IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received January 14, 2019, accepted January 30, 2019, date of publication February 15, 2019, date of current version February 27, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2896878

Underdetermined Speech Blind Identification Based on Spectrum Correction and Phase Coherence Criterion

XIANGDONG HUANG[1], (Member, IEEE), JINGWEN XU[1], AND YU LIU[2], (Member, IEEE)
[1]School of Electrical and Information Engineering, Tianjin University, Tianjin 300072, China
[2]School of Microelectronics, Tianjin University, Tianjin 300072, China

Corresponding author: Yu Liu (liuyu@tju.edu.cn)

This work was supported by the National Natural Science Foundation of China, under Grant 61671012.

ABSTRACT The existing underdetermined speech blind identification (BI) algorithms can hardly possess both high recovery quality and high efficiency. This limitation may lie in the neglection of the phase extraction of speeches, which requires technical innovation of the phase-coherence identification. This paper proposes a BI scheme characterized by a combination of the ratio-interpolation-based spectrum corrector and a phase-coherence criterion (involving the operations of frequency merging, effective candidate pattern screening, and single-active-source (SAS) pattern recognition). Its high recovery quality is due to the combination that yields a set of SAS patterns with accurate harmonic parameters. Its high efficiency arises from two aspects: first, the phase-coherence criterion condenses the original patterns into a small quantity of SAS patterns; and second, an efficient density-based clustering algorithm is adopted to classify these SAS patterns. Essentially, the performance enhancement owns to the fact that the sources' phase information can be effectively extracted from the observations by means of the above technique combinations. Both the theoretical analysis and simulation verified that the proposed BI method outperforms the existing BI algorithms in recovery quality, efficiency, and anti-noise performance, which presents a vast potential in other harmonics-related BSS fields, such as mechanical vibration analysis, and channel estimation in communication.

INDEX TERMS Underdetermined blind identification, phase coherence, spectrum correction, harmonics.

I. INTRODUCTION

Blind source separation (BSS) is to recover the sources from the mixtures without the knowledge of the mixing system. BSS is widely applied in speech signal processing, digital communication, machinery diagnosis and so on [1]–[4]. For a linear and instantaneous mixing system with N sources and M mixtures, the BSS problem can be formulated as

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

where $\mathbf{x}(t) = [x_1(t), \ldots, x_M(t)]^{\mathrm{T}}$ is the mixture vector, $\mathbf{s}(t) = [s_1(t), \ldots, s_N(t)]^{\mathrm{T}}$ is the source vector, $\mathbf{A} = [a_1, \ldots, a_N]$ is an $M \times N$ unknown mixing matrix, and $\mathbf{n}(t) = [n_1(t), \ldots, n_M(t)]^{\mathrm{T}}$ is the additive noise vector. As is known, it is more difficult to tackle the underdetermined BSS problem (UBSS, requiring $M < N$) than the overdetermined or well-determined BSS problem (requiring $M \geq N$). Generally, the realization of BSS can be divided into 2 stages: blind identification (BI) and blind recovery (BR) [5]. Blind identification aims to obtain an accurate estimate of the mixing matix from the mixtures $\mathbf{x}(t)$, and blind recovery is to recover the underlying sources $\mathbf{s}(t)$ using the estimated result of the BI stage. Hence, BI stage imposes direct influence on the performance of the subsequent BR stage. This paper addresses the underdetermined BI problem. If the proposed BI method is combined with the existing BR algorithms addressed in [6]–[9] etc., a high-performance BSS system will be built up.

Blind identification of UBSS is mainly classified into two categories: the statistical property based method and the sparse representation based method [10]. They are based on different assumptions of sources.

The statistical method is based on the premise that the sources are independent and identically distributed and thus

The associate editor coordinating the review of this manuscript and approving it for publication was Adnan Shahid.

allows for the estimation of the mixing matrix in a probabilistic framework [11], like the independent component analysis (ICA) [12]–[14] does. However, ICA can only deal with the overdetermined case. Constructing high-order tensors is an effective approach to obtain feasibility for the under-determined case. For example, the fourth order statistics (FOOBI) algorithm achieves UBSS blind identification through simultaneous matrix diagonalization associated with fourth-order cumulant tensors [15]. Similar algorithms can also be found in [16]–[18].

The sparse representation method is based on the premise that each source exhibits a parsimonious distribution [19] in a transform domain (such as fast Fourier transform (FFT), short time Fourier transform (STFT), Wigner-Ville distribution (WVD)). To be specific, for any source, its energy tends to concentrate in several small regions in this domain. If the sparsity is sufficiently high, which was discussed in the sparse component analysis (SCA) in [20], then UBSS blind identification is feasible. For example, Bofill and Zibulevsky [21] adopted a potential function associated with the FFT representation of two mixtures to estimate the mixing matrix. However, since FFT is merely a 1-D transform which cannot provide a sufficiently sparse representation of speech signals, the blind identification can hardly reach a high accuracy. Compared to FFT, time-frequency analysis (TFA) provides a sparser 2-D representation in which the time information smeared by FFT is preserved. Hence, a series of TFA tools (such as STFT, WVD, Cohen distributions) have been introduced to improve the accuracy of the estimated mixing matrix [10], [18], [22]–[27].

In particular, due to the fact that a speech signal is characterized with sparsity in time-frequency (TF) domain, the sparse representation based method becomes the mainstream of blind identification of speech UBSS systems. Specifically, TFA based blind identification can be further classified into two categories: non time-frequency masking method and time-frequency masking method. However, they seem to suffer from a high complexity or inaccuracy.

For the non time-frequency masking methods (such as Line orientation separation technique (LOST) algorithm [19], nonlinear projection column masking (NPCM) algorithm [28]), it is necessary to construct an objective function considering all the points covering the entire TF plane (each TF position corresponds to an $M \times 1$ mixture vector). By means of some optimization techniques (such as particle swarming optimization [28], expectation-maximization (EM) optimization [19]) on these vectors, the columns of the matrix $\mathbf{A}$ can be determined one by one. However, these enormous patterns lead the optimization to a high computational complexity. Moreover, improper initialization may also prolong the optimization's convergence time.

For the TF masking methods (such as degenerate unmixing estimate technique (DUET) algorithm [22], [23], nondisjoint sources based method [6], UBSS-FAS method [10]), blind identification is divided into two steps: Firstly, pick out those time-frequency points relevant to one active source to comprise a single-active source domain(SSD) [10]; Secondly, cluster those SSD-related patterns to acquire the final estimate of the matrix A. Hence, the accuracy and efficiency of blind identification depend on both the SSD recognition criterion and the clustering technique.

On one hand, the accuracy of the TF masking methods [6], [10], [22], [23] is affected by the neglection of the phase information of a speech, which is hidden beneath multiple TF points in an SSD. As is known, a speech contains several harmonic-like voiced-sound components [29], [30], and each component can be exactly described by 3 parameters: frequency, amplitude and phase. Hence, this paper attempts to take advantage of this ignored phase information to improve the BI accuracy. As elaborated later, affected by TFA tools' inherent spectral leakage, a harmonic component will evolve into multiple time-frequency points. Hence, it is necessary to further condense the region of an SSD [31]. Typically, some novel TFA tools such as synchrosqueezed windowed Fourier transform (SWFT) [31] and synchrosqueezed wavelet transform (SWT) [32] are able to concentrate these TF SSDs. However, as [33] pointed out, these two synchrosqueezed transforms do not seem to imply better resolution properties and thus cannot provide significant performance improvement.

On the other hand, the efficiency of the TF masking methods is affected by two facets. First, these methods have to tackle numerous patterns, since each SSD includes multiple TF points. Secondly, the existing blind identification schemes generally adopt the k-means clustering technique, which actually lacks high efficiency. If this clustering method is replaced by a better one, the complexity of blind identification will surely be further reduced.

Considering the above analysis, this paper puts forward a high-efficiency UBSS blind identification scheme based on spectrum correction and phase coherence criterion. By means of spectrum correction, a lot of consecutive TF points in an SSD can be represented by a parametric triple (frequency, amplitude and phase). Moreover, the phase-coherence criterion can effectively pick out a small number of SAS patterns from these SSD-related triples, while a large quantity of patterns that may degrade the BI accuracy can be expelled. Also, instead of the k-means clustering technique, the data-density based clustering technique further enhances the BI efficiency. Numerical results will verify that the proposed scheme concurrently possesses high efficiency, high accuracy, strong robustness to noise.

The remaining of this paper is organized as follows. In Section II, we propose a harmonic parametric triple based TF representation and introduce a spectrum corrector to achieve this representation. In Section III, we propose a phase coherence criterion to refine the single-active-source patterns from these harmonic parametric triples. In Section IV, we employ an efficient data density based clustering method to acquire the source number estimate and the final BI result and give a summary on the proposed BI procedure. In Section V, we present several simulations to compare the proposed BI

scheme with the existing BI algorithms. Finally, we come to some conclusions in VI.

II. HARMONIC PARAMETRIC TRIPLE BASED REPRESENTATION OF TF POINTS

A. HARMONIC PARAMETRIC TRIPLE BASED REPRESENTATION OF TF POINTS

Suppose that the time-frequency tool is ideal. Considering the system model formulated in (1), if at a moment t_0, only the source s_n is active and it includes a single component with the instantaneous frequency ω_0, the amplitude d_0 and the phase ϕ_0, then the ideal time-frequency analysis of M mixtures is

$$\begin{bmatrix} X_1(t_0,\omega) \\ X_2(t_0,\omega) \\ \vdots \\ X_M(t_0,\omega) \end{bmatrix} = \begin{bmatrix} a_{1,n} \\ a_{2,n} \\ \vdots \\ a_{M,n} \end{bmatrix} d_0 e^{\mathrm{j}\phi_0}\delta(\omega-\omega_0), \tag{2}$$

where $n = 1,\ldots,N$ and '$\delta(\cdot)$' is the dirac function. Thus, the mixture vector $[X_1(t_0,\omega_0),\ldots,X_M(t_0,\omega_0)]^T$ is parallel to $\mathbf{a}_n$, i.e., this single point (t_0,ω_0) in an ideal 2-D time-frequency plane is sufficient to achieve the accurate estimate of $\mathbf{a}_n$, since amplitude uncertainty is allowed in BSS.

