15th European Signal Processing Conference (EUSIPCO 2007), Poznan, Poland, September 3-7, 2007, copyright by EURASIP

A FEEDBACK APPROACH TO OVER COMPLETE BSS AND ITS LEARNING ALGORITHM

Kenji NAKAYAMA, Haruo KATOU, and Akihiro HIRANO

Graduate School of Natural Science and Technology, Kanazawa University
Kakuma-machi, Kanazawa, 920-1192, JAPAN
email: nakayama@t.kanazawa-u.ac.jp
web: http://leo.ec.t.kanazawa-u.ac.jp/~nakayama/

ABSTRACT

When the number of sensors is less than that of the signal sources, this problem is called' Over Complete BSS' (OC-BSS), which is a difficult problem for lack of information about signal sources and a mixing process. A feedback approach has been proposed for the OC-BSS. In this paper, new learning algorithms used in the feedforward separation and the feedback cancelation are proposed. The separated single source is fed back to the inputs of the separation block, and is subtracted from the observations, in order to reduce the number of equivalent signal sources. Signal distortion, which is caused in the subtraction process, is suppressed by a spectral suppression technique. The learning of the feedforward separation is accelerated by using the constraints derived from the estimated mixing block. The proposed method can improve a signal to interference ratio by $6 \sim 10$ *dB compared to the conventional methods.*

1. INTRODUCTION

In many real applications of blind source separation (BSS), it is hard to estimate the number of the signal sources. The number of the sensors is usually different from that of the signal sources. When the number of the sensors is less than that of the signal sources, this problem is called 'Over Complete' BSS (OC-BSS). The OC-BSS is a difficult problem for lack of information about the signal sources and the mixing process. Therefore, the OC-BSS requires another information concerning the mixing process and the signal sources, besides the observed signals. Several kinds of conventional methods have been proposed, which mainly use the histogram of the observed signals as the additional information [1],[2],[6]. Furthermore, a feedback approach has been proposed, in which a separated signal source is fed back to the input nodes and is subtracted from the observations in order to reduce the number of the equivalent signal sources [10].

In this paper, new learning algorithms used in the feedforward separation and the feedback cancelation are are proposed for the feedback OC-BSS. Signal distortion caused in the subtraction is suppressed by using a spectral suppression technique. In the feedforward source separation, an acceleration technique is proposed to make fast and stable convergence possible. Finally, simulation results will be shown in order to confirm usefulness of the proposed method.

2. A FEEDBACK APPROACH TO OVER COMPLETE BSS

For simplicity, 3 signal sources and 2 sensors are used. A block diagram of the proposed feedback structure OC-BSS is shown in Fig.1.

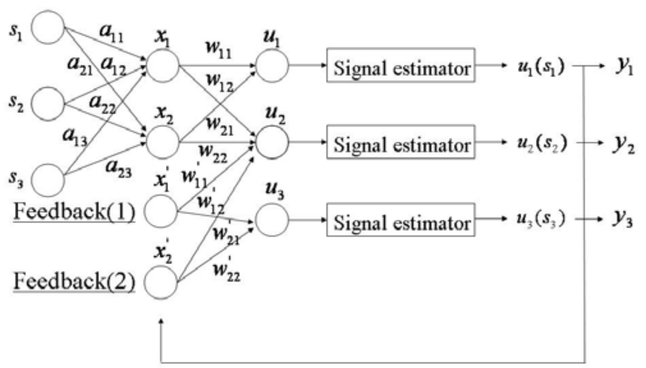

Figure 1: Feedback approach to over complete BSS with 3 signal sources and 2 sensors.

Letting N be the number of the signal sources, the number of the sensors M is set to be $M \geq [N/2+1]$, where $[X]$ means an integer number not exceed X. Under this condition, at least one output can separate a single signal source. Because learning algorithms of the BSS make the output signals of the separation block to be statistically independent to each other [4]. Therefore, one signal source can be separated. On the other hand, the other output signals can include a plural number of the signal sources. For example, u_1 includes s_1 and u_2 includes s_2 and s_3.

