CrossMark

A Time–Frequency Domain Blind Source Separation Method for Underdetermined Instantaneous Mixtures

Tianliang Peng[1] · Yang Chen[1] · Zengli Liu[2]

Abstract We propose a new method for underdetermined blind source separation based on the time–frequency domain. First, we extract the time–frequency points that are occupied by a single source, and then, we use clustering methods to estimate the mixture matrix A. Second, we use the parallel factor (PARAFAC), which is based on nonnegative tensor factorization, to synthesize the estimated source. Simulations using mixtures of audio and speech signals show that this approach yields good performance.

Keywords Underdetermined blind source separation · Time–frequency · PARAFAC · Nonnegative tensor factorization

1 Introduction

Blind source separation (BSS) is a method that reconstructs N unknown sources of M observed signals from an unknown mixture. In this paper, we consider an instantaneous mixture system. Let A be the $M \times N$ mixing matrix, then the observations can be written as $x(t) = As(t)$, where $x(t) = [x_1(t), x_2(t), \ldots x_M(t)]$ and

✉ Yang Chen
   cheny@seu.edu.cn

Tianliang Peng
pengtianliang@seu.edu.cn

Zengli Liu
liuzengli@hotmail.com

[1] School of Information Science and Engineering, Southeast University, Nanjing 210018, China

[2] Faculty of Information Engineering and Automation, Kunming University of Science and Technology, Kunming 650500, China

Birkhäuser

$s(t) = [s_1(t), s_2(t), \ldots s_N(t)]^T$. Our investigation considers the underdetermined case, where there are more sources than observations ($N > M$).

BSS based on the instantaneous mixtures model has been extensively studied since the first papers by Herault and Jutten [25–27]. The early classical BSS method is based on independent component analysis (ICA) theory [10]. The classic ICA theory can only separate stationary non-Gaussian signals. Because of these limitations, it is difficult to apply it to real signals, such as audio signals. Some authors [5,14,15,28,36,37] have proposed different approaches to enhance the classical ICA theory. These methods have better performance than classical ICA methods for nonstationary signals. However, these approaches [5,28,36,37] cannot be applied to the underdetermined case. Many notable works for the underdetermined case have been published in recent years [1–3,6,7,17,22,23,29,30,32,33,38,40,41,46,47]. The basic assumption of these studies is that the mixed signals are sparse in the time–frequency domain. A signal is said to be sparse when it is zero or nearly zero more than might be expected from its variance. A notable sparse-based method called DUET was developed for delay and attenuation mixtures [29,40]. The DUET algorithm requires W-disjoint orthogonal signals or approximate W-disjoint orthogonal signals [41,46] in the time–frequency domain. In fact, the DUET algorithm is a ratio matrix (RM) method that uses clustering by an RM to accomplish source separation. Many well-known algorithms rely on the RM method, such as the SAFIA algorithm [3], the TIFROM algorithm [1], Abdeldjalil Aïssa-El-Bey et al.'s algorithm [2], Yuanqing Li et al.'s algorithm [13], and SangGyun Kim et al.'s algorithm [30]. In [6], Adel Belouchrani et al. presented an overview of source separation in the time–frequency domain.

Two methods have been widely used to make the signal sparse: the short-time Fourier transform (STFT) method [29,41,46] and the Wigner–Ville distribution (WVD) method [2,5]. We use the WVD to represent the time–frequency spectrum to estimate the mixing matrix and the STFT for signals synthesis. The proposed underdetermined BSS algorithm in this paper is based on a two-stage approach. In the first stage, we estimate the mixture matrix using a RM in the time–frequency domain on the spatial Wigner–Ville spectrum. The second stage is the synthesis of the signals: We use a method combining the minimum mean square error (MMSE) and the parallel factor (PARAFAC) technique to reconstruct the estimated sources. PARAFAC is a multilinear algebra tool for tensor decomposition in a sum of rank-1 tensors. PARAFAC is a multidimensional method originating from psychometrics [24] that has slowly found its way into various disciplines. A good overview of the tensor decomposition can be found in [31]. PARAFAC is also a powerful tool for BSS [4,8,11,13,18,20–23,34,35,42,45]. Recently, Pierre Comon provided a good overview of BSS and tensor decomposition in [12]. BSS based on nonnegative tensor factorization can be found in [8,18,20,21,35]. Nonnegative tensor factorization is a decomposition method for multidimensional data, in which all elements are nonnegative. As shown in [13,34], the traditional application of the tensor decomposition technique to BSS is limited to restricted numbers of microphones and mixed sources: The number of sources N and the number of observed mixtures M must satisfy $N \leq 2M - 2$. However, the application of nonnegative tensor factorization to BSS is not restricted by the number of microphones and mixed sources. We therefore use a nonnegative PARAFAC model for source synthesis in this paper. The authors of [45] have proposed a time–frequency

analysis based on the WVD together with the PARAFAC algorithm to separate electroencephalographic data.

There are many differences between the algorithm proposed in this paper and that in [45]. First, the PARAFAC model in [45] is based on the time–frequency domain with the WVD, while the PARAFAC model in this paper is based on the time–frequency domain with a STFT. Second, our PARAFAC model is different from the one in [45] because it has nonnegative elements. Thus, our PARAFAC model is not restricted by the number of microphones and mixed sources for BSS. Finally, the algorithm in this paper is a two-stage technique: We use time–frequency representation with the WVD to estimate the mixing matrix in the first stage, and the second stage is the source recovery with the nonnegative PARAFAC model.

