UNDERDETERMINED CONVOLUTIVE BLIND SOURCE SEPARATION USING A NOVEL MIXING MATRIX ESTIMATION AND MMSE-BASED SOURCE ESTIMATION

Janghoon Cho, Jinho Choi, and Chang D. Yoo

Div. of EE, Korea Advanced Institute of Science & Technology

{bluehawk2k,cjh3836}@kaist.ac.kr, cdyoo@ee.kaist.ac.kr

ABSTRACT

This paper considers underdetermined blind source separation of super-Gaussian signals that are convolutively mixed. The separation is performed in three stages. In the first stage, the mixing matrix in each frequency bin is estimated by the proposed single source detection and clustering (SSDC) algorithm. In the second stage, by assuming complex-valued super-Gaussian distribution, the sources are estimated by minimizing a mean-square-error (MSE) criterion. Special consideration is given to reduce computational load without compromising accuracy. In the last stage, the estimated sources in each frequency bin are aligned for recovery. In our simulations, the proposed algorithm outperformed conventional algorithm in terms of the mixing-error-ratio and the signal-to-distortion ratio.

1. INTRODUCTION

A blind source separation (BSS) system recovers unobserved sources from a number of observed mixtures without knowing the mixing system [1–4]. This paper considers the separation of super-Gaussian source signals that are convolutively mixed, and the number of sources is assumed larger than the number of mixtures. The considered BSS system is characterized as underdetermined and convolutive.

Many convolutive BSS systems have been proposed in the past, and generally, separation is performed in the Fourier domain such that convolutive BSS problem can be considered as an instantaneous BSS problem [2]. In each frequency bin, independent component analysis (ICA) can be employed after which appropriate component alginment must be carried out. This alignment is often referred to as solving the permutation problem [5].

When the number of the sources is larger than that of the mixtures, the ICA cannot be applied since the sources cannot be directly obtained, and most conventional algorithms assume additional constraints [4].

The convolutive and underdetermined BSS problem is recognized as a challenging task. The time-frequency(T-F) masking [6–8] algorithm is widely used in this case. The algorithm assumes that at most one of the sources is active in a particular T-F point (disjoint region), and this assumption is unrealistic since these points are often scarce in reality. As the number of sources increases, such non-disjoint regions also increase and this situation may lead to poor performance. Another popular approach is based on a maximum a posteriori (MAP) estimator [9]. It also assumes that the sources follow Laplacian distribution and separates the sources by l_1-norm minimization.

In this paper, an underdetermined convolutive BSS algorithm which can -to a certain degree- overcome the limitations mentioned above is proposed. A novel framework for estimating the mixing matrix and a minimum mean-square-error (MMSE) based approach for source estimation are involved in the algorithm. The algorithm consists of three stages and performs in the T-F domain. The mixing matrix in each frequency bin is estimated by the considered single source detection and clustering (SSDC) algorithm in the first stage. The sources are estimated by proposed MMSE-based algorithm requiring low computational load in the second stage. The source coefficients in T-F domain are assumed to follow complex-valued super-Gaussian distribution. Since these two stages are performed in each frequency bin, the permutation ambiguities among the frequency bins should be solved. Therefore, the estimated sources in each frequency bin are aligned in last stage.

The rest of the paper is organized as follows. Section 2 briefly describes problem formulation of the underdetermined convolutive BSS. Section 3 describes the proposed algorithm and its subsections describe each of three stages. A number of experimental results are presented and discussed in Section 4. Finally, Section 5 concludes this paper.

2. PROBLEM FORMULATION

In this section, the problem which the underdetermined convolutive BSS considers and the assumptions made to tackle this problem is described.

2.1. Problem description

Let N source signals $s_i[n]$ be convolutively mixed and observed at M sensors $x_j[n]$ for $i = 1, \ldots, N$ and $j = 1, \ldots, M$ such that

$$x_j[n] = \sum_{i=1}^{N} \sum_{l} h_{ji}[l] s_i[n-l] \quad (1)$$

$$= \sum_{i=1}^{N} s_{ji}^{\text{img}}[n], \quad (2)$$

where $h_{ji}[n]$ and $s_{ji}^{\text{img}}[n]$ are the impulse response from the ith source to the jth sensor and the spatial image of the ith source on the jth channel, respectively. Assume that N is known with $M < N$. The objective is to estimate $s_{ji}^{\text{img}}[n]$ from $x_j[n]$ for $j = 1, \ldots, M$.

