ICIC Express Letters
Volume **10**, Number **1**, January **2016**

ICIC International ©2016 ISSN 1881-803X
pp. 191–196

BLIND SOURCE SEPARATION WITHOUT SCALING INDETERMINACY USING AMPLITUDE RATIO OF OBSERVED SIGNALS

Takaaki Ishibashi[1], Kohei Fujimori[2], Hirohito Shintani[1]
Chiharu Okuma[3], Kiyoteru Hayama[1] and Hiromu Gotanda[4]

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

[4]Faculty of Humanity-Oriented Science and Engineering
Kinki University
11-6 Kayanomori, Iizuka, Fukuoka 820-8555, Japan

Received June 2015; accepted August 2015

Abstract. *The paper proposes a real-time blind source separation method without a scaling indeterminacy. The proposed method does not require an iteration processing. The scale of separated signals using the proposed method is decided by transfer functions. This fact means that the separated signals are expressed as the observed signals when one source is only active. The proposed methods have been verified by a simulation.*
Keywords: Blind source separation, Independent component analysis, Scaling indeterminacy, Real-time BSS

1. **Introduction.** BSS (Blind Source Separation) is a method for estimating the sound sources from observed mixture signals without using the information about the sources and the transfer functions. For BSS, ICA (Independent Component Analysis) can estimate original source signals from their mixtures, provided that the sources are statistically independent. For the instantaneous mixtures, the original sources can be completely recovered except for indeterminacy of scale and permutation [1, 2, 3, 4]. The indeterminacy of scale is that the amplitude scale of the separated signals is not equal to that of the source signals. The indeterminacy of permutation is that the order of the separated signals is not equal to that of the source signals. Furthermore, ICA algorithms are iteration methods based on a gradient method or a Newton method. This fact means that these algorithms are not good at a real-time processing.

For a real-time separation, several methods have been proposed. SS (Spectral Subtraction) and SAFIA (sound source Segregation based on estimating incident Angle of each Frequency component of Input signals Acquired by multiple microphones) can estimate the original source signals [5, 6]. In these methods, the musical-noise has been generated depending on the parameter. In order to reduce the musical-noise, a method based on the high-order statistics has been proposed [7]. However, multivariate data are necessary for the method.

Therefore, we propose a new BSS method using an amplitude ratio of observed signals. When the target source signal and the noise signals exist at the same time, the joint distribution made from the observed signals has straight lines depending on transfer functions. Based on this fact, the proposed method estimates a separating matrix using ratios of the straight lines. The proposed method does not require an iteration. Furthermore, the method can estimate separated signals without a scaling indeterminacy.

2. Blind Source Separation. Under the situation that some sound sources are observed by the microphones, BSS (Blind Source Separation) is a method for estimating the sound sources without using the information about the sources and the transfer functions. We assume that the observed mixture signals $\boldsymbol{x} = [x_1, \cdots, x_m, \cdots, x_M]^T$ by M microphones are generating a linear mixture of the sources as

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

where $\boldsymbol{s} = [s_1, \cdots, s_n, \cdots, s_N]^T$ denotes unknown source signals, N denotes the number of the sources and A denotes an unknown mixing matrix whose elements are a_{mn}. The separated signals $\boldsymbol{u} = [u_1, \cdots, u_n, \cdots, u_N]^T$, the estimated of the source signals $\boldsymbol{s}$, are expressed as

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

where W denotes a separating matrix. The matrix W is estimated by ICA algorithms [1, 2, 3, 4].

ICA can estimate the sources $\boldsymbol{s}$ except for indeterminacy of scaling and permutation under the assumption that each component of $\boldsymbol{s}$ is statistically independent. The separated matrix using ICA algorithms has scaling indeterminacy and permutation problem as

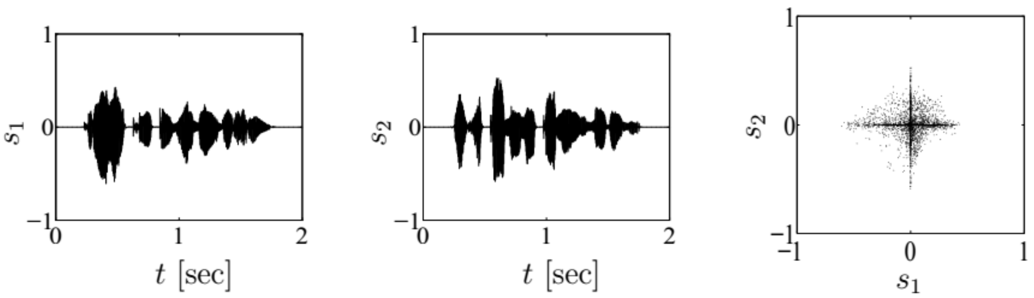
$$WA = PD \tag{3}$$

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

Furthermore, ICA needs statistically information of the sources and these algorithms are iterative methods. It means that ICA is not good at a real-time processing. For a real-time process, we propose a separation method using a distribution of the observed signals.

