*ICIC Express Letters*
*Part B: Applications*
Volume **12**, Number **5**, May **2021**

ICIC International ©2021 ISSN 2185-2766
pp. 435–442

# BLIND SOURCE SEPARATION BASED ON ESTIMATION FOR THE NUMBER OF SOURCES AND TARGET SPEECH SELECTION

Takaaki Ishibashi[1] and Kei Eguchi[2]

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

[2]Department of Information Electronics
Fukuoka Institute of Technology
3-30-1 Wajiro-higashi, Higashi-ku, Fukuoka 811-0295, Japan

Received October 2020; accepted January 2021

Abstract. *This paper proposes an estimation method for the number of source signals under a two-microphone condition. A joint distribution of observed mixture signals by microphones has the same number of lines as the source signals. Therefore, we propose a number estimation method using image processing. The method can estimate using only observed mixture signals without the information of sources. Additionally, we propose a blind source separation method based on the estimated number of the source signals. The proposed methods have been verified by several simulations.*
**Keywords:** Blind source separation, Source number estimation, Target speech selection, Short frame processing, Dynamic acoustic environment

1. **Introduction.** Many applied technologies based on BSS (Blind Source Separation) have been developed. BSS is a technology that restores the original signal from a mixed observation signal without information on the source signal. ICA (Independent Component Analysis) [1, 2, 3], which is one of the BSS technologies, is expected in many fields, such as a speech recognition technology [4], EEG (Electroencephalogram) data analysis, MEG (Magnetoencephalograph) data analysis and image processing [5, 6, 7]. ICA can separate unknown sources from their mixtures without information on the transfer functions, provided that the sources are statistically independent.

The original sources can be completely recovered using ICA, when the number of source signals is equal to that of the observed mixture signals. However, separation performance often deteriorates because the number of the source signals is different from that of the mixture signals. Therefore, ICA is not good at estimating the original source signals when the number of sources is unknown. It is very important that the number of sources is estimated by using only the observed mixture signals before BSS process.

There have been proposed several estimation methods for the number of sources. The methods [8, 9] are based on a clustering of the speech signals. These methods only estimate the number of sound sources and do not consider the separation process. The method [10] functions well when the number of sources is equal to or less than that of the microphones. However, the method depends on the performance of BSS, and it fails if the number of sources is larger than that of microphones. Furthermore, BSS and target speech extraction in a dynamic environment where the number of sources changes is difficult.

DOI: 10.24507/icicelb.12.05.435

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-microphone configuration. The joint distribution has as many linear components as the number of sound sources. Therefore, the number of straight lines is detected by image processing. The proposed method can estimate the number of sources, even if it is larger than the number of microphones. Based on our number estimation, a target source signal is restored under dynamic conditions such as the starting point and end point of an utterance.

This paper is organized as follows. First, Section 1 is the introduction. Next, Section 2 details the blind source separation and the adjustment method for the scale of estimated signals. In Section 3, we propose an estimation method for the number of the source signals and a separation method according to the estimated number of source signals. Then, Section 4 shows the simulation results of the extraction of the target speech signal under the two-microphone configuration. Finally, Section 5 summarizes the results of this work.

2. **Blind Source Separation.** For the BSS processing in the case of instantaneously mixing, the mixture signals $\boldsymbol{x} = [x_1, \ldots, x_m, \ldots, x_M]^T$ by $M$ microphones are expressed as

$$\boldsymbol{x} = A\boldsymbol{s} \tag{1}$$

where $\boldsymbol{s} = [s_1, \ldots, s_n, \ldots, s_N]^T$ denotes unknown source signals, $N$ denotes the number of the sound sources and $A$ denotes an unknown mixing matrix. Under the assumption that each component of $\boldsymbol{s}$ is statistically independent, ICA can estimate the sources $\boldsymbol{s}$ except for indeterminacy of scaling and permutation. The separated signals $\boldsymbol{u} = [u_1, \ldots, u_n, \ldots, u_N]^T$, the estimate of the source signals $\boldsymbol{s}$, are expressed as

$$\boldsymbol{u} = W\boldsymbol{x} \tag{2}$$

where $W = [\boldsymbol{w}_1, \ldots, \boldsymbol{w}_n, \ldots, \boldsymbol{w}_N]^T$ denotes a separating matrix. The matrix $W$ is estimated by ICA algorithms such as natural gradient algorithm [2] and FastICA algorithm [3].

