International Journal of Electrical, Electronic and Communication Sciences
ISSN: 2517-9438
Vol:4, No:9, 2010

Blind Source Separation based on the Estimation for the Number of the Blind Sources under a Dynamic Acoustic Environment

Takaaki Ishibashi

Abstract—Independent component analysis can estimate unknown source signals from their mixtures under the assumption that the source signals are statistically independent. However, in a real environment, the separation performance is often deteriorated because the number of the source signals is different from that of the sensors. In this paper, we propose an estimation method for the number of the sources based on the joint distribution of the observed signals under two-sensor configuration. From several simulation results, it is found that the number of the sources is coincident to that of peaks in the histogram of the distribution. The proposed method can estimate the number of the sources even if it is larger than that of the observed signals. The proposed methods have been verified by several experiments.

Keywords—blind source separation, independent component analysys, estimation for the number of the blind sources, voice activity detection, target extraction.

I. Introduction

THE speech recognition technology has significantly been improved to achieve provision of speech recognition engine with extremely high recognition capabilities for the case of ideal environments, i.e. no surrounding noise. However, it is still difficult to attain a desirable recognition rate in a household or office where there are daily activities noises. Therefore, a certain preprocessing before recognition is needed to reduce the noises and to select the target speech signal.

Many noise reduction methods using ICA (independent component analysis) have been proposed. ICA can separate unknown sources from their mixtures without information on the transfer functions, provided that the sources are statistically independent [1], [2], [3], [4], [5], [6], [7], [8]. For the instantaneous mixtures, the original sources can be completely recovered in the time domain except for indeterminacy of scale and permutation.

In a real environment, the signals observed at microphones are not instantaneous mixtures but are convoluted version of the sound sources. On account of this, there have been reported many trials to separate the convoluted mixtures in the frequency domain. However, the indeterminacy of scale and permutation appears at every frequency bin. In order to recover the sources properly, this indeterminacy problem must be essentially solved before making an inverse transformation from the frequency to the time domain.

T. Ishibashi is with the Department of Information Communication and Electronic Engineering, Kumamoto National College of Technology, Kumamoto, Japan e-mail: ishibashi@knct.ac.jp.

Manuscript received May 31, 2010.

The scale indeterminacy can be solved by use of a decomposed spectrum [5]. We have derived that the decomposed spectrum is uniquely expressed as a product of a source and its transfer function [9], [10]. For the permutation problem, we have proposed a permutation correction in terms of the power of decomposed spectra using prior information about source directions [9], [10].

In the case where the number of source signals is equal to that of the observed mixture signals, from above methods, the original sources can be completely recovered. However, separation performance often deteriorates because the number of the source signals is different from that of the mixture signals. Therefore, it is very important that the sources number is estimated by using only the observed mixture signals before ICA process. There has been proposed an estimation method of the sources number [11]. It functions well if the number of the sources is equal to or less than that of the sensors but it fails if the former is larger than the latter.

In this paper, we propose an estimation method for the number of the source signals based on the joint distributions of the observed signals under two-sensor configuration. And we propose a blind source separation method using the estimated the source number under a dynamic acoustic environment. Several simulation results elucidate that the number of the sources are coincident to that of peaks in the histogram of the distribution. The proposed method can estimate the number of the sources even if it is larger than the sensor number. And the method can also estimate the target source signal under a dynamic acoustic environment.

II. Blind source separation

A. Independent component analysis in frequency domain

Consider the case where the original source signals $\boldsymbol{s}(t) = [s_1(t), \cdots, s_n(t), \cdots, s_N(t)]^T$ are observed by microphones in a convoluted way, the obserbed mixture signals $\boldsymbol{x}(t) = [x_1(t), \cdots, x_m(t), \cdots, x_M(t)]^T$ are represented by

$$\boldsymbol{x}(t) = G(t) * \boldsymbol{s}(t) \tag{1}$$

where $G(t)$ denotes a mixing matrix whose elements are transfer functions $g_{mn}(t)$ from the n-th sources to the m-th microphones, and $*$ denotes the convolutional operator.