In the T-F domain, convolutive mixtures can be approximated as instantaneous mixtures in each frequency bin:

$$\mathbf{X}[\tau, k] = \begin{pmatrix} H_{11}[k] & \cdots & H_{1N}[k] \\ \vdots & \vdots & \vdots \\ H_{M1}[k] & \cdots & H_{MN}[k] \end{pmatrix} \begin{pmatrix} S_1[\tau, k] \\ \vdots \\ S_N[\tau, k] \end{pmatrix}, \quad (3)$$

where $\mathbf{X}[\tau, k] = [X_1[\tau, k], \ldots, X_M[\tau, k]]^T$ is the column vector of the short-time Fourier transform (STFT) coefficients of the M mixtures, and $S_i[\tau, k]$ is the STFT coefficient of $s_i[n]$ at time frame τ and frequency bin k. Here, $H_{ji}[k]$ is the frequency response of $h_{ji}[n]$. To estimate $s_{ji}^{\text{img}}[n]$, Equation (3) can be rewritten as follows :

$$\mathbf{X}[\tau, k] = \begin{pmatrix} \frac{H_{11}[k]}{H_{j1}[k]} & \cdots & \frac{H_{1N}[k]}{H_{jN}[k]} \\ \vdots & \vdots & \vdots \\ \frac{H_{M1}[k]}{H_{j1}[k]} & \cdots & \frac{H_{MN}[k]}{H_{jN}[k]} \end{pmatrix} \begin{pmatrix} H_{j1}[k] S_1[\tau, k] \\ \vdots \\ H_{jN}[k] S_N[\tau, k] \end{pmatrix} \quad (4)$$

$$= \mathbf{A}[k] \mathbf{S}_j^{\text{img}}[\tau, k], \quad (5)$$

where $\mathbf{S}_j^{\text{img}}[\tau, k] = [S_{j1}^{\text{img}}[\tau, k], \ldots, S_{jN}^{\text{img}}[\tau, k]]^T$ and $S_{ji}^{\text{img}}[\tau, k]$ is the STFT of $s_{ji}^{\text{img}}[n]$, since $s_{ji}^{\text{img}}[n]$ is the inverse-STFT of $H_{ji}[k] S_i[\tau, k]$. Henceforth, the subscript and superscript of $\mathbf{S}_j^{\text{img}}[\tau, k]$ are omitted as $\mathbf{S}[\tau, k]$ for simplicity.

2.2. Assumptions

This paper makes two assumptions about the sources :

1. The sources are mutually independent one another such that

$$p(S_1, \cdots, S_N) = \prod_{i=1}^{N} p(S_i). \quad (6)$$

2. The number of sources N is known and is larger than the number of observations M ($M < N$).

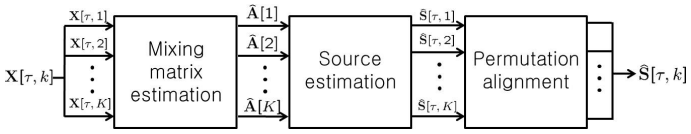

Fig. 1. Block diagram of the proposed algorithm

3. PROPOSED APPROACH

The proposed algorithm consists of three stages as shown in Figure 1. Detailed description of each stage is given in the following subsections.

3.1. Mixing matrix estimation based on SSDC

In the first stage, complex-valued mixing matrix $\mathbf{A}[k]$ is estimated for each frequency bin. Conventional algorithm estimates $\mathbf{A}[k]$ based on hierarchical clustering [9]. The algorithm proposed in [9] assumes that there are many T-F points where only a single source is active for each source under the sparseness assumption. The T-F points where more than one source is active occur as outliers, and these points lower the estimation accuracy of the column vector. To prevent clustering involving these outliers, additional parameter should be selected well. The proposed algorithm detects T-F points where only one source is active. Given these points, a well-known k-means clustering algorithm is applied.

To find a set of T-F points, denoted as $\mathcal{S}_{s,k}$ where only a single source is active for each source in the kth frequency bin, the single source detection (SSD) algorithm based on the ratio of the T-F transforms is described. The algorithm assumes that there exists at least one pair of two consecutive T-F points of single source occupancy (SSO) for each source in every frequency bin. The mixing matrix is estimated based on the ratio calculated at the T-F points of SSO. The detection is conducted as follows.

