

An Approach for Blind Source Separation using the Sliding DFT and Time Domain Independent Component Analysis

Koji Yamanouchi, Masaru Fujieda, Takahiro Murakami, and Yoshihisa Ishida

Abstract—“Cocktail party problem” is well known as one of the human auditory abilities. We can recognize the specific sound that we want to listen by this ability even if a lot of undesirable sounds or noises are mixed. Blind source separation (BSS) based on independent component analysis (ICA) is one of the methods by which we can separate only a special signal from their mixed signals with simple hypothesis.

In this paper, we propose an online approach for blind source separation using the sliding DFT and the time domain independent component analysis. The proposed method can reduce calculation complexity in comparison with conventional methods, and can be applied to parallel processing by using digital signal processors (DSPs) and so on. We evaluate this method and show its availability.

Keywords—Cocktail party problem, blind Source Separation (BSS), independent component analysis, sliding DFT, online processing.

I. INTRODUCTION

THE blind source separation (BSS) [1], [2], [5]-[11] is an approach to estimate the original source signals using only the information of 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) is one of statistical analysis methods and identifies the independent components in the random variables. ICA can be used as a technique for BSS. In acoustic signal separation, the problem of estimating source signals is more complex because the mixing process is not given by liner combination.

In this paper, we propose an approach to solve the blind source separation problem of acoustic signals using the sliding DFT [3], [4] and the online time domain independent component analysis (TDICA) [7], [9], [11]. By combining the sliding DFT and the TDICA, it can cover the fault of several conventional methods. Since this method uses online algorithm, real-time processing by using digital signal processor is expected.

Manuscript received November 30, 2006.

The authors are with the School of Science and Technology, Meiji University 214-8571 Kanagawa, Japan (corresponding author to provide phone: +81-44-934-7307; fax: +81-44-934-7883; e-mail: ce55069@isc.meiji.ac.jp).

This paper is organized as follows. In section II, we show the framework of blind source separation of acoustic signals, and section III shows the framework of independent component analysis, and in section IV, we introduce the proposed method. Finally, in section V, we evaluate this method.

II. BLIND SOURCE SEPARATION OF ACOUSTIC SIGNALS

A. Statement of Problem

In general, the mixture model of acoustic signals is given by a form of finite impulse response filter, which is written as follows:

$$\mathbf{X}(t) = \sum_{\tau=0}^K \mathbf{H}(\tau) \mathbf{S}(t-\tau) \quad (1)$$

where

$$\mathbf{H}(\tau) = \begin{bmatrix} h_{11}(\tau) & \cdots & h_{1m}(\tau) \\ \vdots & \ddots & \vdots \\ h_{n1}(\tau) & \cdots & h_{nm}(\tau) \end{bmatrix}, \quad (2)$$

$\mathbf{X}(t) = [x_1(t), x_2(t), \dots, x_m(t)]^T$ is the observed signal vector, $\mathbf{S}(t) = [s_1(t), s_2(t), \dots, s_n(t)]^T$ is the source signal vector, and $h_{kl}(\tau)$ is a room impulse response between k -th source and l -th sensor. In ordinary, the length of the impulse response is from several hundred milliseconds to a few seconds. Fig. 1 shows the mixture model of acoustic signals (1).

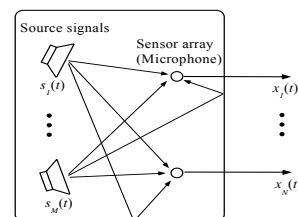


Fig. 1 The mixing model in blind source separation

In this paper, we focus on the mixture of acoustic signals, which consists of discrete acoustic signals.

The goal of blind source separation (BSS) problem is to determine the original source signals $\mathbf{S}(t)$ using only the information obtained by the observation signals $\mathbf{X}(t)$.

B. The Conventional Method of Blind Source Separation

There are several methods to solve the BSS problem. One of these methods is independent component analysis (ICA). This method supposes that the source signals are statistically independent. Another is Time-Frequency Masking Technique [10]. This method supposes that the source signals are mutually sparse. We introduce a method using the ICA.

III. INDEPENDENT COMPONENT ANALYSIS (ICA)

A. Overview of Independent Component Analysis

The independent component analysis (ICA) is the one of the statistical methods estimating unknown original variables from their mixed variables. ICA supposes that the original variables are mutually statistical independent. The general linear mixture model of ICA is given by the following equation:

$$\mathbf{X} = \mathbf{A} \cdot \mathbf{S} \quad (3)$$

where $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m]^T$ is the obtained mixed variables, $\mathbf{S} = [\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_n]^T$ is the original variables, and \mathbf{A} is the mixing matrix.

ICA estimates the original variable by finding the separating matrix \mathbf{W} which makes the output variables $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n]^T$ “mutually independent”.

$$\mathbf{Y} = \mathbf{W} \cdot \mathbf{X} \quad (4)$$

It is known that, the case in which mixing model is written by (3), we can estimate original variables as \mathbf{Y} in (4).