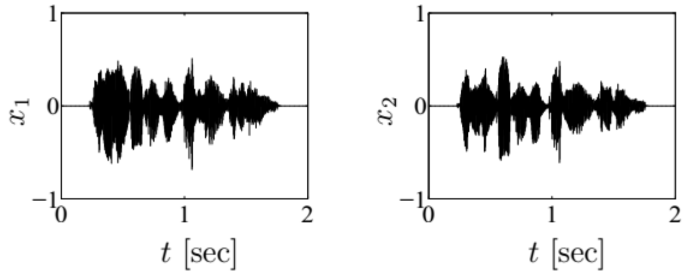
3. Real-Time Blind Source Separation Method. Consider two speakers have uttered. These waveforms are shown in Figure 1(a). Using these waveforms, their joint distribution is shown in Figure 1(b) where the horizontal and the vertical axis are denoted by amplitude of s_1 and s_2, respectively.

The mixture signals $\boldsymbol{x}$ are observed by Equation (1). The waveforms and the joint distribution of $\boldsymbol{x}$ are shown in Figures 2(a) and 2(b), respectively. From these facts, the essence of the BSS is to transform from the distribution of the mixtures as shown in Figure 2(b) to the distribution of the sources as shown in Figure 1(b) [8].

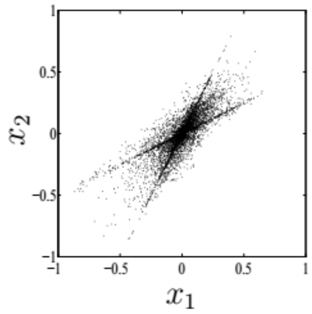


(a) Waveforms (b) Joint distribution

FIGURE 1. Waveforms of the sound source signals and their joint distribution



(a) Waveforms



(b) Joint distribution

FIGURE 2. Waveforms of the observed mixture signals and their joint distribution

Therefore, we propose a new BSS method under the condition of two-microphones. We define an amplitude ratio r of two mixture signals as $r = x_2/x_1$. In Figure 2, two dense crossing lines are recognized. Then we define r_1 calculated as the mode value (the most frequent value) of r. And r_2 is defined as the second most frequent value of r. r_1 and r_2 are estimated value of either $\frac{x_2}{x_1} = \frac{a_{21}s_1}{a_{11}s_1} = \frac{a_{21}}{a_{11}}$ or $\frac{x_2}{x_1} = \frac{a_{22}s_2}{a_{12}s_2} = \frac{a_{22}}{a_{12}}$, respectively, because the human speech has the sparseness such as the silent interval.

Using r_1 and r_2, separated signals without the scaling indeterminacy are estimated as follows (see Appendix).

$$\begin{bmatrix} v_{11} \\ v_{12} \end{bmatrix} = \frac{1}{r_2 - r_1} \begin{bmatrix} -r_1 & 1 \\ -r_1 r_2 & r_2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \tag{4}$$

$$\begin{bmatrix} v_{21} \\ v_{22} \end{bmatrix} = \frac{1}{r_2 - r_1} \begin{bmatrix} r_2 & -1 \\ r_1 r_2 & -r_1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \tag{5}$$

The separated signals are expressed as follows.

$$v_{kl} = a_{mn} s_n \tag{6}$$

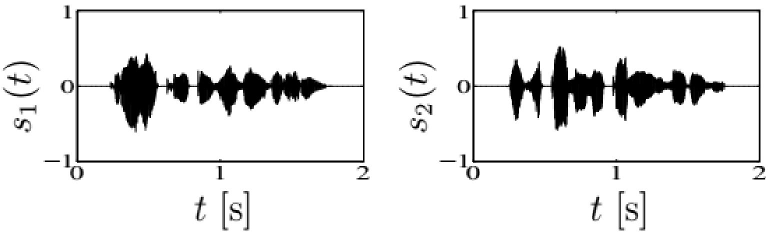
Equation (6) means that the scale of the separated signals is decided by the transfer functions.