For a given $\epsilon > 0$, a set $\mathcal{S}_{s,k}$ of T-F points where only a single source is active is detected such that

$$\mathcal{S}_{s,k} = \left\{ [\tau, k], [\tau+1, k] \,\middle|\, \left\| \text{Im}\left(\frac{X_m[\tau, k] X_m^*[\tau+1, k]}{X_j[\tau, k] X_j^*[\tau+1, k]} \right) \right\| < \epsilon, \text{ for } \forall j, m \in \{1, \ldots, M\} \right\}, \quad (7)$$

where $\text{Im}(x)$ denotes the imaginary part of x and $*$ stands for complex conjugation. When the only i-th source is ac-

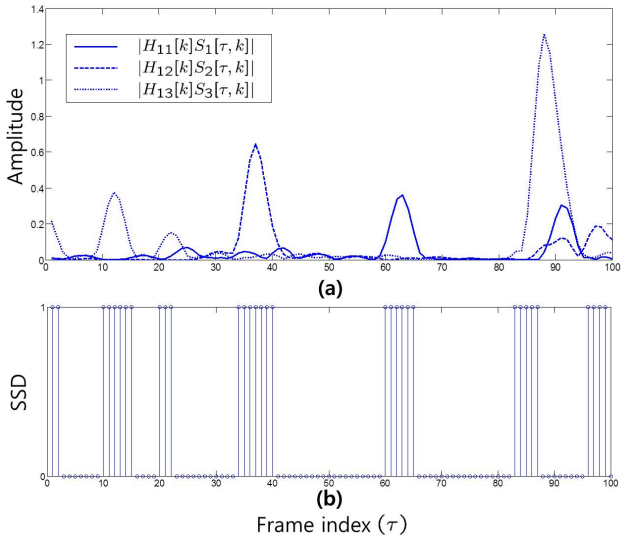

Fig. 2. Detected T-F points of SSO for frequency bin $k = 150$. (a) Amplitude envelope of the spatial image of the ith source ($i = 1, 2, 3$) to the first sensor (b) Detected T-F points

tive at two consecutive points $[\tau_s, k_s], [\tau_s + 1, k_s]$, then

$$\begin{aligned} & \left\| \mathrm{Im}\left(\frac{X_m[\tau_s,k_s]X_m^*[\tau_s+1,k_s]}{X_j[\tau_s,k_s]X_j^*[\tau_s+1,k_s]} \right) \right\| \\ = & \left\| \mathrm{Im}\left(\frac{H_{mi}[k_s]S_i[\tau_s,k_s]H_{mi}^*[k_s]S_i^*[\tau_s+1,k_s]}{H_{ji}[k_s]S_i[\tau_s,k_s]H_{ji}^*[k_s]S_i^*[\tau_s+1,k_s]} \right) \right\| \\ = & \left\| \mathrm{Im}\left(\left| \frac{H_{mi}[k_s]}{H_{ji}[k_s]} \right|^2 \right) \right\| = 0 \ \text{ for } \ \forall j,m \in \{1,\ldots,M\}. \end{aligned}$$

Thus, $[\tau_s, k_s], [\tau_s + 1, k_s]$ are included in $\mathcal{S}_{s,k_s}$.

At any T-F points in $\mathcal{S}_{s,k}$ where the only the ith source is active, the ratio vector of mixtures is as follows :

$$\frac{\mathbf{X}[\tau,k]}{X_j[\tau,k]} = \left[\frac{H_{1i}[k]}{H_{ji}[k]} \cdots \frac{H_{Mi}[k]}{H_{ji}[k]} \right]^T = \mathbf{a}_i[k], \tag{8}$$

where $\mathbf{a}_i[k]$ denotes the ith column vector of $\mathbf{A}[k]$.

To estimate $\mathbf{a}_i[k]$, the T-F points in $\mathcal{S}_{s,k}$ are clustered into N classes based on following ratio vector

$$\frac{\mathbf{X}[\tau,k]}{X_j[\tau,k]}, \forall [\tau,k] \in \mathcal{S}_{s,k}. \tag{9}$$

Here, a well-known k-means clustering algorithm is used to cluster the T-F points in $\mathcal{S}_{s,k}$ and the set of the T-F points in ith cluster is denoted as $\mathcal{S}_{\mathcal{C}_i,k}$ for $i = 1, \ldots, N$. It should be noted that the k-means algorithm is sensitive to the initial condition. Considering the continuity of mixing matrices between adjacent frequency bins, initial centroids are set to the column vectors of the mixing matrix estimated in previous frequency bin.

