ICIC Express Letters
Volume **12**, Number **3**, March **2018**

ICIC International ©2018 ISSN 1881-803X
pp. **229–236**

TWO-CHANNEL MICROPHONE SYSTEM WITH VARIABLE ARBITRARY DIRECTIONAL PATTERN

Chiharu Okuma[1], Kiyoteru Hayama[2] and Takaaki Ishibashi[2]

[1]Department of Human-Oriented Information Systems Engineering
[2]Department of Information, Communication and Electronic Engineering
National Institute of Technology, Kumamoto College
2659-2, Suya, Koshi, Kumamoto 861-1102, Japan
chiharu@kumamoto-nct.ac.jp

Received September 2017; accepted December 2017

Abstract. *In this paper, we propose a two-channel microphone system with variable arbitrary directional pattern for multiple sound signals. The proposed method can change the direction of unidirectional pattern and can select multiple directivity patterns. The method can reduce noise even when the number of sound sources is larger than that of observed mixture signals. In addition, the noise reduction system is developed using two-channel microphone arrays. The system can be processed in real time. By several simulations, it is found that the proposed method can extract the target speech signals from multi speaker's utterance sounds.*
Keywords: Noise reduction, Variable arbitrary directional pattern, Two-channel microphone system

1. **Introduction.** Many noise reduction methods such as ICA (Independent Component Analysis) [1], FDICA (Frequency Domain ICA) [2], SS (Spectral Subtraction) [3], SAFIA (sound source Segregation based on estimating incident Angle of each Frequency component of Input signals Acquired by multiple microphones) [4], MUSIC (MUltiple SIgnal Classification) [5] and NMF (Nonnegative Matrix Factorization) [6] have been proposed. These methods can estimate the original source signals. However, these algorithms need a lot of data points since the methods are based on the stochastic theory and use the frequency domain information using Fourier transform. In addition, the methods take much time to estimate because the methods use an iteration scheme to estimate the parameter. Therefore, these methods are not suitable for processing in real time. As far as the authors know, there is no development of a device to be applied for real-time processing based on these methods.

In order for real-time processing, we have already proposed some systems to reduce noise from mixture signals observed by two-channel microphones [7, 8]. These systems are implemented with a microcontroller and can suppress noise in real time. The systems extract only one target sound, and all other sounds were deleted as noise. In actual application, when multiple speakers utter in a noisy environment, the multi-speaker's speech becomes target sounds. It means that we need multiple speaker's voice extraction methods under a noisy environment.

In this paper, we propose a simple two-channel microphone system with variable arbitrary directional pattern. A typical microphone forms unidirectional directivity pattern. Therefore, when multiple speakers speak at the same time, it is necessary to place as many microphones as the number of speakers. Our method works well even when the number of sound sources is larger than that of microphones. The target speeches of one person or plural persons are emphasized by our method. And other speech and noise are reduced.

Since the proposed method is a very simple algorithm and the amount of calculation is small, it can be expected to be applied to industries.

This paper consists of five sections. In Section 2, a principle of controlling a single directivity by a program is provided. In Section 3, we propose a microphone system with multiple directivity patterns. In Section 4, we experimentally evaluate the estimation performance of the proposed method. Finally, the paper is concluded in Section 5.

2. **Control of Directional Pattern.** When some sound sources $s_n(t)$ $(n = 1, 2, \dots, N)$ are observed by some microphones, the observed mixture signals $x_m(t)$ $(m = 1, 2, \dots, M)$ are expressed as follows:

$$x_m(t) = \sum_{n=1}^{N} a_{mn} s_n(t), \tag{1}$$

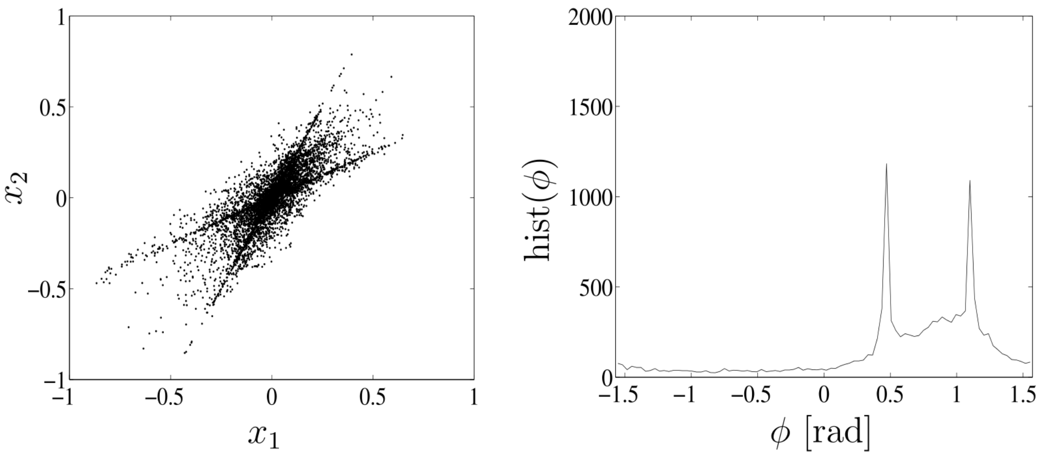
where a_{mn} denote mixing parameters, t denotes index of time series, and N and M are the numbers of sound sources and mixture signals, respectively.

Using two mixture signals observed by two microphones in the case of two speaker's utterances at the same time, a joint distribution is plotted as Figure 1(a). The horizontal and vertical axes denote the amplitudes of $x_1(t)$ and $x_2(t)$, respectively. In the following, we consider the audio signals as stochastic varieties and omit the time series t.

From Figure 1(a), it is found that the joint distribution has two linear components. For clarity of the linear components, directions ϕ of distribution are calculated as follows:

$$\phi = \tan^{-1} \frac{x_2}{x_1}. \tag{2}$$

Using all ϕ, its histogram $\mathrm{hist}(\phi)$ as shown in Figure 1(b) means that the components of the joint distribution are concentrated on two directions. The reason is that since the human speech has a silent interval, it becomes only one source in many data points.

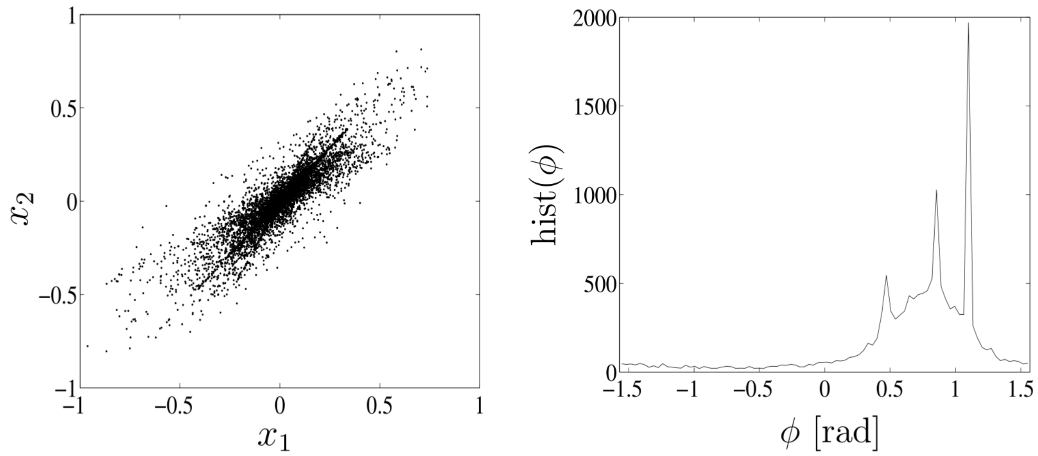


(a) Joint distribution (b) Histogram

FIGURE 1. Joint distribution and histogram ($N = 2$)

In the case of three active sources, a joint distribution and a histogram observed two microphones are shown in Figures 2(a) and 2(b), respectively. In these figures, the dense crossing lines are still discernible and three peaks are recognizable. These values of the inclinations of lines depend on the transfer functions from the sources to the microphones.