The separated signal $u_n$ using ICA algorithms has scaling indeterminacy and permutation problem as

$$WA = PD \tag{3}$$

where $P$ is a permutation matrix, which all elements of each column and row are 0 except for one element with value 1, and $D = \mathrm{diag}[d_1, \ldots, d_n, \ldots, d_N]$ a diagonal matrix, of which elements $d_n$ denotes the scaling factors.

In order to solve the scaling indeterminacy, a method using the inverse of the separating matrix $W^{-1}$ has been proposed as follows [11].

$$\boldsymbol{v}_n = W^{-1}[0, \ldots, 0, u_n, 0, \ldots, 0]^T \tag{4}$$

Then the final output signals $\boldsymbol{v}_n = [v_{n1}, \ldots, v_{nm}, \ldots, v_{nM}]^T$ are uniquely expressed as a product of the source signal $s_n$ and the transfer function $a_{mn}$ as follows [12].

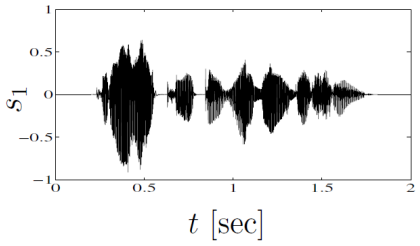
$$\boldsymbol{v}_n = [a_{1n}s_n, \ldots, a_{mn}s_n, \ldots, a_{Mn}s_n]^T \tag{5}$$

This means that $v_{nm}$ is the observation of the $n$-th source $s_n$ through the $m$-th microphone. These output signals by this approach are the same as [13]. It is also clarified that every $v_{nm}$ has no ambiguity of scale in that the scaling factor is a transfer function itself, while the scale factor $d_n$ for the separated signal $u_n$ varies arbitrarily.

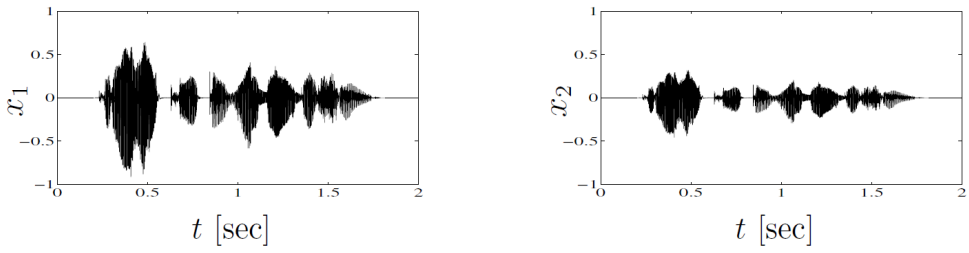
3. **Blind Source Separation Based on a Number Estimation Method.** The original source signals can be recovered using ICA when the number of the source signals $N$ is equal to that of the observed signals $M$. However, the separation performance of ICA often deteriorates when $N \neq M$. Therefore, we propose an estimation method for the number $N$ of the source signals under the two-microphone configuration.

3.1. **A number estimation based on Hough transform.** When there is no active source, it is clear that the observed signals do not have power. Therefore, we estimate $N = 0$ in the case where the power of the observed signals is very small.

When only $s_1$ is active, the waveforms $x_1$ and $x_2$ observed at the microphones are depicted as in Figure 1. From these waveforms, we generate their joint distribution as shown in Figure 2 where the horizontal and the vertical axes are denoted by amplitude of $x_1$ and $x_2$, respectively. Since $x_1$ and $x_2$ are completely similar, the joint distribution is expressed by a straight line.




