blind Identification and Source Separation in 2×3 Underdetermined Mixtures," *IEEE Trans. Signal Process.*, vol. 52, no. 1, pp. 11-22, Jan. 2004.
- [7] N. Cho and C. C. J. Kuo, "Enhanced speech separation in room acoustic environments with selected binaural cues," *IEEE Trans. Consumer Electron.*, vol. 55, no. 4, pp. 2163-2171, Nov. 2009.
- [8] P. Georgiev, F. Theis and A. Cichocki, "Sparse Component Analysis and Blind Source Separation of Underdetermined Mixtures," *IEEE Trans. Neural Networks*, vol. 16, no. 4, pp. 992-996, Jul. 2005.
- [9] Y. Li, S. Amari, A. Cichocki, D.W.C. Ho and S. Xie, "Underdetermined Blind Source Separation Based on Sparse Representation," *IEEE Trans. Signal Process.*, vol. 54, no.2, pp. 423-437, Feb. 2006.
- [10] P. Comon and M. Rajih, "Blind Identification of Under-determined Mixtures Based on the Characteristic Function," *Signal Processing*, vol. 86, no. 9, pp. 2271-2281, Sep. 2006.
- [11] L. De Lathauwer, J. Castaing and J. Cardoso, "Four-order Cumulant-based Blind Identification of Under-determined Mixtures," *IEEE Trans. Signal Process.*,

vol. 55, no. 6, pp. 2965-2973, Jun. 2007.

- [12] X. Luciani, Andre L.F. de Almeida and P. Comon, "Blind Identification of Underdetermined Mixtures Based on the Characteristic Function: the complex case," *IEEE Trans. Signal Process.*, vol. 59, no. 2, pp. 540-553, Feb. 2011.
- [13] R. Saab, O. Yilmaz, M. J. McKeown, and R. Abugharbieh, "Underdetermined Anechoic Blind Source Separation via l_q -basis-pursuit with $q \leq 1$," *IEEE Trans. Signal Process.*, vol. 55, no. 8, pp. 4004-4017, Aug. 2007.
- [14] F.B. Lu, Z.T. Huang, and W.L. Jiang, "Underdetermined Blind Separation of Non-disjoint Signals in Time-frequency Domain Based on Matrix Diagonalization," *Signal Processing*, vol. 91, no. 7, pp. 1568-1577, Jul. 2011.
- [15] A. Aissa-El-Bey, N. Linh-Trung, K. Abed-Meraim, A. Belouchrani, and Y. Grenier, "Underdetermined Blind Separation of Nondisjoint Sources in the Time-frequency Domain," *IEEE Trans. Signal Process.*, vol. 55, no. 3, pp. 897-907, Mar. 2007.
- [16] V. G. Reju, S. N. Koh and I. Y. Soon, "Underdetermined Convulsive Blind Source Separation via Time-Frequency Masking," *IEEE Trans. Audio, Speech, and Language Process.*, vol. 18, no. 1, pp. 101-116, Jan. 2010.
- [17] V. G. Reju, S. N. Koh and I. Y. Soon, "An Algorithm for Mixing Estimation in Instantaneous Blind Source Separation," *Signal Processing*, vol. 89, no. 9, pp. 1762-1773, Sep. 2009.
- [18] H. Li, Y. H. Shen, J. G. Wang and X. S. Ren, "Estimation of the Complex-valued Mixing matrix by Single-source-points Detection with Less Sensors than Sources," *Trans. Emerging Tel. Tech.*, vol. 23, no. 2, pp. 137-147, Mar. 2012.
- [19] F. Abrard and Y. Deville, "A Time-frequency Blind Signal Separation Method Applicable to Underdetermined Mixtures of Dependent Sources," *Signal Processing*, vol. 85, no. 7, pp. 1389-1403, Jul. 2005.
- [20] M. Puigt and Y. Deville, "Time-frequency Ratio-based Blind Separation Methods for Attenuated and Time-delayed Sources," *Mechanical Systems and Signal Processing*, vol. 19, no. 6, pp. 1348-1379, Nov. 2005.
- [21] J. Yang, L. J. Zhang, K. W. Lu, and Q. N. Zhang, "Underdetermined Blind Source Separation Using Modified Subspace Method Based on Convex Model," in Proc. IEEE International Conference on Consumer Electronics, Las Vegas, USA, pp. 135-136, Jan. 2014.



Liangjun Zhang He received his BS degree in Electronic Information Engineering from Wuhan University of Technology, China, in 2009 and MS degree in Signal and Information Processing in 2012. He is currently a PHD candidate in Information and Communication Engineering in Wuhan University of Technology, Wuhan, China. His current research includes signal processing on image and passive radar, and blind signal processing.



Jie Yang She received her BS degree in Communication Engineering from XiDian University, China, in 1982 and MS degree in Computer and Automation from Wuhan Transportation University, China, in 1988. She received Ph.D. degree in Electronic Engineering from the Shanghai Jiao Tong University, China, in 1999. She has been a Professor of School of Electronics and Information at Wuhan University of Technology since 1999. Her current research interests include image processing, information hiding, cryptography and multimedia communication.



Zhiqiang Guo He received the PhD degree in electrical engineering from Wuhan University of Technology (WHUT), China, in 2010. He is an associate professor of communication department of WHUT. His research interests include the signal processing, image processing, machine learning and pattern recognition. He has published more than 20 technical papers in these areas.



Yanwei Zhou He received his BS degree from Gui Lin Air Force Academy in 2002, majored in the military command. He is currently studying for MS degree in Information and Communication Engineering in Wuhan University of Technology since 2012. His main research includes the Signal and Information Processing.