From the above discussions, in the case that the target source signal and the noise signals exist at the same time, the joint distributions made from the observed signals have straight lines depending on transfer functions of each source. Then, it means that the target source signal can be estimated by extracting the straight line.



(a) Joint distribution (b) Histogram

Figure 2. Joint distribution and histogram ($N = 3$)

To extract the target speech using the joint distribution means to reduce the straight line components generated by the noise, and to extract the straight line components created by the target signals. Based on the fact, the authors have proposed a noise reduction method based on the ratio of observed signals by two microphones [7]. We define the error ε as

$$\varepsilon = \Phi - \phi, \tag{3}$$

where Φ denotes the direction of the target source. From the above discussion, Φ is estimated as the mode value (the most frequent value) of ϕ as follows:

$$\Phi = \arg\max_{\phi} \mathrm{hist}(\phi). \tag{4}$$

In the case that the target source signal is multiple speakers' speech, we estimate another peak Φ_k again excluding the mode value Φ and neighbor value.

In order to reduce sounds in a direction away from the target source, we use some functions $f_i(\varepsilon)$ that gain decreases as the ε increases as

$$f_1(\varepsilon) = \frac{1}{1 + \alpha\varepsilon^2} \tag{5}$$

$$f_2(\varepsilon) = -0.5 \tanh\left(\beta\varepsilon^2 - \gamma\right) + 0.5 \tag{6}$$

$$f_3(\varepsilon) = \begin{cases} 1 & \text{if } \varepsilon^2 \le \delta \\ 0 & \text{if } \varepsilon^2 > \delta \end{cases} \tag{7}$$

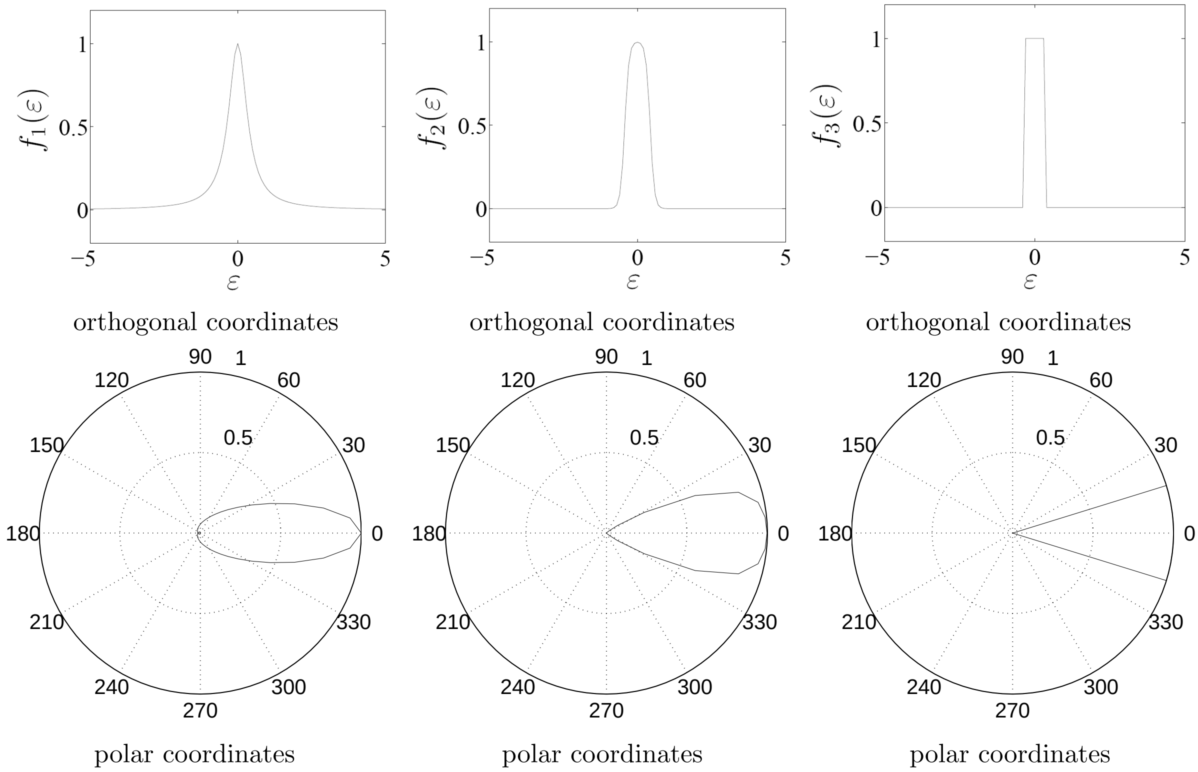
where α, β, γ and δ denote learning parameters for determining the amount of noise reduction. Figures 3(a), 3(b) and 3(c) show orthogonal coordinates and polar coordinates of $f_1(\varepsilon)$, $f_2(\varepsilon)$ and $f_3(\varepsilon)$, respectively.