(a) Source signal $s_1$




(b) Mixture signals $x_1$ and $x_2$

FIGURE 1. Observed signals in the case of $N = 1$

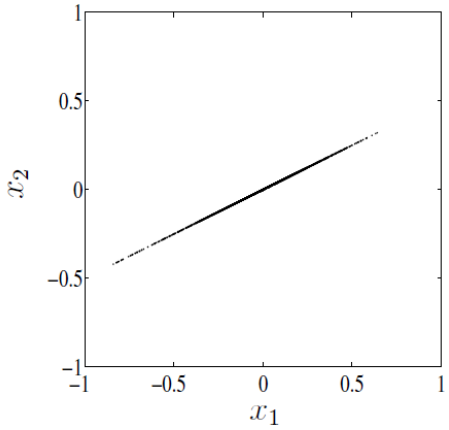
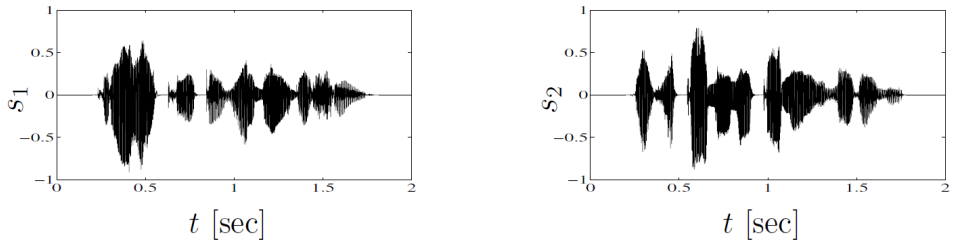



FIGURE 2. Joint distribution in the case of $N = 1$

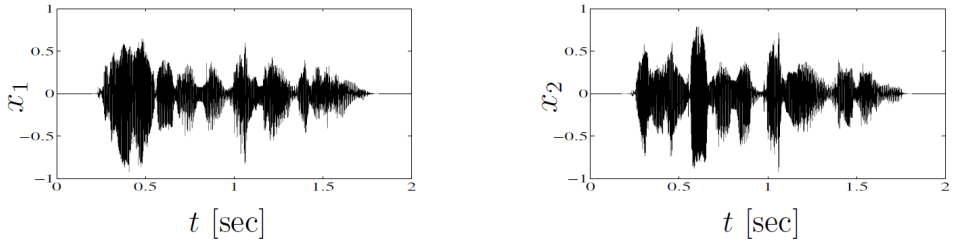
In the case of two active sources, $s_1$ and $s_2$, the observed mixture signals $x_1$ and $x_2$ are shown in Figure 3. As shown in Figure 4, their joint distribution is scattered around but is characterized by two dense crossing lines.

In the case of three active sources, $s_1$, $s_2$ and $s_3$, the observed mixture signals $x_1$ and $x_2$ are shown in Figure 5. Their joint distribution is shown in Figure 6. In this figure, the dense crossing lines are still discernible.

From these facts, in the case which there are active sources ($N \geq 1$), the joint distribution of the observed signals has the same number of straight lines as the sources.




(a) Source signals $s_1$ and $s_2$




(b) Mixture signals $x_1$ and $x_2$

FIGURE 3. Observed signals in the case of $N = 2$

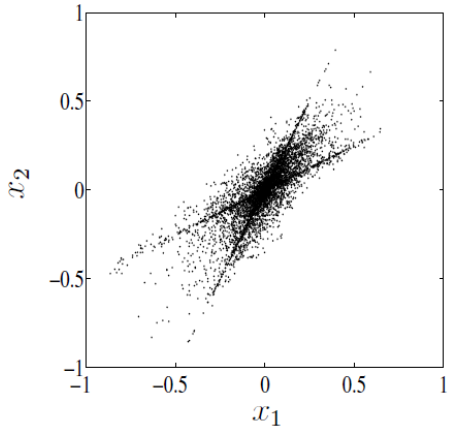
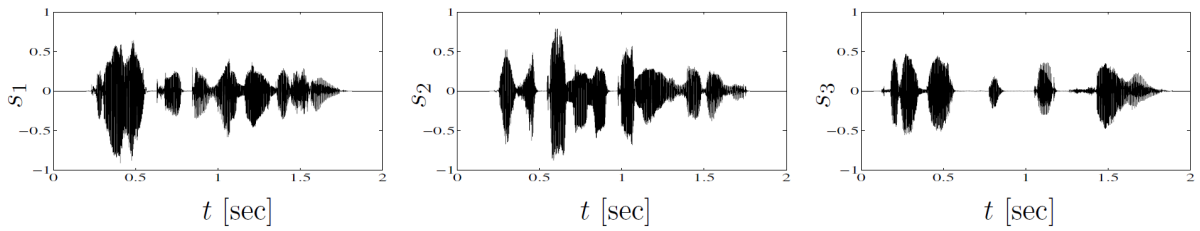
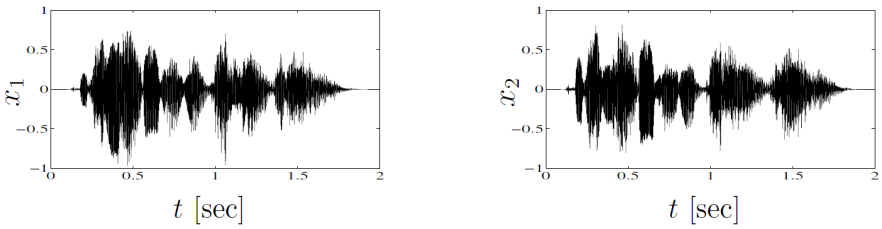