The mixtures $x_m(t)$ are transformed into the short time spectra by the discrete Fourier transform to perform separation in frequency domain.

$$x_m(\omega, k) = \sum_t e^{-j\omega t} x_m(t) w(t - k\tau) \tag{2}$$

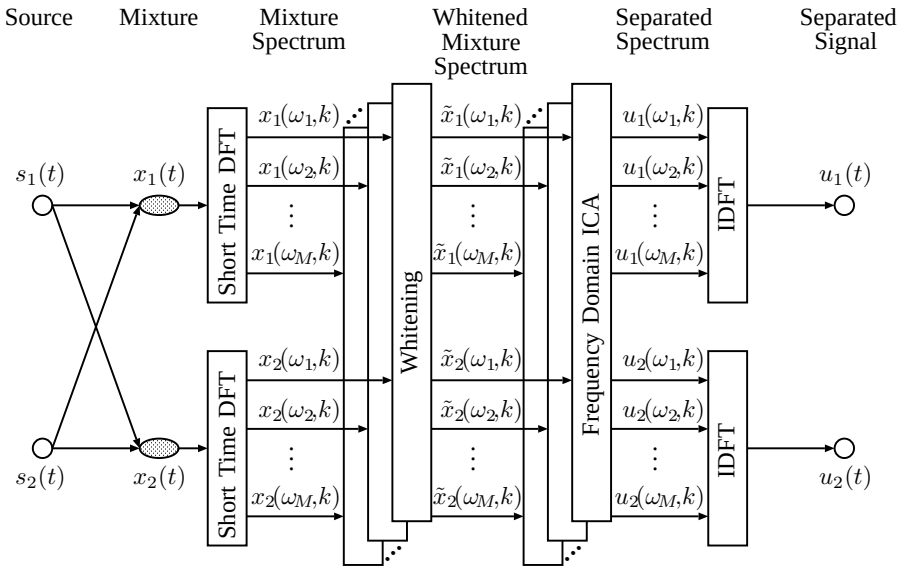


Fig. 1. Frequency domain independent component analysis.

where ω denotes a frequency, k the frame number, τ the frame shift time and $w(t)$ a window function.

In the frequency domain, the mixtures are approximated as

$$\boldsymbol{x}(\omega,k) = G(\omega)\boldsymbol{s}(\omega,k) \tag{3}$$

where $\boldsymbol{x}(\omega,k) = [x_1(\omega,k),\cdots,x_M(\omega,k)]^T$, $\boldsymbol{s}(\omega,k) = [s_1(\omega,k),\cdots,s_N(\omega,k)]^T$ and $G(\omega)$ are the discrete Fourier transformed representation of the mixtures, the sources and the transfer function matrix, respectively.

The mixtures are generally whitened as

$$\tilde{\boldsymbol{x}}(\omega,k) = Q(\omega)\boldsymbol{x}(\omega,k) \tag{4}$$

where $Q(\omega)$ is a whitening matrix.

By applying ICA under the assumption that each components $s_n(t)$ of $\boldsymbol{s}(t)$ are statistically independent of each other, the separated spectra $\boldsymbol{u}(\omega,k) = [u_1(\omega,k),\cdots,u_N(\omega,k)]^T$ can be obtained as

$$\boldsymbol{u}(\omega,k) = H(\omega)\tilde{\boldsymbol{x}}(\omega,k) \tag{5}$$

where $H(\omega) = [\overline{\boldsymbol{h}}_1(\omega),\cdots,\overline{\boldsymbol{h}}_N(\omega)]^T$ is a demixing matrix and " $^-$ " denotes a conjugation operator.

The separated signal $u_n(t)$ in the time domain for the source $s_n(t)$ can be obtained by applying the inverse Fourier transform of spectrograms $\{u_n(\omega,k)|k=0,1,\cdots,K-1\}$.

$$u_n(t) = \frac{1}{2\pi}\frac{1}{W(t)}\sum_k\sum_\omega e^{j\omega(t-k\tau)}u_n(\omega,k) \tag{6}$$

where $W(t) = \sum_k w(t-k\tau)$. Fig.1 shows the frequency domain ICA under N-source and M-sensor configuration.