However, there always exists deviation between the ideal time-frequency representation in (2) and the commonly-used time-frequency analysis tools. As a result, an ideal TF point will spread over several frequency positions in a single-active source domain. The reasons are as follows.

Firstly, limited by the uncertainty principle [34], i.e., the time resolution and the frequency resolution of any time-frequency analysis tool cannot be high simultaneously. Besides, affected by noise, it is impossible that the information of a column of the mixing matrix is entirely concentrated in a single time-frequency point, like (2).

Secondly, a mixture contains abundant components and the interferences among these components inevitably expand the TF covering region, also.

Thirdly, different types of time-frequency analysis tools have distinct deficiencies. For example, the WVD is likely to incur cross-term interferences. Although these cross-term interferences can be removed by means of smoothing measures in ambiguity domain [35], it is at the cost of expanding auto-term TF regions. As for the STFT, it is merely a translated and windowed Fourier transform, i.e.,

$$X_m(t,\omega) = \int_{-\infty}^{+\infty} x_m(\tau)h(\tau-t)e^{-j\omega\tau}d\tau, \tag{3}$$

thus when STFT is practically implemented by translated FFT, its inherent spectral leakage and picket fence effect will degrade the spectrum quality.

This paper aims to improve the performance of STFT-based BI. Thus it is necessary to investigate the STFT time-frequency distribution influenced by FFT spectral leakage effect.

Example I: Consider a signal composed up of 3 harmonics as $x(t) = \cos(2\pi f_1 t + 10°) + 3\cos(2\pi f_2 t + 60°) + 2\cos(2\pi f_3 t + 90°)$, $f_1 = 152$Hz, $f_2 = 2f_1 = 304$Hz, $f_3 = 3f_1 = 456$Hz. Specify the sampling rate $f_s = 16$kHz and implement L-point FFT ($L = 512$ and thus the frequency unit $\Delta f = f_s/L = 31.25$Hz). Hence, their fractional frequency offsets $\delta_1 = f_1/\Delta f - [f_1/\Delta f] = -0.1360$, $\delta_2 = f_2/\Delta f - [f_2/\Delta f] = -0.2720$, $\delta_3 = f_3/\Delta f - [f_3/\Delta f] = -0.4080$, ('$[\cdot]$' refers to the rounding operation). The FFT amplitude spectrum $|X(k)|$ is illustrated in Fig. 1.

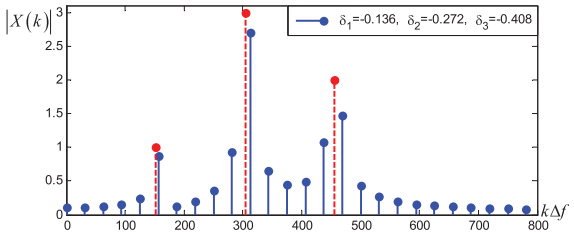

FIGURE 1. Spectrum of $x(t)$, $\delta_1 = -0.136$, $\delta_2 = -0.272$, $\delta_3 = -0.408$.

Fig. 1 shows that, due to the effect of spectral leakage, each component f_i evolves into a cluster of FFT bins emerging around f_i. Clearly, the larger the frequency offset $|\delta_i|$ is, the heavier the spectral leakage exhibits. Further, since each cluster of spectral bins only corresponds to a single component, there always exists a stronger unobservable spectral bin (marked in red dotted line), which is ideally located at the centrobaric position of this cluster. Therefore, it is desired to recover these hidden ideal bins. In other words, each cluster of FFT bins illustrated in Fig. 1 can be refinedly represented by its ideal harmonic parametric triple (f_i, d_i, ϕ_i).

As [29] and [30] pointed out, a speech signal mainly contains harmonics-like voiced sounds and noise-like unvoiced sounds. In particular, voiced sounds occupy most of the speech energy. Hence, voiced sounds can be easily distinguished from large-amplitude clusters of STFT spectrogram, as Fig. 2 depicts (the sampling rate f_s=16kHz and 50% hanning window overlapping is considered).

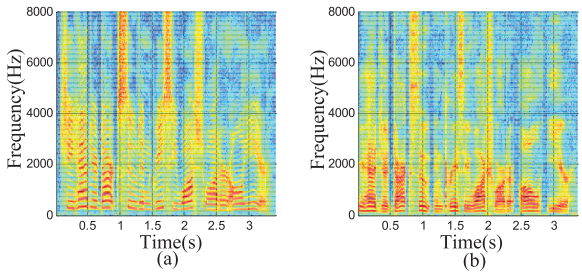

FIGURE 2. STFT spectrograms of two speech signals: (a) female; (b) male.

Fig. 2 presents the STFT spectrograms of two speech signals from a female and a male. One can notice that, in both sub-figures, there are a lot of conspicuous and almost parallel flat stripes over some small time intervals. These flat stripes actually refer to slowly varying spectrum of harmonics-like voiced sounds. In fact, at some fixed moment of active voiced sound, the vertical STFT spectrogram within a small frequency band resembles the FFT spectrum illustrated in Fig. 1,

in which the thickness of a flat stripe depends on the number of the leaked FFT bins.

Hence, if each centrobaric position (i.e., the parametric triple (f_i, d_i, ϕ_i) of the ideal harmonic illustrated in Fig. 1) along the frequency axis of all these flat stripes can be estimated, STFT representation will be greatly simplified, which helps to develop a more efficient blind identification scheme. This simplification can be effectively realized by means of the technique of spectrum correction.

B. SPECTRUM CORRECTION BASED HARMONICS EXTRACTION

The objective of spectrum correction is to accurately estimate all the harmonic parametric triples of each mixture, which helps to further recover ideal parametric triples (f_i, d_i, ϕ_i).

In the selection of spectrum corrector, two factors should be considered. One is the windowing operation, which helps to suppress the spectral leakage; The other is that the numbers of consumed spectral bins should be as small as possible, since harmonics exhibit a dense distribution in a speech spectrum (as Fig. 2 shows).

Here we employ a ratio-interpolation based spectrum corrector [36], since it can accurately provide 3 corrected parameters only consuming two hanning-windowed spectral bins (the peak bin and its adjacent sub-peak bin).

Given the hanning-windowed STFT spectrograms of M mixtures $X_m(t_0, k\Delta\omega)$, $m = 1, \ldots, M$, $\Delta\omega = 2\pi/L$ (denoting $X_m(t_0, k)$ for simplicity), at some moment $t = t_0$, the spectrum correction consists of the following steps:

Step 1 Collect all the large-amplitude peak indices k_p of $X_m(t_0, k)$. For each index k_p, calculate the amplitude ratio v_p between $X_m(t_0, k_p)$ and its sub-peak neighbor, i.e.,

$$v_p = \frac{|X_m(t_0, k_p)|}{\max\{|X_m(t_0, k_p - 1)|, |X_m(t_0, k_p + 1)|\}}. \tag{4}$$

Further, a variable u_p can be calculated as

$$u_p = (2 - v_p)/(1 + v_p). \tag{5}$$

Step 2 Estimate the aforementioned frequency offset δ_p as

$$\hat{\delta}_p = \begin{cases} u_p, & \text{if } |X_m(t_0, k_p + 1)| \\ & \quad > |X_m(t_0, k_p - 1)| \\ -u_p, & \text{else,} \end{cases} \tag{6}$$

then, the frequency estimate is $\hat{\omega}_{m,p} = (k_p + \hat{\delta}_p)2\pi/L$.

Step 3 Acquire the amplitude estimate $\hat{d}_{m,p}$ and phase estimate $\hat{\phi}_{m,p}$ as

$$\hat{d}_{m,p} = 2\pi\hat{\delta}_p(1 - \hat{\delta}_p^2)|X_m(t_0, k_p)|/\sin(\pi\hat{\delta}_p). \tag{7}$$

$$\hat{\phi}_{m,p} = \text{ang}[X_m(t_0, k_p)] - \pi\hat{\delta}_p(L - 1)/L, \tag{8}$$

where ang(·) refers to the angle operation.

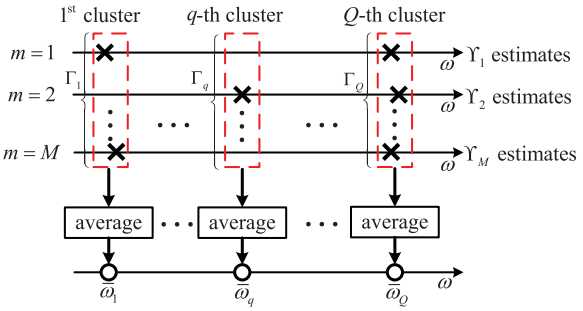

FIGURE 3. Distribution of M mixtures' frequency estimates and their merged frequencies (note that: $\Gamma_q \le M, q = 1, \cdots, Q$).