This paper is organized as follows. In Sect. 2, we describe how the mixing matrix is estimated using the time–frequency ratio of the spatial Wigner–Ville spectrum. Then, we synthesize the estimated sources using MMSE and the PARAFAC method in Sect. 3. Section 4 provides the simulation results, and Sect. 5 draws various conclusions from this investigation.

2 Mixing Matrix Estimation

2.1 Spatial Time–Frequency Distributions

The WVD of $x(t)$ is defined as:

$$D_{xx}^{WV}(t, f) = \int_{-\infty}^{+\infty} x(t + \tau/2)x^H(t - \tau/2)\mathrm{e}^{-j2\pi f\tau}\,\mathrm{d}\tau, \tag{1}$$

where t and f represent the time index and the frequency index, respectively. The signal $x(t)$ of Cohen's class of spatial time–frequency distributions (STFD) is written as [9]:

$$D_{xx}(t, f) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \phi(u - t, v - f)D_{xx}^{WV}(t, f)\,\mathrm{d}u\mathrm{d}v, \tag{2}$$

where $\phi(u, v)$ is the kernel function of both the time and lag variables. In this paper, we assume that $\iint \phi(t, f)\mathrm{d}t\mathrm{d}f = 1$. Under the linear instantaneous mixing model of $x(t) = As(t)$, the STFD of $x(t)$ becomes:

$$D_{xx}(t, f) = AD_{ss}(t, f)A^H. \tag{3}$$

We note that $D_{xx}(t, f)$ is an $M \times M$ matrix, whereas $D_{ss}(t, f)$ is an $N \times N$ matrix.

2.2 Selecting a Single Source Active in the Time–Frequency Plane

If we assume that two signals x_1 and x_2 share the same frequency f_0 at the time t_0, then the cross-time–frequency distribution (cross-TFD) between x_1 and x_2 at (t_0, f_0) is nonzero. Hence, if $D_{x_1x_1}(t_0, f_0)$ and $D_{x_2x_2}(t_0, f_0)$ are nonzero, it is likely $D_{x_1x_2}(t_0, f_0)$ and $D_{x_2x_1}(t_0, f_0)$ are nonzero. Therefore, if $D_{xx}(t, f)$ is diagonal, it is likely that it has only one nonzero diagonal entry [9]. The proposed algorithm is based on single autoterms (SATs) theory [19]; we can detect the points of single-source occupancy in every time–frequency plane by SATs. In this paper, we use a mask-TFD to compute the SATs. We define the mask-TFD as $D_{xx}^{\mathrm{mask}}(t, f) = D_{xx}(t, f) * X(t, f)$, where $X(t, f)$ is the matrix of a STFT with x, as the same process in [19]; the mask-SATs satisfy:

$$\begin{cases} C(t, f) \approx \dfrac{\max\left|\mathrm{eig}(D_{xx}^{\mathrm{mask}}(t,f))\right|}{\sum \left|\mathrm{eig}(D_{xx}^{\mathrm{mask}}(t,f))\right|} \\ \|\mathrm{Grad}_C(t, f)\|_2 \leq \varepsilon_{\mathrm{Grad}} \\ H_C(t, f) < 0 \end{cases}, \tag{4}$$

where $\mathrm{eig}(D_{xx}^{\mathrm{mask}}(t, f))$ denote the eigenvalues of $D_{xx}^{\mathrm{mask}}(t, f)$, and $\mathrm{Grad}_C(t, f)$ and $H_C(t, f)$ are the gradient function and the Hessian matrix of $C(t, f)$, respectively.

2.3 Mixing Matrix Estimation Using the Time–Frequency Ratio

If we assume that there exist two observations, $x_1 = [x_1(t_1) \ldots x_1(t_{T_0})]^T$ and $x_2 = [x_2(t_1) \ldots x_2(t_{T_0})]^T$, and the two observations are mixed from N sources, $s_1 \ldots s_n \ldots s_N$ and $s_n = [s_n(t_1) \ldots s_n(t_{T_0})]^T$, then we can compute the mask-TFDs for $x_1(t)$ and $x_2(t)$. They become:

$$D_{x_1x_1}^{\mathrm{mask}}(t, f) = \sum_n D_{s_n s_n}^{\mathrm{mask}}(t, f) a_{1n}^2 a_{1n}^H \tag{5}$$

$$D_{x_2x_2}^{\mathrm{mask}}(t, f) = \sum_n D_{s_n s_n}^{\mathrm{mask}}(t, f) a_{2n}^2 a_{2n}^H. \tag{6}$$

If we have extracted all the time–frequency points that are occupied by a single source via (4), and if furthermore we assume that s_n occupies frequency f_0 at time t_0, then (5) and (6) become:

$$D_{x_1x_1}^{\mathrm{mask}}(t_0, f_0) = D_{s_n s_n}^{\mathrm{mask}}(t_0, f_0) a_{1n}^2 a_{1n}^H \tag{7}$$

$$D_{x_2x_2}^{\mathrm{mask}}(t_0, f_0) = D_{s_n s_n}^{\mathrm{mask}}(t_0, f_0) a_{2n}^2 a_{2n}^H. \tag{8}$$