From these figures, the noise reduction functions $f_i(\varepsilon)$ mean spatial filters on the ratio of the transfer functions. By the noise reduction function, only the signal having the specific direction according to the transfer functions is extracted. Signals having other directions of the transfer functions are suppressed.

The target signal y is estimated based on the ϕ using observed signals x_m mixed multiple source signals as follows:

$$y = f_i(\varepsilon)x_m. \tag{8}$$

3. **Variable Arbitrary Directional Pattern.** The noise reduction functions proposed by the authors have been unidirectional [7, 8]. It means that the functions $f_i(\varepsilon)$ emphasize only one sound. In some applications, it is often shown that the target signal is a speech



(a) $f_1(\varepsilon)$ (b) $f_2(\varepsilon)$ (c) $f_3(\varepsilon)$

FIGURE 3. Noise reduction functions

among multiple speakers. Therefore, we propose a method that can record multiple speakers' speech reducing other noises.

As a function with multiple directivity pattern, we adopt the Rose curve g_l as follows,

$$g_l = \begin{cases} \sin l(\theta - \varphi) & \text{if } l : \text{odd} \\ \sqrt{\sin l(\theta - \varphi)} & \text{if } l : \text{even} \end{cases} \tag{9}$$

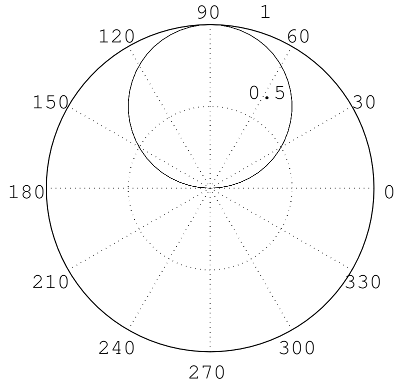
where θ denotes the angle from 0 to 360 degrees and φ denotes the phase. The Rose curves are shown in Figure 4. From the figures, the directivity can change shape and direction by using g_l.

Using the noise reduction functions $f_i(\varepsilon)$ and the Rose curve g_l, a variable arbitrary directional pattern is proposed as follows and some directivity patterns are shown in Figure 5.

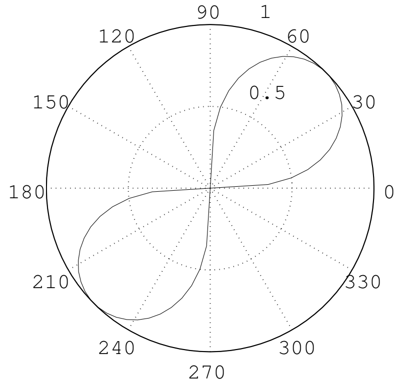
$$y = g_l f_i(\varepsilon) x_m. \tag{10}$$

Even when the number of sound sources is larger than the number of observed mixture signals, our proposed method can reduce noise signals. In addition, the algorithm of our noise reduction method is very simple. Then the method works in real time. From the same way, the method can also reduce only one source signal. It means that the method has a function to reduce only the target source signal.