4. **Simulation.** In order to verify our proposal, a simulation was carried out. Sources were 2 female speakers' speech signals [9] in 2 seconds. The mixed signals were sampled at a rate of 8000Hz with 16bit resolution. The mixed signals were calculated by Equation (1) in which the diagonal components have $0.9 \pm \eta$ and non-diagonal components have $0.6 \pm \eta$, and η is a random value from 0 to 0.1.

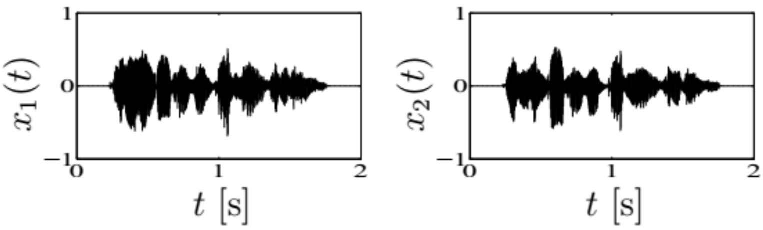
Figure 3(a) are two female speakers' uttered source signals, Figure 3(b) are mixed signals using Figure 3(a), and Figure 3(c) are estimated signals by the proposed method. From the waveforms, it is found that the estimated signals can restore the source signals.

Sources were 6 speakers' (3 females and 3 males) speech signals in 2 seconds. The simulations were carried out using 30 ($= {}_6\mathrm{P}_2$) mixing patterns. The average value of the mean square error of the proposed method was 9.31×10^{-7}. For comparison, the mean value of the mean square error was 1.79×10^{-6} when using a natural gradient method of ICA. From these results, it is found that our proposed method has the high estimation accuracy.

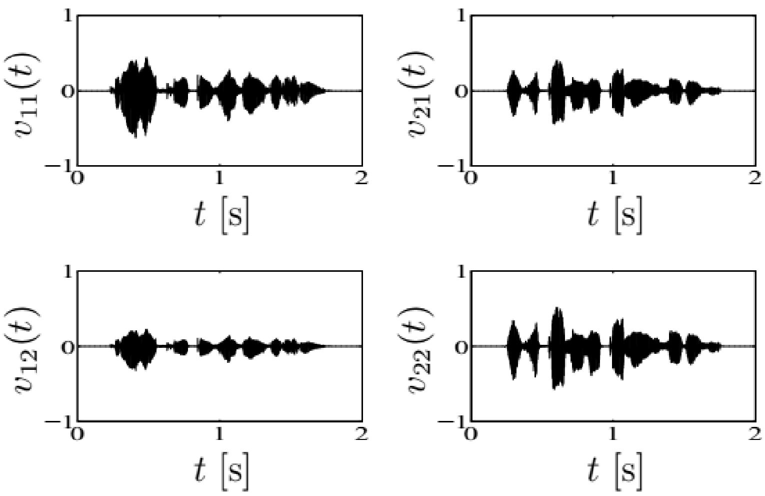
5. **Conclusions.** We propose a new blind source separation method without the scaling indeterminacy using amplitude ratios of observed mixture signals. The proposed method can estimate the separation matrix based on the amplitude ratio of the joint distribution of the observed mixture signals. By using the proposed method, we can propose a real-time BSS introduced to a short-time process.



(a) Source signals



(b) Mixture signals



(c) Estimated signals

FIGURE 3. Simulation results when two female speakers' uttered: (a) two female speakers' uttered source signals, (b) mixture signals, (c) estimated signals by the proposed method

Several directions for future research are pointed out. Separation performance for sampling frequency, frame length and frame period should be evaluated. The separation method under the condition of the multi-source must be discussed. It is necessary to propose the separation method for the convolution mixing process. In addition, we will build a noise reduction device.

REFERENCES

[1] T. W. Lee, *Independent Component Analysis, Theory and Applications*, Kluwer Academic Publishers, 1998.

[2] A. Hyvärinen, J. Karhunen and E. Oja, *Independent Component Analysis*, John Wiley & Sons, Ltd, 2001.

[3] A. Cichocki and S. Amari, *Adaptive Blind Signal and Image Processing, Learning Algorithm and Applications*, John Wiley & Sons, Ltd, 2002.