After spectrum correction, the original enormous STFT points are substituted with many harmonic parametric triples $(\hat{\omega}_{m,p}, \hat{d}_{m,p}, \hat{\phi}_{m,p})$. If these parametric triples can be refined, the efficiency of blind identification will be further enhanced. The following refinement procedure based on phase coherence criterion (involving the operations of frequency merging, candidate pattern screening and single-active-source (SAS) pattern recognition) can achieve this goal.

III. PHASE COHERENCE CRITERION BASED PATTERN REFINEMENT

A. REFINEMENT PROCEDURE

1) FREQUENCY MERGING

It should be noted that, due to the noise effect and interference effect, even for a same component of a single active source, its frequency estimates $\{\hat{\omega}_{m,p}\}$ resulting from spectrum correction still exhibit tiny difference between different mixtures. Hence, these frequency estimates should be merged.

Suppose the numbers of $\{\hat{\omega}_{1,p}\},\ldots,\{\hat{\omega}_{M,p}\}$ are $\Upsilon_1,\ldots,\Upsilon_M$. If we put all these frequency estimates together and sort them in an ascending order, the aforementioned frequency estimates of tiny difference tend to form a cluster (marked with a dashed box, as Fig. 3 illustrates). Assume altogether Q clusters are formed ($Q \ge \Upsilon_m, m = 1, \ldots, M$). Without loss of generality, denote the q-th ($q = 1, \ldots, Q$) cluster as $\{\tilde{\omega}_{q,p'}, p' = 1, \ldots, \Gamma_q\}$ (i.e., Γ_q elements are included, $\Gamma_q \le M$). Then, we have $\Upsilon_1 + \cdots + \Upsilon_M = \Gamma_1 + \cdots + \Gamma_Q$.

For the q-th cluster, since its Γ_q frequency estimates are of tiny difference, they can be merged by their average (as Fig. 3 depicts)), i.e.,

$$\bar{\omega}_q = \frac{1}{\Gamma_q}\sum_{p'=1}^{\Gamma_q}\tilde{\omega}_{q,p'}. \tag{9}$$

2) EFFECTIVE CANDIDATE PATTERN SCREENING

Note that, although the Q merged frequencies $\bar{\omega}_1 \sim \bar{\omega}_Q$ are acquired from all mixtures' frequency estimates, however, as Fig. 3 depicts, for an individual merged frequency estimate $\bar{\omega}_q$, it is very likely that $\bar{\omega}_q$ is not included by every mixture.

In particular, in terms of the BSS model (1), as long as the mixing matrix $\mathbf{A}$ does not contain zero element, any source

component is bound to be incorporated by every mixture. Hence, it is necessary to judge whether an individual merged frequency $\bar{\omega}_q$ belongs to all M mixtures (or only small deviation exists). In other words, if the quantity Γ_q of the q-th cluster $\{\tilde{\omega}_{q,p'}, p' = 1, \ldots, \Gamma_q\}$ equals M, $\bar{\omega}_q$ is regarded as an effective candidate component.

Assume altogether $\bar{Q}$ components are effective candidates. For mathematical simplicity, it does not matter to denote them as

$$\mathbf{z}_{\bar{q}} = \begin{bmatrix} \hat{d}_{1,\bar{q}} e^{j\hat{\phi}_{1,\bar{q}}} \\ \vdots \\ \hat{d}_{m,\bar{q}} e^{j\hat{\phi}_{m,\bar{q}}} \\ \vdots \\ \hat{d}_{M,\bar{q}} e^{j\hat{\phi}_{M,\bar{q}}} \end{bmatrix}, \quad \bar{q} = 1, \ldots, \bar{Q}. \tag{10}$$

3) RECOGNIZING SINGLE-ACTIVE-SOURCE PATTERNS

As [10] pointed out, only those mixture patterns corresponding to a single active source make contribution to the estimation of the mixing matrix $\mathbf{A}$. The reason lies in: if only the n-th source is active at some moment, the BSS model (1) can be simplified as

$$\begin{aligned} \mathbf{x} &= \mathbf{A}[0, \cdots, s_n, \cdots, 0]^T \\ &= s_n \mathbf{a}_n, \end{aligned} \tag{11}$$

then, the mixture $\mathbf{x}$ is theoretically parallel to the n-th column $\mathbf{a}_n$ of $\mathbf{A}$.

Furthermore, suppose that $\mathbf{z}_{\bar{q}}$ is a single-active-source pattern. Considering that the matrix $\mathbf{A}$ is real-valued, one can conclude from (2) that the phases of $X_1(t_0, \omega), \ldots, X_M(t_0, \omega)$ (i.e., the M phases included in $\mathbf{z}_{\bar{q}}$) originate from the same phase ϕ_0 of some component of the n-th source. Hence, the M angles $\hat{\phi}_{1,\bar{q}}, \ldots, \hat{\phi}_{M,\bar{q}}$ included in $\mathbf{z}_{\bar{q}}$ have to uniformly point to a same direction (or allowing any two angles are in opposite directions, since amplitude uncertainty is allowed in BSS). In other words, the projection between any two angle-related unit vectors should be close to 1, i.e.,

$$|\langle e^{j\hat{\phi}_{r,\bar{q}}}, e^{j\hat{\phi}_{l,\bar{q}}} \rangle| \ \to 1, \tag{12}$$

where $1 \le r, l \le M$, $r \ne l$ and '$\langle\cdot\rangle$' represents inner product operation.

Further, if all the combination cases of projection are taken into account (altogether $C_M^2 = M(M-1)/2$ cases), one can calculate the average projection $\bar{P}_{\bar{q}}$ as

$$\bar{P}_{\bar{q}} = \frac{1}{C_M^2} \sum_{r,l,r\ne l} |\langle e^{j\hat{\phi}_{r,\bar{q}}}, e^{j\hat{\phi}_{l,\bar{q}}} \rangle|. \tag{13}$$

Thus, given a threshold ζ, $\mathbf{z}_{\bar{q}}$ can be regarded as an SAS pattern if the following inequality holds

$$|\bar{P}_{\bar{q}} - 1| < \zeta. \tag{14}$$

Assume that altogether $\tilde{Q}$ patterns are SAS patterns denoted as

$$\mathbf{z}_{\tilde{q}} = \begin{bmatrix} \hat{d}_{1,\tilde{q}} e^{j\hat{\phi}_{1,\tilde{q}}} \\ \vdots \\ \hat{d}_{m,\tilde{q}} e^{j\hat{\phi}_{m,\tilde{q}}} \\ \vdots \\ \hat{d}_{M,\tilde{q}} e^{j\hat{\phi}_{M,\tilde{q}}} \end{bmatrix}, \quad \tilde{q} = 1, \ldots, \tilde{Q}. \tag{15}$$

Remind that the phases $\hat{\phi}_{1,\tilde{q}}, \ldots, \hat{\phi}_{M,\tilde{q}}$ are of consistence. Therefore, to construct real-valued SAS patterns, $\hat{\phi}_{1,\tilde{q}}$ can be utilized as a reference and thus a real-valued version of $z_{\tilde{q}}$ is

$$\mathbf{z}_{\tilde{q}} = \begin{bmatrix} d_{1,\tilde{q}} & z_{2,\tilde{q}} & \ldots & z_{m,\tilde{q}} & \cdots & z_{M,\tilde{q}} \end{bmatrix}^T,$$

where

$$z_{m,\tilde{q}} = \begin{cases} d_{m,\tilde{q}}, & \text{if } \cos(\hat{\phi}_{1,\tilde{q}} - \hat{\phi}_{m,\tilde{q}}) \to 1 \\ -d_{m,\tilde{q}}, & \text{if } \cos(\hat{\phi}_{1,\tilde{q}} - \hat{\phi}_{m,\tilde{q}}) \to -1 \end{cases} \tag{16}$$

Further, in the consideration that amplitude uncertainty is allowed in BSS, $\mathbf{z}_{\tilde{q}}$ needs to be normalized.

To achieve a complete and accurate estimate of the mixing matrix $\mathbf{A}$, it is necessary to implement the above refinement procedure frame by frame. As a result, an SAS pattern set $\Omega = \{\mathbf{z}_i, i = 1, \ldots, P\}$ will be generated. Finally, the mixing matrix $\mathbf{A}$ can be estimated by clustering these P SAS patterns, as will be elaborated in the Section IV.

B. PERFORMANCE ANALYSIS OF PHASE COHERENCE CRITERION BASED PATTERN REFINEMENT

The above procedure of phase coherence criterion based pattern refinement possesses high efficiency, high noise robustness, high anti-interference.

The high efficiency lies in the following aspects: Firstly, frequency merging only requires sorting operation and averaging operation; Secondly, constructing effective candidate patterns only needs to collect those merged clusters with M frequency estimates; Thirdly, it can be inferred from (13)∼(14) that the recognition criterion of SAS patterns is also very simple, only involving several inner product and averaging operations. Lastly, the above 3 steps gradually reduce the quantity of patterns (i.e., $Q \to \bar{Q} \to \tilde{Q}$).

The high noise robustness lies in: In the process of constructing effective candidate patterns, it is almost impossible that the noise can yield a conspicuous component occupied by all M mixtures.