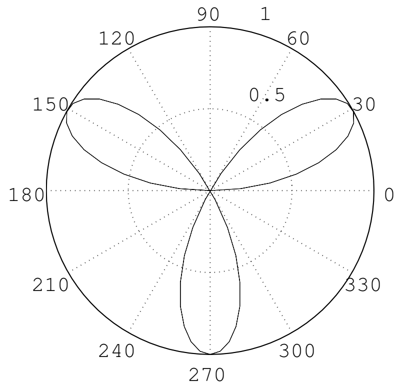
4. **Simulation.** In order to verify our proposals, several simulations were carried out. At first, we evaluate the estimation performance of the proposed method with unidirectional pattern. The source signals $s_1(t)$, $s_2(t)$, $s_3(t)$ and $s_4(t)$ were human speech signals in the database [9]. Figure 6(a) shows these source signals. The signals were sampled at the rate of 8 [kHz] with 16 [bit] resolution. In order to clarify the speech segment, the target signal $s_1(t)$ was active from 0 to 6 [s] and the noises $s_2(t)$, $s_3(t)$ and $s_4(t)$ are active from 2 to 8 [s]. Using these sources, the mixture signals generated by Equation (1) are shown



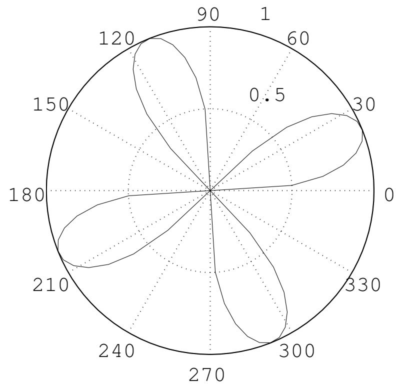
(a) g_1



(b) g_2

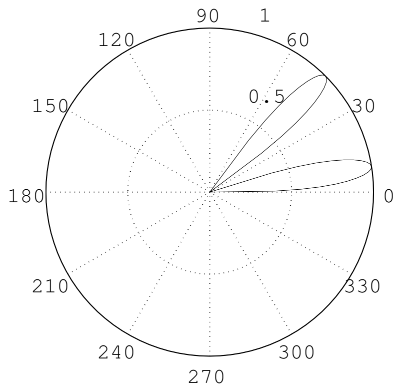


(c) g_3

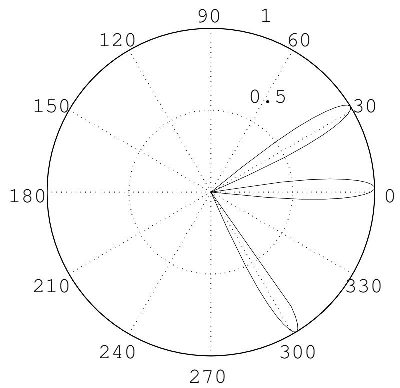


(d) g_4

FIGURE 4. Rose curves