[4] 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.

[5] 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.
[6] 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.
[7] Y. Takahashi, H. Saruwatari, K. Shikano and K. Kondo, Musical-noise analysis in methods of integrating microphone array and spectral subtraction based on higher-order statistics, *EURASIP Journal on Advances in Signal Processing*, vol.2010, 2010.
[8] T. Ishibashi, Y. Tajiri, K. Inoue and H. Gotanda, An approach to blind source separation based on rotation of joint distribution of observed mixture signals, *2014 RISP International Workshop on Nonlinear Circuits, Communications and Signal Processing*, pp.21-24, 2014.
[9] Acoustical Society of Japan, *ASJ Continuous Speech Corpus – Japanese Newspaper Article SXentences (JNAS)*, 1997.

Appendix A. **Derivation of Equation (6).** First, consider the case where r_1 is estimated $\frac{a_{21}}{a_{11}}$ and r_2 is estimated $\frac{a_{22}}{a_{12}}$, the estimated signals v_{kl} are generated by Equation (4) as

$$\begin{bmatrix} v_{11} \\ v_{12} \end{bmatrix} = \frac{1}{r_2 - r_1} \begin{bmatrix} -r_1 & 1 \\ -r_1 r_2 & r_2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

$$= \frac{1}{r_2 - r_1} \begin{bmatrix} -r_1 & 1 \\ -r_1 r_2 & r_2 \end{bmatrix} \begin{bmatrix} a_{11}s_1 + a_{12}s_2 \\ a_{21}s_1 + a_{22}s_2 \end{bmatrix} \quad (7)$$

$$= \frac{1}{\frac{a_{22}}{a_{12}} - \frac{a_{21}}{a_{11}}} \begin{bmatrix} -\frac{a_{21}}{a_{11}} & 1 \\ -\frac{a_{21}}{a_{11}}\frac{a_{22}}{a_{12}} & \frac{a_{22}}{a_{12}} \end{bmatrix} \begin{bmatrix} a_{11}s_1 + a_{12}s_2 \\ a_{21}s_1 + a_{22}s_2 \end{bmatrix} \quad (8)$$

$$= \frac{a_{11}a_{12}}{a_{11}a_{22} - a_{12}a_{21}} \begin{bmatrix} -\frac{a_{21}}{a_{11}}(a_{11}s_1 + a_{12}s_2) + (a_{21}s_1 + a_{22}s_2) \\ -\frac{a_{21}}{a_{11}}\frac{a_{22}}{a_{12}}(a_{11}s_1 + a_{12}s_2) + \frac{a_{22}}{a_{12}}(a_{21}s_1 + a_{22}s_2) \end{bmatrix} \quad (9)$$

$$= \frac{a_{11}a_{12}}{a_{11}a_{22} - a_{12}a_{21}} \begin{bmatrix} -a_{21}s_1 - \frac{a_{12}a_{21}}{a_{11}}s_2 + a_{21}s_1 + a_{22}s_2 \\ -\frac{a_{21}a_{22}}{a_{12}}s_1 - \frac{a_{21}a_{22}}{a_{11}}s_2 + \frac{a_{21}a_{22}}{a_{12}}s_1 + \frac{a_{22}^2}{a_{12}}s_2 \end{bmatrix} \quad (10)$$

$$= \frac{a_{11}a_{12}}{a_{11}a_{22} - a_{12}a_{21}} \begin{bmatrix} \left(-\frac{a_{12}a_{21}}{a_{11}} + a_{22}\right)s_2 \\ \left(-\frac{a_{21}a_{22}}{a_{11}} + \frac{a_{22}^2}{a_{12}}\right)s_2 \end{bmatrix} \quad (11)$$

$$= \frac{a_{11}a_{12}}{a_{11}a_{22} - a_{12}a_{21}} \begin{bmatrix} \left(-\frac{a_{12}^2a_{21}}{a_{11}a_{12}} + \frac{a_{11}a_{12}a_{22}}{a_{11}a_{12}}\right)s_2 \\ \left(-\frac{a_{12}a_{21}a_{22}}{a_{11}a_{12}} + \frac{a_{11}a_{22}^2}{a_{11}a_{12}}\right)s_2 \end{bmatrix} \quad (12)$$