The high anti-interference lies in the follows: Compared with large-amplitude components, small-amplitude components tend to suffer from more serious interference. As a result, this interference inevitably brings phase distortion on these small-amplitude components. On the contrary, for a large-amplitude component, it suffers from small phase distortion and thus its M observation phases tend to exhibit the coherence (i.e., satisfying (13) and (14)). In this way, its

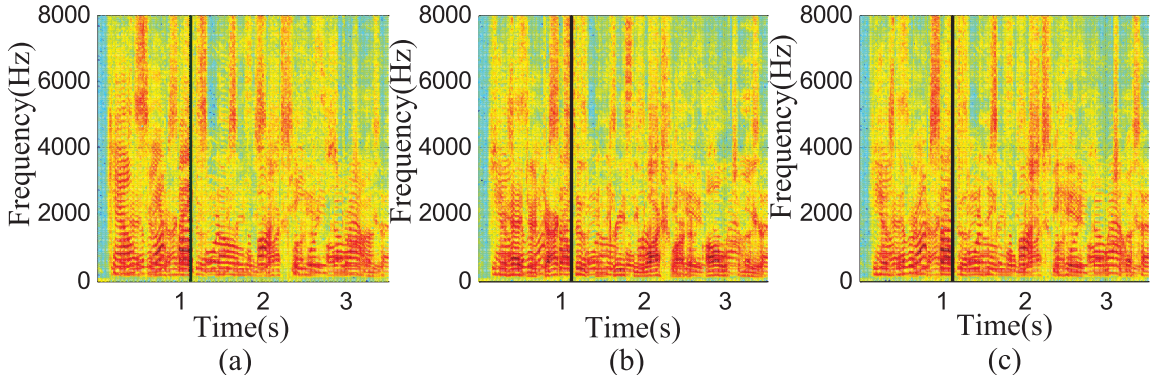

FIGURE 4. The STFT spectrograms of 3 mixtures (hanning windowed, 50% frame overlapping). (a) mixture 1. (b) mixture 2. (c) mixture 3.

corrected amplitudes of multiple mixtures are sure to uniformly approximate the ideal amplitude, which also ensures a high-accuracy BI performance.

C. EXAMPLE OF PATTERN REFINEMENT

We present an example to demonstrate the above procedure of pattern refinement.

Example II: Consider a speech mixing system with $N = 4$ sources and $M = 3$ mixtures (speeches are chosen from TIMIT). The normalized mixing matrix $\mathbf{A}$ is specified as

$$\mathbf{A} = \begin{bmatrix} 0.9356 & 0.6354 & 0.2813 & 0.2858 \\ 0.2433 & 0.2102 & 0.4571 & 0.9147 \\ 0.2557 & 0.7430 & 0.8438 & 0.2858 \end{bmatrix}. \quad (17)$$

The sampling rate of these speeches is $f_s = 16$kHz and the window length of STFT is $L = 512$ (hanning windowing, 50% frame overlapping, 219 time frames altogether). Fig. 4 illustrates the STFT spectrograms of the 3 mixtures. For each mixture, Fig. 5 presents its vertical section (i.e., the FFT spectrum) along the frequency axis at the 71-th time frame ($t_0 = 1.14$s), which corresponds to the black line in Fig. 4. Tab. 1 lists the results of spectrum correction for $\bar{Q} = 5$ effective candidate patterns (including $\tilde{Q} = 2$ SAS patterns, the threshold ζ is set as 0.0008).

From Fig. 4, one can find that, due to spectral leakage and interference between components, the STFT spectrogram of each mixture appears chaos and it is impossible to recognize any source. Meanwhile, there are a lot of peaks in the 3 FFT spectra in Fig. 5. Following the procedure of the ratio-interpolation based spectrum corrector summarized in Section II-B and the frequency merging operation addressed in Section III-A.1, one can acquire $Q = 66$ merged frequencies. Further, following the screening procedure addressed in Section III-A.2, among these 66 components, only $\bar{Q} = 5$ components are eligible to construct the effective candidate patterns, whose merged frequencies $\bar{\omega}_{\bar{q}}$, corrected amplitudes $\hat{d}_{1,\bar{q}} \sim \hat{d}_{3,\bar{q}}$ and corrected phases $\hat{\phi}_{1,\bar{q}} \sim \hat{\phi}_{3,\bar{q}}$, $\bar{q} = 1, \ldots, \bar{Q}$, are listed in Tab. 1.

Next, following the procedure of recognizing single-active-source patterns addressed in Section III-A.3, only the

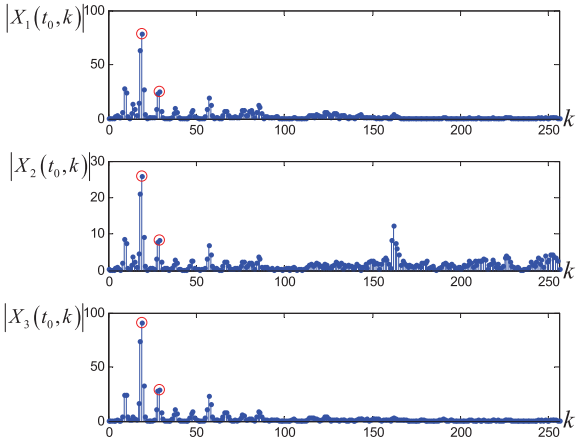

FIGURE 5. The FFT spectra of 3 mixtures at the 71-th time frame.

former 2 components are selected as the reliable SAS patterns (in gray background), since their maximum phase differences are about 1° while others range between 5° ∼ 10°. Finally, these two components' corrected results listed in Tab. 1 are combined to generate the following 2 vectors:

$$\mathbf{z}_1 = \begin{bmatrix} 84.5548 \\ 28.1076 \\ 97.8573 \end{bmatrix}, \quad \mathbf{z}_2 = \begin{bmatrix} 28.3142 \\ 9.4284 \\ 33.6466 \end{bmatrix}, \quad (18)$$

and their normalized versions are

$$\mathbf{z}_1 = \begin{bmatrix} 0.6389 \\ 0.2124 \\ 0.7394 \end{bmatrix}, \quad \mathbf{z}_2 = \begin{bmatrix} 0.6296 \\ 0.2096 \\ 0.7481 \end{bmatrix}. \quad (19)$$

Clearly, these two refined patterns $\mathbf{z}_1, \mathbf{z}_2$ are very close to the 2-nd column of matrix $\mathbf{A}$ in (17).

Furthermore, as Tab. 1 lists, one can find that the above refined two components are exactly the strongest two components among $\bar{Q} = 5$ candidates. The reasons are as follows: For any component, not only it suffers from other components' interferences but also it will exert interferences over other components. Thus, only those strong components tend to be less influenced by others. As a result, M mixtures'

TABLE 1. The effective candidate patterns and their corresponding parameters.

	$\bar{q}$	1	2	3	4	5
Merged frequencies	$\bar{\omega}_{\bar{q}}(\Delta\omega)$	18.6578	28.5477	38.1990	57.1885	66.5404
Mixture 1	$\hat{d}_{1,\bar{q}}$	84.5548	28.3142	9.9833	20.4351	8.5568
	$\hat{\phi}_{1,\bar{q}}(^\circ)$	74.4978	168.1535	44.6779	177.9293	334.3651
Mixture 2	$\hat{d}_{2,\bar{q}}$	28.1076	9.4284	3.1117	6.9339	2.9455
	$\hat{\phi}_{2,\bar{q}}(^\circ)$	74.8359	168.6693	49.6109	179.9742	323.3826
Mixture 3	$\hat{d}_{3,\bar{q}}$	97.8573	33.6466	11.4650	24.2463	9.7226
	$\hat{\phi}_{3,\bar{q}}(^\circ)$	73.4826	169.5666	49.7044	175.7009	342.2555

corrected phases of some strong component stemming from a single active source are likely to be consistent (i.e., they uniformly point to a same direction or opposite directions). Therefore, the recognition criterion involved in (10)∼(14) is suitable to detect this phase consistency. In other words, this recognition criterion is of high robustness to interferences.

What's more, combining Fig. 4 with Fig. 5, we can find that, in each STFT spectrogram, these $\bar{Q} = 5$ candidate components are individually located on different flat stripes, which actually correspond to voiced sounds. As aforementioned, since the spectrum corrector is well suitable for extracting harmonic information, the refined SAS patterns arise from harmonic-like voiced sounds rather than noise-like unvoiced sounds.

IV. DATA DENSITY BASED CLUSTERING FOR SAS PATTERNS

A. CONSIDERATIONS IN THE SELECTION OF THE CLUSTERING ALGORITHM

To effectively cluster the aforementioned SAS pattern set $\Omega = \{\mathbf{z}_i, i = 1, \ldots, P\}$, the following three points should be considered.

Firstly, the clustering algorithm should have the ability to determine the number N of categories, which is generally unknown in the UBSS blind identification problem.

Secondly, the clustering algorithm should concurrently possess high efficiency and high accuracy.

Thirdly, the clustering algorithm should be insensitive to the initialization or other specified parameters.

Thus, we employ the data density based clustering algorithm recently proposed in [37] rather than the mainstream k-means algorithm [3], [6], [10], since the former is superior to the latter in the above 3 aspects [37].

B. DATA DENSITY BASED CLUSTERING AND MIXING MATRIX ESTIMATION

The data density based clustering algorithm takes full advantages of the following two characteristics of sample distributions [37]:

a) Cluster centers are surrounded by neighbors with lower local density;

b) Cluster centers are at a relatively large distance from any points with a higher local density.

Hence, this clustering algorithm is able to determine the source number N of a BSS system. The procedure is as follows.