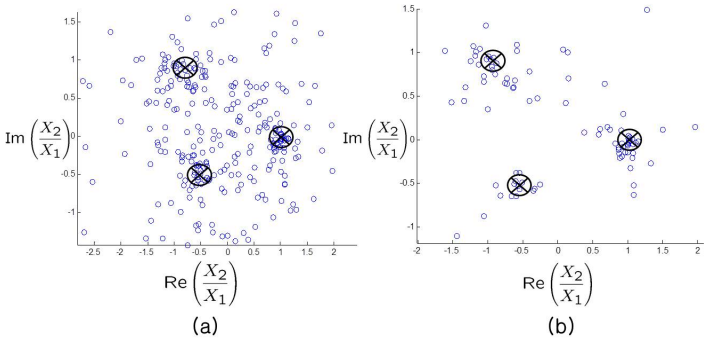

Fig. 3. Scatter plot of ratio vectors in $k = 150$. The cross in the circle indicates $\frac{H_{2i}}{H_{1i}}, i = 1, 2, 3$ (a) Whole T-F points (b) T-F points included in $\mathcal{S}_{s,150}$

Given the ratio vectors in $\mathcal{S}_{\mathcal{C}_i,k}$, $\mathbf{a}_i[k]$ can be estimated to be the centroid of the ith cluster as follows:

$$\hat{\mathbf{a}}_i[k] = \frac{1}{|\mathcal{S}_{\mathcal{C}_i,k}|} \sum_{[\tau,k] \in \mathcal{S}_{\mathcal{C}_i,k}} \frac{\mathbf{X}[\tau,k]}{X_j[\tau,k]}, \tag{10}$$

where $|\mathcal{S}_{\mathcal{C}_i}|$ represents the number of the points in the class for $i = 1, \ldots, N$.

The amplitude envelope of the three sources and detected T-F points of SSO in the frequency bin $k = 150$ are shown in Figure 2 (a) and (b), respectively. It shows that when one source is considerably more dominant than the others, these points are considered as $\mathcal{S}_s$, in other words, SSD $= 1$. Figure 3 (a) illustrates the scatter plot of ratio vectors of the whole T-F points in $k = 150$. In Figure 3 (b), the ratio vectors of the T-F points included in $\mathcal{S}_{s,150}$ are illustrated. The scatters are concentrated around true ratio of $\mathbf{a}_i[k]$ when the SSD algorithm is applied. It leads to a good estimation of $\mathbf{A}[k]$.

When SSO does not exist in a particular frequency bin, the algorithm cannot estimate the mixing matrix, but this rarely happens and the conventional algorithm proposed in [9] is employed in those frequency bins.

3.2. MMSE-based source estimation

Given the estimated mixing matrix $\hat{\mathbf{A}}[k]$, we estimate $S_i[\tau,k]$ in the second stage. Here, we assume that $S_i[\tau,k]$ follows complex-valued generalized Gaussian distribution. The phase of $S_i[\tau,k]$ is assumed to be uniformly distributed in $[-\pi, \pi]$ and its magnitude is assumed to be distrubuted as follows :

$$p(|S_i[\tau,k]|) = c\frac{\beta^{1/c}}{\Gamma(1/c)} e^{-\beta|S_i[\tau,k]|^c}, \tag{11}$$

where the parameters c and β are the shape and the variance of the distribution, respectively. Also, $c, \beta > 0$. When $c = 1$, the distribution is Laplacian and when $c = 2$, it is Gaussian. With decreasing c sparsity increases. In this paper, the source prior is assumed to follow the super-Gaussian, $c < 1$.

Even with the true mixing matrix $\mathbf{A}[k]$, the sources cannot be recovered directly since $\mathbf{A}[k]$ is not square and there are infinitely many solutions satisfying Equation (5). Based on subspace representation, the sources can be decomposed as follows [1]:

$$\mathbf{S} = \mathbf{A}^{\dagger}\mathbf{X} + \mathbf{V}\mathbf{z}, \tag{12}$$

where $\mathbf{A}^{\dagger}$ is the Moore-Penrose pseudo inverse matrix of $\mathbf{A}$, $\mathbf{V}$ is an $N \times (N-M)$ matrix whose columns are bases of the nullspace of $\mathbf{A}$ and $\mathbf{z}$ is an $(N-M)\times 1$ arbitrary vector, respectively. As shown in [4], $\mathbf{A}^{\dagger}\mathbf{X}$ indicates the rowspace component of $\mathbf{S}$ which can be directly recovered and $\mathbf{V}\mathbf{z}$ indicates the nullspace component of $\mathbf{S}$ which should be inferred. Here, the problem of estimating $\mathbf{S}$ boils down to the problem of estimating $\mathbf{z}$.