Then, it follows that $\dfrac{D_{x_1x_1}^{\mathrm{mask}}(t_0, f_0)}{D_{x_2x_2}^{\mathrm{mask}}(t_0, f_0)} = \dfrac{D_{s_n s_n}^{\mathrm{mask}}(t_0, f_0) a_{1n} a_{1n}^H}{D_{s_n s_n}^{\mathrm{mask}}(t_0, f_0) a_{2n} a_{2n}^H} = \dfrac{a_{1n}^2 * a_{1n}^H}{a_{2n}^2 * a_{2n}^H}$. For the general case of M observations and if we define a vector, $D'_{xx}(t_0, f_0) = [D_{x_1x_1}^{\mathrm{mask}}(t_0, f_0), \ldots, D_{x_M x_M}^{\mathrm{mask}}(t_0, f_0)]^T$, then we can obtain a ratio vector:

$$\frac{D'_{xx}(t_0, f_0)}{D^{\text{mask}}_{x_1 x_1}(t_0, f_0)} = \left[1, \frac{a_{2n}^2 * a_{2n}^H}{a_{1n}^2 * a_{1n}^H}, \dots, \frac{a_{Mn}^2 * a_{Mn}^H}{a_{1n}^2 * a_{1n}^H} \right]^T. \tag{9}$$

We denote the set of the above ratio as C_{s_n} for $n = 1, \dots, N$. Then, we can apply a clustering method to estimate the mixing matrix combining SATs and (9). We assume that the mixtures are normalized to have unit l_2-norm. We denote the nth column vector of A as a_n, which is estimated as follows:

$$\hat{a}_n = \frac{1}{|C_{s_n}|} \sum_{[t, f] \in C_{s_n}} D'_{xx}(t, f), \tag{10}$$

where $|C_{s_n}|$ denotes the number of the points in the class for $n = 1, \dots, N$.

3 Source Synthesis

Nonnegative tensor factorization (NTF) of multichannel spectrograms under the PARAFAC structure has been widely applied to BSS of multichannel signals [18,20,21]. In this paper, we use the IS-NTF method [18] to synthesize sources. First, we apply the STFT to time-domain observations $x(t)$; we further assume that $s(t, f)$ obeys the following distribution [18]:

$$\begin{cases} x_m(t, f) = \sum_n a_{mn} s_n^m(t, f) \\ s_n^m(t, f) \sim N_c(0 | w_{fn} h_{tn}) \end{cases}, \tag{11}$$

where m is the number of observations, n is the mixed-source index, N_c denotes the complex Gaussian distribution subject to $N_c(x | u, \Sigma) = |\pi \Sigma|^{-1} \exp -(x - u)^H \Sigma^{-1}(x - u)$, w_{fn} is the (f, n)th element of matrix W (its size is $F \times N$), and h_{tn} is the (t, n)th element of matrix H (its size is $T \times N$). Set $V = |X|^2$ and $Q = |A|^2$, where X is the time–frequency matrix with elements $x(t, f)$. We note that (11) is equivalent to $\min_{Q, W, H} \sum_{mft} d_{IS}(v_{mft} | \hat{v}_{mft})$, where $Q, W, H \geq 0$, $d_{IS}(x | y) = \frac{x}{y} - \log \frac{x}{y} - 1$. Then, we can obtain [18]:

$$W = W \cdot \frac{\langle G_-, QoH \rangle_{\{1,3\},\{1,2\}}}{\langle G_+, QoH \rangle_{\{1,3\},\{1,2\}}} \tag{12}$$

$$H = H \cdot \frac{\langle G_-, QoW \rangle_{\{1,2\},\{1,2\}}}{\langle G_+, QoW \rangle_{\{1,2\},\{1,2\}}}, \tag{13}$$

where G is the $M \times F \times T$ derivatives tensor with $g_{mft} = d'(v_{mft} | \tilde{v}_{mft})$ and $\langle A, B \rangle_{\{1,\dots,M\},\{1,\dots,M\}} = \sum_{i_1}^{I_1} \dots \sum_{i_M}^{I_M} a_{i_1,\dots,i_M, j_1,\dots,j_N} b_{i_1,\dots,i_M, k_1,\dots,k_O}$. Then, we can obtain the MMSE reconstruction as:

$$s_n^m(t, f) = \frac{a_{mn}^2 w_{fn} h_{fn}}{\sum_n a_{mn}^2 w_{fn} h_{fn}} x_m(t, f). \tag{14}$$

Finally, applying the inverse STFT to $s_n^m(t, f)$, we can obtain the time-domain source. Our source separation route is distinct from the method in [18,20]. Our algorithm has two stages. The first is to estimate the mixing matrix. Then, source reconstruction is the inverse problem in the second stage. We fix the mixing matrix in the source reconstruction stage, which is equivalent to fixing Q in (12) and (13).

4 Simulation

To show the validity of our technique of mixing matrix estimation, we conducted four experiments for mixing matrices of orders 2×3, 2×4, 3×4, and 3×6. For each order of the mixing matrix, we used random values. Each experiment was run 30 times. We used two methods to select the single signal active in the time–frequency plane: traditional SATs and mask-SATs, and we set $\varepsilon_{\text{Grad}}$ randomly in the range $0.005 \leq \varepsilon_{\text{Grad}} \leq 0.05$. The mixing sources used in the experiments were music signals and speech. The length of the STFT was 1024, the window overlap was 0.5, and all signals were sampled at 16 kHz with a sample length of 160,000.