Step 1: Calculate the distances $d_{i,j}$ of all pattern pairs, i.e.,

$$d_{i,j} = \left\| z_i - z_j \right\|, \quad 1 \le i, j \le P, \ i \ne j; \tag{20}$$

Step 2: Calculate each pattern point's local data density ρ_i, i.e.,

$$\rho_i = \sum_j \chi(d_{i,j} - d_c), \tag{21}$$

where d_c is a given cutoff distance of neighbourhood and '$\chi(\cdot)$' is a threshold function as

$$\chi(t) = \begin{cases} 1, & t \le 0 \\ 0, & t > 0; \end{cases} \tag{22}$$

Step 3: Sort $\rho_1, \ldots, \rho_P$ in a descending order and thus yield a subscript set $\{q_i, i = 1, \ldots, P\}$ satisfying $\rho_{q_1} \ge \rho_{q_2} \ge \ldots \ge \rho_{q_P}$;

Step 4: Calculate each pattern point's characteristic distance δ_i defined as

$$\delta_i = \begin{cases} \max\limits_{j \ge 2}(d_{q_i,q_j}), & i = 1; \\ \min\limits_{q_j j<i}(d_{q_i,q_j}), & i > 1. \end{cases} \tag{23}$$

Step 5: Calculate the products $\gamma_i = \rho_i \delta_i, i = 1, \ldots, P$, from which the number N of sources arises intuitively. In other words, there exists a subscript set $\Theta = \{\dot{q}_1, \ldots, \dot{q}_N\}$ satisfying $\gamma_{\dot{q}_1} \ge \gamma_{\dot{q}_2} \ge \ldots \ge \gamma_{\dot{q}_N} \gg \gamma_j, j \notin \Theta$. Hence, the N cluster centers are $z_{\dot{q}_1}, z_{\dot{q}_2}, \ldots, z_{\dot{q}_N}$.

Step 6: Based on these cluster centers $z_{\dot{q}_n}, n = 1, \ldots, N$, classify all the P patterns into N categories. Then, take the average pattern of each category as the estimate of a column of the matrix $\mathbf{A}$.

In this procedure, only a threshold parameter d_c needs to be specified. Meanwhile, as [37] pointed out, this clustering algorithm is insensitive to the choice of d_c.

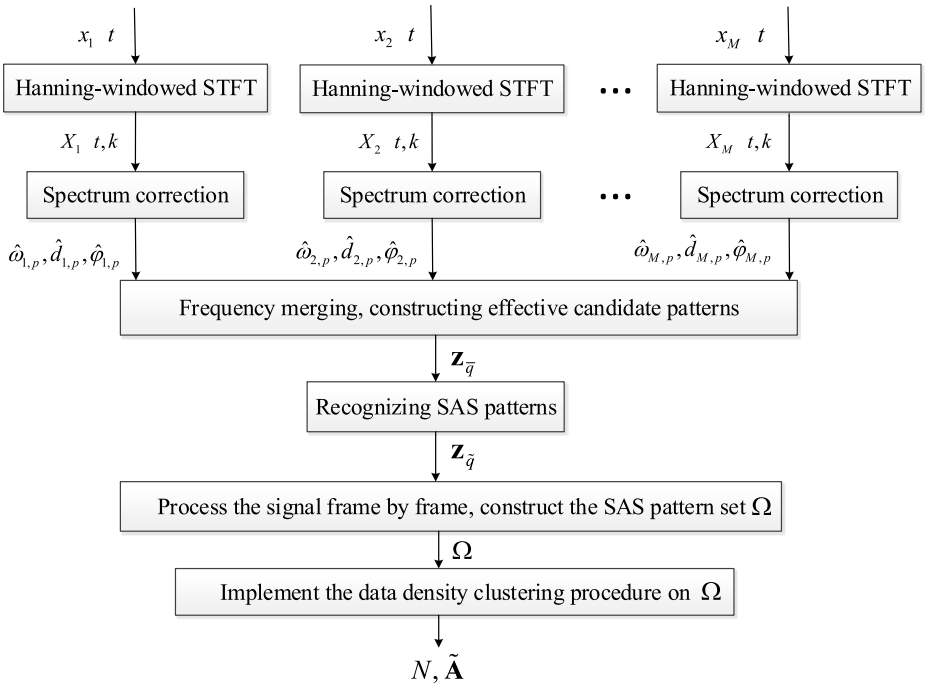

FIGURE 6. The flow diagram of the proposed blind identification algorithm).

To sum up, the proposed blind identification algorithm is illustrated in Fig. 6.

V. SIMULATIONS

A. CONTRIBUTION OF THE SAS PATTERN RECOGNITION

In this section, both qualitative demonstration and quantitative analysis are presented to investigate the contribution of the SAS pattern recognition to the whole BI scheme summarized in Fig. 6(the parameter d_c is set as 0.05).

Consider the same speech signals involved in *Example 2* and the same parameter settings. Besides, to verify the SAS pattern recognition criterion's robustness, the mixing matrix $\mathbf{A}$ in *Example 2* is substituted with a matrix containing negative elements, i.e.,

$$\mathbf{A} = \begin{bmatrix} 0.9356 & 0.6354 & 0.2813 & 0.2858 \\ -0.2433 & -0.2102 & 0.4571 & 0.9147 \\ 0.2557 & -0.7430 & -0.8438 & 0.2858 \end{bmatrix}. \tag{24}$$

Following our proposed BI procedure in Fig. 6, we can obtain the final matrix estimate $\tilde{\mathbf{A}}$ as

$$\tilde{\mathbf{A}} = \begin{bmatrix} 0.2832 & 0.2839 & 0.9349 & 0.6344 \\ 0.4594 & 0.9146 & -0.2459 & -0.2106 \\ -0.8416 & 0.2877 & 0.2556 & -0.7435 \end{bmatrix}. \tag{25}$$

1) QUALITATIVE DEMONSTRATION

The scatter diagrams of the patterns directly from STFT, candidate screening and SAS recognition are illustrated in Fig. 7(a), Fig. 7(b) and Fig. 7(c), respectively (all the patterns are normalized).

From the scatter diagram directly from STFT in Fig. 7(a), one can see that there are numerous pattern points (altogether $(512/2+1)\times 219 = 56283$ ones) disorderly spread on the spherical surface that it is impossible to observe any apparent clusters, let alone their corresponding cluster centers. After candidate screening, $\bar{Q} = 911$ patterns are preserved and illustrated in Fig. 7(b), which exhibits a denser distribution (besides several outliers) than Fig. 7(a) does. Further, among these $\bar{Q} = 911$ candidates, only $P = 190$ SAS patterns satisfying the phase coherence criterion are illustrated in Fig. 7(c), in which $N = 4$ clusters can be distinctly identified. Moreover, the SAS pattern recognition also makes these $N = 4$ cluster centers intuitively identified in the density-distance plane (see red stars in Fig. 8).

Hence, the above illustrations show that, the redundancy of time-frequency representation is greatly reduced and the pattern distribution apparently gets denser, indicating that the proposed phase coherence based refinement procedure greatly improves the efficiency and the accuracy of BI.

2) QUANTITATIVE ANALYSIS

Here, to find out which part of our proposed phase coherence criterion based pattern refinement algorithm is responsible for the final result, we made a minor revision on our algorithm. Specifically, we bypass the core SAS pattern recognition (i.e., the data density based clustering directly follows the effective candidate pattern screening). Then, two quantitative performance indexes (recovery SNR of the matrix estimate

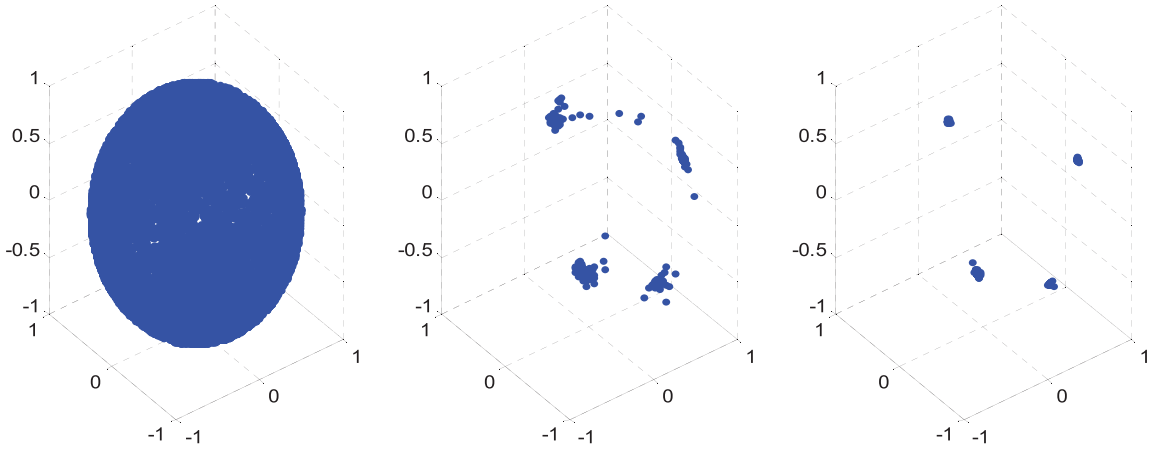

FIGURE 7. (a) Scatter plot directly from STFT, (b) Scatter plot from candidate screening, (c) Scatter plot from SAS recognition (219 time frames, FFT size: $L = 512$).

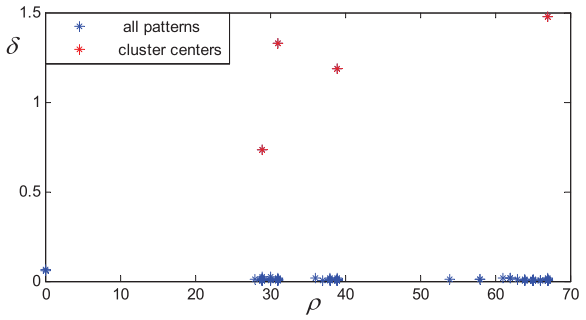

FIGURE 8. Distribution of the densities and distances of all SAS patterns ($d_c = 0.05$).

and averaged dispersion measure of N categories [38]) will be compared between our proposed algorithm and this SAS pattern recognition excluded version.