We estimate the sources by minimizing the following MSE cost function:

$$\hat{\mathbf{S}}_{\text{MS}} = \arg\min_{\hat{\mathbf{S}}} E_{p(\mathbf{S}|\mathbf{X})}[||\mathbf{S} - \hat{\mathbf{S}}||^2], \quad \text{s.t. } \mathbf{X} = \hat{\mathbf{A}}\mathbf{S}. \tag{13}$$

As shown in [4], the cost function of $\mathbf{S}$ can be expressed as that of $\mathbf{z}$ as follows :

$$\hat{\mathbf{z}}_{\text{MS}} = \arg\min_{\hat{\mathbf{z}}} E_{p(\mathbf{z}|\mathbf{X})}[||\mathbf{z} - \hat{\mathbf{z}}||^2]. \tag{14}$$

This cost function will be minimized when $\hat{\mathbf{z}}$ is equal to the posterior mean,

$$\hat{\mathbf{z}}_{\text{MS}} = \int \mathbf{z} p(\mathbf{z}|\mathbf{X}) d\mathbf{z}. \tag{15}$$

In [4], it is approximated by Monte Carlo integration as follows:

$$\hat{\mathbf{z}}_{\text{MS}} \approx \frac{1}{J}\sum_{l=1}^{J} \mathbf{z}^{(l)}, \tag{16}$$

where $\mathbf{z}^{(1)}, \ldots, \mathbf{z}^{(J)}$ are J drawn samples from $p(\mathbf{z}|\mathbf{X})$. However, when we use the sampling method, the computational load increases exponentially as the dimension of $\mathbf{z}$ increases. Rather than using the sampling method, an approximation is used to reduce the computational load.

When $c < 1$, $p(\mathbf{z}|\mathbf{X})$ has ${}_N\mathrm{C}_M$ non-differentiable local maximums [2]. These maximums are located at $\mathbf{z}^{\star}_m, m = 1, \ldots, {}_N\mathrm{C}_M$ where $(N-M)$ components of $\mathbf{S}$ are zero. To reduce the computational load of the algorithm, $p(\mathbf{z}|\mathbf{X})$ can be approximated as follows :

$$p(\mathbf{z}|\mathbf{X}) \approx \frac{1}{Z_p}\sum_{m=1}^{{}_N\mathrm{C}_M} p(\mathbf{S} = \mathbf{A}^{\dagger}\mathbf{X} + \mathbf{V}\mathbf{z})\delta(\mathbf{z} - \mathbf{z}^{\star}_m), \tag{17}$$

where $Z_p = \sum_{m=1}^{{}_N\mathrm{C}_M} p(\mathbf{S} = \mathbf{A}^{\dagger}\mathbf{X} + \mathbf{V}\mathbf{z}^{\star}_m)$ and $\delta(\cdot)$ denotes the Dirac delta function.

By Equation (15) and (17), the MMSE estimates of $\mathbf{z}$ can be derived as follows :

$$\hat{\mathbf{z}}_{\text{MS}} \approx \frac{1}{Z_p}\sum_{m=1}^{{}_N\mathrm{C}_M} p(\mathbf{S} = \mathbf{A}^{\dagger}\mathbf{X} + \mathbf{V}\mathbf{z}^{\star}_m)\mathbf{z}^{\star}_m. \tag{18}$$

Using $\hat{\mathbf{z}}_{\text{MS}}$, the sources satisfying the MSE criterion, $\hat{\mathbf{S}}_{\text{MS}}$, can be expressed as

$$\hat{\mathbf{S}}_{\text{MS}} = \mathbf{A}^{\dagger}\mathbf{X} + \mathbf{V}\hat{\mathbf{z}}_{\text{MS}}. \tag{19}$$

3.3. Permutation alignment and scaling ambiguity

Now the estimated sources in each frequency bin must be algined. Clustering-based method using the correlation between adjacent frequency bins [7] is considered.