FIGURE 4. Joint distribution in the case of $N = 2$




(a) Source signals $s_1$, $s_2$ and $s_3$




(b) Mixture signals $x_1$ and $x_2$

FIGURE 5. Observed signals in the case of $N = 3$

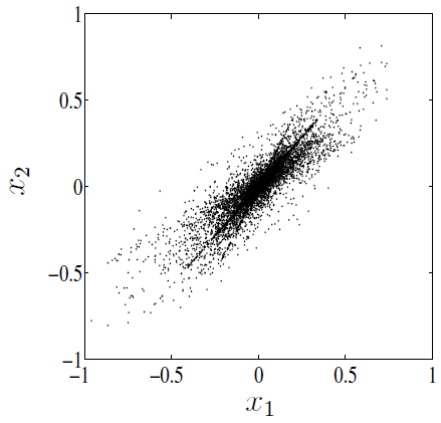



FIGURE 6. Joint distribution in the case of $N = 3$

Therefore, we propose a number estimation method based on Hough transform. The image plane by the joint distribution $(x_1, x_2)$ is transformed to Hough space $(\rho, \theta)$ by

$$\rho = x_1 \cos\theta + x_2 \sin\theta \quad (0 \leq \theta < \pi) \tag{6}$$

where $\rho$ denotes the distance between the straight line and the original point and $\theta$ denotes the angle of the vector from the original point to this closest point of the straight line. Using the Hough transform, we can estimate the number of the sources by majority rule.

The result by Hough transform from Figure 2 is shown in Figure 7. It is found that all curved lines generated from Figure 2 are getting through $(\rho, \theta) = (0, 2.1)$. In the case of $N = 2$ as shown in Figure 3, Figure 8 is calculated by Hough transform. From Figure 8, it