The performance of the mixing matrix estimation was evaluated using the normalized mean square error (NMSE), which is defined in [39]:

$$\text{NMSE} = 10 \log \left(\frac{\sum_{mn} (\hat{a}_{mn} - a_{mn})^2}{\sum_{mn} a_{mn}^2} \right), \tag{15}$$

where $\hat{a}_{mn}$ is the (m, n)th element of the estimated mixing matrix, and a_{mn} is the (m, n)th element of the mixing matrix. Smaller NMSEs indicate better performance.

We must note that matrix A_{TFD} was obtained similarly to $A_{\text{mask-TFD}}$. The only difference is that in (4), $D_{xx}^{\text{mask}}(t, f)$ is replaced by $D_{xx}(t, f)$. To show the validity of our method, we compared it with the LI-TIFROM algorithm in [1] and the method in [39], as shown in Fig. 1. We see that the mask-TFD algorithm achieves a better mixing matrix estimation than that in [39]. The performance of the algorithm in [39] is better than that of the LI-TIFROM and TFD algorithms. As pointed out in [16], the TIFROM algorithm requires at least two adjacent windows in the time–frequency domain for each source to ensure the degree of sparsity needed. If this condition is not satisfied, the TIFROM algorithm cannot find the single-source points and thus cannot estimate the mixing matrix correctly. This is why the performance of the TIFROM algorithm is not very good in Fig. 1.

For the source synthesis stage, we performed two simulation experiments using 2×3 and 3×4 mixing matrices. The sources were taken from the Signal Separation Evaluation Campaign (SiSEC 2008) [43]. We used some "development data" from the "underdetermined speech and music mixtures task":

1. For the music sources including drums: a linear instantaneous stereo mixture (with positive mixing coefficients) of two drum sources and one bass line.
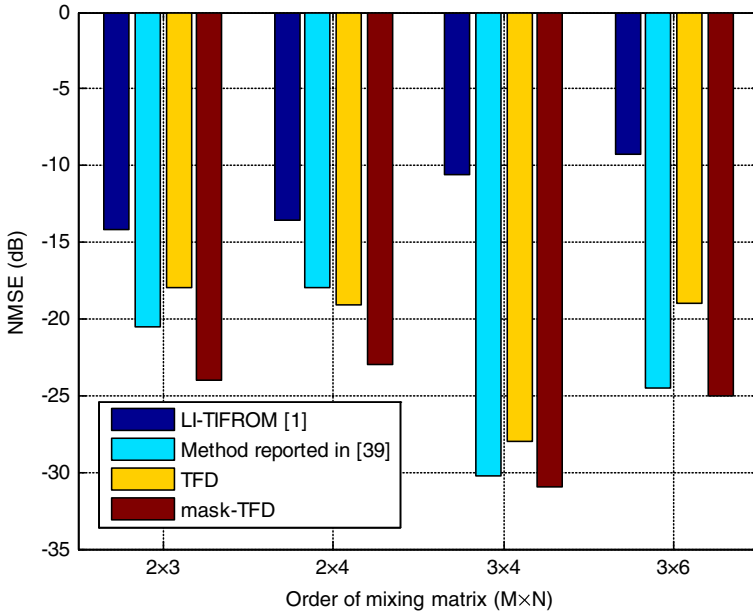
Fig. 1 Comparison of proposed algorithm with that in [1] and [39]

2. For the nonpercussive music sources: a linear instantaneous stereo mixture (with positive mixing coefficients) of one rhythmic acoustic guitar, one electric lead guitar, and one bass line.
3. Four female sources.

The original mixing matrix of data 1 and data 2 had positive coefficients: $A_{\text{nodrum}} = \begin{bmatrix} 0.4937 & 0.6025 & 0.8488 \\ 0.7900 & 0.6575 & 0.4232 \end{bmatrix}$ and $A_{\text{wdrum}} = \begin{bmatrix} 0.5846 & 0.7135 & 1.0053 \\ 0.9356 & 0.7786 & 0.5012 \end{bmatrix}$. For the four female sources, we used the original mixing matrix:

$A_{\text{female}} = \begin{bmatrix} 0.6547 & 0.6516 & 0.8830 & -0.5571 \\ 0.3780 & -0.5923 & 0.3532 & 0.7428 \\ -0.6547 & 0.4739 & 0.3091 & 0.3714 \end{bmatrix}$. We used the proposed

mask-TFD algorithm to estimate the three matrixes: A_{nodrum}, A_{wdrum}, and A_{female}. The results of the estimations are $A^e_{\text{nodrum}} = \begin{bmatrix} 0.5345 & 0.5739 & 0.8318 \\ 0.7836 & 0.6825 & 0.4194 \end{bmatrix}$, $A^e_{\text{wdrum}} = \begin{bmatrix} 0.5595 & 0.7480 & 1.0053 \\ 0.8957 & 0.8109 & 0.5012 \end{bmatrix}$, and

$A^e_{\text{female}} = \begin{bmatrix} 0.6472 & 0.6500 & 0.8715 & -0.5601 \\ 0.4010 & -0.6013 & 0.3541 & 0.7626 \\ -0.6936 & 0.4732 & 0.2987 & 0.3619 \end{bmatrix}$. The NMSEs of A^e_{nodrum},

A^e_{wdrum}, and A^e_{female} are -35.10, -50.88, and -33.90 dB, respectively.