Letting s_1 be separated in u_1, and u_1 includes only a single source, it is selected as the final output y_1 through a 'Signal estimator'. When the signal sources are speech signals, a single speech signal is estimated by using a pitch frequency. Details are omitted here. y_1 is fed back to the inputs of the separation block, and is subtracted from the input signals x_1 and x_2, in order to eliminate the s_1 component in x_1 and x_2. Let the resulting x_1 and x_2 be x'_1 and x'_2, respectively. x'_1 and x'_2, which include only s_2 and s_3, are separated through another separation block denoted w'_{ji}. In this case, the number of the sensors is the same as that of the signal sources, then two signal sources can be separated by the conventional learning algorithms.

3. SIGNAL SOURCE SEPARATION IN FIRST PHASE

3.1 Theoretical Analysis of Source Separation

The network shown in Fig.1 is taken into account here. Furthermore, the mixing process is assumed to be an instantaneous process, that is a_{ji} do not include any time delay. The signal sources, the mixing block, the observed signals and the outputs of the separation block are related by

$$\begin{bmatrix} x_1(n) \\ x_2(n) \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix} \begin{bmatrix} s_1(n) \\ s_2(n) \\ s_3(n) \end{bmatrix} \quad (1)$$

$$\begin{aligned} \begin{bmatrix} u_1(n) \\ u_2(n) \end{bmatrix} &= \begin{bmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \end{bmatrix} \begin{bmatrix} x_1(n) \\ x_2(n) \end{bmatrix} & (2) \\ &= \begin{bmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \end{bmatrix} \begin{bmatrix} s_1(n) \\ s_2(n) \\ s_3(n) \end{bmatrix} & (3) \end{aligned}$$

Furthermore, the above equations are expressed by using vectors and matrices as follows:

$$\begin{aligned} \boldsymbol{x}(n) &= \boldsymbol{A}\boldsymbol{s}(n) & (4) \\ \boldsymbol{u}(n) &= \boldsymbol{W}\boldsymbol{x}(n) = \boldsymbol{W}\boldsymbol{A}\boldsymbol{s}(n) = \boldsymbol{P}\boldsymbol{s}(n) & (5) \end{aligned}$$

Assume $s_1(n)$ is separated in $u_1(n)$, and $s_2(n)$ and $s_3(n)$ are separated in $u_2(n)$. A signal to interference ratio in the 1st and the 2nd outputs, $u_1(n)$ and $u_2(n)$, are evaluated by

$$\frac{p_{11}^2}{p_{12}^2 + p_{13}^2} \qquad \frac{p_{22}^2 + p_{23}^2}{p_{21}^2} \quad (6)$$

3.2 Learning Algorithm in First Phase

Conventional learning algorithms can be basically applied to the group separation, that is separating $s_1(n)$ and ($s_2(n)$, $s_3(n)$). The learning algorithm using a mutual information as a cost function, and adjusts the weights following the natural gradient method [4],[5] is applied to this problem.

$$\begin{aligned} l(\boldsymbol{W}(n)) &= -\log|\det(\boldsymbol{W}(n))| - \sum_{k=1}^{M} \log p(u_k(n)) & (7) \\ \boldsymbol{W}(n+1) &= \boldsymbol{W}(n) + \eta[\boldsymbol{\Lambda}(n) \\ &\quad - < \phi(\boldsymbol{u}(n))\boldsymbol{u}^T(n) >]\boldsymbol{W}(n) & (8) \\ \phi(u_k(n)) &= \frac{1 - e^{-u_k(n)}}{1 + e^{-u_k(n)}} & (9) \end{aligned}$$

The operation $< >$ is time averaging. $p(u_k(n))$ is a probability density function of $u_k(n)$. $\boldsymbol{\Lambda}$ is a diagnal matrix.