B. Frequency domain FastICA algorithm

Under the assumption that all the spectra of whitened mixtures $\tilde{\boldsymbol{x}}(\omega,k)$ are zero mean and have unit variances with uncorrelated real and imaginary parts of equal variances. FastICA algorithm is formulated in the frequency domain as follows [2].

$$\boldsymbol{h}_n^+(\omega) = \frac{1}{K}\sum_{k=0}^{K-1}\{\tilde{\boldsymbol{x}}(\omega,k)\overline{u_n}(\omega,k)f(|u_n(\omega,k)|^2) - \Big[f(|u_n(\omega,k)|^2)+|u_n(\omega,k)|^2 f'(|u_n(\omega,k)|^2)\Big]\boldsymbol{h}_n(\omega)\} \tag{7}$$

$$\boldsymbol{h}_n(\omega) = \frac{\boldsymbol{h}_n^+(\omega)}{\parallel \boldsymbol{h}_n^+(\omega) \parallel} \tag{8}$$

where $\boldsymbol{h}_n(\omega)$ is a demixing weight vector, $f(\cdot)$ a nonlinear function and $f'(\cdot)$ is its differential.

At each frequency ω, the convergence condition is

$$|\overline{\boldsymbol{h}}_{n,old}^T(\omega)\boldsymbol{h}_{n,new}(\omega)| \simeq 1 \tag{9}$$

where $\boldsymbol{h}_{\cdot,old}(\cdot)$ and $\boldsymbol{h}_{\cdot,new}(\cdot)$ denote the demixing weight before and that after update, respectively.

In addition, $\boldsymbol{h}_{n+1}(\omega)$ is orthogonalized as

$$\boldsymbol{h}_{n+1}(\omega) = \boldsymbol{h}_{n+1}(\omega) - \sum_{i=1}^{n}\boldsymbol{h}_i(\omega)\overline{\boldsymbol{h}}_i^T(\omega)\boldsymbol{h}_{n+1}(\omega) \tag{10}$$

and $\boldsymbol{h}_{n+1}(\omega)$ is again regularized by Eq.(8).

The separated spectra $\boldsymbol{u}_n(\omega,k)$ are yielded by substituting $\boldsymbol{h}_n(\omega)$ to Eq.(5). The separated signal $\boldsymbol{u}(t)$ in the time domain can be obtained by Eq.(6).

C. Indeterminacy of scale and permutation

In the frequency domain, the indeterminacy of scale and permutation occur at every frequency ω:

$$H(\omega)Q(\omega)G(\omega) = P(\omega)D(\omega) \tag{11}$$

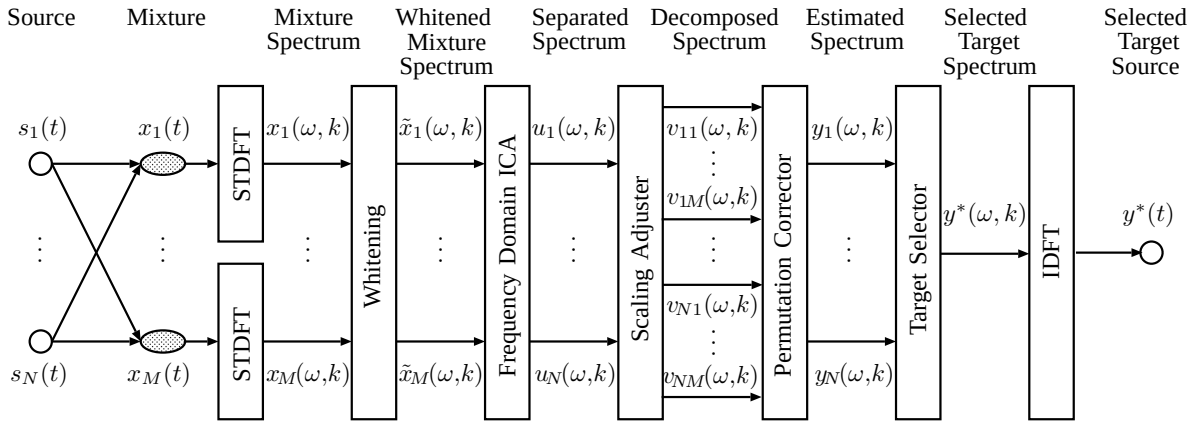


Fig. 2. Target source signal estimation based on frequency domain independent component analysis.