To measure the performance, we decompose an estimated source image as [44]:

$$\hat{s}^{\text{img}}_{mn}(t) = s^{\text{img}}_{mn}(t) + e^{\text{spat}}_{mn}(t) + e^{\text{interf}}_{mn}(t) + e^{\text{artif}}_{mn}(t), \tag{16}$$
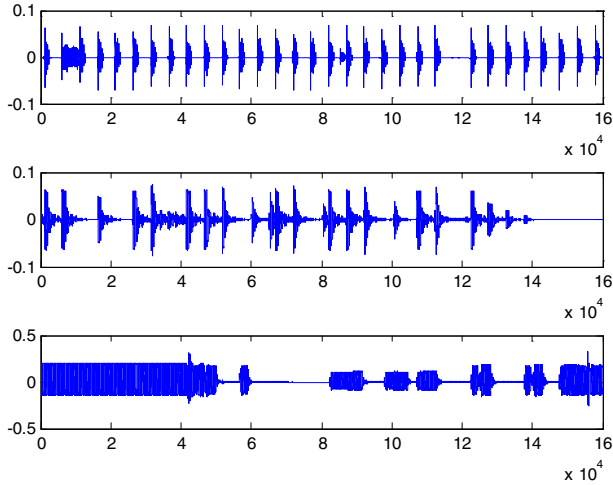
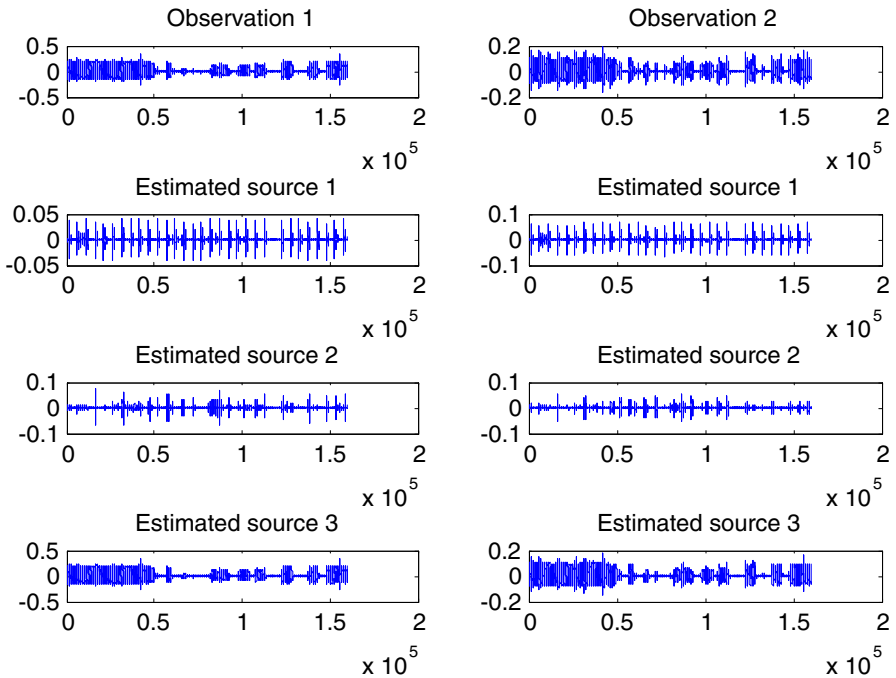**Fig. 2** Three original source signals with hi-hat, drums, and bass

Fig. 3 Three estimated signals with IS-MMSE

where $s_{mn}^{\mathrm{img}}(t)$ is the true source image, and $e_{mn}^{\mathrm{spat}}(t)$, $e_{mn}^{\mathrm{interf}}(t)$, and $e_{mn}^{\mathrm{artif}}(t)$ are distinct error components representing spatial (or filtering) distortion, interference, and artifacts, respectively. As performance measures, we use the source to distortion ratio (SDR):
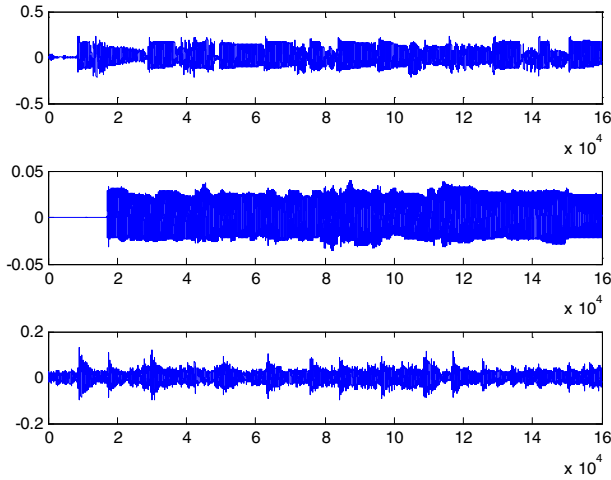
Fig. 4 Three original source signals with lead guitar, rhythm guitar, and bass

$$\text{SDR}_n = 10 \log_{10} \frac{\sum_{m=1}^{M} \sum_t s_{mn}^{\text{img}}(t)^2}{\sum_{m=1}^{M} \sum_t \left(e_{mn}^{\text{spat}}(t) + e_{mn}^{\text{interf}}(t) + e_{mn}^{\text{artif}}(t) \right)^2}, \qquad (17)$$