As the criterion for independence of variables, we use the mutual information method via Kullback-Leibler divergence introduced in [12].

To apply ICA to mixed acoustic signals, time-frequency domain approach (FDICA) and time domain approach (TDICA) are proposed as conventional methods.

B. Time Frequency Domain ICA (FDICA)

FDICA [1], [2], [5]-[7] is a method using the frequency domain transformation in observation signals; thus (1) turns into the following model:

$$\mathbf{X}(\omega, t) = \mathbf{H}(\omega)\mathbf{S}(\omega, t) \quad (5)$$

For fixed ω in (5), it is regarded as the model in (3). Fig. 2 shows the model of FDICA in the case with two source signals and two observation signals.

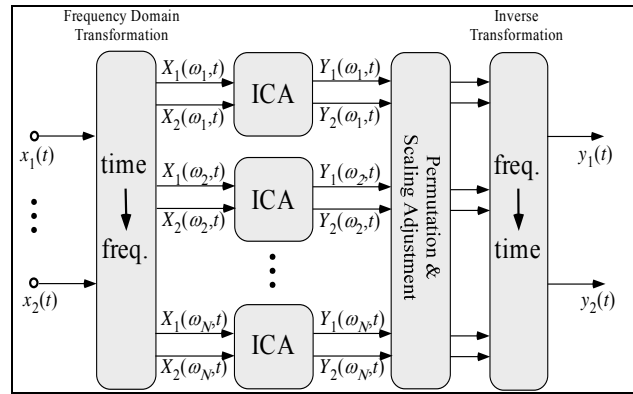


Fig. 2 The overview of FDICA

The advantage of this method is that a calculation is simple. On the other hand, the disadvantage of this method is inexactness of approximation in the case K is larger than DFT window length. Furthermore, it is necessary to consider ambiguity of amplitude and permutation.

C. Time Domain ICA (TDICA)

TDICA [7]-[9], [12] is a method which estimate separating filter. In this case, the separating process is written as follows:

$$\mathbf{Y}(t) = \sum_{\tau=0}^L \mathbf{W}(\tau)\mathbf{X}(t-\tau) \quad (6)$$

where $\mathbf{Y}(t) = [y_1(t), y_2(t), \dots, y_n(t)]^T$ is the output signals in time t , and $\mathbf{W}(\tau)$ is the τ -th tap of the separating filter which makes output signals mutually independent.

In strictly, because the separating filter is inversion of mixing filter, it is written by the form of infinite impulse response filter; thus $L \rightarrow \infty$. However, in ordinary, the approximation by finite impulse response filter has been used for instance and stability of separating filter.

TDICA can be applicable even in the case that the length K of mixing filter is large. On the other hand, the problem of TDICA is the complexity of calculation. There is the case that the coefficient of separation filter does not converge when the mixing filter length K is large.

IV. PROPOSED METHOD

A. Overview of the Proposed Method

In this section we present our method combining the sliding DFT (SDFT) [3], [4] as band division filter and the online TDICA algorithm as blind source separation technique. Fig. 3 shows the model of this method in the same case with Fig. 2.

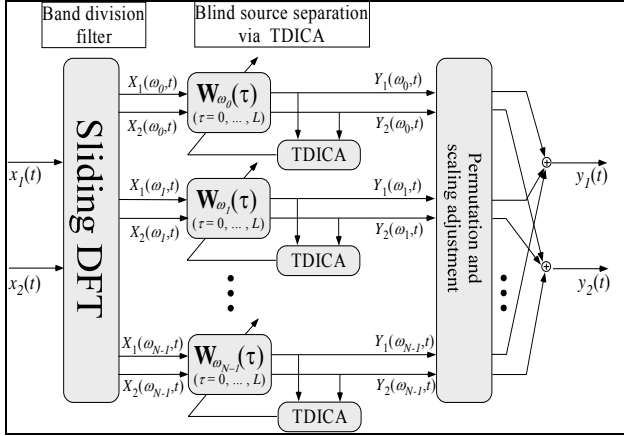


Fig. 3 The overview of proposed method

It is considerable that there are two dominant advantages in this method. One is the availability in long mixing filter taps. By using band division filter, it is considered that the length of mixing filter in a single band is smaller than original mixing filter length. Thus, it can reduce the complexity of calculation in comparison with the case without band division filters. The other is that on-line processing is available. However, in this method, the scaling and the permutation problem occur. To solve these problems for FDICA, several methods have been proposed. However, they cannot apply to this method directly. In this paper, we adjust them manually. Indeed, adjustment of scaling and permutation remains future research topics.