where $P(\omega)$ is a permutation matrix, which all elements of each column and row 0 except for one element with value 1, and $D(\omega) = \mathrm{diag}[d_1(\omega), \cdots, d_N(\omega)]$ a diagonal matrix, of which elements $d_n(\omega)$ denote the scaling factors determined by the whitening. This means that not only the amplitude but also the phase are indeterminate. Therefore, the indeterminacy of permutation, the amplitude and the phase must be settled to get a meaningful signal $\boldsymbol{u}(t)$ before inversely transforming $\boldsymbol{u}(\omega, k)$ from the frequency to the time domain.

D. Solution of scale indeterminacy

Fig.2 shows a target source signal estimation process based on the frequency domain ICA under N-source and M-sensor configuration.

In order to solve the scale indeterminacy, a decomposed spectrum $\boldsymbol{v}_n(\omega, k) = [v_{n1}(\omega, k), \cdots, v_{nM}(\omega, k)]^T$ is calculated as follows [5].

$$\boldsymbol{v}_n(\omega, k) = B(\omega)^{-1}[0, \cdots, 0, u_n(\omega, k), 0, \cdots, 0]^T \tag{12}$$

where $B(\omega) = H(\omega)Q(\omega)$. The sum of the decomposed spectra is equal to the mixture $\boldsymbol{x}(\omega, k)$.

The decomposed spectrum $v_{nm}(\omega, k)$ is uniquely determined as a product of the source spectrum $s_n(\omega, k)$ and the transfer function $g_{mn}(\omega)$, although the combination of the source spectrum and the transfer function differ depending on whether permutation occur or not [9], [10]. It means that the scaling factor of the decomposed spectrum is the transfer function itself and the decomposed spectrum has no scale indeterminacy.

E. Solution of permutation indeterminacy

We assume that the source $s_n(t)$ is closer to the n-th sensor than to others. From this assumption, gain and phase inequalities on the transfer functions obtained, respectively, by

$$|g_{nn}(\omega)| > |g_{mn}(\omega)|, \tag{13}$$

$$\angle g_{nn}(\omega) > \angle g_{mn}(\omega). \tag{14}$$

From the above relations, we adopt the estimated spectrum of the source $s_n(\omega, k)$ as $g_{nn}(\omega)s_n(\omega, k)$, since it may be less affected by ambient noise than the others. To do this, we calculate the absolute value of every component in $\boldsymbol{v}_n(\omega, k)$ and compare them to choose the maximum value, since every component contains the same $s_n(\omega, k)$ and differs from the others by its transfer function.

Therefore, if $|v_{nm}(\omega, k)|$ takes the maximum value at $m = n$, the number n indicates the one corresponding to the source $s_n(\omega, k)$ and $v_{n,m=n}(\omega, k)$ becomes the estimate of $s_n(\omega, k)$. This discussion is formulated as a permutation correction rule [9], [10] with respect to the gain inequality Eq.(13):

$$\hat{n} = \arg\max_m |v_{nm}(\omega, k)|. \tag{15}$$

Similarly, another permutation correction rule is derived from the phase inequality Eq.(14):

$$\hat{n} = \arg\max_m \angle v_{nm}(\omega, k). \tag{16}$$

F. Target source signal selection

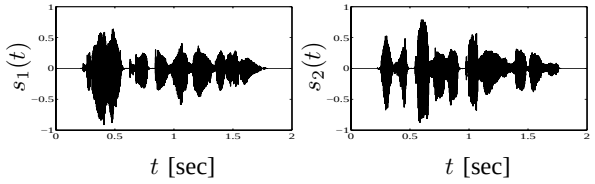
There still remains a target source selection problem even if the scale and permutation problem can be settled. After permutation correction using Eq.(15) or Eq.(16), the separated spectrum $y_n(\omega, k)$ can be expressed as

