International Journal of Electrical, Electronic and Communication Sciences
ISSN: 2517-9438
Vol:2, No:6, 2008

An Approach to Solving a Permutation Problem of Frequency Domain Independent Component Analysis for Blind Source Separation of Speech Signals

Masaru Fujieda, Takahiro Murakami, and Yoshihisa Ishida

***Abstract*—**Independent component analysis (ICA) in the frequency domain is used for solving the problem of blind source separation (BSS). However, this method has some problems. For example, a general ICA algorithm cannot determine the permutation of signals which is important in the frequency domain ICA. In this paper, we propose an approach to the solution for a permutation problem. The idea is to effectively combine two conventional approaches. This approach improves the signal separation performance by exploiting features of the conventional approaches. We show the simulation results using artificial data.

***Keywords*—**Blind source separation, Independent component analysis, Frequency domain, Permutation ambiguity.

I. Introduction

BLIND source separation (BSS) [1], [4] is an approach to estimating original source signals by using only the information of the mixed signals observed at a sensor array. This technique is applicable to the realization of the noise-robust speech recognition, the hearing aid which can enhance the specific sound that a user wants to listen, and so on. Independent component analysis (ICA) [4] is one of the statistical analysis methods and identifies the independent components in the random variables. ICA can be used for BSS of linear (instantaneous) mixtures. To achieve BSS of convolutive mixtures, several methods have been proposed [1], [2], [3], [5]. Although there are a number of applications for BSS to mixed speech signals in realistic acoustical environments, the separation performance is still not good enough [6].

In this paper, we propose an approach to solving a permutation problem of frequency domain ICA (FDICA) for BSS. In FDICA, it is necessary to solve the problem of ambiguity of scaling factors and permutation. Several methods have already been proposed for such problems [1], [2] but each of them has imperfection. For example, the method [1] is impossible to estimate the permutation when the envelopes of source signals are similar. On the other hand, the method [2] is not dependent on the envelope of source signals. The method [1] performs better than the method [2] when the envelopes of source signals are different from each other. To improve the performance, we propose a new method for combining the features of these methods.

The following chapter explains the formulation of the general ICA problems and FDICA. In Section III, the algorithm to solve scaling and permutation problems is explained. In Section IV, the source separation experiments are performed and the simulation results are illustrated. Finally, Section V shows the conclusion of this paper.

Manuscript received September 27, 2006.

The authors are with the School of Science and Technology, Meiji University, Kanagawa, Japan (corresponding author to provide phone: +81-44-934-7307; e-mail: ce66050@isc.meiji.ac.jp).

II. Independent Component Analysis

A. Statement of the Problem

In general, the liner mixture model of ICA is given by the following equation:

$$\boldsymbol{x}(t) = \boldsymbol{A} \cdot \boldsymbol{s}(t), \tag{1}$$

where $\boldsymbol{x}(t) = [x_1(t), x_2(t), \ldots, x_n(t)]^T$ is a set of observed signals, $\boldsymbol{A}$ is an unknown mixing matrix, and $\boldsymbol{s}(t) = [s_1(t), s_2(t), \ldots, s_m(t)]^T$ is a set of source signals which is assumed to be mutually independent. The purpose of ICA is to find a separation matrix $\boldsymbol{W}$ so that the output signals

$$\boldsymbol{y}(t) = \boldsymbol{W} \cdot \boldsymbol{x}(t), \tag{2}$$

where $\boldsymbol{y}(t) = [y_1(t), y_2(t), \ldots, y_m(t)]^T$ is mutually independent. However, there are still ambiguity of scaling factors and permutation of output signals because of lack of information about the amplitude and permutation of the source signals, that is, $\boldsymbol{W}$ is allowed to satisfy the following equation:

$$\begin{aligned} \boldsymbol{y}(t) &= \boldsymbol{W} \cdot \boldsymbol{x}(t) \\ &= \boldsymbol{Q} \cdot \boldsymbol{D} \cdot \boldsymbol{s}(t), \end{aligned} \tag{3}$$

where $\boldsymbol{D}$ is a diagonal matrix which represents scaling factors and $\boldsymbol{Q}$ is a permutation matrix.

