HUI LI[1], YUE-HONG SHEN[1], MIN CAO[2], JIANG-GONG WANG[1]

Institute of Communications Engineering, PLA University of Science and Technology (1), Xi'an Communication Institute(2)

Underdetermined Blind Separation using modified subspace-based algorithm in the time-frequency domain

***Abstract**. To solve the underdetermined blind separation (UBSS) problem, Aissa-El-Bey et al. have proposed the significant subspace-based algorithms in the time-frequency (TF) domain, where a fixed (maximum) value of K, i.e., the number of active sources overlapping at any TF point, is considered for simplicity. In this paper, based on the principle component analysis (PCA) technology, we propose a modified algorithm by estimating the number K for selected frequency bins where most energy is concentrated. Improved performances are obtained without increasing complexity.*

***Streszczenie.** Do rozwiązania problem nieokreślonej ślepej separacji (UBSS) Aissa-El_Bey zaproponował algorytm czasowo-częstotliwościowy gdzie ustalono liczbę aktywnych źródeł pokrywających każdy punkt TF. W artykule zaproponowano zmodyfikowany algorytm bazujący na analizie składowej głównej PCA. Otrzymano poprawę parametrów bez powiększania skomplikowania metody. (**Nieokreślona ślepa separacja przy wykorzystaniu algorytmu w dziedzinie czasu i częstotliwości**)*

Keywords: underdetermined blind separation; sparsity; time-frequency subspace projection; principle component analysis.
Słowa kluczowe: ślepa separacja, analiza składowej głównej PCA

Introduction

Blind source separation (BSS) aims to extract the original source signals from their mixtures observed by a set of sensors with no, or very limited, knowledge about the source signals and the mixing channel. Potential applications of BSS include speech processing, telecommunications, biomedical signal processing, analysis of astronomical data or satellite images, etc.

If there are fewer mixtures than sources, we have a challenging UBSS problem, where estimating the mixing channel is not sufficient for the recovery of sources. Nowadays, a large amount of algorithms for UBSS start from the assumption that the sources are sparse [1-6], i.e., the signals are mostly close to zero with the exception of several large values. Then the algorithms consist of two steps [1]: first estimating the mixing matrix and then estimating the sources.

Many natural signals can achieve satisfied sparsity in transform domains, such as by wavelet packet transform or by short-time Fourier transform (STFT) [7]. Yilmaz et al. exploit sparsity in the STFT domain, and assume that the sources are disjoint, i.e., there exists only one source at any TF point [5,8]. Later, Aissa-El-Bey et al. relax the condition and propose an efficient linear algorithm based on subspace projection using STFT, which assumes that the sources can be non-disjoint in the TF domain, i.e., the number of the active sources that coexist at any TF point (denoted by K in this paper) is less than that of the mixtures [6]. Both algorithms work well on audio signals. However, [6] does not estimate the number K of active sources at a given TF point and fixes the value of K instead (usually K is the maximum). Though doing this has the advantage of simpler computation and lower complexity, the authors admit that the estimation error of the source signals increases, especially at low signal-noise-ratio (SNR).

To enhance the separation accuracy, in this paper we propose to tackle the estimation of the number K by adopting the PCA technology [9], which mainly performs the eigen value decomposition (EVD) on the covariance matrix of observed mixtures in the TF domain. Readers can also refer to other methods such as information theoretic-based criterion [10]. If we do the estimation job for each frequency bin, undoubtedly the added computation complexity may be excessive or even unbearable, which is not our desired result. Paper [11] discovers the fact that most of the signal energy will be concentrated in nearly 10% of the frequency bins. Therefore, we decide to sort the frequency bins in the descending order according to their variance. By selecting the much fewer frequency bins with the most energy to do the estimation of K and fixing K at other unselected frequency bins with poor energy (usually K is the minimum), unnecessary computation can be avoided.

Since several algorithms have been successfully applied in the identification of the mixing matrix [12-14] for the UBSS problem, we are not going to repeat the same procedure, and focus on the recovery of the source signals, supposing that the mixing matrix is already known.

The rest of this paper is organized as follows. In section II, the signal model and assumptions are briefly introduced. We review the subspace-based algorithm in section III. Section VI presents our proposed modified algorithm. Numerical simulation results and discussion are provided in section V. At the end of the paper, a concise conclusion is given.