$$y_n(\omega, k) = g_{nn}(\omega)s_n(\omega, k). \tag{17}$$

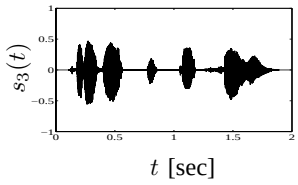
This shows that $y_n(\omega, k)$ is an estimate of $s_n(\omega, k)$. If we know the target source number n or the target source location in advance, we can extract the target source easily as

$$y^*(\omega, k) = y_n(\omega, k). \tag{18}$$

In the case where the location of human speech is not known, another selecting method has been proposed as follows. We assume that one source is human speech and the others are noises, but we do not know the location of human speech. Human speech is usually larger in non-Gaussianity than noises. The FastICA [2] proposed by Hyvärinen generates



(a) source signal $s_1(t)$ (b) source signal $s_2(t)$



(c) source signal $s_3(t)$

Fig. 3. Source signals.

the separated signal in order of large non-Gaussianity. Under the above situation, therefore, the separated spectra for the speech are most frequently yielded at the first output channel. In other word, the first output channel gives the spectral estimate for the speech at the highest probability [9], [10].

III. Blind source separation under a dynamic acoustic environment

A. Estimation for the number of the sources

The original source signals can be recovered using ICA if the number of the source signals N is equal to that of the observed signals M. However, the separation performance of ICA often deteriorates if $N \neq M$. Here, we propose an estimation method for N under the two-sensor configuration.

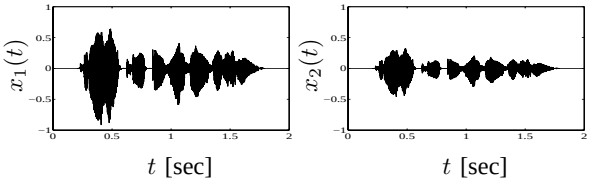
Consider three source signals $s_n(t)$ $(n=1,2,3)$ which are human speeches shown in Fig.3. If there is no active source, it is clearly that the observed signals doesn't have power. Therefore, we estimate $N=0$ in the case where the power of the observed signals is very small.

When only $s_1(t)$ is active, the waveforms $x_1(t)$ and $x_2(t)$ observed at the sensors are depicted as in Fig.4. Their joint distribution is shown in Fig.5(a) where the horizontal and the vertical axis are denoted by $x_1(t)$ and $x_2(t)$, respectively. Since $x_1(t)$ and $x_2(t)$ are completely similar, the joint distribution is expressed by a straight line. This fact implies that the distribution is of one-dimensional structure in the case of one active source. The histograms $\theta(t)$ of the joint distributions as

$$\theta(t) = \tan^{-1} \frac{x_2(t)}{x_1(t)} \tag{19}$$

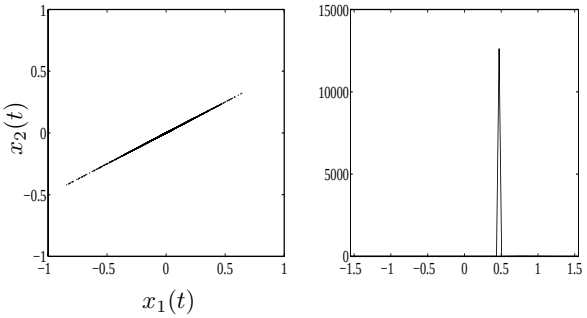
has only one peak as shown in Fig.5(b). The horizontal axis denotes the source arrival direction from $-\frac{\pi}{2}$ to $\frac{\pi}{2}$ and the vertical axis denotes the frequency.

In the case of two active sources, $s_1(t)$ and $s_2(t)$, the observed mixture signals $x_1(t)$ and $x_2(t)$ are shown in Fig.6. As shown in Fig.7(a), their joint distribution is scattered around



(a) observed signal $x_1(t)$ (b) observed signal $x_2(t)$

Fig. 4. Observed signals in the case of $N=1$.



(a) joint distribution (b) histogram

Fig. 5. Joint distribution and histogram in case of $N=1$.