In order to stabilize a learning process, the nonlinear function $\phi()$ must satisfy Eq.(10), where p' is a 1st derivative of p [4]. Several methods have been proposed for this purpose [7],[8]. In this paper, ϕ is controlled by Eqs.(11) and (12), where κ_4 is kurtosis.

$$\begin{aligned} \phi(u_k(n)) &= \frac{p'(u_k(n))}{p(u_k(n))} & (10) \\ \phi(u_k(n)) &= a\tanh(u_k(n)) + (1-a)u_k^3(n) & (11) \\ a &= \frac{1 - \exp(-2.1\kappa_4 - 2.5)}{1 + \exp(-2.1\kappa_4 - 2.5)} & (12) \end{aligned}$$

κ_4 is estimated by using the output $u_k(n)$ following the recurrence formula [7].

3.3 Learning Acceralation by Using Histogram of Observations

The observation $\boldsymbol{x}(n)$ is normalized.

$$\boldsymbol{v}(n) = \frac{\boldsymbol{x}(n)}{||\boldsymbol{x}(n)||} \quad (13)$$

The histogram of the observations is formulated by using distribution of $\boldsymbol{v}(n)$ along an angle. An example is shown in Fig.2. The angles of three peaks denoted S_1, S_2 and S_3 are related to $\tan\theta_1 = a_{21}/a_{11}$, $\tan\theta_2 = a_{22}/a_{12}$ and $\tan\theta_3 = a_{23}/a_{13}$, respectively.

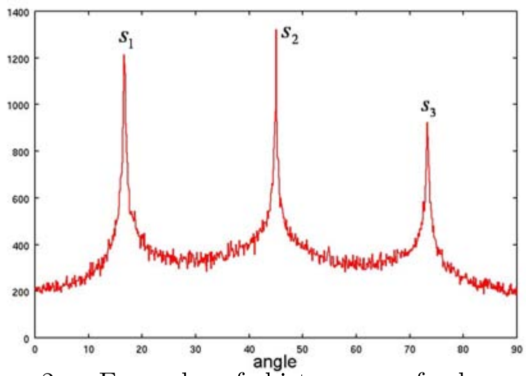

Figure 2: Example of histogram of observations $\boldsymbol{x}(n)$. Horizontal axis indicates angle θ of $\boldsymbol{x}(n) = [x_1(n), x_2(n)]^T$, that is $\tan\theta = x_2(n)/x_1(n)$, and vertical axis is histogram.

The learning algorithm is modified as follows:

$$\begin{aligned} \boldsymbol{w}^*(n) &= \arg\max_{\boldsymbol{w}_j(n)} \{\boldsymbol{u}^T(n)\boldsymbol{w}_j(n)\} & (14) \\ \boldsymbol{w}^*(n+1) &= \boldsymbol{w}^*(n) \\ &\quad -\eta[< \phi(\boldsymbol{u}(n))\boldsymbol{u}^T(n) >]\boldsymbol{w}^*(n) & (15) \end{aligned}$$

$\boldsymbol{w}^*(n)$ is the winner column in $\boldsymbol{W(n)}$. An idea behind the above learning algorithm is to frequently update the weight vector, whose angle is close to that of the peak of the histogram. In Eqs.(14) and (15), the norm of the weights $\boldsymbol{w}_j(n)$, $\boldsymbol{w}^*(n)$ and $\boldsymbol{w}^*(n+1)$ are normalized to be unity.

4. ELIMINATION OF A SINGLE SOURCE THROUGH FEEDBACK

4.1 Estimation of Mixing Block

An estimation method for the mixed process by using the histogram of the observations has been proposed [2]. This approach is taken into the learning process.