B. The Sliding DFT (SDFT)

In this paper, we use the sliding DFT [3], [4] as band division filter. The SDFT calculates an N-point DFT within a sliding window.

For single frequency bin ω_k , the new spectral component of $x(t)$ is derived by using the previous spectral component:

$$X(\omega_k, t) = e^{j2\pi k/N} [X(\omega_k, t-1) + x(t) - x(t-N)], \quad (7)$$

where $X(\omega_k, t)$ is the new spectral component and $X(\omega_k, t-1)$ is the previous spectral component. The feature of SDFT is simplicity of calculating frequency component. To calculate $X(\omega_k, t)$, the sliding DFT requires only one complex multiplication and two real adds. Fig. 4 shows the single-bin SDFT structure.

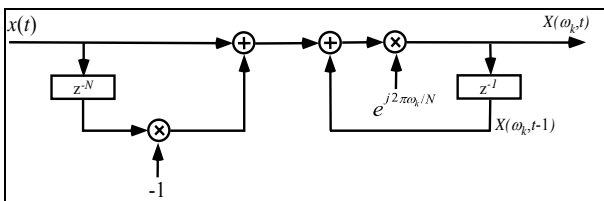


Fig. 4 SDFT structure of a single frequency bin

C. TDICA Algorithm

As mentioned in section II, we use the feed-forward TDICA algorithm based on the minimization of mutual information via Kullback-Leibler divergence in [12]. For a single frequency bin ω_k , the update rule of separating filter is written as follows:

$$\mathbf{W}_{\omega_k}(\tau, t+1) = \mathbf{W}_{\omega_k}(\tau, t) + \eta \left[\mathbf{W}_{\omega_k}(\tau, t) - \Phi(\mathbf{Y}_{\omega_k}(t-L)) \left(\sum_{q=0}^L \mathbf{W}_{\omega_k}^H(L-q, t) \mathbf{Y}_{\omega_k}(t-q) \right)^H \right] \quad (8)$$

where $\mathbf{W}_{\omega_k}(\tau, t)$ is the τ -th tap of separating filter in t -th iteration number for ω_k , $\mathbf{Y}_{\omega_k}(t)$ is the output signals in time t for ω_k , η is the step size parameter, $\Phi(\bullet)$ is a nonlinear function, where $\Phi(\bullet) = 2 \tanh(G \cdot \text{Re}(\bullet)) + 2j \tanh(G \cdot \text{Im}(\bullet))$, (G : gain for nonlinear function), and \bullet^H is the Hermitian transpose. Note that, in the case of $L=0$, this algorithm can apply for FDICA.

V. EXPERIMENTS

A. Experiment Conditions

A signal separation experiment uses the mixture of speech signals. The signals are two Japanese speech sounds recorded separately, and the conditions of mixing process are as follows,

$$\begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} = \sum_{\tau=0}^K \begin{bmatrix} h_{11}(\tau) & h_{12}(\tau) \\ h_{21}(\tau) & h_{22}(\tau) \end{bmatrix} \begin{bmatrix} s_1(t-\tau) \\ s_2(t-\tau) \end{bmatrix}. \quad (9)$$

where the value of mixing filter $h_{kl}(\tau)$ is hypothesized by 16 taps filter under sampling frequency $f_s = 16$ [kHz] as shown in Fig. 5.

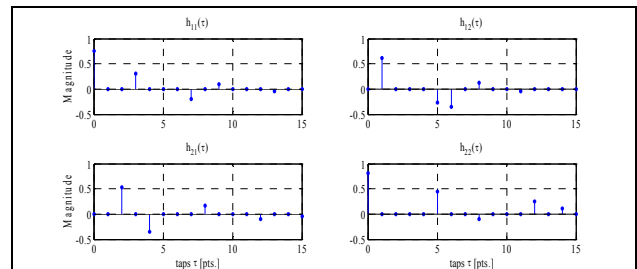


Fig. 5 Hypothesized room impulse response value

The parameters of separation processing are shown in Table I.

TABLE I
THE PARAMETER OF SEPARATION PROCESSING

Length of speech data		12.5[s] = 2.0×10^5 [pts.]
Number of band division (proposed method)	N	8
Number of separation filter taps (proposed method)	K	8
Number of separation filter length (conventional TDICA)		32
Gain for nonlinear function	G	300
Step size parameter	η	1.0×10^{-6}