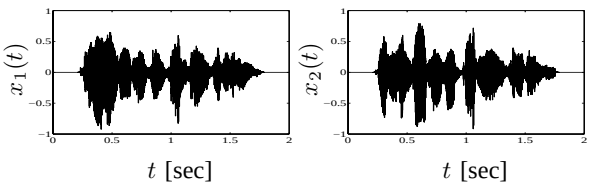
but is characterized by two dense crossing lines. Fig.7(b) shows the histograms in the case of two active sources. In the figure, two peaks are clearly seen.

In the case of three active sources, $s_1(t)$, $s_2(t)$ and $s_3(t)$, the observed mixture signals $x_1(t)$ and $x_2(t)$ are shown in Fig.8. Their joint distribution and the histgram are shown in Fig.9(a) and (b), respectively. In these figure, the dense crossing lines are still discernible and three peaks are recognizable.

From these results, the joint distribution is considered to take on a peculiar structure depending on the number of the sources. Their histograms have the same number of peaks as the sources signals. Therefore, we can estimate the number of the source signals by finding the number of the peaks.

B. Blind source separation under a dynamic acoustic environment

From the above discussions, the histgrams of the observed signals have the same peaks of the sources. Therefore, we



(a) observed signal $x_1(t)$ (b) observed signal $x_2(t)$

Fig. 6. Observed signals in the case of $N=2$.

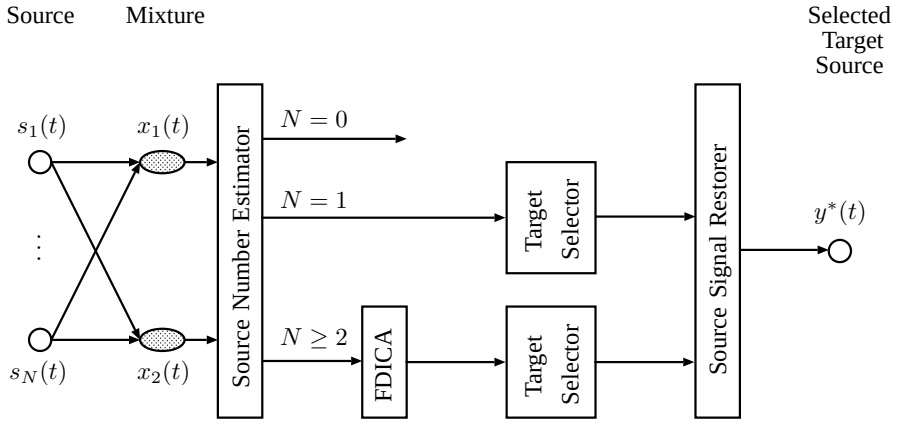
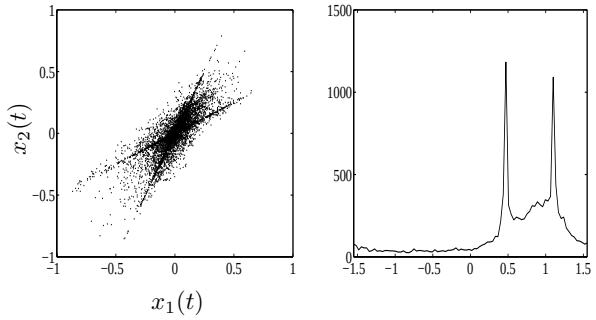
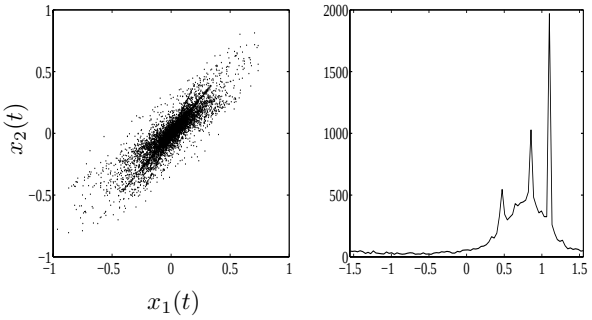


Fig. 10. Blind source separation based on the estimation for the number of the source signals.