$\hat{\boldsymbol{a}}_i(n)$ is the estimation of $\boldsymbol{a}_i = [a_{1i}, a_{2i}]^T$, which follows the distribution of $\boldsymbol{x}(n)$, that is $\boldsymbol{v}(n)$. $\hat{\boldsymbol{a}}_i(n)$ is updated following

$$\hat{\boldsymbol{a}}_i(n+1) = \hat{\boldsymbol{a}}_i(n) + \eta(\boldsymbol{v}(n) - \hat{\boldsymbol{a}}_i(n)) \quad (16)$$

The initial guess of $\hat{\boldsymbol{a}}_i$ is very important. The histogram is roughly estimated, and the angle of the peak is used for the initial guess of $\hat{\boldsymbol{a}}_i$. In the example shown in Fig.2, there are three peeks. Their angle are assigned to $\hat{\boldsymbol{a}}_i, i = 1, 2, 3$, that is $a_{2i}/a_{1i} = \tan\theta_i$. Furthermore, $||\hat{\boldsymbol{a}}_i||$ is normalized to unity.

4.2 Direct Subtraction of Separated Single Source - Feedback(1) -

Suppose $u_1(n)$ includes a single source, that is $s_1(n)$. Although it is impossible to remove $s_2(n)$ and $s_3(n)$ from $u_1(n)$ completely, for simplicity we assume $u_1(n)$, that is $y_1(n)$ includes only $s_1(n)$. $u_1(n)$ is selected as the final output $y_1(n)$ as shown in Fig.1. $x_1(n)$ and $y_1(n)$ are expressed by

$$\begin{aligned} x_1(n) &= a_{11}s_1(n) + a_{12}s_2(n) + a_{13}s_3(n) && (17)\\ y_1(n) &= u_1(n) = (a_{11}w_{11} + a_{21}w_{12})s_1(n) && (18)\end{aligned}$$

Furthermore, $s_1(n)$ can be approximated by using the estimated mixing block $\hat{a}_{ji}$ as follows:

$$\hat{s}_1(n) = \frac{y_1(n)}{\hat{a}_{11}w_{11} + \hat{a}_{21}w_{12}} \tag{19}$$

Following these relations, $s_1(n)$ is subtracted from $x_1(n)$, in which $a_{11}s_1(n)$ is included as shown in Eq.(17), as follows:

$$\begin{aligned} x'_1(n) &= x_1(n) - \hat{a}_{11}\hat{s}_1(n)\\ &= x_1(n) - \frac{\hat{a}_{11}y_1(n)}{\hat{a}_{11}w_{11} + \hat{a}_{21}w_{12}} && (20)\end{aligned}$$

As a result, the $s_1(n)$ component in $x'_1(n)$ is cancelled. Actually, since the assumption is not complete, $s_1(n)$ still remains in $x'_1(n)$. This point will be investigated through simulation.

4.3 Subtraction of Separated Single Source Based on Histogram - Feedback(2) -

$s_1(n)$ included in $x_2(n)$ is subtracted based on the histograms of $y_1(n)$ and $x_2(n)$. The histogram generated by using $x_1(n)$ and $x_2(n)$ is denoted $F_{xx}(\theta)$, which is shown in Fig.2. The other histogram generated by using $y_1(n)$ and $x_2(n)$ is denoted $F_{xy}(\theta)$, which includes $s_1(n)$, $s_2(n)$ and $s_3(n)$ components. Since $y_1(n)$ mainly includes $s_1(n)$ and $x_2(n)$ contains all sources with the same probability, $F_{xy}(\theta)$ has the maximum value at the angle corresponding to $s_1(n)$. The histogram, in which $s_1(n)$ is not included, can be approximated by

$$F(\theta) = F_{xx}(\theta)\left(1 - \frac{F_{xy}(\theta)}{F_{max}}\right)^5 \tag{21}$$

F_{max} indicates is the maximum value of $F_{xy}(\theta)$. The 5th power is determined by experience. $F(\theta)$ takes zero at $F_{xy}(\theta) = F_{max}$. In the other part, $F(\theta)$ is reduced according to $F_{xy}(\theta)$. Since, the histogram of $s_1(n)$ is dominant in $F_{xy}(\theta)$, the $s_1(n)$ component is well reduced.