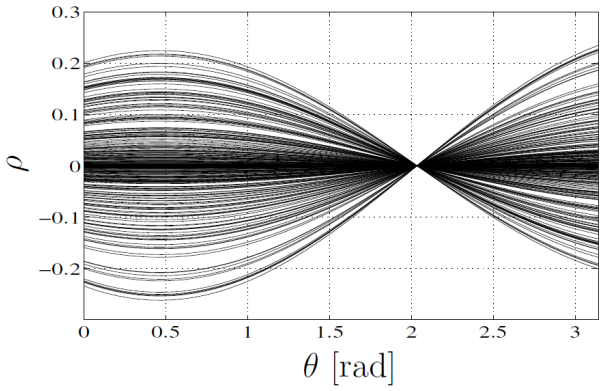



FIGURE 7. Hough transform in the case of $N = 1$

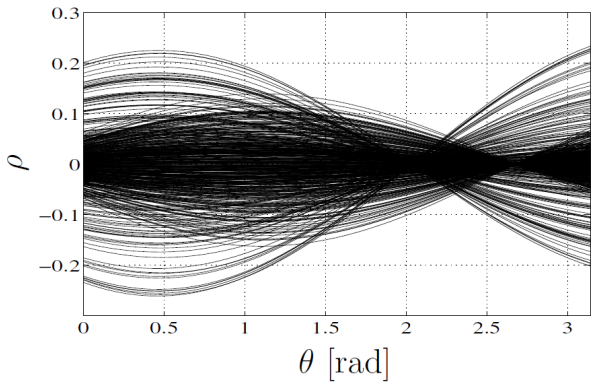



FIGURE 8. Hough transform in the case of $N = 2$

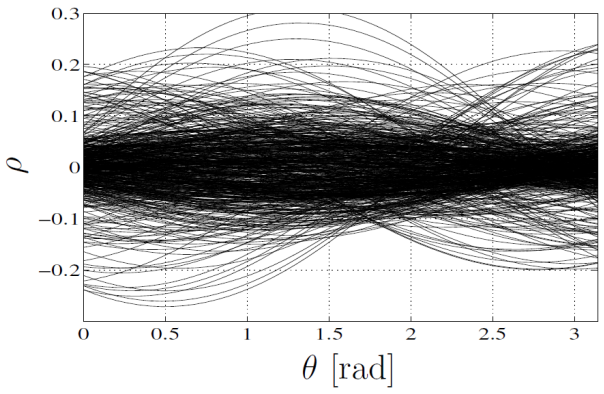



FIGURE 9. Hough transform in the case of $N = 3$

is found that the curved lines are concentrated on two points of $(\rho, \theta) = (0, 2.0), (0, 2.7)$. When the number of sources is three ($N = 3$), curved lines pass through three points of $(\rho, \theta) = (0, 1.7), (0, 2.0), (0, 2.7)$ in Figure 9. Using these results, we can estimate the number of the source signals by majority rule.

In order to calculate simply, we discuss about the joint distributions. The straight lines of joint distribution in acoustic signals are passed through original point, because a sound source is $E[x_m(t)] = 0$. Therefore, $\rho$ is equal to 0 in the $(\rho, \theta)$ space by Hough transform. Furthermore, in the acoustic field which a transfer function can approximate as a damping coefficient, a gradient of a straight line becomes a positive value. From these discussions, search space in the $(\rho, \theta)$ is reduced when we use conditions of $\rho = 0$ and $\frac{\pi}{2} \leq \theta \leq \pi$.

3.2. **Blind source separation based on a number estimation method.** From the above discussions, we can estimate the number of the blind sources from only the observed signals. And a new blind source separation method under a dynamic acoustic environment is proposed as shown in Figure 10. The proposed method is based on the source number estimation using Hough transform, the target source signal selection, and our BSS method [14]. Namely, when we estimate $N = 0$, 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 BSS and the target selection. The ICA can separate the original sources under the condition of $N = M$. The target signal is selected by [15].

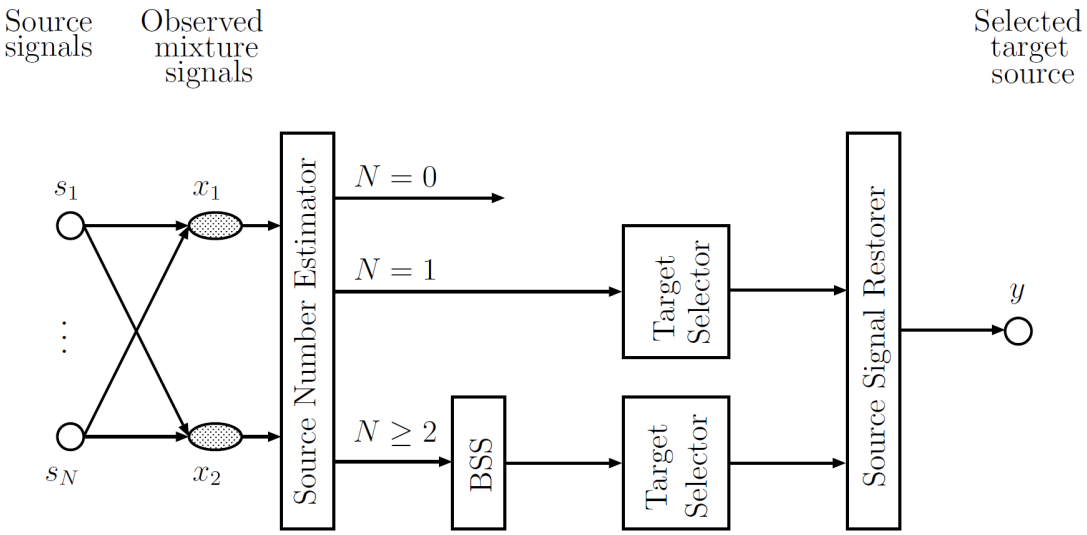
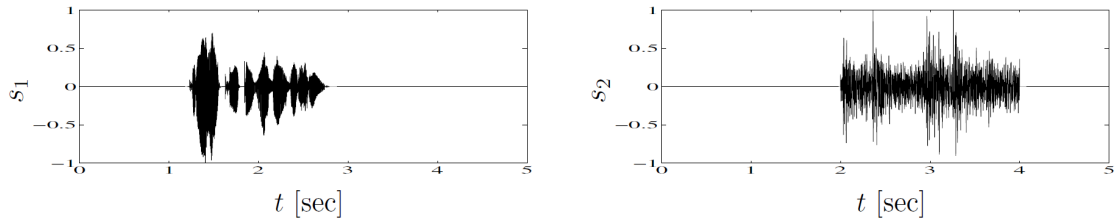