In the case of acoustical signals, the mixing model is expressed by the convolutive mixture as

$$\begin{aligned}\boldsymbol{x}(t) &= \sum_{\tau} \boldsymbol{A}(\tau)\cdot \boldsymbol{s}(t-\tau) \\ &= \boldsymbol{A}(\tau) * \boldsymbol{s}(t),\end{aligned} \tag{4}$$

where operator $*$ denotes convolution and $\boldsymbol{A}(\tau)$ is an unknown mixing filter. In the convolutive mixture case, (2) is rewritten by using a separation filter $\boldsymbol{W}(\tau)$ as

$$\begin{aligned}\boldsymbol{y}(t) &= \boldsymbol{W}(\tau) * \boldsymbol{x}(t) \\ &= \boldsymbol{Q}\cdot\boldsymbol{D}\cdot\boldsymbol{s}(t),\end{aligned} \tag{5}$$

B. Frequency Domain Independent Component Analysis (FDICA)

By using the short time Fourier transform (STFT), we can transpose the convolutive model to the instantaneous model with complex values as in the form

$$\hat{\boldsymbol{x}}(\omega,t_s) = \hat{\boldsymbol{A}}(\omega)\cdot\hat{\boldsymbol{s}}(\omega,t_s), \tag{6}$$

where $\hat{\boldsymbol{x}}(\omega,t_s)$ and $\hat{\boldsymbol{s}}(\omega,t_s)$ denote STFT of $\boldsymbol{x}(t)$ and $\boldsymbol{s}(t)$, respectively and $\hat{\boldsymbol{A}}(\omega)$ denotes the discrete Fourier transform (DFT) of the mixing filter $\boldsymbol{A}(\tau)$. $\hat{\boldsymbol{x}}(\omega,t_s)$ and $\hat{\boldsymbol{s}}(\omega,t_s)$ are often referred as spectrograms. For any fixed frequency ω in (6), $\hat{\boldsymbol{x}}(\omega,t_s)$ can be regarded as the linear mixture model. Therefore, we can apply ICA for the linear mixture model at each frequency independently.

In FDICA, ambiguity of scaling factors and the permutation of output signals become problematic when time domain signals are reconstructed.

We propose an approach to solving the permutation problem. For the scaling problem, the method presented in [1] is applied.

III. Algorithm

A. Solving Scaling Problem

When a separating matrix $\hat{\boldsymbol{W}}(\omega)$ is obtained from the spectrogram $\hat{\boldsymbol{x}}(\omega,t_s)$, the corresponding independent components are given by:

$$\hat{\boldsymbol{u}}(\omega,t_s) = \hat{\boldsymbol{W}}(\omega)\hat{\boldsymbol{x}}(\omega,t_s). \tag{7}$$

Then we can obtain the scaling factor as

$$\hat{\boldsymbol{v}}(\omega,t_s;i) = \hat{\boldsymbol{W}}^{-1}(\omega)\begin{pmatrix} 0 \\ \vdots \\ \hat{u}_i(\omega,t_s) \\ \vdots \\ 0 \end{pmatrix}, \tag{8}$$

where $\hat{u}_i(\omega,t_s)$ denotes the i-th independent component of $\hat{\boldsymbol{u}}(\omega,t_s)$ and $\hat{\boldsymbol{v}}(\omega,t_s;i)$ denotes the i-th independent components observed at each sensor [1].

B. Solving Permutation Problem

The proposed method is to combines the features of two conventional techniques presented in [1], [2]. In this section, we review these methods briefly.

1) Method based on temporal structure of speech signals (MTS)

This algorithm was proposed by Murata et al. [1], utilizes a property of speech signals, that is, signals are intrinsically non-stationary in a long range mainly because of amplitude modulation.