(a) $g_{10}f_3(\varepsilon)$

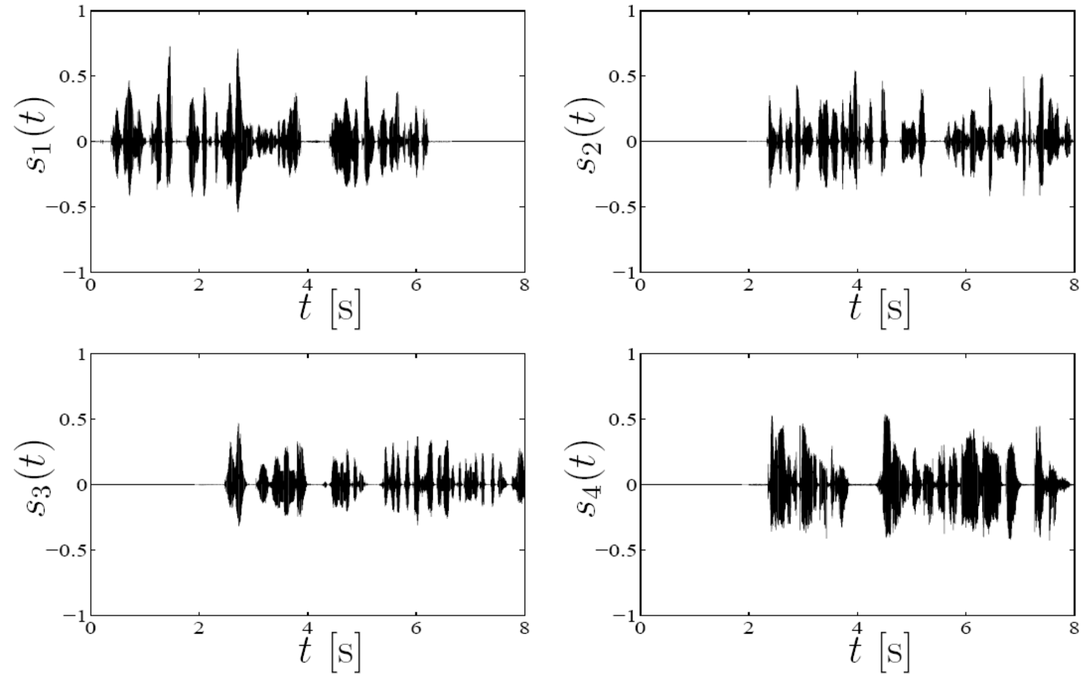


(b) $g_{12}f_3(\varepsilon_1)f_3(\varepsilon_2)$

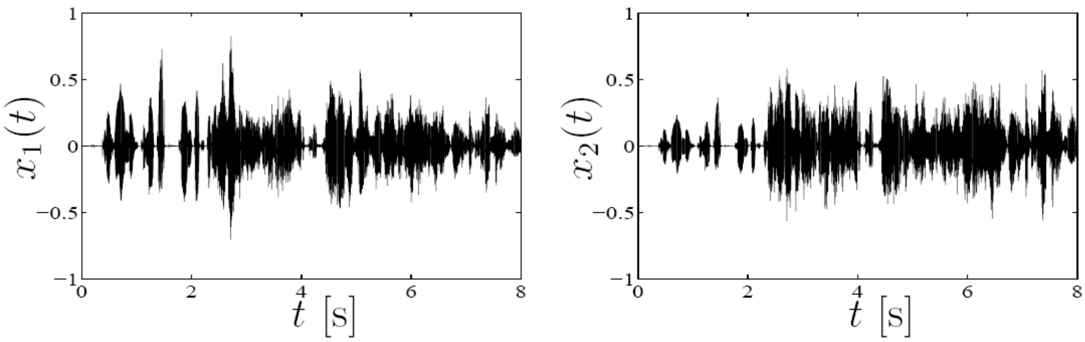
FIGURE 5. Multiple target extractable functions

in Figure 6(b). From these figures, it is recognized that the observed signals from 2 to 8 [s] are covered with noises.

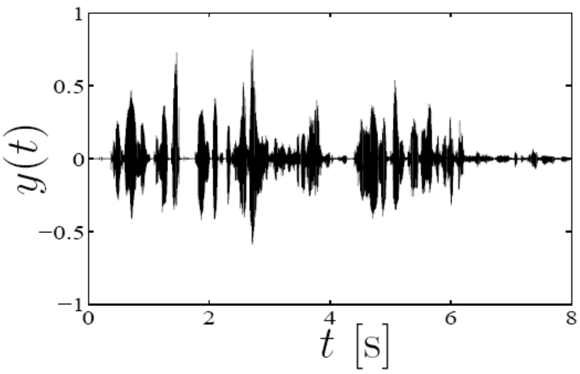
The estimated signal as shown in Figure 6(c) has amplitude in 0 to 6 [s]. The target signal $s_1(t)$ is extracted and the noises are reduced for both the period in which the target signal and noises exist from 2 to 6 [s], and the period in which only noises exist from 6 to 8 [s]. The estimated signal is similar to the target signal. From the results, it is found that the proposed method can extract the target signal in the case that the number of sound sources is larger than that of observed signals.