The scaling ambiguity can be ignored since the components of the jth row of $\mathbf{A}$ are set to 1 in Equation (5).

4. EXPERIMENT

In this section, the evaluation criteria is explained and simulation results for benchmark dataset are reported.

4.1. Evaluation criteria

We used the evaluation criteria introduced in [10] and [11] for measuring the performance of estimating the mixing matrix and sources.

The estimated ith column vector of $\mathbf{A}$ can be decomposed as

$$\hat{\mathbf{a}}_i = \mathbf{a}_i^{\text{coll}} + \mathbf{a}_i^{\text{orth}}, \tag{20}$$

where $\mathbf{a}_i^{\text{coll}}$ and $\mathbf{a}_i^{\text{orth}}$ denote collinear and orthogonal components of $\mathbf{a}_i$, respectively. They can be computed by least squares projection. Based on the decomposition of $\hat{\mathbf{a}}$, the mixing-error-ratio (MER) is defined as

$$\text{MER}_i = 10\log\frac{||\mathbf{a}_i^{\text{coll}}||^2}{||\mathbf{a}_i^{\text{orth}}||^2}. \tag{21}$$

For convolutive mixtures, the MER in each frequency bin k is computed and averaged over all frequency bins.

The estimated source image can be decomposed as

$$\hat{s}_{ji}^{\text{img}}[n] = s_{ji}^{\text{img}}[n] + e_{ji}^{\text{spat}}[n] + e_{ji}^{\text{interf}}[n] + e_{ji}^{\text{artif}}[n], \tag{22}$$

[1] Henceforth, the indexes of $\mathbf{S}[\tau, k]$, $\mathbf{X}[\tau, k]$ and $\mathbf{A}[k]$ are omitted as $\mathbf{S}$, $\mathbf{X}$ and $\mathbf{A}$ for simplicity.

[2] ${}_N\mathrm{C}_M = \frac{N!}{(N-M)!M!}$

Table 1. Experimental conditions

Number of microphones	$M = 2$
Number of sources	$N = 3$ or 4
Mic spacing	5cm or 1m
Source signals	Speeches of 10s
Reverberation time	130ms or 250ms
Sampling rate	16kHz
STFT frame size	2048samples (128ms)
STFT frame shift	256samples (16ms)

where $s_{ji}^{\text{img}}[n]$ is the true source image and $e_{ji}^{\text{spat}}[n]$, $e_{ji}^{\text{interf}}[n]$ and $e_{ji}^{\text{artif}}[n]$ are error components caused by spatial distortion, interference and artifacts, respectively. Performance measures of estimating sources, the source image-to-spatial distortion-ratio (ISR), the source-to-interference ratio (SIR) and the source-to-artifacts ratio (SAR) are defined as follows :

$$\text{ISR}_i = 10\log\frac{\sum_{j=1}^{M}\sum_n s_{ji}^{\text{img}}[n]^2}{\sum_{j=1}^{M}\sum_n e_{ji}^{\text{spat}}[n]^2}, \quad (23)$$

$$\text{SIR}_i = 10\log\frac{\sum_{j=1}^{M}\sum_n (s_{ji}^{\text{img}}[n] + e_{ji}^{\text{spat}}[n])^2}{\sum_{j=1}^{M}\sum_n e_{ji}^{\text{interf}}[n]^2}, \quad (24)$$

$$(25)$$

and

$$\text{SAR}_i = 10\log\frac{\sum_{j=1}^{M}\sum_n (s_{ji}^{\text{img}}[n]+e_{ji}^{\text{spat}}[n]+e_{ji}^{\text{interf}}[n])^2}{\sum_{j=1}^{M}\sum_n e_{ji}^{\text{artif}}[n]^2}. \quad (26)$$

The overall evaluation criterion, the signal-to-distortion ratio (SDR) is defined as

$$\text{SDR}_i = 10\log\frac{\sum_{j=1}^{M}\sum_n s_{ji}^{\text{img}}[n]^2}{\sum_{j=1}^{M}\sum_n (e_{ji}^{\text{spat}}[n]+e_{ji}^{\text{interf}}[n]+e_{ji}^{\text{artif}}[n])^2}. \quad (27)$$

4.2. Experimental setups and SiSEC 2008 dataset

Experiment is performed on publicly available benchmark dataset organized in the Signal Separation Evaluation Campaign (SiSEC 2008) [11]. The first development data-set in "Under-determined speech and music mixtures" that includes 20 different sets of (synthetic/live recording) mixtures with varying type, reverberation time and microphone spacing is used. Here, we used 16 sets of speech sources. In Table 1, the experimental conditions are described.