Let us rewrite $\hat{s}_i(\omega,t_s)$, i-th element of $\hat{\boldsymbol{s}}(\omega,t_s)$ in (6):

$$\hat{s}_i(\omega,t_s) = \left|\hat{s}_i(\omega,t_s)\right| e^{j\phi_i(\omega,t_s)}, \tag{9}$$

where $\left|\hat{s}_i(\omega,t_s)\right|$ and $\phi_i(\omega,t_s)$ are an absolute value, a phase of source signals in time-frequency domain, respectively. Since source signals are mutually independent, they satisfy the following equations:

$$\operatorname{corr}\left(\left|\hat{s}_i(\omega,t_s)\right|,\left|\hat{s}_j(\omega,t_s)\right|\right) = 0, \quad i \neq j, \tag{10}$$

$$\operatorname{corr}\left(\left|\hat{s}_i(\omega,t_s)\right|,\left|\hat{s}_j(\omega',t_s)\right|\right) = 0,\ i \neq j,\ \omega \neq \omega', \tag{11}$$

where

$$\begin{aligned}&\operatorname{corr}\left(\left|\hat{s}_i(\omega,t_s)\right|,\left|\hat{s}_j(\omega,t_s)\right|\right) \\ &\quad= \left(\left|\hat{s}_i(\omega,t_s)\right|\cdot\left|\hat{s}_j(\omega,t_s)\right|\right) + \overline{\left|\hat{s}_i(\omega,t_s)\right|}\cdot\overline{\left|\hat{s}_j(\omega,t_s)\right|},\end{aligned} \tag{12}$$

$$\overline{\left|\hat{s}_i(\omega,t_s)\right|} = \frac{1}{T}\sum_{s=1}^{T}\left|\hat{s}_i(\omega,t_s)\right|, \tag{13}$$

We assume that T is sufficiently large. On the other hand, for different frequency components corresponding to the same source signal, we can assume

$$\operatorname{corr}\left(\left|\hat{s}_i(\omega,t_s)\right|,\left|\hat{s}_i(\omega',t_s)\right|\right) \neq 0, \quad \omega \neq \omega'. \tag{14}$$

Therefore, the correlation coefficient of their envelopes shown as

$$= \frac{r\left(\left|\hat{s}_i(\omega,t_s)\right|,\left|\hat{s}_j(\omega',t_s)\right|\right)}{\sqrt{\mathrm{corr}\left(\left|\hat{s}_i(\omega,t_s)\right|,\left|\hat{s}_i(\omega,t_s)\right|\right)\mathrm{corr}\left(\left|\hat{s}_j(\omega',t_s)\right|,\left|\hat{s}_j(\omega',t_s)\right|\right)}}, \quad (15)$$

may be available for estimating an appropriate combination of frequency elements.

Then, let us define a moving average operator ε for estimating the envelope of time series as

$$\varepsilon\hat{\boldsymbol{v}}(\omega,t_s;i) = \frac{1}{2M+1}\sum_{t_s'=t_s-M}^{t_s+M}\sum_{j=1}^{n}\left|\hat{v}_j(\omega,t_s;i)\right|, \quad (16)$$

where M is a positive constant and $\hat{v}_j(\omega,t_s;i)$ denotes the j-th element of $\hat{\boldsymbol{v}}(\omega,t_s;i)$.

We solve the permutation problem by the sorting based on the correlation of envelopes as follows:

1) Sort ω in order of independent components with the low correlation. This is done by sorting them in ascending order of similarity defined as

$$\mathrm{sim}(\omega) = \sum_{i\neq j} r\left(\varepsilon\hat{\boldsymbol{v}}(\omega,t_s;i),\varepsilon\hat{\boldsymbol{v}}(\omega,t_s;j)\right), \quad (17)$$

$$\mathrm{sim}(\omega_1) \le \mathrm{sim}(\omega_2) \le \cdots \le \mathrm{sim}(\omega_N). \quad (18)$$