The observation $\boldsymbol{x}(n) = [x_1(n), x_2(n)]^T$ is reduced based on the histogram $F(\theta)$ at the angle $\theta = \tan^{-1}(x_2(n)/x_1(n))$. If $F(\theta)$ is small, then $\boldsymbol{x}(n)$ is well reduced. After subtraction of $s_1(n)$ from $x_2(n)$, $x_2(n)$ is denoted $x'_2(n)$.

Simulation results for the above method of subtracting $s_1(n)$ following the histogram $F(\theta)$ is shown here. The histogram of $(x_1(n), x_2(n))$ before the $s_1(n)$ subtraction is shown in Fig.2 in Sec.3.3. Furthermore, the histogram of $(x'_1(n), x'_2(n))$ after the $s_1(n)$ subtraction is shown in Fig.3. From these simulation results, $s_1(n)$

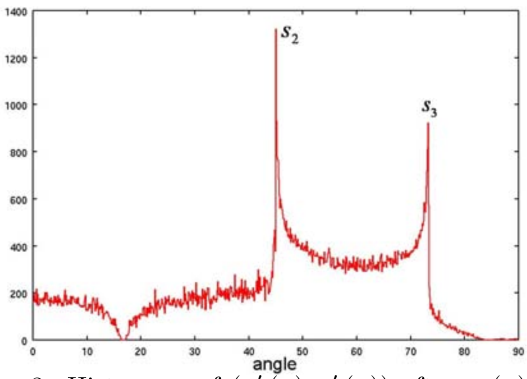

Figure 3: Histogram of $(x'_1(n), x'_2(n))$ after $s_1(n)$ subtraction.

is well subtracted from $x_1(n)$ and $x_2(n)$. Furthermore, the histogram related to $s_2(n)$ and $s_3(n)$ still remain.

4.4 Signal Distortion Reduction by Spectral Suppression

Even though $s_1(n)$ is subtracted from $x_2(n)$ based on the histogram, the samples to be reduced are somewhat randomly selected. This causes signal distortion. The signal distortion can be regarded as 'additive noise'. Therefore, a spectral suppression method is applied to suppress the signal distortion in this paper. The Joint MAP method [9], in which distribution of speech is assumed to be 'Super Gaussian' is employed.

It is assumed that $x'_1(n)$, in which $s_1(n)$ component is well reduced by 'Feedback(1)', is the clean speech, and the signal distortion components caused by 'Feedback(2)' is the noisy speech. A spectral gain $G(k)$ is given by [9]

$$\begin{aligned} G(k) &= q(k) + \sqrt{q^2(k) + \frac{\tau}{2\gamma(k)}} && (22)\\ q(k) &= \frac{1}{2} - \frac{\mu}{4\sqrt{\gamma(k)\xi(k)}} && (23)\end{aligned}$$

τ and μ are parameters, which determine a probability density function of the speech. $\xi(k)$ is a power ratio of the clean speech $A^2(k)$ and the noise $\lambda(k)$. $\gamma(k)$ is a power ratio of the noisy speech $R^2(k)$ and the noise $\lambda(k)$. They are expressed by

$$\xi(k) = \frac{A^2(k)}{\lambda(k)}, \qquad \gamma(k) = \frac{R^2(k)}{\lambda(k)} \tag{24}$$

Actually, $A^2(k)$ is determined by using the power spectrum of $x'_1(n)$, and $R^2(k)$ is determined by using the power spectrum of $x'_2(n)$. The noise power spectrum $\lambda(k)$ is estimated by $(R(k) - A(k))^2$.

The conventional Joint MAP method [9] is modified in this paper as follows: The probability density function of the speech signal is not always the super Gaussian. Therefore, the parameters τ and μ are controlled based on kurtosis of $x'_2(n)$. In order to avoid over suppression for low $\xi(k)$, the lower bound of $G(k)$ is determined. Furthermore, when $\xi(k)$ is high, $G(k)$ is desirable to be 1, however, sometime, $G(k)$ becomes small values. Therefore, we also employ the idea of 'Wiener Filter'. When $\xi(k)$ is large, $G(k)$ is determined by

$$G(k) = \frac{\xi(k)}{\xi(k) + \epsilon} \tag{25}$$

ϵ is a constant less than 1.