(a) Source signals



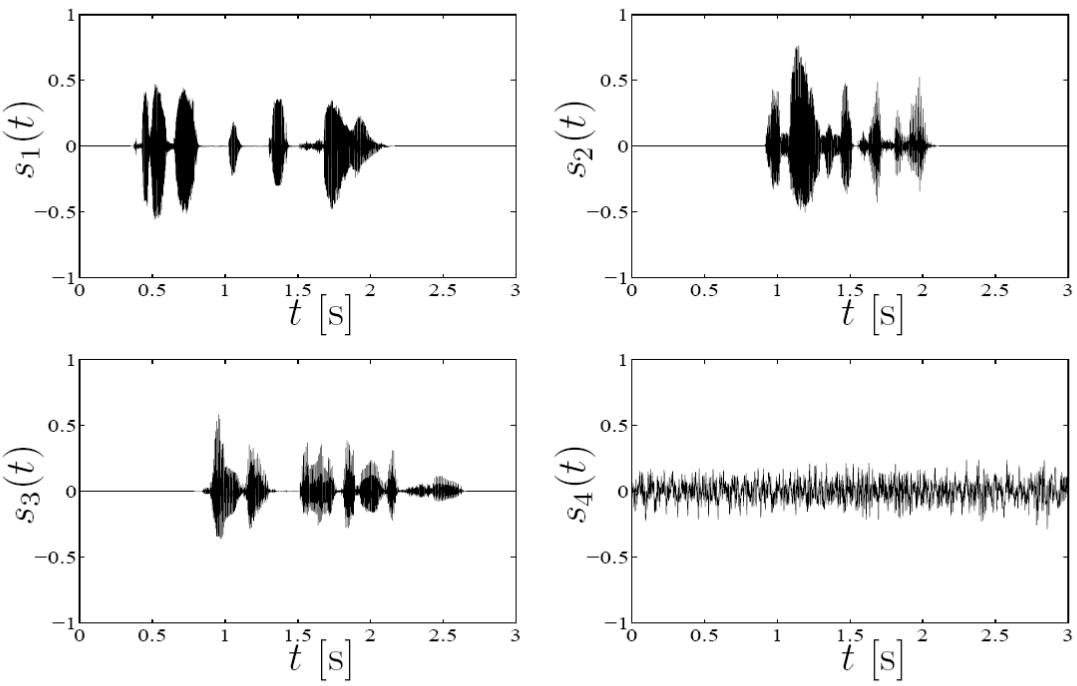
(b) Mixture signals



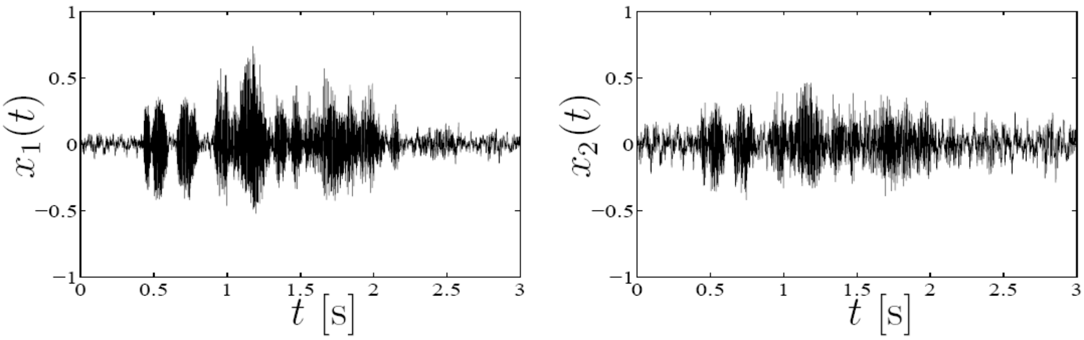
(c) Estimated signals

FIGURE 6. Simulation result of unidirectional pattern

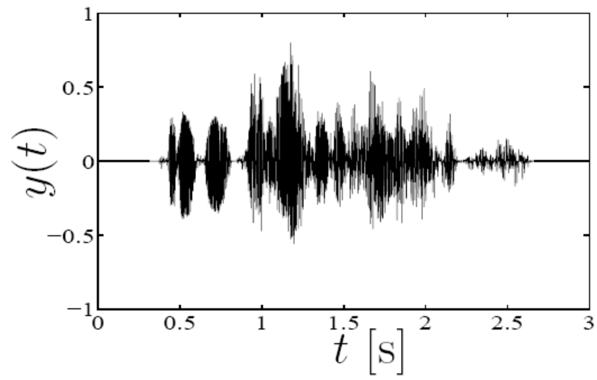
Next, we evaluate the estimation performance of the proposed method with multi-directional pattern. Three target signals were human speech signals [9] and the noise was car engine sound in the database [10]. Figure 7(a) shows these target signals and noise. The mixture signals are shown in Figure 7(b). The estimated signal is shown in Figure 7(c). From these figures, it can be confirmed that the estimated signal has the signal mixing three target sounds without noise. It means that our method can extract



(a) Source signals



(b) Mixture signals



(c) Estimated signals

FIGURE 7. Simulation result of multi-directional pattern

multi-speaker's speeches under the noisy environment. We confirmed that the method can obtain similar results with other combination of speech data and noise. It is confirmed that the proposed method works well even when the number of sound sources increases.

5. **Conclusions.** The two-channel microphone system with the variable arbitrary directional pattern is proposed. The proposed method can reduce the noise by deleting the distribution except a linear component depending on the target speech. The proposed method can extract the target speech, even when the number of source signals is larger than that of observed mixture signals. Furthermore, the method can estimate multi-speaker's speech without noise making multi-directional pattern, when multiple speaker's utterance under a noisy environment.

Acknowledgment. This work was supported by JSPS KAKENHI Grant Number JP 60455178. The authors also gratefully acknowledge the helpful comments and suggestions of the reviewers, which have improved the paper.

REFERENCES