the source image to spatial distortion ratio (ISR):

$$\text{ISR}_n = 10 \log_{10} \frac{\sum_{m=1}^{M} \sum_t s_{mn}^{\text{img}}(t)^2}{\sum_{m=1}^{M} \sum_t e_{mn}^{\text{spat}}(t)^2}, \qquad (18)$$

the source to interference ratio (SIR):

$$\text{SIR}_n = 10 \log_{10} \frac{\sum_{m=1}^{M} \sum_t \left(s_{mn}^{\text{img}}(t) + e_{mn}^{\text{spat}}(t) \right)^2}{\sum_{m=1}^{M} \sum_t e_{mn}^{\text{interf}}(t)^2}, \qquad (19)$$

and the sources to artifacts ratio (SAR):

$$\text{SAR}_n = 10 \log_{10} \frac{\sum_{m=1}^{M} \sum_t \left(s_{mn}^{\text{img}}(t) + e_{mn}^{\text{spat}}(t) + e_{mn}^{\text{interf}}(t) \right)^2}{\sum_{m=1}^{M} \sum_t e_{mn}^{\text{artif}}(t)^2}. \qquad (20)$$

Higher values indicate better results for SDR, ISR, SIR, and SAR. We note that in the MMSE method, we do not reconstruct the single-channel sources but their multichannel contribution to the multichannel data.

Figure 3 shows the three estimated signals with IS-MMSE. The two mixed signals in Fig. 3 are obtained using the three original source signals in Fig. 2 with mixing matrix A_{nodrum}. Figure 5 shows the three estimated signals with IS-MMSE. The two
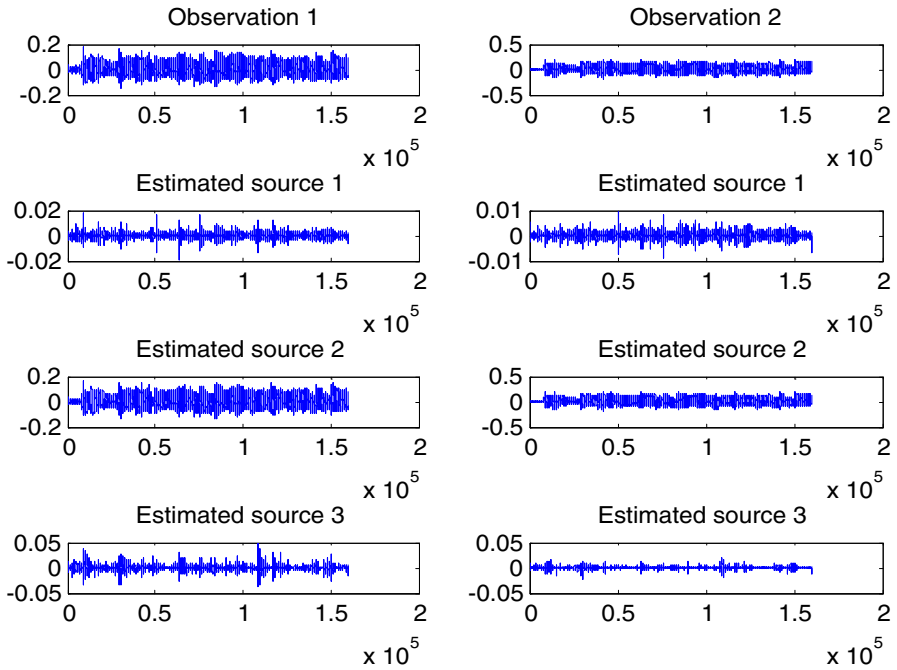
Fig. 5 Two mixed signals with lead guitar, rhythm guitar, and bass separated by IS-MMSE
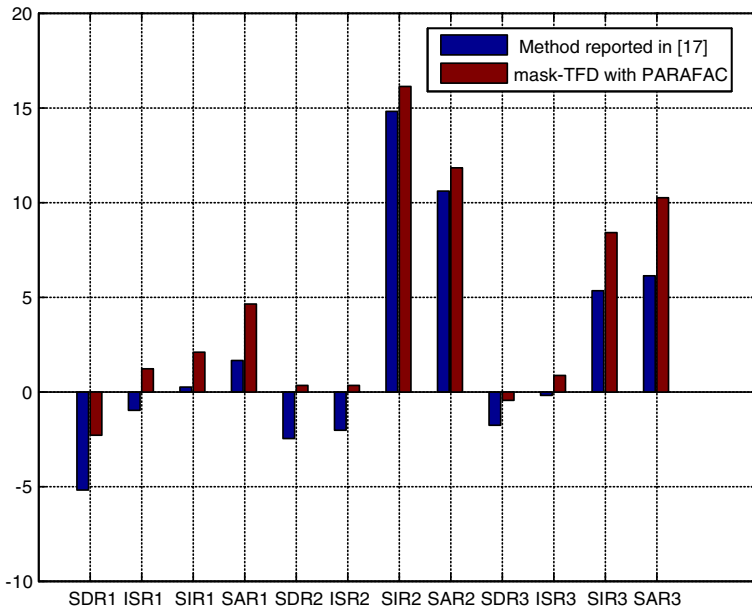

Fig. 6 Performance of the PARAFAC algorithm [18] and mask-TFD with PARAFAC in three examples (two microphones and three sources with no drums, two microphones and three sources with drums, and three microphones mixed with four female sources)

mixed signals in Fig. 5 are obtained using the three original source signals in Fig. 4 with mixing matrix A_{wdrum} (we use $A^{\text{mask-TFD}}$ in the IS-MMSE algorithm).