To calculate the recovery SNR of the matrix estimate $\tilde{\mathbf{A}}$, since order uncertainty is allowed in BSS (as (24) and (25) exhibit), it is necessary to adjust $\tilde{\mathbf{A}}$ by taking into account all the column re-arrangements, i.e.,

$$\hat{\mathbf{A}} = \tilde{\mathbf{A}}\tilde{\mathbf{M}}, \tag{26}$$

where

$$\tilde{\mathbf{M}} = \arg\min_{\mathbf{M}\in\mathcal{P}} \left\|\mathbf{A} - \tilde{\mathbf{A}}\mathbf{M}\right\|.$$

In (26), $\mathcal{P}$ is the set of all invertible real $N \times N$ matrices where only one entry is nonzero in each column [28]. Hence, the *recovery signal to noise ratio* (SNR) can be defined as

$$\text{SNR} = 10\lg\left[\sum_{i=1}^{M}\sum_{j=1}^{N} a_{i,j}^2 \Big/ \sum_{i=1}^{M}\sum_{j=1}^{N}\left(a_{i,j} - \hat{a}_{i,j}\right)^2\right]. \tag{27}$$

where $a_{i,j}$ and $\hat{a}_{i,j}$ are the elements of real mixing matrix $\mathbf{A}$ and the adjusted matrix estimate $\hat{\mathbf{A}}$, respectively. The recovery SNR values of the proposed BI scheme and its SAS pattern recognition excluded version are listed in Tab.2.

TABLE 2. Performance indices of two BI schemes.

BI scheme	Recovery SNR	$\bar{D}$
proposed BI	51.23	0.0167
SAS recognition excluded BI	37.57	0.0783

To evaluate the general dispersion of N clusters, it is necessary to calculate an individual dispersion measure D_n as

$$D_n = \sqrt{\frac{1}{R_n}\sum_{i=1}^{R_n}\|\mathbf{z}_i - \mathbf{c}_n\|^2}, \quad n = 1, \ldots, N \tag{28}$$

where $\mathbf{c}_n$ is the center of the n-th cluster $\{\mathbf{z}_i, i = 1, \ldots, R_n\}$, and R_n is the element number of the cluster. Then, the average dispersion measure $\bar{D}$, i.e.,

$$\bar{D} = \frac{1}{N}\sum_{n=1}^{N} D_n, \tag{29}$$

reflects the clustering effect. In other words, the smaller the $\bar{D}$ is, the denser the pattern distribution is, thereby leading to a higher-quality BI. The specific values of $\bar{D}$ of these two BI schemes are also listed in Tab.2.

From Tab. 2, one can find that if the SAS pattern recognition is excluded, the recovery SNR deteriorates with 13.66dB and the average dispersion measure $\bar{D}$ increases with about 3.7 times. Therefore, the SAS pattern recognition plays an essential role in the whole BI scheme.

B. TRACING THE ORIGINS OF SAS PATTERNS

Recall that once an SAS pattern $\mathbf{z}_p$ is recognized, its time-frame index and DFT-bin index can be recorded. Moreover, since the individual $\mathbf{z}_p$ only origins from a single source, its source index c_p can be traced by the following formula

$$c_p = \arg\min_{n=1,\ldots,N} \left|\mathbf{z}_p - \hat{\mathbf{a}}_n\right|. \tag{30}$$

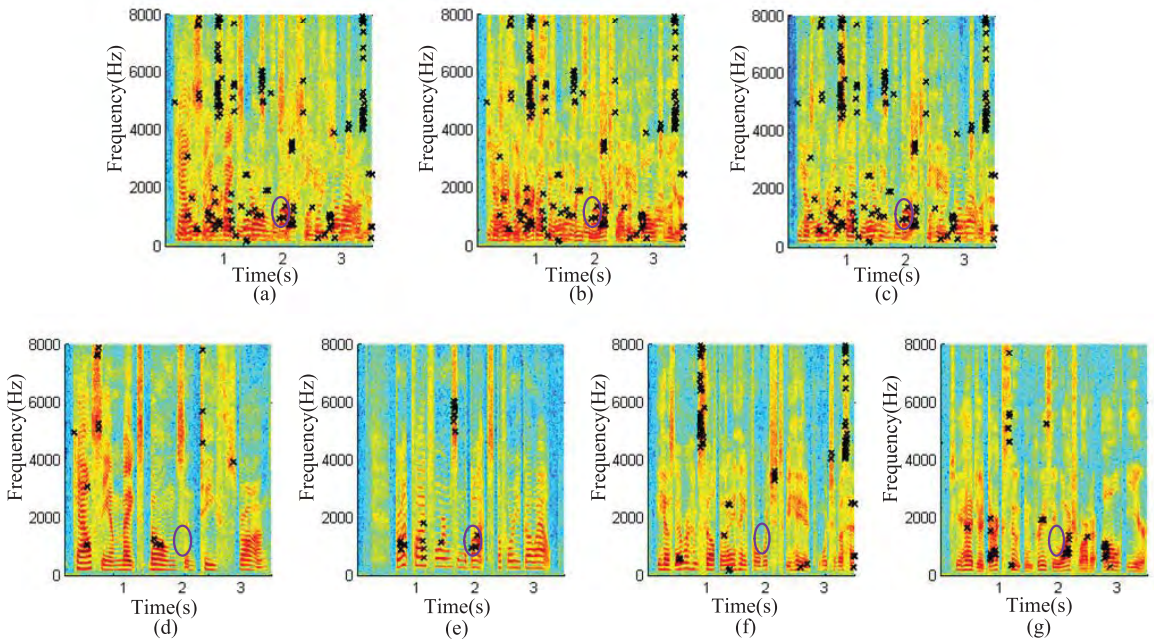

FIGURE 9. SAS distribution: (a) mixture 1; (b) mixture 2; (c) mixture 3; (d) source 1; (e) source 2; (f) source 3; (g) source 4.

Hence, one can mark $\mathbf{z}_p$ on the TF planes of all mixtures and the c_p-th source. Further, repeating this operation on all P SAS patterns yields the SAS distribution diagrams illustrated in Fig. 9(a)-(g) (marked in black stars), which exhibit the following distribution features:

a) As Fig. 9(a)-(c) depict, all these SAS pattern points fall in large-amplitude areas (i.e., those regions dyed in deep red), which verified that SAS patterns correspond to a strong components and are less likely to be disturbed by other components.

b) As Fig. 9(d)-(g) depict, one can notice that, for any individual pattern $\mathbf{z}_p$, there exists only a single TF plane (i.e., the c_p-th source it belongs to) in which its surrounding region is in deep red. For example, for any pattern point bounded by the specified purple ellipses, only the surrounding region of the 2-nd source TF plane appears active (i.e., in deep color) while the surrounding regions of the other 3 TF planes appear inactive (i.e., in light color). This implies that these P single-active-source patterns are correctly searched out.

c) As Fig. 9(d)-(g) depict, most of the strong SAS pattern points are located on flat red stripes in low-frequency regions, which correspond to the harmonic-like voiced-sound components. This reflects that the proposed phase coherence criterion based pattern refinement is in accordance with the speech generation mechanism.

C. ROBUSTNESS TO THE RANDOMNESS OF THE MIXING MATRIX

In this section, the proposed algorithm is compared with 3 other UBSS blind identification algorithms including TIFROM [39], LOST [19] and NPCM [28]. Unlike the case of Section V-A only involving a fixed mixing matrix $\mathbf{A}$, this section aims to investigate the robustness to different mixing matrices. Specifically, we conducted 100 comparison trials and each trial deals with a randomly generated normalized matrix $\mathbf{A}$. In consideration that, the performance of any BI algorithm might deteriorate when matrix $\mathbf{A}$ is in ill condition, each $\mathbf{A}$ is required to satisfy the following conditions

a) The intersection angle between any two columns of $\mathbf{A}$ is larger than 10°;

b) The absolute value of any element of $\mathbf{A}$ is greater than 0.1.

To enhance the proposed BI scheme's robustness to the randomness of the mixing matrix, we present a minor improvement measure on the data density based classifier. Specifically, following Step 6 of the clustering procedure addressed in Section IV-B, one can further compress the cluster space using another specified distance threshold $\tilde{d}_c$ ($\tilde{d}_c < d_c$, here $\tilde{d}_c$ is set to be $0.65d_c$). Thus, around the center $\mathbf{z}_{\dot{q}_n}$ of an individual cluster Θ_n, only a portion of neighbour patterns $\mathbf{z}_j$ satisfying

$$d_{\dot{q}_n,j} = \left\|\mathbf{z}_{\dot{q}_n} - \mathbf{z}_j\right\| < \tilde{d}_c, \quad \mathbf{z}_j \in \Theta_n,\ n = 1, \ldots, N, \tag{31}$$

need to be collected. Thus, the average of $\mathbf{z}_j$ in Θ_n can be treated as an accurate estimate of a column $\mathbf{a}_n$ of the matrix $\mathbf{A}$. In this way, $\mathbf{a}_1, \ldots, \mathbf{a}_N$ will be estimated one by one.

Among these 4 BI algorithms, it seems a bit difficult and inconvenient for TIFROM and LOST to estimate the matrix $\mathbf{A}$ without knowing the source number *a priori* [28]. Hence, the correct source number N was directly specified in both of them. In contrast, since both NPCM algorithm and the proposed algorithm do not need any knowledge of the source

number, N was estimated by themselves. Other parameter settings are as follows: The mixtures are transformed into STFT domain (time frame length $L = 512$, hanning windowed, 50% frame overlapping). For the NPCM algorithm, the parameter ρ is set to be $\rho = 10^4$, and the population size and the generation number of PSO are respectively configured as $P = 30$, $K = 30$. Besides, the termination parameter $\varepsilon_T = 0.4$ and the masking threshold $\alpha_0 = 12°$. For LOST algorithm, the Laplacian density parameter β is initialized as 1. For the TIFROM, the length of a TF window is set as 10. Their recovery SNRs are listed in Tab. 3.