[1] S. Makino, T.-W. Lee and H. Sawada, *Blind Speech Separation*, Springer, 2007.
[2] Y. Mizuno, K. Kondo, T. Nishino, N. Kitaoka and K. Takeda, Effective frame selection for blind source separation based on frequency domain independent component analysis, *IEICE Trans. Fundamentals of Electronics, Communications and Computer Sciences*, vol.E97-A, no.3, pp.784-791, 2014.
[3] S. F. Boll, Suppression of acoustic noise in speech using spectral subtraction, *IEEE Trans. Acoustics, Speech and Signal Processing*, vol.ASSP-27, no.2, pp.113-120, 1979.
[4] M. Aoki, M. Okamoto, S. Aoki, H. Matsui, T. Sakurai and Y. Kaneda, Sound source segregation based on estimating incident angle of each frequency component of input signals acquired by multiple microphones, *Acoustical Science and Technology*, vol.22, no.2, pp.149-157, 2001.
[5] Y. Sugimoto, S. Miyabe, T. Yamada, S. Makino and B.-H. Juang, An extension of MUSIC exploiting higher-order moments via nonlinear mapping, *IEICE Trans. Fundamentals of Electronics, Communications and Computer Sciences*, vol.E99-A, no.6, pp.1152-1162, 2016.
[6] A. Cichocki, R. Zdunek, A. H. Phan and S. Amari, *Nonnegative Matrix and Tensor Factorizations, Applications to Exploratory Multi-Way Data Analysis and Blind Source Separation*, John Wiley & Sons, Ltd., 2009.
[7] C. Okuma, T. Ishibashi, K. Hayama and H. Gotanda, Variable arbitrary directional characteristic pattern and its application to two-channel microphone system, *Proc. of Life Engineering Symposium*, pp.239-241, 2015.
[8] K. Hayama, T. Ishibashi, C. Okuma and H. Gotanda, Implementation of directional characteristics by real-time processing of sounds observed by two microphones, *ICIC Express Letters*, vol.10, no.1, pp.251-254, 2016.
[9] Acoustical Society of Japan, ASJ continuous speech corpus Japanese newspaper article sentences, *JNAS*, vols.1-16, 1997.
[10] NTT Advanced Technology Corporation, *Ambient Noise Database for Telephonometry*, 1996.