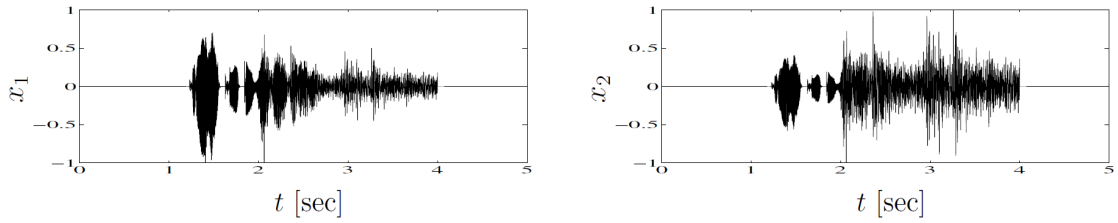

FIGURE 10. Blind source separation based on the estimation for the number of the source signals

4. **Simulation.** In order to verify our proposals, several simulations were carried out. The target source signals $s_1$ were 6 speaker's (3 females and 3 males) speech signals of the database [16]. The noise source signals $s_2$ were 5 pattern roaring train noises recorded at a station premise [17]. These signals were sampled at a rate of 8000Hz with 16bit resolution. The mixture signals ware calculated by Equation (1) which the diagonal components have $0.9 \pm \eta$ and non-diagonal components have $0.6 \pm \eta$, $\eta$ is a random value from 0 to 0.1. The simulations were carried out using 30 mixture signals.

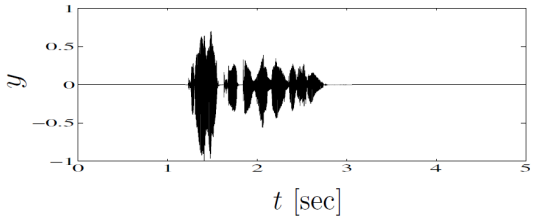
A result of the simulation is shown in Figure 11(a) shows the source signals. These data were set for the number of sound sources to become dynamically each time. Namely, from 0 to 1 second, the number $N$ is equal to 0. From 1 to 2 seconds, $N = 1$, only the target source signal is active. $N = 2$ from 2 to 3 seconds. From 3 to 4 seconds, $N = 1$, the target source signal is not active. From 4 to 5 second, the number $N$ is equal to 0 again. Using these sources, the mixture signals are generated in Figure 11(b).




(a) Source signals




(b) Observed mixture signals




(c) Selected target source signal

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

In the source number estimation, the sampled data were processed with a frame length 0.5 seconds. For the BSS, we use our BSS and target human speech extraction method [14]. Figure 11(c) shows the selected target source signal. It is found that the selected signal is estimated for the source $s_1(t)$.

The average value of RMSE (Root Mean Squared Error) and processing time of the 30 patterns of separated signals by the proposed method were 0.0287 and 0.0330 seconds, respectively. The average of RMSE and processing time using the natural gradient algorithm [2] were 0.1588 and 3.8643 seconds, respectively. From the simulation results, it is clarified that our proposed method works well under a dynamic acoustic environment.