5. SOURCE SEPARATION IN SECOND PHASE

5.1 Learning Algorithm

$x'_1(n)$ and $x'_2(n)$ mainly include $s_2(n)$ and $s_3(n)$ components. They are further separated through the new separation block w'_{kj}. The learning algorithm proposed for the first phase is applied to the second phase. Furthermore, an acceleration technique is proposed.

5.2 Acceleration of Learning Process

Assuming an equivalent mixing block from $s_i(n)$ to $x'_j(n)$ to be a'_{ji}, $u_2(n)$ and $u_3(n)$ can be expressed by

$$\begin{aligned} u_2(n) &= (w'_{11}a'_{12} + w'_{12}a'_{22})s_2(n) \\ &\quad +(w'_{11}a'_{13} + w'_{12}a'_{23})s_3(n) \end{aligned} \tag{26}$$

$$\begin{aligned} u_3(n) &= (w'_{21}a'_{12} + w'_{22}a'_{22})s_2(n) \\ &\quad +(w'_{21}a'_{13} + w'_{22}a'_{23})s_3(n) \end{aligned} \tag{27}$$

Assuming $s_2(n)$ and $s_3(n)$ be separated at $u_2(n)$ and $u_3(n)$, respectively, the following conditions can be held.

$$w'_{11}a'_{13} + w'_{12}a'_{23} = 0 \tag{28}$$

$$w'_{21}a'_{12} + w'_{22}a'_{22} = 0 \tag{29}$$

Furthermore, we obtain

$$w'_{11} = -\frac{w'_{12}a'_{23}}{a'_{13}} \qquad w'_{21} = -\frac{w'_{22}a'_{22}}{a'_{12}} \tag{30}$$

w'_{ji} satisfying these conditions are the ideal solutions. a'_{23}/a'_{13} and a'_{22}/a'_{12} are estimated based on the histogram generated by using the modified observations $(x'_1(n), x'_2(n))$. Letting angle, at which peaks of the histogram appear, be θ_1 and θ_2, respectivey, they are related by

$$\tan\theta_1 = \frac{a'_{23}}{a'_{13}} \qquad \tan\theta_2 = \frac{a'_{22}}{a'_{12}} \tag{31}$$

$$w'_{11} = -w'_{12}\tan\theta_1 \qquad w'_{21} = -w'_{22}\tan\theta_2 \tag{32}$$

Furthermore, letting the correction terms determined by Eqs.(14) and (15) be $\Delta w'_{kj}$, the proposed learning algorithm is expressed by
Winner: First column= $[w'_{11}, w'_{21}]^T$

$$\begin{aligned} w'_{11}(n+1) &= w'_{11}(n) + (1-\zeta)\Delta w'_{11}(n) \\ &\quad - \zeta\Delta w'_{12}\tan\theta_1 \end{aligned} \tag{33}$$

$$\begin{aligned} w'_{21}(n+1) &= w'_{21}(n) + (1-\zeta)\Delta w'_{21}(n) \\ &\quad - \zeta\Delta w'_{22}\tan\theta_2 \end{aligned} \tag{34}$$

Winner: Second column= $[w'_{12}, w'_{22}]^T$

$$\begin{aligned} w'_{12}(n+1) &= w'_{12}(n) + (1-\zeta)\Delta w'_{12}(n) \\ &\quad - \zeta\frac{\Delta w'_{11}(n)}{\tan\theta_1} \end{aligned} \tag{35}$$

$$\begin{aligned} w'_{22}(n+1) &= w'_{22}(n) + (1-\zeta)\Delta w'_{22}(n) \\ &\quad - \zeta\frac{\Delta w'_{21}(n)}{\tan\theta_2} \end{aligned} \tag{36}$$

ζ is a constant less than 1. In the above update equations, the conditions satisfied by the ideal solutions are combined to accelerate the learning process.