2) For ω_1, assign $\hat{\boldsymbol{v}}(\omega_1,t_s;i)$ to $\hat{\boldsymbol{y}}(\omega_1,t_s;i)$ as follows:

$$\hat{\boldsymbol{y}}(\omega_1,t_s;i) = \hat{\boldsymbol{v}}(\omega_1,t_s;i), \quad i=1,\ldots,n. \quad (19)$$

3) For ω_k, find a permutation $\sigma(i)$ which maximizes the correlation between the envelope of ω_k and the summed envelope from ω_1 through ω_{k-1}. This is achieved by maximizing the sum of correlation coefficients

$$\sum_{i=1}^{m} r\left(\varepsilon\hat{\boldsymbol{v}}(\omega_k,t_s;\sigma(i)),\sum_{j=1}^{k-1}\varepsilon\hat{\boldsymbol{y}}(\omega_j,t_s;i)\right), \quad (20)$$

for all the possible permutations of $i=1,\ldots,m$.

4) Assign the appropriate permutation to $\hat{\boldsymbol{y}}(\omega_k,t_s;i)$:

$$\hat{\boldsymbol{y}}(\omega_k,t_s;i) = \hat{\boldsymbol{v}}(\omega_k,t_s;\sigma(i)), \quad i=1,\ldots,m. \quad (21)$$

5) Go to 3) until $k=N$.

As a result, we obtain separated spectrograms

$$\hat{\boldsymbol{y}}(\omega_k,t_s;i), \quad i=1,\ldots,m. \quad (22)$$

2) Method based on relation of the mixing matrix at the adjacent frequencies (MRM)

Alternative approach, which was proposed by Asano et al. [2], utilizes a relation of the location vectors in the mixing matrix at adjacent frequencies. The relation between the mixing matrices at frequencies ω_k and ω_{k-1} are written as

$$\hat{\boldsymbol{A}}(\omega_k) = \boldsymbol{T}(\omega_k,\omega_{k-1})\cdot\hat{\boldsymbol{A}}(\omega_{k-1}) \quad (23)$$

where $\boldsymbol{T}(\omega_k,\omega_{k-1})$ is a rotation matrix. Absolute values of mixing matrices $\hat{\boldsymbol{A}}(\omega_k)$ and $\hat{\boldsymbol{A}}(\omega_{k-1})$ are assumed to be unity for the sake of simplicity. If the difference between two frequencies $\Delta\omega=\omega_k-\omega_{k-1}$ is sufficiently small, the relation can be written as

$$\hat{\boldsymbol{A}}(\omega_k)\approx\hat{\boldsymbol{A}}(\omega_{k-1}),\ \boldsymbol{T}(\omega_k,\omega_{k-1})\approx\boldsymbol{I}, \quad (24)$$

where $\boldsymbol{I}$ is a unit matrix. Then, an angle between the location vectors is small. The location vector $\hat{\boldsymbol{a}}_i(\omega_k)$ is obtained by the column vector in the mixing matrix $\hat{\boldsymbol{A}}(\omega)$ as follows:

$$\hat{\boldsymbol{A}}(\omega) = \left[\hat{\boldsymbol{a}}_1(\omega)\quad\hat{\boldsymbol{a}}_2(\omega)\quad\cdots\quad\hat{\boldsymbol{a}}_m(\omega)\right]. \quad (25)$$

$\hat{\boldsymbol{a}}_i(\omega)$ is defined as

$$\hat{\boldsymbol{a}}_i(\omega) = \begin{pmatrix} e^{-j\omega\tau_{1i}} \\ \vdots \\ e^{-j\omega\tau_{ni}} \end{pmatrix}, \quad (26)$$

where $\tau_{j,i}$ denotes the propagation time from the i-th source signal to the j-th sensor. A symbol θ_i is defined as the angle between $\hat{\boldsymbol{a}}_i(\omega_k)$ and $\hat{\boldsymbol{a}}_i(\omega_{k-1})$. Then, θ_i is expected to be the smallest for the correct permutation.