From Tab. 3, one can see that the proposed BI algorithm has highest recovery quality (SNR=51.14dB), and the NPCM algorithm takes the second place (SNR=41.58dB), compared to that the LOST algorithm and the TIFROM exhibit relatively poor quality (SNR=23.65dB and SNR=21.39dB, respectively).

D. COMPARISON OF COMPUTATIONAL COMPLEXITY

Tab. 3 also lists the elapsed time (Equipped with Intel Core i3 CPU 2.0GHz, 4GB RAM) consumed by different BI algorithms. Specifically, the average elapsed time of the TIFROM (0.70s) is comparable to the proposed algorithm (0.85s), both of which are much shorter than that of NPCM (15.22s) and the LOST (8.38s). This experimental result can be well explicated by the following computational complexity analysis.

For the TIFROM, although the computational complexity seems to be a bit lower than our proposed algorithm, it also suffers nearly 30 dB of recovery SNR deterioration compared to our proposed method. This loss arises from its block-wise (each block includes multiple time slots) SAS identification operation, which only adopts coarse statistical analysis to judge whether those observed TF points among a chosen time slot origin from a single source or not. To achieve a high BI efficiency, the time slot has to be very long, which in turn inevitably increases roughness of BI performance.

For the LOST algorithm, the computational load is much heavier than our proposed algorithm, which is mainly caused by 2 aspects. On one hand, all the TF points across the STFT plane are utilized to search for the line orientation close to a column of the mixing matrix, whereas only a small portion of TF points (i.e., those TF points inside an SAS region) are used by the proposed algorithm; On the other hand, the LOST algorithm is implemented in an iterative way consisting of two alternative steps (i.e., the Expectation step and the Maximization step), whereas the proposed algorithm works in a forward way.

For the NPCM algorithm, the computational load is heavier than the other 3 BI algorithms, which arises from its evolution-based optimization mode for searching the columns of the mixing matrix. Specifically, this scheme introduces the PSO algorithm to solve a multivariate optimization problem. To achieve a high BI accuracy, both the population size and the number of evolved generations have to be very large, which inevitably increases the computational load.

E. COMPARISON OF ANTI-NOISE PERFORMANCE

To explore the anti-noise performance of these 4 BI algorithms, two types of noise disturbances (additive gaussian white noise and pink noise) were considered. For each noisy case, we conducted 100 Monte Carlo trials (in each trial, the matrix $\mathbf{A}$ was still randomly generated) under different SNR levels. The averaged recovery SNR curves of the gaussian noisy case and the pink noisy case are plotted in Fig. 10 and Fig. 11, respectively.

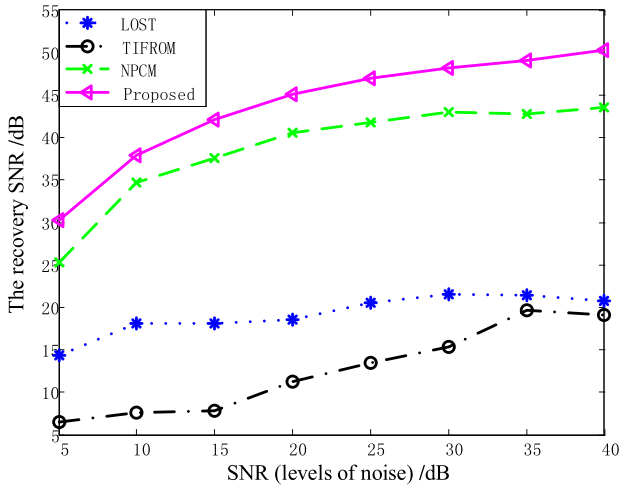

FIGURE 10. Recovery SNR curves across different Gaussian white noise levels.

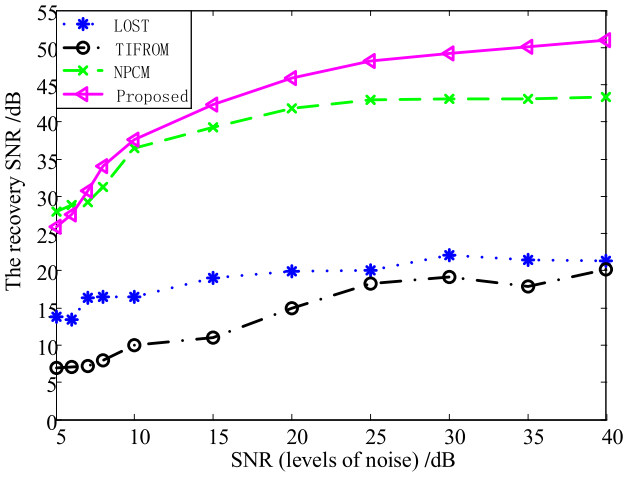

FIGURE 11. Recovery SNR curves across different pink noise levels.

Note that, in both Fig. 10 and Fig. 11, the SNR at the horizontal coordinate and the SNR at the vertical coordinate have different physical meanings. The SNR at the horizontal coordinate refers to the signal-to-noise ratio at each observation rather than at each source. In other words, in order to investigate the anti-noise robustness of the proposed BI algorithm across different noisy cases, we purposely contaminate any mixture $x_m(t)$ formulated in (1) with some noise at the level of SNR. The SNR at the vertical coordinate reflects

TABLE 3. Comparison of robustness to alternating mixing matrices.

Algorithms	TIFROM	LOST	NPCM	Proposed
Average recovery SNR (dB)	21.39	23.65	41.58	51.14
Average elapsed time (s)	0.70	8.38	15.22	0.85

the recovery quality of the mixing matrix, which is explicitly defined in (27).

1) GAUSSIAN WHITE NOISE

From Fig. 10, one can see that, across the entire SNR range of the mixtures, the recovery SNR curve of proposed BI algorithm (marked in '$\triangleleft$') is higher than the NPCM curve (marked in '$\times$'), the LOST curve (marked in '$*$') and the TIFROM curve (marked in '$\bigcirc$'). This is due to the fact that, the other 3 algorithms directly use the STFT results to construct patterns, whereas our proposed algorithm use the results of spectrum correction of large-amplitude STFT results. Further, as aforementioned, any corrected amplitude of each recognized SAS pattern actually stands for an approximation to the amplitude of an ideal peak spectral bin (as the red dotted line in Fig. 1 illustrates), which is stronger than other surrounding leaked spectral bins. Hence, the proposed BI algorithm exhibits higher anti-noise performance than others.

2) PINK NOISE

Pink noise, whose power spectrum's shape resembles the inverse proportional function $1/f$ [40] (i.e., the energy mainly concentrates on the low frequency regions), is one of the most common behavior of natural noise. Therefore, the recovery SNR curve versus pink noisy levels of a BI scheme reflects the robustness to the disturbance arising from the natural world. From Fig.11, two conclusions can be drawn.

a) Similar to the case of Gaussian noise, across most of SNR levels (especially the high SNR levels), the recovery SNR curve of the proposed BI algorithm is higher than other 3 BI schemes.

b) Only when SNR$\leq$6dB, the recovery SNR curve of the proposed BI scheme is a little lower than that of the NPCM. The reason lies in the following: As Fig. 9(d)-(g) depicts, large-amplitude SAS pattern points are mainly distributed over the low-frequency regions of N source STFT spectrograms, which overlaps the bandwidth occupied by most of the pink noise energy. Hence, this bandwidth overlapping may exert a tiny influence to our proposed BI scheme when the pink-noise disturbance gets serious.

VI. CONCLUSION

This paper proposes a UBSS blind identification method incorporating the ratio-interpolation based spectrum correction and the phase coherence criterion based pattern refinement. The main characteristics of this work lie in 3 aspects: First, this paper discovers that the performance degradation of TF masking based BI schemes arises from the deviation between the ideal time-frequency representation and the commonly-used TF analysis tool, thus a spectrum corrector is employed to calibrate this deviation. Second, a phase coherence based criterion is proposed to refine the single-active-source patterns, which greatly enhances the efficiency and the accuracy. Finally, the data density based clustering is introduced to estimate the source number and enhance the recovery quality. Both theoretical analysis and simulation verified that, the proposed BI scheme outperforms other BI algorithms in accuracy, anti-noise performance and efficiency.

Due to the fact that the spectrum corrector incorporated in the proposed BI scheme does well in harmonic extraction, this work actually possesses a vast potential in non-speech BSS fields (such as mechanical vibration analysis, channel estimation in communication, accoustic signal separation), since harmonic components also occupy overwhelming majority of the energy of signals in these fields.

REFERENCES

[1] T. Xu, W. Wang, and W. Dai, "Sparse coding with adaptive dictionary learning for underdetermined blind speech separation," *Speech Commun.*, vol. 55, no. 3, pp. 432–450, 2013.

[2] G. Bao, Z. Ye, X. Xu, and Y. Zhou, "A compressed sensing approach to blind separation of speech mixture based on a two-layer sparsity model," *IEEE Trans. Audio, Speech, Language Process.*, vol. 21, no. 5, pp. 899–906, May 2013.

[3] Z.-C. Sha, Z.-T. Huang, Y.-Y. Zhou, and F.-H. Wang, "Frequency-hopping signals sorting based on underdetermined blind source separation," *IET Commun.*, vol. 7, no. 14, pp. 1456–1464, Sep. 2013.