Signal model and assumptions

Consider the noiseless instantaneous mixing model:

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

where $\mathbf{s}(t) = [s_1(t), s_2(t), \ldots, s_N(t)]^T$ is the N-dimensional vector of source signals, $\mathbf{x}(t) = [x_1(t), x_2(t), \ldots, x_M(t)]^T$ is the M-dimensional vector of mixture signals observed by M sensors, $\mathbf{A} = [\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_N]$ is the mixing matrix of size $M \times N$ and $(\ )^T$ denotes the transpose operator. For UBSS, $M < N$.

Combing the analysis in [6] and [15], we make the following two assumptions:

Assumption 1: Any $M \times M$ submatrix of the mixing matrix $\mathbf{A}$ is of full rank.

Assumption 2: The number K of active sources that contribute their energy at any TF point is less than the number M of sensors, i.e., $K < M$.

Review of the Subspace-based algorithm

The TF processing provides effective tool for nonstationary signals, whose frequency spectra vary in time. In this paper, we choose to use STFT as the preferred TF representation. The reasons for this choice are: i) unlike the Wigner-Ville distribution, STFT is linear and free of cross terms; ii) STFT of speech or audio signals are sparse; iii) STFT is easy to invert and implement, which is important for computational speed.

The STFT of the m-th mixture $x_m(t)$ is defined by:

$$X_m(t,f)=\int_{-\infty}^{\infty}x_m(\tau)h(t-\tau)e^{-j\omega\tau}d\tau \tag{2}$$

where $h(t)$ is a weighting window function. Let $\mathbf{X}(t,f)=[X_1(t,f),\ldots,X_M(t,f)]^T$ and $\mathbf{S}(t,f)=[S_1(t,f),\ldots,S_N(t,f)]^T$, applying the STFT in (2) to (1) yields:

$$\mathbf{X}(t,f)=\mathbf{AS}(t,f)=\sum_{n=1}^{N}\boldsymbol{a}_n S_n(t,f) \tag{3}$$

Paper [6] assumes $\alpha_1,\ldots,\alpha_K\in\mathbb{N}$ are the indexes of the K active sources present at a given TF point (t^*,f^*), and denotes $\tilde{\mathbf{s}}(t)=[s_{\alpha_1}(t),s_{\alpha_2}(t),\ldots,s_{\alpha_K}(t)]^T$ and $\widetilde{\mathbf{A}}=[\boldsymbol{a}_{\alpha_1},\boldsymbol{a}_{\alpha_2},\ldots,\boldsymbol{a}_{\alpha_K}]$. Based on Assumption 2, then (3) is reduced to

$$\mathbf{X}(t^*,f^*)=\widetilde{\mathbf{A}}\widetilde{\mathbf{S}}(t^*,f^*)=\sum_{k=1}^{K}\boldsymbol{a}_{\alpha_k}S_{\alpha_k}(t^*,f^*) \tag{4}$$

Taking the effect of noise into account, let Q represent the orthogonal projection matrix onto the noise subspace of $\widetilde{\mathbf{A}}$, which is given by:

$$Q=\mathrm{I}-\widetilde{\mathbf{A}}(\widetilde{\mathbf{A}}^H\widetilde{\mathbf{A}})^{-1}\widetilde{\mathbf{A}}^H \tag{5}$$

where I is the identity matrix. When $\mathbf{A}$ is available, then:

$$\{\alpha_1,\ldots,\alpha_K\}=\min_{\beta_1,\ldots,\beta_K}\left\{\left\|Q\mathbf{X}(t^*,f^*)\right\|\left\|\widetilde{\mathbf{A}}_\beta\right\}\right. \tag{6}$$

where $\widetilde{\mathbf{A}}_\beta=[\boldsymbol{a}_{\beta_1},\boldsymbol{a}_{\beta_2},\ldots,\boldsymbol{a}_{\beta_K}]$. And the STFT value of the K active sources at (t^*,f^*) can be estimated by:

$$\tilde{\mathbf{S}}(t^*,f^*)\approx\widetilde{\mathbf{A}}^{\#}\mathbf{X}(t^*,f^*) \tag{7}$$

where $(\ )^{\#}$ is the Moore-Penrose's pseudo-inversion operator.