5. **Conclusions.** In this paper, based on the distributions of the observed signals, the estimation method for the number of the source signals is proposed under a two-microphone

configuration. Our method using Hough transform can estimate the number of the sources in the case where the number of the source signals is larger than that of the observed signals. BSS based on our number estimation method works well under a dynamic acoustic environment. The proposed method has been verified by several experiments.

In this study, experiments were carried out under the conditions of two or less sound sources. A separation algorithm should be considered for multiple sound sources that have three or more sound sources in the conditions of two microphones.

**Acknowledgment.** This work was supported by JSPS KAKENHI Grant Number JP 20K12763. The authors also gratefully acknowledge the helpful comments and suggestions of the reviewers.

## REFERENCES

[1] T. W. Lee, *Independent Component Analysis, Theory and Applications*, Kluwer Academic Publishers, 1998.
[2] A. Cichocki and S. Amari, *Adaptive Blind Signal and Image Processing, Learning Algorithm and Applications*, John Wiley & Sons, Ltd., 2002.
[3] A. Hyvärinen, J. Karhunen and E. Oja, *Independent Component Analysis*, John Wiley & Sons, Ltd., 2001.
[4] H. M. S. Naing, R. Hidayat, B. Winduratna and Y. Miyanaga, Psychoacoustical masking effect-based feature extraction for robust speech recognition, *International Journal of Innovative Computing, Information and Control*, vol.15, no.5, pp.1641-1654, 2019.
[5] S. Ikeda and K. Toyama, Independent component analysis for noisy data: MEG data analysis, *Neural Networks*, vol.13, no.10, pp.1063-1074, 2000.
[6] J. Cao, N. Murata, S. Amari, A. Cichocki and T. Takeda, Independent component analysis for unaveraged single-trial MEG data decomposition and single-dipole source localization, *Neurocomputing*, vol.49, pp.255-277, 2002.
[7] A. Hyvärinen, J. Hurri and P. O. Hoyer, *Natural Image Statistics, A Probabilistic Approach to Early Computational Vision*, Springer, 2009.
[8] Y. Sun, Y. Xian, P. Feng, J. Chambers and S. M. Naqvi, Estimation of the number of sources in measured speech mixtures with collapsed Gibbs sampling, *Proc. of the 2017 Sensor Signal Processing for Defence Conference*, pp.1-5, 2017.
[9] I. Jafari, N. Ito, M. Souden, S. Araki and T. Nakatani, Source number estimation based on clustering of speech activity sequences for microphone array processing, *Proc. of the 2013 IEEE International Workshop on Machine Learning for Signal Processing*, pp.1-6, 2013.
[10] H. Sawada, R. Mukai, S. Araki and S. Makino, Estimating the number of sources using independent component analysis, *Acoustical Science and Technology*, vol.26, no.5, pp.450-452, 2005.
[11] N. Murata, S. Ikeda and A. Ziehe, An approach to blind source separation based on temporal structure of speech signals, *Neurocomputing*, vol.41, nos.1-4, pp.1-24, 2001.
[12] T. Ishibashi, K. Inoue, H. Gotanda and K. Kumamaru, Frequency domain independent component analysis without permutation and scale indeterminacy, *Proc. of the 41st ISCIE International Symposium on Stochastic Systems Theory and Its Applications*, pp.190-195, 2009.
[13] K. Matsuoka, Elimination of filtering indeterminacy in blind source separation, *Neurocomputing*, vol.71, nos.10-12, pp.2113-2126, 2008.
[14] T. Ishibashi H. Shintani and K. Nagata, Fast blind source separation and target human speech extraction method for acoustic signals, *ICIC Express Letters*, vol.11, no.12, pp.1715-1721, 2017.
[15] T. Ishibashi, K. Inoue, H. Gotanda and K. Kumamaru, A solution of permutation problem inherent in frequency domain ICA, *Proc. of the 36th ISCIE International Symposium on Stochastic Systems Theory and Its Applications*, pp.259-264, 2004.
[16] Acoustical Society of Japan, *ASJ Continuous Speech Corpus Japanese Newspaper Article Sentences (JNAS)*, 1997.
[17] NTT Advanced Technology Corporation, *Ambient Noise Database for Telephonometry 1996*, 1996.