Based on the above discussion, the permutation problem is solved by minimizing the sum of the angels $\{\theta_1,\ldots,\theta_m\}$ between the location vectors in the adjacent frequencies. An estimate of the mixing matrix can be obtained by the inverse of the separation matrix $\hat{\boldsymbol{W}}(\omega)$ as

$$\tilde{\boldsymbol{A}}(\omega) = \hat{\boldsymbol{W}}^{-1}(\omega). \quad (27)$$

Let us define the mixing matrix multiplied by the arbitrary permutation matrix $\boldsymbol{P}$ as

$$\overline{\boldsymbol{A}}^T(\omega) = \boldsymbol{P}\cdot\tilde{\boldsymbol{A}}^T(\omega). \quad (28)$$

The permutation $\boldsymbol{P}\tilde{\boldsymbol{A}}^T(\omega)$ exchanges the row vectors of

$\tilde{\boldsymbol{A}}^{T}(\omega)$. The column vectors of $\bar{\boldsymbol{A}}(\omega)$ are denoted as $\bar{\boldsymbol{A}}(\omega)=\left[\bar{\boldsymbol{a}}_1(\omega),\ldots,\bar{\boldsymbol{a}}_m(\omega)\right]$. The cosine of the angle θ_i between the two vectors, $\bar{\boldsymbol{a}}_i(\omega_k)$ and $\bar{\boldsymbol{a}}_i(\omega_{k-1})$, is defined as

$$\cos\theta_i=\frac{\bar{\boldsymbol{a}}_i^{H}(\omega_k)\bar{\boldsymbol{a}}_i(\omega_{k-1})}{\left\|\bar{\boldsymbol{a}}_i(\omega_k)\right\|\cdot\left\|\bar{\boldsymbol{a}}_i^{H}(\omega_{k-1})\right\|}. \tag{29}$$

By using this, the permutation matrix is determined as

$$\hat{\boldsymbol{P}}=\arg\max_{\boldsymbol{P}}F(\boldsymbol{P}), \tag{30}$$

where the cost function $F(\boldsymbol{P})$ is defined as

$$F(\boldsymbol{P})=\frac{1}{m}\sum_{i=1}^{m}\cos\theta_i. \tag{31}$$

The above method assumes that the estimate of the mixing matrix $\tilde{\boldsymbol{A}}(\omega)$ is a good approximation of the true mixing matrix $\boldsymbol{A}(\omega)$. Therefore, if $\tilde{\boldsymbol{A}}(\omega_{k-1})$ is a bad approximation, we may fail to estimate the correct permutation at the frequency ω_k.

To prevent this, the reference frequency is extended to the following frequency range:

$$\omega_{k-l}=\omega_k-l\cdot\Delta\omega_k, \text{ for } l=1,\ldots,L. \tag{32}$$

The cost function (31) is calculated at all L frequencies within this range. Let us define the value of the cost function at $\omega_{k-l}=\omega_k-l\cdot\Delta\omega_k$ as $F(\boldsymbol{P},l)$. Next, a confidence measure for $F(\boldsymbol{P},l)$ is considered. When the largest value of $\max F(\boldsymbol{P},l)$ closes to other cost functions $F(\boldsymbol{P},l)$, it may be difficult to determine the correct permutation, and then the value of $F(\boldsymbol{P},l)$ is not reliable. Based on this assumption, the following confidence measure is defined:

$$C(l)=\max_{\boldsymbol{P}\in\Omega}\left[F(\boldsymbol{P},l)\right]-\max_{\boldsymbol{P}\in\Omega'}\left[F(\boldsymbol{P},l)\right] \tag{33}$$

where Ω denotes a set of all possible $\boldsymbol{P}$ and Ω' denotes Ω with $\hat{\boldsymbol{P}}=\arg\max_{\boldsymbol{P}\in\Omega}\left[F(\boldsymbol{P},l)\right]$. An appropriate reference frequency is determined as

$$\hat{l}=\max_{l}C(l). \tag{34}$$

The permutation is then solved by using the information at this reference frequency as

$$\hat{\boldsymbol{P}}=\arg\max_{\boldsymbol{P}}F\left(\boldsymbol{P},\hat{l}\right). \tag{35}$$

3) Proposed method which combines two conventional methods (MCC)

In this paper, we propose a method which combines MTS and MRM. MTS has a disadvantage, that is, it is impossible to estimate the permutation when the envelopes of source signals are similar. On the other hand, MRM is not dependent on the envelope of source signals. However, MTS provides better performance than MRM when the envelopes of source signals are different each other. Considering these facts, we propose to combine two methods in order to cover each disadvantage.