Experimental results of synthetically mixed data and live recording data are shown in Table 2 and 3, respectively. The proposed algorithm was compared to the conventional algorithm based on the MAP-based approach [9]. Five averaged measures of both proposed and conventional algorithms are shown in Table 2. Since there is no information about the mixing parameters in the live recording data, four averaged measures except MER are shown in Table 3. As shown, the proposed algorithm estimated the mixing matrices and separated the sources better than the conventional algorithm. As for computational loads, the proposed algorithm coded in MATLAB separated the mixtures in 4 minutes ($N = 3$ sources) for 10 second mixture. Considering that the sampling method takes more than an hour, the computational time has been dramatically reduced.

5. CONCLUSION

This paper considers the problem of blindly separating the sources of super-Gaussian distribution from underdetermined convolutive mixtures. In this paper, a novel algorithm which consists of three stage is proposed. In the first stage, the mixing matrix in each frequency bin is estimated by proposed SSDC algorithm. The proposed algorithm has fewer parameters to tune and its performance is better than that of conventional algorithm. In second stage, given the estimated mixing matrix, the sources which follow super-Gaussian prior are estimated by minimizing the MSE criteria. In the last stage, the permutation between frequency bins is algined.

Simulation results show that the proposed algorithm estimated the mixing matrix with higher MER and separates super-Gaussian sources with higher SDR than the conventional algorithm.

6. ACKNOWLEDGEMENT

This work was supported (National Robotics Research Center for Robot Intelligence Technology, KAIST) by Ministry of Knowledge Economy under Human Resources Development Program for Convergence Robot Specialists.

7. REFERENCES

[1] T.W. Lee and T. Sejnowski, *Independent component analysis: theory and applications*, vol. 474, Kluwer academic publishers Boston, MA:, 1998.

[2] S. Makino, T.W. Lee, and H. Sawada, *Blind speech separation*, Springer Verlag, 2007.

[3] S. Haykin, "Unsupervised Adaptive Filtering vol. 1: Blind Source Separation ," 2000.

[4] S.G. Kim and C.D. Yoo, "Underdetermined blind source separation based on subspace representation," *Signal Processing, IEEE Transactions on*, vol. 57, no. 7, pp. 2604–2614, 2009.

Table 2. Performance measure on synthetically mixed data

Type		female3				male3				female4				male4			
RT_{60}		130ms		250ms		130ms		250ms		130ms		250ms		130ms		250ms	
mic spacing		5cm	1m	5cm	1m	5cm	1m	5cm	1m	5cm	1m	5cm	1m	5cm	1m	5cm	1m
SDR	Proposed	**7.21**	**7.09**	**3.93**	**2.85**	**5.35**	**4.59**	**3.72**	**2.85**	**3.58**	**2.75**	**2.79**	**2.05**	**2.53**	**2.69**	**2.15**	**1.51**
	MAP	2.77	2.88	0.19	0.74	1.10	1.63	-0.60	0.03	0.11	0.33	-0.89	-0.51	0.02	0.63	-0.84	-0.45
ISR	Proposed	**11.95**	**11.82**	**7.75**	**6.76**	**10.33**	**9.90**	**7.70**	**7.14**	**7.19**	**6.07**	**3.77**	**4.13**	**6.06**	**5.92**	**5.44**	**4.72**
	MAP	7.38	7.74	4.07	3.82	5.00	5.88	3.19	2.78	4.36	3.86	3.11	2.72	4.10	5.03	3.43	3.17
SIR	Proposed	**11.11**	**10.86**	**6.72**	**4.83**	**8.54**	**7.10**	**6.95**	**4.85**	**5.82**	**4.11**	**4.35**	**3.63**	**3.32**	**4.59**	**2.56**	**1.56**
	MAP	6.84	5.59	3.94	2.46	4.77	4.55	0.73	1.37	1.60	0.96	0.42	-1.00	1.29	1.78	0.00	-0.68
SAR	Proposed	**11.08**	**10.58**	**8.01**	**6.48**	**8.42**	**8.18**	**6.16**	**5.90**	**6.75**	**7.33**	**5.81**	**4.79**	**5.68**	**5.16**	**5.06**	**4.08**
	MAP	4.81	5.20	1.87	2.55	2.53	3.54	0.86	0.79	3.38	3.46	3.56	1.56	1.86	2.48	1.44	0.84
MER	Proposed	**20.19**	**16.99**	**14.84**	**12.39**	**19.41**	**16.27**	**14.61**	**11.93**	**17.33**	**14.47**	**13.86**	**11.46**	**16.97**	**14.38**	**13.70**	**11.39**
	MAP	17.20	14.24	13.26	10.85	16.50	13.83	12.91	10.43	15.38	13.00	12.67	10.56	15.14	13.08	12.57	10.42