The proposed modified algorithm

In [6], a fixed (maximum) value of the number K $(K=M-1)$ for all the TF points is carried out. Though the computation procedure is simplified, the estimation error is sacrificed. To keep balance between the estimation accuracy and the computation complexity, in this section we propose the modified subspace-based algorithm, which choose to evaluate the number K of active sources for selected frequency bins with most of the signal energy, and consider a fixed (minimum) value of the number K (typically $K=1$) for the rest of frequency bins unselected.

Since the unselected frequency bins cover poor signal energy, there is no necessity to assume more than one active source there. If no source is effectively contributing, we would estimate a close-to-zero source STFT coefficient.

The Selection of Frequency Bins with Most Energy

By exploiting the STFT sparsity in speech signals, [16] comes to a conclusion that the STFT provides slightly higher sparsity with the proper window choice, when compared with the wavelet transform. Later, [11] does additional experiments on fifty speech sources from the TIMIT data base, each comsisting of 50000 samples sampled at 16kHz. The results indicate that the STFT with a 64ms window-size demonstrates superior performance in terms of sparsity, capturing 98% of the total signal power with approximate 10% of the STFT coefficients only.

Let $\Gamma=\{f_1,f_2,\ldots,f_l\}$ denote the aggregate of all the frequency bins after performing STFT, where l is the number of total frequency bins. Here we sort the frequency bins corresponding to one mixture, $X_1(t,f)$, in the descending order of their variance, and the order of the frequency bins of other mixtures are modified according to that of $X_1(t,f)$. Hence the frequency bins whose orders are ranking at the head contain most of the signal energy, and they are the goal of our selection step. In this paper when we refer to the percent of frequency bins p, it also means that those frequency bins, with the first p percent orders ranking at the head of the total orders, are selected and used for the next estimation step. Let Γ' be the aggregate of all the selected frequency bins. This process can be formulated as:

$$\Gamma'=\left\{f_{\lambda_1},f_{\lambda_2},\ldots,f_{\lambda_{pl}}\right\}\quad\text{if} \tag{8}$$

$$\mathrm{var}(X_1(t,f_{\lambda_1}))\geq\mathrm{var}(X_1(t,f_{\lambda_2}))\geq\ldots\geq\mathrm{var}(X_1(t,f_{\lambda_{pl}})) \tag{9}$$

where $\mathrm{var}(\ )$ denotes the variance.

The Estimation of Active Source Number for Selected Frequency Bins

In order to estimate $K\big|_{f=f^*}$, i.e., the number of active sources overlapping at a given selected frequency bin f^*, the PCA technology is adopted, which is based on the EVD of the covariance matrix of observed mixtures in the STTF domain.

At the beginning, a pre-treatment should be executed, that is, centralization by subtracting the mean of observed mixtures as follows:

$$\mathbf{X}(t,f^*)\leftarrow\mathbf{X}(t,f^*)-E\left[\mathbf{X}(t,f^*)\right] \tag{10}$$

Then the covariance matrix can be expressed by:

$$\mathbf{R}_{\mathbf{X}}(t,f^*)=E\left[\mathbf{X}(t,f^*)\mathbf{X}(t,f^*)^H\right] \tag{11}$$

where $(\ )^H$ denotes the complex conjugate transpose operator.

Considering the background noise which is modelled as additive white Gaussian noise (AWGN), statistically independent from the source signals with its mean and variance being 0 and σ^2 respectively, (1) should be rewritten as:

$$\mathbf{x}'(t)=\mathbf{As}(t)+\mathbf{n}(t) \tag{12}$$

Consequently, (3) becomes:

$$\mathbf{X}'(t,f)=\mathbf{AS}(t,f)+\mathbf{N}(t,f) \tag{13}$$

Then

$$\begin{aligned}\mathbf{R}_{\mathbf{X}'}(t,f^*)&=E\left[(\mathbf{AS}(t,f^*)+\mathbf{N}(t,f^*))(\mathbf{AS}(t,f^*)+\mathbf{N}(t,f^*))^H\right]\\&=E[\mathbf{AS}(t,f^*)\mathbf{S}(t,f^*)^H\mathbf{A}^H+\mathbf{AS}(t,f^*)\mathbf{N}(t,f^*)^H\\&\quad+\mathbf{N}(t,f^*)\mathbf{S}(t,f^*)^H\mathbf{A}^H+\mathbf{N}(t,f^*)\mathbf{N}(t,f^*)^H]\\&=\mathbf{R}_{\mathbf{X}}(t,f^*)+\mathbf{R}_{\mathbf{N}}(t,f^*)\\&\quad+\mathbf{A}E\left[\mathbf{S}(t,f^*)\mathbf{N}(t,f^*)^H\right]+E\left[\mathbf{N}(t,f^*)\mathbf{S}(t,f^*)^H\right]\mathbf{A}^H\end{aligned} \tag{14}$$