We consider two situations. In the case that the envelopes of separated signals are similar, MTS is employed. In the case that the envelopes of separated signals are different, MRM is utilized. This gives better performance than using only MTS or MRM.

The proposed procedure is summarized as follows:

1) Set an appropriate threshold α ($|\alpha|\le 1$).
2) For ω_1, assign $\hat{\boldsymbol{v}}(\omega_1,t_s;i)$ to $\hat{\boldsymbol{y}}(\omega_1,t_s;i)$ as

$$\hat{\boldsymbol{y}}(\omega_1,t_s;i)=\hat{\boldsymbol{v}}(\omega_1,t_s;i),\quad i=1,\ldots,n. \tag{36}$$

3) For ω_k, use MTS methods (3) and (4).
4) $k=k+1$, and go to 3) while $\mathrm{sim}(\omega_k)<\alpha$.
5) In the ascending frequency order, apply MRM within the reference frequency range $[-L,L]$.

IV. Experimental Results

In this section, we compare the performance of the conventional method with the proposed method.

A. Conditions for Experiments

For the source signals, three Japanese speech sounds recorded separately are used.

Intonation of Japanese is given by the change of the fundamental frequency. Thus, the envelopes of the Japanese speech can be often similar to each other. The sets are prepared as follows:

Set 1: Envelopes are mutually different,
Set 2: Envelopes are mutually similar.

The waveforms are shown in Fig. 1.

The mixing process is assumed to be

$$\begin{aligned}x_1(t)&=0.4249s_1(t)+0.3902s_2(t-\tau)\\ x_2(t)&=0.3322s_1(t-\tau)+0.3401s_2(t),\end{aligned} \tag{37}$$

where $s_1(t)$ is "aoiie," which means a blue house in English. $s_2(t)$ is "sakuragasaita," that means Japanese cherries blossomed in English, or "aiueo," five vowels in Japanese. τ is a delay factor. The other parameters are shown in Table I.

B. Evaluation

Since the true sources and mixing process are available, we can evaluate the performance using the noise-reduction rate (NRR) [3] that is defined by output signal-to-noise ratio (SNR) minus input SNR as

$$\begin{aligned}\mathrm{SNR}_{out} &= 10\log_{10}\left(\frac{\mathrm{var}[s(t)]}{\mathrm{var}[y(t)-s(t)]}\right)\\ \mathrm{SNR}_{in} &= 10\log_{10}\left(\frac{\mathrm{var}[s(t)]}{\mathrm{var}[x(t)-s(t)]}\right),\\ \mathrm{NRR} &= \mathrm{SNR}_{out}-\mathrm{SNR}_{in}\\ &= 10\log_{10}\left(\frac{\mathrm{var}[x(t)-s(t)]}{\mathrm{var}[y(t)-s(t)]}\right)\end{aligned} \tag{38}$$

where $\mathrm{var}[\bullet]$ denotes the variance of the signal. In practice, we use the following equation:

$$\mathrm{NRR} = \frac{1}{nm}\sum_{i=1}^{n}\sum_{j=1}^{m}10\log_{10}\left(\frac{\mathrm{var}[x_i(t)-s_j(t)]}{\mathrm{var}[y_{ij}(t)-s_j(t)]}\right), \tag{39}$$

$i = 1,\ldots n,\ \ j = 1,\ldots m$

where $n = m = 2$. Table II shows the experimental results.