Figure 6 shows the performance of the PARAFAC algorithm [18] and mask-TFD with PARAFAC in three examples with music, audio, and speech sources. As shown in Fig. 6, neither the method reported in [18] nor mask-TFD with PARAFAC yields good performance for the nonpercussive music sources. This may be because the number of single signals active in the time–frequency plane is very small. The performances of mask-TFD with the PARAFAC algorithm are better than that of the PARAFAC algorithm reported in [18] in all cases. As shown in [18,35], the performance of the nonnegative PARAFAC model is heavily related to its initialization. If the initialization is poor, then it is difficult for the nonnegative PARAFAC algorithm to achieve global convergence. In fact, the first stage of mask-TFD with PARAFAC is the initialization of the nonnegative PARAFAC algorithm.

5 Conclusion

We have proposed a two-stage approach to solve the underdetermined instantaneous BSS problem. We used a cluster method for single time–frequency active points in the mixing matrix estimation stage. Methods using joint diagonalization [5,19] are not suitable for underdetermined mixtures. In this paper, we have used a new cluster method for underdetermined mixing matrix estimation and then applied NTF to the source synthesis. Numerical simulations have illustrated the effectiveness of the new approach for the audio nonstationary signals of the Signal Separation Evaluation Campaign (SiSEC 2008) public data.

References

1. F. Abrard, Y. Deville, A time–frequency blind signal separation method applicable to underdetermined mixtures of dependent sources. Signal Process. **85**(7), 1389–1403 (2005)
2. A. Aissa-El-Bey, N. Linh-Trung, K. Abed-Meraim et al., Underdetermined blind separation of nondisjoint sources in the time–frequency domain. IEEE Trans. Signal Process. **55**(3), 897–907 (2007)
3. M. Aoki, M. Okamoto, S. Aoki et al., Sound source segregation based on estimating incident angle of each frequency component of input signals acquired by multiple microphones. Acoust. Sci. Technol. **22**(2), 149–157 (2001)
4. H. Becker, P. Comon, L. Albera et al., Multi-way space–time–wave-vector analysis for EEG source separation. Signal Process. **92**(4), 1021–1031 (2012)
5. A. Belouchrani, M.G. Amin, Blind source separation based on time–frequency signal representations. IEEE Trans. Signal Process. **46**(11), 2888–2897 (1997)
6. A. Belouchrani, M.G. Amin, N. Thirion-Moreau et al., Back to results source separation and localization using time–frequency distributions: a overview. IEEE Signal Process. Mag. **30**(6), 97–107 (2013)
7. S. Chen, D.L. Donoho, M.A. Saunders, Atomic decomposition by basis pursuit. SIAM J. Sci. Comput. **20**(1), 33–61 (1998)
8. A. Cichocki, R. Zdunek, A.H. Phan et al., *Nonnegative matrix and tensor factorizations: applications to exploratory multi-way data analysis and blind source separation* (Wiley, New Jersey, 2009)