Due to the independence between the source and the noise, the following properties are satisfied:

$$E\left[\mathbf{S}(t,f^*)\mathbf{N}(t,f^*)^H\right]=0,\ E\left[\mathbf{N}(t,f^*)\mathbf{S}(t,f^*)^H\right]=0 \tag{15}$$

Hence, (12) can be simplified as:

$$\mathbf{R}_{\mathbf{X}'}(t,f^*)=\mathbf{R}_{\mathbf{X}}(t,f^*)+\mathbf{R}_{\mathbf{N}}(t,f^*) \tag{16}$$

Obviously under high SNR circumstances, (16) can be further reduced into:

$$\mathbf{R}_{\mathbf{X}'}(t,f^*)\approx\mathbf{R}_{\mathbf{X}}(t,f^*) \tag{17}$$

Perform EVD on $\mathbf{R}_{\mathbf{X}'}(t,f^*)$:

$$\mathbf{R}_{\mathbf{X}'}(t,f^*)=\boldsymbol{UDU}^H \tag{18}$$

where the unitary matrix $\boldsymbol{U}\in\mathbb{C}^{M\times M}$ is the so-called eigen vector matrix, $\boldsymbol{D}\in\mathbb{R}^{M\times M}$ is a diagonal matrix, i.e., $\boldsymbol{D}=diag(\sigma_{11}^2,\sigma_{22}^2,\ldots,\sigma_{MM}^2)$, and $\sigma_{11}^2\geq\sigma_{22}^2\geq\cdots\geq\sigma_{MM}^2\geq0$ are the so-called eigen values of $\mathbf{R}_{\mathbf{X}'}(t,f^*)$.

Definition: Supposing $\sigma_{11}^2\geq\sigma_{22}^2\geq\cdots\geq\sigma_{nn}^2\geq0$ are eigen values of the matrix $\mathbf{Z}$, if there exists a threshold $\varepsilon_\sigma\in\mathbb{N}$, making $\sigma_{jj}^2\gg\sigma_{ii}^2$ for $i>\varepsilon_\sigma$ and $j\leq\varepsilon_\sigma$, then ε_σ is called the main eigen number of $\mathbf{Z}$.

If the m-th source signal contributes at the frequency bin f^*, then:

(19) $$\sigma_{mm}^2(t,f^*) \gg 0$$

It is easy to recognize the fact that the number of active sources at the frequency bin f^* equals to the main eigen value number of the covariance matrix $\mathbf{R}_{\mathbf{X}'}(t,f^*)$. Therefore, for the M eigen values of $\mathbf{R}_{\mathbf{X}'}(t,f^*)$, and let

(20) $$\gamma_k = \sigma_{k,k}^2 / \sigma_{k+1,k+1}^2, \quad k=1,2,\ldots,M-1$$

If $\gamma_k = \max(\gamma_1,\gamma_2,\ldots,\gamma_{M-1})$, then

(21) $$K\big|_{f=f^*} = k$$

The estimation method described above has the advantage of simple calculation as well as satisfied accuracy.

Summary

Table 1 provides a summary of our proposed modified subspace-based UBSS algorithm.

Table 1. The proposed modified subspace-based UBSS algorithm

Step 1: STFT computation by (2). Step 2: Frequency bins selection by (8) and (9). Step 3: For selected frequency bins, perform evaluation of active source number by (20) and (21). Step 4: For unselected frequency bins, fix the active source number to be 1. Step 5: Perform subspace-based algorithm by (5) and (6). Step 6: Source TF synthesis by (7).