B. Evaluation

For evaluation of experimental results, the noise reduction rate (NRR) [5] of short time window is considered. The NRR is defined by follows:

$$\text{NRR}(t_s) = \text{SNR}_{out}(t_s) - \text{SNR}_{in}(t_s) \quad , \quad (10)$$

where $t_s = [t_0, t_0 + \Delta t, t_0 + 2\Delta t, \dots]$ and $\text{SNR}_{out}(t_s)$ are respectively calculated as:

$$\text{SNR}_{out}(t_s) = \frac{\sum_{t_k=t_s-P/2}^{t_s+P/2} (s(t_k) - \bar{s}_{t_s})^2}{\sum_{t_k=t_s-P/2}^{t_s+P/2} ((y(t_k) - s(t_k)) - (\bar{y}_{t_s} - s_{t_s}))^2} \quad (11)$$

$$\text{SNR}_{in}(t_s) = \frac{\sum_{t_k=t_s-P/2}^{t_s+P/2} (s(t_k) - \bar{s}_{t_s})^2}{\sum_{t_k=t_s-P/2}^{t_s+P/2} ((x(t_k) - s(t_k)) - (\bar{x}_{t_s} - s_{t_s}))^2} \quad (12)$$

where P is the size of short time window, and over line $\bar{\bullet}$ denotes the average of signals.

In this experiment, we apply $P = 8.0 \times 10^3$ (= 500[ms]) which is the size of short time window, and the amount of window shift $\Delta t = 1.6 \times 10^3$ (=100[ms]).

C. Experimental Results

Fig. 6 shows a result of the experiments. The proposed method obtains a comparatively good result in NRR with only 8-taps separating filters.

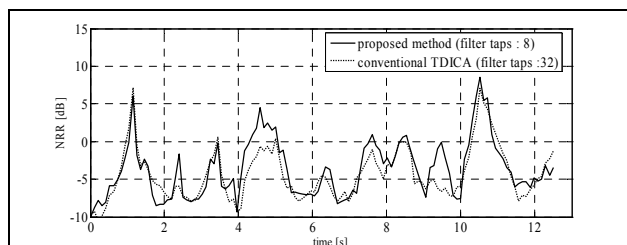


Fig. 6 Noise reduction rate of short time window

VI. CONCLUSION

In this paper, an approach to blind source separation using the sliding DFT and the time domain independent component analysis has been proposed. From the experimental result, the proposed method has obtained a level equivalent to the conventional method in noise reduction rate by using shorter filter taps. Considering these facts, the proposed method can reduce complexity of a calculation. Furthermore, it has been shown that the approach combining the sliding DFT and online time domain independent component analysis algorithm is available for solving the blind source separation problem of acoustic signals.

REFERENCES

- [1] N. Murata, S. Ikeda and A. Ziehe "An approach to blind source separation based on temporal structure of speech signals" *Neurocomputing*, 41, 1-24 (2001)
- [2] F. Asano, S. Ikeda, M. Ogawa, H. Asoh and N. Kitawaki "Combined Approach of Array Processing and Independent Component Analysis for Blind Separation of Acoustic Signals" *IEEE Trans. Speech and Audio Processing*, 11, No.3, 204-214 (May. 2003)
- [3] E. Jacobsen and R. Lyons "The Sliding DFT", *IEEE Signal Processing Magazine*, 74-80 (Mar. 2003)
- [4] E. Jacobsen and R. Lyons "An Update to the Sliding DFT", *IEEE Signal Processing Magazine*, 110-111 (Jan. 2004)
- [5] S. Kurita, H. Saruwatari, S. Kajita, K. Takeda, and F. Itakura "Evaluation of blind signal separation method using directivity pattern under reverberant conditions", *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*, SAM-P2-5, 3140-3143 (Jun. 2000)
- [6] S. Araki, R. Mukai, S. Makino, T. Nishikawa, and H. Saruwatari "The Fundamental Limitation of Frequency Domain Blind Source Separation for Convolutional Mixture of Speech", *IEEE Trans. Speech and Audio Processing*, 11, No.2, 109-116 (Mar. 2003)
- [7] T. Nishikawa, H. Saruwatari, K. Shikano, S. Araki, S. Makino "Multistage ICA for Blind Source Separation of Real Acoustic Convolutional Mixture", *Proceedings of 4th International Symposium on Independent Component Analysis and Blind Source Separation (ICA2003)* 523-528 (Apr. 2003)
- [8] R. Aichner, H. Buchner, S. Araki, S. Makino "On-line time-domain blind source separation of nonstationary convolved signals", *Proceedings of 4th International Symposium on Independent Component Analysis and Blind Source Separation (ICA2003)* 987-992 (Apr. 2003)
- [9] N. Murata, S. Ikeda "An On-line Algorithm for Blind Source Separation on Speech Signals", *Proceedings of NOLTA'98* 923-926 (Sep.1998)
- [10] O. Yilmaz, S. Rickard, "Blind Separation of Speech Mixtures via Time-Frequency Masking", *IEEE Trans. on Signal Processing*, vol.52, No.7 1830-1847 (Jul. 2004)
- [11] Te-Won Lee "Independent Component Analysis - Theory and Applications -" Kluwer Academic Publishers
- [12] A. Hyvarinen, J. Karhunen, E. Oja "Independent Component Analysis" Wiley-Interscience