9. L. Cohen, Time–frequency distributions—a review. Proc. IEEE **77**(7), 941–981 (1989)
10. P. Comon, Independent component analysis, a new concept? Signal Process. **36**(3), 287–314 (1994)
11. P. Comon, Blind identification and source separation in 2×3 under-determined mixtures. IEEE Trans. Signal Process. **52**(1), 11–22 (2004)
12. P. Comon, Tensors: a brief introduction. Signal Process. Mag. **31**(3), 44–53 (2014)
13. L. De Lathauwer, J. Castaing, Blind identification of underdetermined mixtures by simultaneous matrix diagonalization. IEEE Trans. Signal Process. **56**(3), 1096–1105 (2008)
14. Y. Deville, M. Benali, Differential source separation: concept and application to a criterion based on differential normalized kurtosis, in *Proceedings of EUSIPCO*, Tampere, Finland, 4–8 Sept 2000
15. Y. Deville, S. Savoldelli, A second-order differential approach for underdetermined convolutive source separation, in: *Proceedings of ICASSP 2001*, Salt Lake City, USA, 2001
16. T. Dong, Y. Lei, J. Yang, An algorithm for underdetermined mixing matrix estimation. Neurocomputing **104**, 26–34 (2013)
17. D.L. Donoho, M. Elad, Maximal sparsity representation via l_1 minimization. Proc. Nat. Acad. Sci. **100**, 2197–2202 (2003)
18. C. Févotte, A. Ozerov, *Notes on nonnegative tensor factorization of the spectrogram for audio source separation: statistical insights and towards self-clustering of the spatial cues. Exploring Music Contents* (Springer, Heidelberg, Berlin, 2011), pp. 102–115
19. C. Fevotte, C. Doncarli, Two contributions to blind source separation using time–frequency distributions. IEEE Signal Process. Lett. **11**(3), 386–389 (2004)
20. D. FitzGerald, M. Cranitch, E. Coyle, Non-negative tensor factorization for sound source separation, in *Proceedings of Irish Signals and Systems Conference*, pp. 8–12 (2005)
21. D. FitzGerald, M. Cranitch, E. Coyle, Extended nonnegative tensor factorization models for musical sound source separation. Computational Intelligence and Neuroscience, Article ID 872425 (2008)
22. S. Ge, J. Han, M. Han, Nonnegative mixture for underdetermined blind source separation based on a tensor algorithm. Circuits Syst. Signal Process. (2015). doi:10.1007/s00034-015-9969-8
23. F. Gu, H. Zhang, W. Wang et al., PARAFAC-based blind identification of underdetermined mixtures using Gaussian mixture model. Circuits Syst. Signal Process. **33**(6), 1841–1857 (2014)
24. R.A. Harshman, Foundations of the PARAFAC procedure: models and conditions for an "explanatory" multimodal factor analysis. UCLA Working Papers in Phonetics, 16 (1970)
25. J. Herault, C. Jutten, Space or time adaptive signal processing by neural network models, in *International Conference on Neural Networks for Computing*, Snowbird, USA, 1986
26. J. Herault, C. Jutten, Blind separation of sources. Part 1: an adaptive algorithm based on neuromimetic architecture. Signal Process. **24**(1), 1–10 (1991)
27. J. Herault, C. Jutten, B. Ans, Détection de grandeurs primitives dans un message composite par une architecture de calcul neuromimétique en apprentissage non supervisé. In 10° Colloque sur le traitement du signal et des images, FRA. GRETSI, Groupe d'Etudes du Traitement du Signal et des Images (1985)
28. A. Hyvarinen, Blind source separation by nonstationarity of variance: a cumulant-based approach. IEEE Trans. Neural Netw. **12**(6), 1471–1474 (2001)
29. A. Jourjine, S. Rickard, O. Yilmaz, Blind separation of disjoint orthogonal signals: demixing n sources from 2 mixtures, in *Proceedings of ICASSP 2000*, Turkey, vol. 6, pp. 2986–2988 (2000)
30. S. Kim, C.D. Yoo, Underdetermined blind source separation based on subspace representation. IEEE Trans. Signal Process. **57**(7), 2604–2614 (2009)
31. T.G. Kolda, B.W. Bader, Tensor decompositions and applications. SIAM Rev. **51**(3), 455–500 (2009)
32. M.S. Lewicki, T.J. Sejnowski, Learning overcomplete representations. Neural Comput. **12**, 337–365 (2000)
33. Y. Li, S.I. Amari, A. Cichocki et al., Underdetermined blind source separation based on sparse representation. IEEE Trans. Signal Process. **54**(2), 423–437 (2006)
34. D. Nion, K.N. Mokios, N.D. Sidiropoulos et al., Batch and adaptive PARAFAC-based blind separation of convolutive speech mixtures. IEEE Trans. Audio Speech Lang. Process. **18**(6), 1193–1207 (2010)
35. A. Ozerov, C. Févotte, R. Blouet, et al., Multichannel nonnegative tensor factorization with structured constraints for user-guided audio source separation, in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2011*. IEEE, pp. 257–260 (2011)
36. L. Parra, C. Spence, Convolutive blind separation of nonstationary sources. IEEE Trans. Audio Speech Lang. Process. **8**(3), 320–327 (2000)
37. D.T. Pham, J.F. Cardoso, Blind separation of instantaneous mixtures of non-stationary sources. IEEE Trans. Signal Process. **49**(9), 1837–1848 (2001)

38. R. Qi, Y. Zhang, H. Li, Overcomplete blind source separation based on generalized Gaussian function and sl0 norm. Circuits Syst. Signal Process. (2014). doi:10.1007/s00034-014-9952-9

39. V.G. Reju, S.N. Koh, I.Y. Soon, An algorithm for mixing matrix estimation in instantaneous blind source separation. Signal Process. **89**(9), 1762–1773 (2009)

40. S. Rickard, R. Balan, J. Rosca, Real-time time-frequency based blind source separation, in *Proceedings of ICA 2001*, San Diego, CA, 9–13 Dec 2001

41. S. Rickard, O. Yilmaz, On the approximate w-disjoint orthogonality of speech, in *ICASSP*, Orlando, Florida, 13–17 May 2002

42. P. Tichavsky, Z. Koldovsky, Weight adjusted tensor method for blind separation of underdetermined mixtures of nonstationary sources. IEEE Trans. Signal Process. **59**(3), 1037–1047 (2011)

43. E. Vincent, S. Araki, P. Bofill, Signal separation evaluation campaign. In (SiSEC 2008)/Underdetermined speech and music mixtures task results (2008), http://www.irisa.fr/metiss/SiSEC08/SiSEC_underdetermined/dev2_eval.html

44. E. Vincent, *First stereo audio source separation evaluation campaign: data, algorithms and results. Independent Component Analysis and Signal Separation* (Springer, Berlin Heidelberg, 2007), pp. 552–559

45. M. Weis, F. Romer, M. Haardt, et al., Multi-dimensional space–time–frequency component analysis of event related EEG data using closed-form PARAFAC, in *IEEE International Conference on Acoustics, Speech and Signal Processing. ICASSP 2009*, IEEE 2009, pp. 349–352 (2009)

46. O. Yilmaz, S. Rickard, Blind separation of speech mixtures via time–frequency masking. IEEE Trans. Signal Process. **52**(7), 1830–1847 (2004)

47. M. Zibulevsky, B.A. Pearlmutter, Blind source separation by sparse decomposition. Neural Comput. **13**(4), 863–882 (2001)