In the subspace algorithm proposed by [6], at each frequency bin the dimension of $\widetilde{\mathbf{A}}_\beta$ is $M\times(M-1)$, and the minimization in (6) should take the Q in (5) to be calculated $\binom{N}{M-1}$ times, where $\binom{m}{n}$ denotes $\frac{m!}{(m-n)!n!}$. However, in our modified subspace algorithm the extra complexity produced by the evaluation of the active source number can be compensated by two reasons. First, at the $p\times l$ selected frequency bins, the dimension of $\widetilde{\mathbf{A}}_\beta$ is $M\times K$, which is no more larger than $M\times(M-1)$ since $K\le M$, resulting in less complexity when calculating Q. Secondly and much more importantly, at the $(1-p)\times l$ unselected frequency bins, the dimension of $\widetilde{\mathbf{A}}_\beta$ is only $M\times 1$, besides, (5) need to perform only $\binom{N}{1}$ times, which is smaller than $\binom{N}{M-1}$ due to the fact that $M<N$. And the calculation complexity reduction can be rather considerable especially when p is small.

Simulation results and discussion

In the simulations, we use a uniform linear array of $M=4$ sensors to receive $N=5$ independent speech signals, which are not sparse in the time domain. With the given angles of arrival (i.e., $\theta_1=10^o$, $\theta_2=20^o$, $\theta_3=30^o$, $\theta_4=40^o$, and $\theta_5=50^o$), the true mixing matrix is

(22) $$\mathbf{A}=\begin{bmatrix} e^{-j\pi 0\sin(\theta_1)} & \cdots & e^{-j\pi 0\sin(\theta_N)} \\ \cdots & \cdots & \cdots \\ e^{-j\pi(M-1)\sin(\theta_1)} & \cdots & e^{-j\pi(M-1)\sin(\theta_N)} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 0.85-0.52\mathrm{i} & 0.48-0.88\mathrm{i} & -\mathrm{i} & -0.43-0.90\mathrm{i} & -0.74-0.67\mathrm{i} \\ 0.47-0.89\mathrm{i} & -0.55-0.84\mathrm{i} & -1 & -0.62+0.78\mathrm{i} & 0.10+0.99\mathrm{i} \\ -0.07-0.99\mathrm{i} & -0.99+0.08\mathrm{i} & \mathrm{i} & 0.97+0.22\mathrm{i} & 0.59-0.81\mathrm{i} \end{bmatrix}$$

Other experimental conditions are: sampling frequency 16kHz, STFT size 1024, Hanning window as the weighting function and total sample size 10240. And we apply the normalized mean square error (NMSE) criterion as the separation evaluation of the proposed algorithm, which is defined as:

(23) $$\mathrm{NMSE} = \frac{1}{N_r}\sum_{k=1}^{N_r}\frac{\left\|\mathbf{s}'^{(k)}-\mathbf{s}\right\|^2}{\left\|\mathbf{s}\right\|^2}$$

where N_r is the number of Monte Carlo simulation runs and $\mathbf{s}'^{(k)}$ is the estimation of source signal $\mathbf{s}$ in the k-th run.

Fig.1 shows the waveforms of the original source signals, the mixture signals and the estimated source signals in time domain, assuming that additive noise is absent. Clearly, all the source signals have been successfully recovered by the proposed modified algorithm.

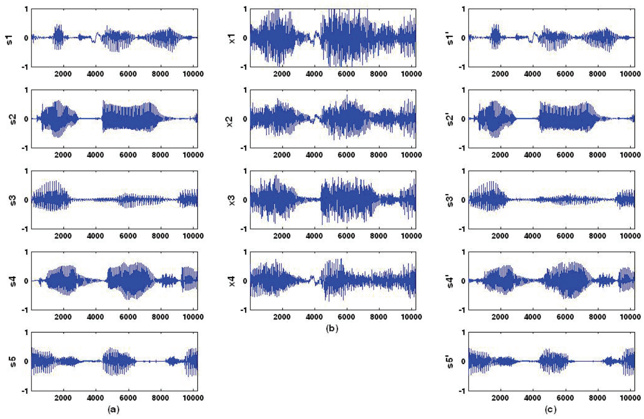

Fig.1. The waveforms in time domain for five speech sources and four sensors case. The left plots (a) represent the original source signals, the middle plots (b) represent the mixture signals, and the right plots (c) represent the source estimates.