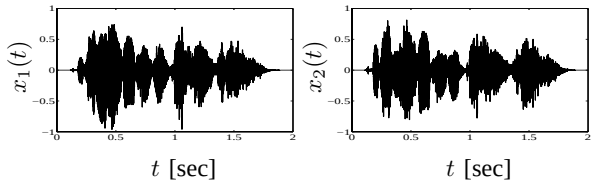
(a) joint distribution (b) histogram

Fig. 7. Joint distribution and histogram in the case of $N = 2$.



(a) joint distribution (b) histogram

Fig. 9. Joint distribution and histogram in the case of $N = 3$.



(a) observed signal $x_1(t)$ (b) observed signal $x_2(t)$

Fig. 8. Observed signals in the case of $N = 3$.

can estimate the number of the blind sources from only the ovserved signals. And a new blind source separation method under a dynamic acoustic environment is proposed as shown in Fig.10. The proposed method is based on the source number estimation, the target source signal selection and frequency domain ICA. Namely, in the case of $N = 0$, we do not output anything. In the case of $N = 1$, the observed signal is selected the target signal or not. In the case of $N \geq 2$, we use the frequency domain ICA and the target selection.

IV. Simulation

In order to verify our proposals, several simulations were carried out. The target source signal $s_1(t)$ was speech signal of the database [12]. The noise source $s_2(t)$ was a roaring train noise recorded at a station premises [13]. Fig.11(a) shows the source signals. These data was set for the number of sound sources to become dynamic each time. Namely, from 0 to 1 sec, the number N is equal to 0. From 1 to 2 sec, $N = 1$ and the target source signal is only active. $N = 2$ from 2 to 3 sec. From 3 to 4 sec, $N = 1$ and the target source signal is not active. From 4 to 5 sec, the number N is equal to 0 again. Using these sources, the mixture signals are generated in Fig.11(b).

The signals were sampled at a rate of 8000Hz with 16bit resolution. In the source number estimation, the sampled data were processed with a frame length 500ms. In the frequency domain ICA, the sampled data were windowed by the Hamming window with a frame length 128ms and a frame shift time 32ms. In the FastICA aigorithm, the weight were initialized by complex random values such that $||\boldsymbol{h}_n(\omega)|| = 1$, the nonlinear function was specified as

$$f(|u_n(\omega, k)|^2) = 1 - 2/(e^{2|u_n(\omega,k)|^2} + 1). \quad (20)$$

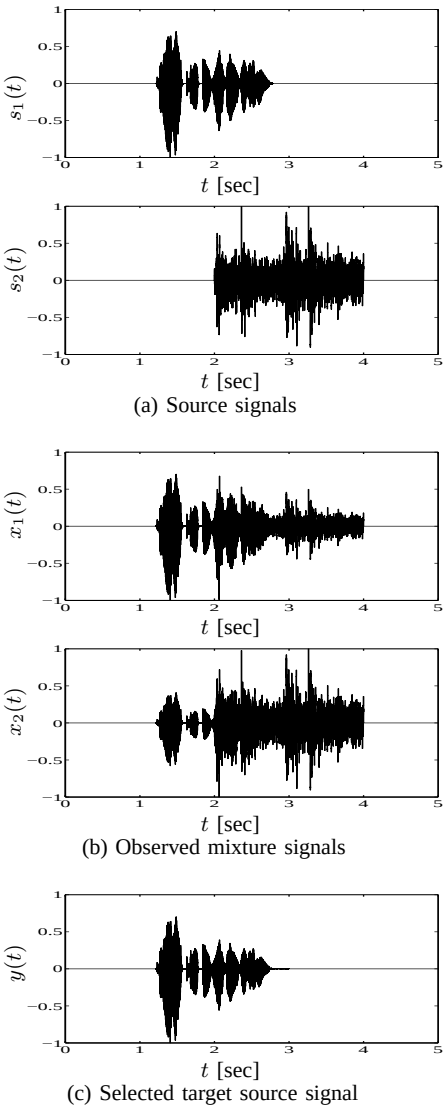


Fig. 11. Experimental results on blind source separation under a dynamic acoustic environment.

The algorithm was repeated until the convergence criterion

$$|\bar{\boldsymbol{h}}_{n,old}^{T}(\omega)\boldsymbol{h}_{n,new}(\omega)| > 0.999999 \tag{21}$$

is satisfied.