6. SIMULATIONS AND DISCUSSIONS

6.1 Simulation Setup

Two male speeches and one female speech are used as the signal sources. The mixing process is assumed to be a instantaneous process, and is determined as follows: $a_{11} = a_{23} = 1, a_{13} = a_{21} = 0.3, a_{12} + a_{22} = 1.4$. Furthermore, a ratio of a_{12} and a_{22} is defined by

$$\alpha = \frac{a_{12}}{a_{22}}, \qquad 0 < \alpha \le 1 \tag{37}$$

When $\alpha = 1$, s_2 locates at the middle point between s_1 and s_3, then separating s_1 and (s_2, s_3) in the first phase is very difficult. On the other hand, when α takes a small value, s_2 locates close to s_3, then s_1 and (s_2, s_3) are easily separated. However, separating s_2 and s_3 in the second phase becomes difficult. Thus, the difficulty of source separation can be controlled by α, performance of the proposed method can be evaluated in detail.

6.2 Evaluation of Signal to Interference Ratio

The output signals are compared to the sources.

$$SIR_i = 10\log_{10}\left(\frac{\sum s_i^2(n)}{\sum(s_i(n) - y_i(n))^2}\right) \text{ [dB]} \tag{38}$$

6.3 Simulation Results and Discussions

Figures 4 through 6 show SIR_i. 'Conventional Method' means the shortest path method [2]. 'Proposed Method I' employs the accelerated learning algorithm given by Eqs.(14) and (15) in the first source separation. 'Proposed Method IIa' does not employ the spectral suppression method proposed in Sec.4.4 and the acceleration method proposed in Sec.5.2. 'Proposed Method I' and 'Proposed Method IIa' are included in [10]. 'Proposed Method IIb', which employs all of them, are newly proposed in this paper.

Since $\alpha = a_{12}/a_{22}$, $\alpha = 1$ means the second source $s_2(n)$ locates at the middle point between $s_1(n)$ and $s_3(n)$. This condition is severe for the separation of $s_1(n)$ and $(s_2(n), s_3(n))$. For this reason, SIR_1 decreases as α increases. On the other hand, in the second phase, $s_2(n)$ and $s_3(n)$ are separated. Therefore, their separation becomes easy for a large α, with which $s_2(n)$ and $s_3(n)$ locate far from the other. However, when α is small, that is s_2 and s_3 locate close to each other, their separation becomes difficult.

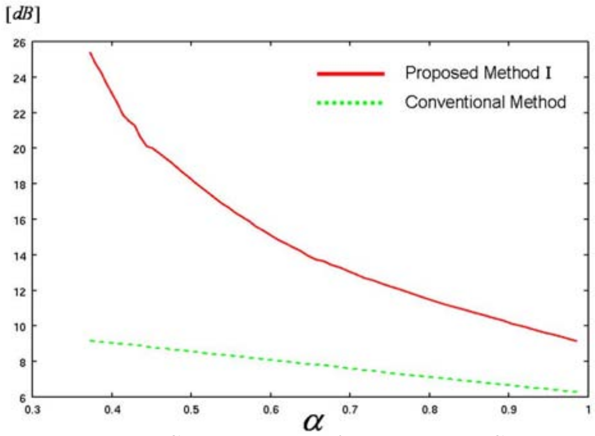

Figure 4: Signal to interference ratio SIR_1.

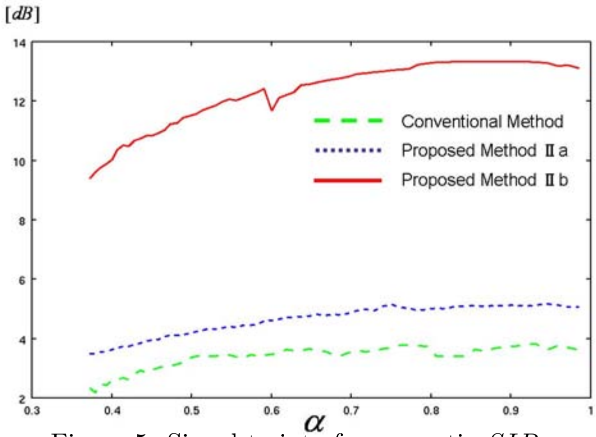

Figure 5: Signal to interference ratio SIR_2.