Fig.2 illustrates the effects of p (i.e., the percent of frequency bins used) on NMSE under different SNR cases. For lower SNRs (for example, SNR=5dB and SNR=15dB), the NMSE performance achieve its best when p is around 10%, and NMSE degrades as p increases when $p>10\%$. However, for higher SNRs (for example, SNR=25dB and SNR=35dB), the more frequency bins are used, the better NMSE performance can be obtained, and the NMSE becomes stable when $p>10\%$. As we state before, approximate 10% frequency bins capture most of the signal power. When the SNR is low, using other frequency bins except the 10% frequency bins will introduce and aggravate the effect of the background noise, so it is not difficult for us to understand the deterioration of the performance. Otherwise, when the SNR is moderate or high, the contribution of the noise is relatively negligible due to its energy spreading, using other frequency bins can prohibit energy loss.

In Fig.3, we compare the NMSE performance with the subspace algorithm in [6] for $K=2$ as well as $K=3$, and the proposed modified algorithm for 10%, 40%, 70% and 100% frequency bins used respectively. One can observe that, for low SNRs, the proposed algorithm with 10% frequency bins can achieve much better separation performance than the subspace-based algorithm, but it becomes a litter worse for high SNRs. On the contrary, the proposed algorithm with 100% frequency bins can achieve much better separation performance than the subspace-based algorithm, but it is not satisfied enough for low SNRs. Then 40% frequency bins and 70% frequency bins seem to be ideal options since their performances improve obviously for both low SNRs and high SNRs. Considering the computation complexity, using 40% frequency bins is prefered.

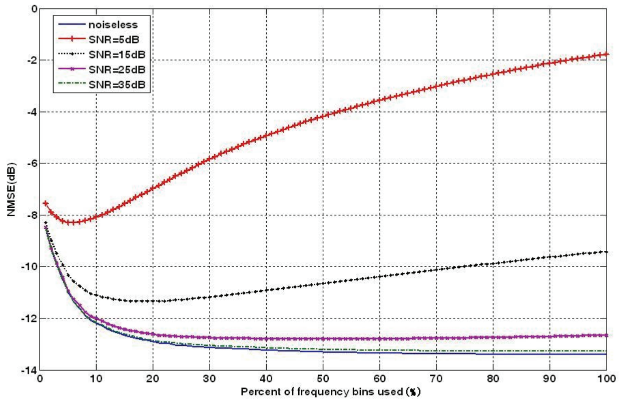

Fig.2. The effects of p (i.e., the percent of frequency bins used) on NMSE under different SNRs for the five speech sources and four sensors case.

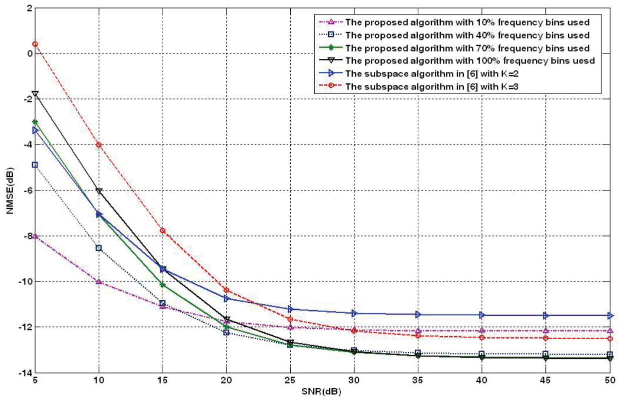

Fig.3. Comparison between the subspace algorithm in [6] and the proposed modified algorithm: NMSE versus SNR for the five speech sources and four sensors case.

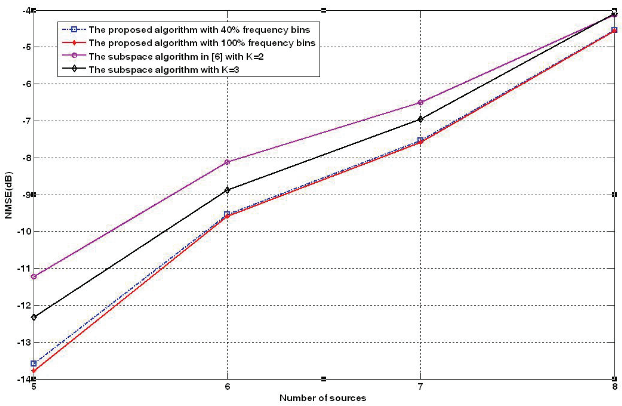

Fig.4.Comparison between the subspace algorithm in [6] and the proposed modified algorithm: NMSE versus number of sources.