Fig.11(c) shows the selected target source signal. It is found that the selected signal is estimated for the source $s_1(t)$. From the simuration result, it is clarified that our proposed method works well under a dynamic acoustic environment.

V. Conclusion

Based on the distributions of the observed signals, the method for the number of the source signals estimation is proposed under two-sensor configuration. The proposed method can estimate the number of the sources in the case that the number of the source signals is larger than that of the observed signals. And it can be applied in many fields such as the speech recognition and radio communication. From these simulation results, it is found that the number of the sources are coincident to that of peaks in the histogram. And it is clarified that our proposed method works well under a dynamic acoustic environment.

Acknowledgment

This research has been supported by the Tateishi Science and Technology Foundation and the Kayamori Foundation of Informational Science Advancement.

References

[1] A. Cichocki and S. Amari, *Adaptive blind signal and image processing, Learning algorithm and applications* John Wiley & Sons, Ltd, 2002.
[2] A. Hyvärinen, J. Karhunen and E. Oja, *Independent component analysis* John Wiley & Sons, Ltd, 2001.
[3] S. Amari, *Natural gradient works efficiently in learning* Neural Computation, Vol. 10, pp. 251-276, 1998.
[4] T. W. Lee, M. Girolami and T. J. Sejnowski, *Independent component analysis using an extended informax algorithm for mixed subgaussian and supergaussian sources* Neural Computation, Vol. 11, No. 2, pp. 417-441, 1999.
[5] N. Murata, S. Ikeda and A. Ziehe, *An approach to blind source separation based on temporal structure of speech signals* Neurocomputing, Vol. 41, Issue 1-4, pp. 1-24, 2001.
[6] S. Ikeda and N. Murata, *A method of ICA in time-frequency domain* International Workshop on Independent Component Analysis and Blind Signal Separation (ICA'99), pp. 365-371, 1999.
[7] H. Saruwatari, H. Yamajo, T. Takatani, T. Nishikawa, and K. Shikano, *Blind separation and deconvolution for convolutive mixture of speech combining SIMO-model-based ICA and multichannel inverse filtering* IEICE Trans. Fundamentals, Vol. E88-A, No. 9, pp. 2387-2400, 2005.
[8] T. Koya, T. Ishibashi, H. Shiratsuchi and H. Gotanda, *Blind source deconvolution based on frequency domain convolution model under highly reverberant environments* Transactions of the Institute of Systems, Control and Information Engineers, Vol. 22, No. 8, pp. 287-294, 2009.
[9] H. Gotanda, K. Nobu, T. Koya, K. Kaneda, T. Ishibashi and N. Haratani, *Permutation correction and speech extraction based on split spectrum through FastICA* 4th International Symposium on Independent Component Analysis and Blind Signal Separation (ICA2003), pp. 379-384, 2003.
[10] T. Ishibashi, K. Inoue, H. Gotanda and K. Kumamaru, *Frequency domain independent component analysis without permutation and scale indeterminacy* Proceedings of the 41st ISCIE International Symposium on Stochastic Systems Theory and Its Applications, pp. 190-195, 2009.
[11] H. Sawada, R. Mukai, S. Araki and S. Makino, *Estimating the number of sources using independent component analysis* Acoustical Science and Technology, the Acoustical Society of Japan, Vol. 26, No. 5, pp. 450-452, 2005.
[12] Acoustical Society of Japan, *ASJ continuous speech corpus Japanese newspaper article sentences* JNAS Vols.1-16, 1997.
[13] NTT Advanced Technology Corporation, *Ambient noise database for telephonometry 1996* 1996.

Takaaki Ishibashi received the Ph.D. degree in Engineering from Kyushu Institute of Technology, in 2007.

He is Associate Professor of the Department of Information Communication and Electronic Engineering, Kumamoto National College of Technology. And his research area of interest is Digital Signal Processing, and Human Interface.

Dr. Ishibahshi is a member of the Institute of Electronics, Information and Communication Engineers (IEICE), the Institute of Systems, Control and Information Engineers (ISCIE), The Society of Instrument and Control Engineers (SICE) and the Astronomical Society of Japan (ASJ).