The performance of MCC was the best. When the envelope of source signals is mutually different, MTS performs better. However, in the frequency bin where the separation by ICA does not succeed, the estimate of permutation is incorrect because the envelope of spectrogram is mutually similar. In this case, the reliability of MRM is higher than that of MTS. On the other hand, when the envelope of source signals is mutually similar, MRM performs better. In the case that $\mathrm{sim}(\omega_k)$ in (17) is sufficiently large, the reliability of MTS is higher than that of MRM. MCC can select the most appropriate approach for every frequency bin by setting the optimum threshold. As the result, MCC shows the best performance.

V. Conclusion

In this paper, a method to combine two conventional methods for solving the permutation problem has been proposed. Our approach is to overcome the drawbacks in the conventional algorithms and the effect of a new method is well-recognized in the experiments. The proposed method works well and gives better performance than the conventional methods.

References

[1] N. Murata, S. Ikeda, A. Ziehe, "An approach to blind source separation based on temporal structure of speech signals," in *Neurocomputing*, 41, 2001, pp. 1–24.

[2] F. Asano, S. Ikeda, M. Ogawa, H. Asoh, N. Kitawaki, "Combined approach of array processing and independent component analysis for blind separation of acoustic signals," in *IEEE trans. Speech and Audio Processing*, 11, No. 3, 2003, pp. 204–214.

[3] S. Kurita, H. Saruwatari, S. Kajita, K. Takeda, F. Itakura, "Evaluation of blind signal separation method using directivity pattern under reverberant conditions," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*, SAM-P2-5, Jun. 2000, pp. 3140-3143

[4] A. Hyvärinen, E. Oja, "Independent component analysis: a tutorial," in *Helsinki University of Technology, Laboratory of computer and Information Science*, Apr. 1999

[5] N. Murata, S. Ikeda, "An on-line algorithm for blind source separation on speech signals," in *Proceedings NOLTA'98*, Sep. 1998, pp. 923-926.

[6] S. Araki, R. Mukai, S. Makino, T. Nishikawa, H. Saruwatari, "The fundamental limitation of frequency domain blind source separation for convolutive mixtures of speech," in *IEEE Trans. Speech and Audio Processing*, 11, No. 2, 2003, pp. 109–116.

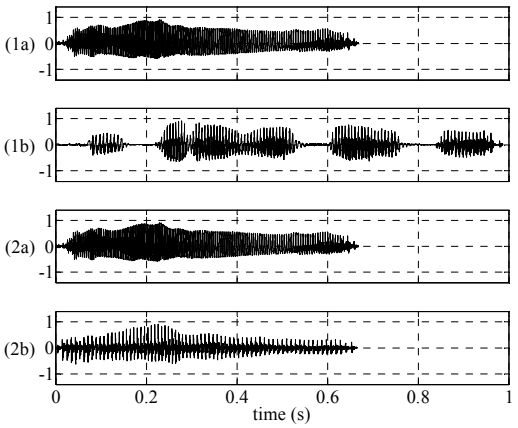

Fig. 1 Waveforms of source signals. (1a) is "aoiie," which means a blue house in English. (1b) is "sakuragasaita," that means Japanese cherries blossomed in English. (2a) is the same (1a). (2b) is "aiueo," five vowels in Japanese. (1a) and (1b) are a set of different envelopes. (2a) and (2b) are a set of similar envelopes

TABLE I
Parameters of Experiment

Parameter Name	Value
Sampling frequency	16 kHz
Delay τ	0.3125 ms (5 points)
Window length	8 ms (128 points)
Window function	Hamming window
Shifting time of analysis frames	1.25 ms (20 points)
Threshold α (for Set 1)	0.70
Threshold α (for Set 2)	-0.06
Window length of moving average M (16)	15
Number of reference frequency range L	3

TABLE II
Noise-Reduction Rates for Convolutive Mixture

Method	Envelope is mutually different	Envelope is mutually similar
MTS	18.22 dB	0.93 dB
MRM	12.86 dB	7.07 dB
MCC	18.74 dB	8.84 dB