[4] Z. Li, X. Yan, Z. Tian, C. Yuan, Z. Peng, and L. Li, "Blind vibration component separation and nonlinear feature extraction applied to the nonstationary vibration signals for the gearbox multi-fault diagnosis," *Measurement*, vol. 46, no. 1, pp. 259–271, 2013.

[5] B. Liu, V. G. Reju, and A. W. H. Khong, "A linear source recovery method for underdetermined mixtures of uncorrelated AR-model signals without sparseness," *IEEE Trans. Signal Process.*, vol. 62, no. 19, pp. 4947–4958, Oct. 2014.

[6] A. Aïssa-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.

[7] R. Saab, Ö. Yılmaz, M. J. McKeown, and R. Abugharbieh, "Underdetermined anechoic blind source separation via ℓ^q-basis-pursuit with $q < 1$," *IEEE Trans. Signal Process.*, vol. 55, no. 8, pp. 4004–4017, Aug. 2007.

[8] Z. He *et al.*, "Improved FOCUSS method with conjugate gradient iterations," *IEEE Trans. Signal Process.*, vol. 57, no. 1, pp. 399–404, Jan. 2009.

[9] G. Mohimani, M. Babaie-Zadeh, and C. Jutten, "A fast approach for overcomplete sparse decomposition based on smoothed ℓ^0 norm," *IEEE Trans. Signal Process.*, vol. 57, no. 1, pp. 289–301, Jan. 2009.

[10] S. Xie, L. Yang, J.-M. Yang, G. Zhou, and Y. Xiang, "Time-frequency approach to underdetermined blind source separation," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 23, no. 2, pp. 306–316, Feb. 2012.

[11] M. S. Lewicki and T. J. Sejnowski, "Learning overcomplete representations," *Neural Comput.*, vol. 12, no. 2, pp. 337–365, 2000.

[12] A. Hyvärinen and E. Oja, "A fast fixed-point algorithm for independent component analysis," *Neural Comput.*, vol. 9, no. 7, pp. 1483–1492, Jul. 1997.
[13] H. Shen, M. Kleinsteuber, and K. Hüper, "Local convergence analysis of FastICA and related algorithms," *IEEE Trans. Neural Netw.*, vol. 19, no. 6, pp. 1022–1032, Jun. 2008.
[14] L. Gu and G. Liu, "Application of ICA-R algorithm in weak signal extraction," *Comput. Sci.*, vol. 43, no. 3, pp. 122–126, 2016.
[15] L. De Lathauwer, J. Castaing, and J.-F. Cardoso, "Fourth-order cumulant-based blind identification of underdetermined mixtures," *IEEE Trans. Signal Process.*, vol. 55, no. 6, pp. 2965–2973, Jun. 2007.
[16] L. De Lathauwer and J. Castaing, "Blind identification of underdetermined mixtures by simultaneous matrix diagonalization," *IEEE Trans. Signal Process.*, vol. 56, no. 3, pp. 1096–1105, Mar. 2008.
[17] A. Karfoul, L. Albera, and G. Birot, "Blind underdetermined mixture identification by joint canonical decomposition of HO cumulants," *IEEE Trans. Signal Process.*, vol. 58, no. 2, pp. 638–649, Feb. 2010.
[18] S. Ge, J. Han, and M. Han, "Nonnegative mixture for underdetermined blind source separation based on a tensor algorithm," *Circuits Syst. Signal Process.*, vol. 34, no. 9, pp. 2935–2950, 2015.
[19] P. D. O'Grady and B. A. Pearlmutter, "The LOST algorithm: Finding lines and separating speech mixtures," *EURASIP J. Adv. Signal Process.*, vol. 2008, no. 1, pp. 1–17, 2008, Art. no. 784296.
[20] P. Georgiev, F. Theis, and A. Cichocki, "Sparse component analysis and blind source separation of underdetermined mixtures," *IEEE Trans. Neural Netw.*, vol. 16, no. 4, pp. 992–996, Jul. 2005.
[21] P. Bofill and M. Zibulevsky, "Underdetermined blind source separation using sparse representations," *Signal Process.*, vol. 81, no. 11, pp. 2353–2362, 2001.
[22] A. Jourjine, S. Rickard, and Ö. Yılmaz, "Blind separation of disjoint orthogonal signals: Demixing N sources from 2 mixtures," in *Proc. ICASSP*, Jun. 2000, pp. 2985–2988.
[23] Ö. Yılmaz and S. Rickard, "Blind separation of speech mixtures via time-frequency masking," *IEEE Trans. Signal Process.*, vol. 52, no. 7, pp. 1830–1847, Jul. 2004.
[24] J. Sun, Y. Li, J. Wen, and S. Yan, "Novel mixing matrix estimation approach in underdetermined blind source separation," *Neurocomputing*, vol. 173, pp. 623–632, Jan. 2016.
[25] Z. Liang, C. Xun, X. Ji, and Z. J. Wang, "Underdetermined joint blind source separation of multiple datasets," *IEEE Access*, vol. 5, no. 5, pp. 7474–7487, 2017.
[26] C. Liu, Y. Li, and N. Wei, "A new underdetermined blind source separation algorithm under the anechoic mixing model," in *Proc. IEEE Int. Conf. Signal Process.*, Nov. 2017, pp. 1799–1803.
[27] G. Qiang, G. Ruan, and L. Qi, "A complex-valued mixing matrix estimation algorithm for underdetermined blind source separation," *Circuits Syst. Signal Process.*, vol. 37, no. 8, pp. 3206–3226, 2018.
[28] G. Zhou, Z. Yang, S. Xie, and J.-M. Yang, "Mixing matrix estimation from sparse mixtures with unknown number of sources," *IEEE Trans. Neural Netw.*, vol. 22, no. 2, pp. 211–221, Feb. 2011.
[29] L. Siegel and A. Bessey, "Voiced/Unvoiced/Mixed excitation classification of speech," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-30, no. 3, pp. 451–460, Jun. 1982.
[30] S. V. Vaseghi, *Advanced Digital Signal Processing and Noise Reduction*. Hoboken, NJ, USA: Wiley, 2008.
[31] F. Auger *et al.*, "Time-frequency reassignment and synchrosqueezing: An overview," *IEEE Signal Process. Mag.*, vol. 30, no. 6, pp. 32–41, Nov. 2013.
[32] I. Daubechies, J. Lu, and H. T. Wu, "Synchrosqueezed wavelet transforms: An empirical mode decomposition-like tool," *Appl. Comput. Harmon. Anal.*, vol. 30, no. 2, pp. 243–261, Mar. 2011.
[33] D. Iatsenko, P. V. E. McClintock, and A. Stefanovska, "Linear and synchrosqueezed time–frequency representations revisited: Overview, standards of use, resolution, reconstruction, concentration, and algorithms," *Digit. Signal Process.*, vol. 42, pp. 1–26, Jul. 2015.
[34] S. Qian, *Introduction to Time-Frequency and Wavelet Transforms*. Beijing, China: Machine Press, 2005.
[35] B. Boashash, N. A. Khan, and T. Ben-Jabeur, "Time–frequency features for pattern recognition using high-resolution TFDs: A tutorial review," *Digit. Signal Process.*, vol. 40, pp. 1–30, May 2015.
[36] F. Zhang, Z. Geng, and W. Yuan, "The algorithm of interpolating windowed FFT for harmonic analysis of electric power system," *IEEE Trans. Power Del.*, vol. 16, no. 2, pp. 160–164, Apr. 2001.
[37] A. Rodriguez and A. Laio, "Clustering by fast search and find of density peaks," *Science*, vol. 344, no. 6191, pp. 1492–1496, Jun. 2014.
[38] D. L. Davies and D. W. Bouldin, "A cluster separation measure," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-1, no. 2, pp. 224–227, Apr. 1979.
[39] F. Abrard and Y. Deville, "A time–frequency blind signal separation method applicable to underdetermined mixtures of dependent sources," *Signal Process.*, vol. 85, no. 7, pp. 1389–1403, 2005.
[40] F. N. Hooge, "1/f noise sources," *IEEE Trans. Electron Devices*, vol. 41, no. 11, pp. 1926–1935, Nov. 1994.

XIANGDONG HUANG was born in 1979. He received the M.S. and Ph.D. degrees from Tianjin University, Tianjin, China, in 2004 and 2007, respectively, where he is currently an Associate Professor with the School of Electrical and Information Engineering. In 2009, he was with The University of Hong Kong, as a Visiting Scholar; was a Research Assistant with the University of Macau, in 2011; and was a Visiting Scholar with the University of Delaware, in 2013. His research interests include filter design and spectral analysis.

JINGWEN XU was born in 1994. She received the bachelor's degree from Tianjin University, Tianjin, China, in 2017, where she is currently pursuing the M.S. degree with the School of Electrical and Information Engineering. Her research interests include filter design and blind signal processing.

YU LIU was born in 1976. He received the bachelor's degree in electronic engineering; the master's degree in information and communication engineering; and the Ph.D. degree in signal and information processing from Tianjin University, Tianjin, China, in 1998, 2002, and 2005, respectively, where he is currently a Full Processor with the School of Electronics. He was an Electronic Engineer with Nantian Electronics Information Corporation, Shenzhen, China, from 1998 to 2000. From 2008 to 2016, he was an Associate Professor with the School of Electronic and Information, Tianjin University. From 2011 to 2012, he was a Visiting Fellow with the Department of Electrical Engineering, Princeton University, Princeton, NJ, USA. His research interests include signal/video processing, medical signal processing, multimedia systems, compressed sensing, and indoor positioning systems.