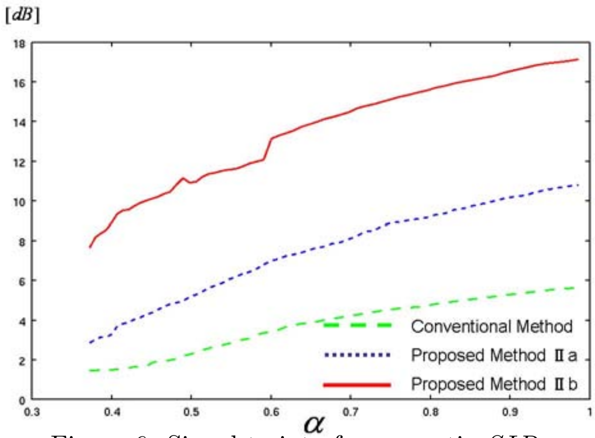

Figure 6: Signal to interference ratio SIR_3.

Comparing 'Proposed Method IIa' and 'Proposed Method IIb', usefulness of the signal distortion reduction method in Sec.4.4 and the accelerated learning algorithm in Sec.5.2 can be confirmed. Furthermore, compared to the results of the conventional method, the proposed method can improve SIR_i by $6 \sim 10$ dB.

7. CONCLUSIONS

New learning algorithms are proposed for the feedback OC-BSS. The signal distortion reduction method in the feedback cancelation and the accelerated learning method for the feedforward separation are proposed. The signal to interference ratio is improved by $6 \sim 10$ dB compared to the conventional methods.

REFERENCES

[1] M.S.Lewicki and T.J.Sejnowski, "Learning overcomplete representations," Neural Computation, vol.12, no.2, pp.337-365, Feb. 2000.

[2] A.Jung, F.J.Theis, C.G.Puntonet and E.W.Lang, "FastGeo - A histogram based approach to linear geometric ICA," Proc. of ICA2001, pp.418-423, April 2001.

[3] J.Van de Laar and P.C.W.Sommen,"Closed-form instantaneous blind signal separation based on second order statics and non-stationarity", 2002.

[4] S. Amari, T. Chen and A. Cichocki, "Stability analysis of learning algorithms for blind source separation", Neural Networks, vol.10, no8, pp.1345-1351, 1997.

[5] S.Amari, "Natural gradient works efficiently in learning," Neural Computation, vol.10, no2, pp.251-275, Feb. 1998.

[6] K.Waheed and F.M.Salem, "Algebraic overcomplete independent component analysis," ICA2003, Nara, Japan, April 2003.

[7] K.Nakayama, A.Hirano and T.Sakai, "An adaptive nonlinear function controlled by kurtosis for blind source separation," IEEE INNS, Proc. IJCNN'2002, Honolulu, pp.1234-1239, May 2002.

[8] K. Nakayama, A.Hirano and T.Sakai, "An adaptive nonlinear function controlled by estimated output pdf for blind source separation," Proc. ICA2003, Nara, Japan, pp.427-432, April 2003.

[9] T.Lotter, P.Vary "Speech Enhancement by MAP Spectral Amplitude Estimation Using a Super-Gaussian Speech Model", EURASIP Journal on Ap. Sig. Proc., pp.1110-1126, July 2005.

[10] K.Nakayama, H.Katou and A.Hirano, "A feedback approach and its learning algorithm for over complete blind source separation", Proc. IEEE, ISPACS2006, Yonago, Japan, pp.459-462, Dec. 2006.