Table 3. Performance measure on live recording data

Type		female3				male3				female4				male4			
RT_{60}		130ms		250ms		130ms		250ms		130ms		250ms		130ms		250ms	
mic spacing		5cm	1m	5cm	1m	5cm	1m	5cm	1m	5cm	1m	5cm	1m	5cm	1m	5cm	1m
SDR	Proposed	**6.00**	**7.57**	**5.45**	**4.25**	**5.79**	**5.40**	**4.29**	**4.47**	**4.18**	**2.90**	**3.23**	**1.94**	**2.84**	**1.87**	**2.14**	**3.22**
	MAP	1.00	2.17	0.63	2.71	3.02	2.11	0.81	2.22	0.67	0.34	0.94	0.78	0.06	0.47	0.71	1.92
ISR	Proposed	**10.37**	**12.86**	**9.12**	**9.26**	**11.57**	**10.29**	**8.86**	**9.25**	**4.45**	**6.94**	**4.43**	**5.46**	**6.27**	**5.36**	**5.54**	5.51
	MAP	5.89	6.62	4.70	7.72	8.17	7.09	5.66	6.97	3.84	3.87	3.96	4.28	3.20	4.21	4.11	**5.60**
SIR	Proposed	**9.58**	**12.30**	**8.43**	**7.10**	**9.21**	**9.33**	**7.25**	**7.72**	**6.46**	**3.73**	**4.27**	**2.72**	**4.31**	**2.06**	**1.13**	**4.84**
	MAP	1.38	4.49	1.28	5.49	6.97	3.97	0.91	4.56	0.47	-0.36	0.33	0.20	-1.18	-0.35	0.11	3.17
SAR	Proposed	**9.97**	**10.25**	**7.30**	**7.79**	**9.01**	**8.08**	**6.99**	**6.94**	**5.84**	**6.62**	**5.41**	**5.44**	**5.61**	**5.16**	**5.42**	**4.36**
	MAP	4.88	5.69	4.62	6.42	5.21	5.57	3.62	4.45	4.27	5.04	5.12	4.78	3.87	3.90	3.19	4.04

[5] H. Sawada, R. Mukai, S. Araki, and S. Makino, "A robust and precise method for solving the permutation problem of frequency-domain blind source separation," *Speech and Audio Processing, IEEE Transactions on*, vol. 12, no. 5, pp. 530–538, 2004.

[6] S. Araki, H. Sawada, R. Mukai, and S. Makino, "Underdetermined blind sparse source separation for arbitrarily arranged multiple sensors," *Signal Processing*, vol. 87, no. 8, pp. 1833–1847, 2007.

[7] V.G. Reju, S.N. Koh, and Y. Soon, "Underdetermined convolutive blind source separation via time–frequency masking," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 18, no. 1, pp. 101–116, 2010.

[8] H. Sawada, S. Araki, and S. Makino, "Underdetermined convolutive blind source separation via frequency bin-wise clustering and permutation alignment," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 19, no. 3, pp. 516–527, 2011.

[9] S. Winter, W. Kellermann, H. Sawada, and S. Makino, "Map-based underdetermined blind source separation of convolutive mixtures by hierarchical clustering and l1-norm minimization," *EURASIP Journal on Applied Signal Processing*, vol. 2007, no. 1, pp. 1–12, 2007.

[10] E. Vincent, H. Sawada, P. Bofill, S. Makino, and J. Rosca, "First stereo audio source separation evaluation campaign: data, algorithms and results," *Independent Component Analysis and Signal Separation*, pp. 552–559, 2007.

[11] E. Vincent, S. Araki, and P. Bofill, "The 2008 signal separation evaluation campaign: A community-based approach to large-scale evaluation," *Independent Component Analysis and Signal Separation*, pp. 734–741, 2009.