In Fig.4, assuming no additive noise, comparision is made between the subspace algorithm in [6] and the proposed modified algorithm, to illustrate the relationship between the NMSE and the source number. Clearly, our proposed algorithm maintain significant advantage over the subspace-based algorithm, however, rapid and serious degradation of the separation quality occurs when the number of source is increased from $N=5$ to $N=8$. As N increases, the level of source interference increases and the sources are less likely to be sparse.

Conclusion

In this paper we have proposed a modified version of the classical subspace-based algortihm for the recovery of source signals in UBSS. Instead of fixing the active source number as a constant at all the frequency bins, the proposed algorithm chooses to estimate it by the PCA technology only at selected frequency bins, which capture most signal energy, and set the active source number fixed at other unselected frequency bins. Simulation results demonstrate that the proposed algorithm has much better NMSE performance (i.e., the separation accuracy) than the subspace-based algorithm. Besides, anlysis tells us that with proper number of selected frequency bins, the computation complexity will not suffer at all.

REFERENCES

[1] P. Bofill and M. Zibulevsky, Underdetermined blind source separation using sparse representations, *Signal Processing*, vol.81, pp.2353-2362, 2001.
[2] N. Linh-Trung, A. Belouchrani, K. Abed-Meraim, et al., Separating more sources than sensors using time-frequency distributions, *EURASIP J. Appl. Signals Process.*, vol.17, pp. 2828-2847, 2005.
[3] D. Luengo, I. Stantamaria, and L. Vielva, A general soulution to blind inverse problems for sparse input signals: deconvolution, equalization and source separation, *Neurocomputing*, vol.69, pp.198-215, 2005.
[4] Y. Li, S. Amari, A.Cichocki, et.al, Underdetermined blind source separation based on sparse representation, *IEEE Trans. Signal Process.*, vol.54, no.2, pp.423-437, 2006.
[5] O. Yilmaz 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.
[6] 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.*, vol.55, no.3, pp.897-907, Mar. 2007.
[7] L. Cohen, Time-Frequency Analysis, *Prentice-Hall PTR, Englewood Cliffs*, NJ, 1995.
[8] A. Jourjine, S. Rickard and O. Yilmaz, Blind separation of disjoint orthogonal signals: Demixing N sources form 2 mixtures, in *Proc. ICASSP2000*, Istanbul, Turkey, Jun. 2000.
[9] A. Hyvarinen, J. Karhunen and E. Oja, Independent component anlysis, *John Wiley & Sons, Inc.*, 2001.
[10] M. Wax and T. Kailath, Detection of signals by information theoretic criteria, *IEEE Trans. Acoust., Speech, Signal Process.*, vol.ASSP-33, no.2, pp.387-392, 1985.
[11] R. Saab, O. Yilmaz, M.J. Mckeown, et al., Underdetermined anechoic blind source separation via l^q –basis-pursuit with $q<1$, *IEEE Trans. Signal Process.*, vol.55, no.8, pp.4004-4017, 2007.
[12] R. Xu, D. Wunsch II, Survey of clustering algorithm, *IEEE Trans. Neural Networks*, vol.16, no.3, pp:645-678, 2005.
[13] P. Bofill, Identifying single source data for mixing matrix estimation in instantaneous blind source separation, in: *ICANN(1)*, pp:759-767, 2008.
[14] V. G. Reju, S. N. Koh and I. Y. Soon, An algorithm for mixing estimation in instantaneous blind source separation, *Signal Processing*, vol.89, pp.1762-1773, 2009.
[15] D. Peng and Y. xiang, Underdetermined blind separation based on relaxed sparsity condition of sources, *IEEE Trans. Signal Process.*, vol.57, no.2, pp.809-814, 2009.
[16] S. Rickard and M. Fallon, The GINI index of speech, in *Conf. Inf. Sci. Syst.*, 2004.

***Authors**: Dr. Hui Li, Wireless Communication Faculty, Institute of Communications Engineering, E-mail: leehoo86@163.com; Prof. Yue-hong Shen, Wireless Communication Faculty, Institute of Communications Engineering, E-mail: chunfeng22259@126.com; M.S. Min Cao,* Xi'an Communication Institute, *E-mail: shrana@126.com; Dr. Jian-gong Wang, Wireless Communication Faculty, Institute of Communications Engineering, E-mail: hw12xian@hotmail.*