$$= \frac{a_{11}a_{12}}{a_{11}a_{22} - a_{12}a_{21}} \begin{bmatrix} \left(\frac{-a_{12}a_{21}+a_{11}a_{22}}{a_{11}a_{12}}\right)a_{12}s_2 \\ \left(\frac{-a_{12}a_{21}+a_{11}a_{22}}{a_{11}a_{12}}\right)a_{22}s_2 \end{bmatrix} \quad (13)$$

$$= \begin{bmatrix} a_{12}s_2 \\ a_{22}s_2 \end{bmatrix} \quad (14)$$

In the same way, v_{kl} are generated by Equation (5) as

$$\begin{bmatrix} v_{21} \\ v_{22} \end{bmatrix} = \frac{1}{r_2 - r_1} \begin{bmatrix} r_2 & -1 \\ r_1 r_2 & -r_1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

$$= \frac{1}{\frac{a_{22}}{a_{12}} - \frac{a_{21}}{a_{11}}} \begin{bmatrix} \frac{a_{22}}{a_{12}} & -1 \\ \frac{a_{21}}{a_{11}}\frac{a_{22}}{a_{12}} & -\frac{a_{21}}{a_{11}} \end{bmatrix} \begin{bmatrix} a_{11}s_1 + a_{12}s_2 \\ a_{21}s_1 + a_{22}s_2 \end{bmatrix} \quad (15)$$

$$= \frac{a_{11}a_{12}}{a_{11}a_{22} - a_{12}a_{21}} \begin{bmatrix} \frac{a_{11}a_{22}}{a_{12}}s_1 + a_{22}s_2 - a_{21}s_1 - a_{22}s_2 \\ \frac{a_{21}a_{22}}{a_{12}}s_1 + \frac{a_{21}a_{22}}{a_{11}}s_2 - \frac{a_{21}^2}{a_{11}}s_1 - \frac{a_{21}a_{22}}{a_{11}}s_2 \end{bmatrix} \quad (16)$$

$$= \frac{a_{11}a_{12}}{a_{11}a_{22} - a_{12}a_{21}} \begin{bmatrix} \left(\frac{a_{11}^2 a_{22}}{a_{11}a_{12}} - \frac{a_{11}a_{12}a_{21}}{a_{11}a_{12}}\right) s_1 \\ \left(\frac{a_{11}a_{21}a_{22}}{a_{11}a_{12}} - \frac{a_{12}a_{21}^2}{a_{11}a_{12}}\right) s_1 \end{bmatrix} \tag{17}$$

$$= \frac{a_{11}a_{12}}{a_{11}a_{22} - a_{12}a_{21}} \begin{bmatrix} \left(\frac{a_{11}a_{22}-a_{12}a_{21}}{a_{11}a_{12}}\right) a_{11}s_1 \\ \left(\frac{a_{11}a_{22}-a_{12}a_{21}}{a_{11}a_{12}}\right) a_{21}s_1 \end{bmatrix} \tag{18}$$

$$= \begin{bmatrix} a_{11}s_1 \\ a_{21}s_1 \end{bmatrix} \tag{19}$$

Next, consider the case where $r_1 = \frac{a_{22}}{a_{12}}$ and $r_2 = \frac{a_{21}}{a_{11}}$, the estimated signals v_{kl} are generated by Equation (4) and Equation (5), respectively, as

$$\begin{bmatrix} v_{11} \\ v_{12} \end{bmatrix} = \frac{1}{r_2 - r_1} \begin{bmatrix} -r_1 & 1 \\ -r_1 r_2 & r_2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

$$= \frac{1}{\frac{a_{21}}{a_{11}} - \frac{a_{22}}{a_{12}}} \begin{bmatrix} -\frac{a_{22}}{a_{12}} & 1 \\ -\frac{a_{21}}{a_{11}}\frac{a_{22}}{a_{12}} & \frac{a_{21}}{a_{11}} \end{bmatrix} \begin{bmatrix} a_{11}s_1 + a_{12}s_2 \\ a_{21}s_1 + a_{22}s_2 \end{bmatrix} \tag{20}$$

$$= \frac{a_{11}a_{12}}{a_{12}a_{21} - a_{11}a_{22}} \begin{bmatrix} -\frac{a_{11}a_{22}}{a_{12}} s_1 - a_{22}s_2 + a_{21}s_1 + a_{22}s_2 \\ -\frac{a_{21}a_{22}}{a_{12}} s_1 - \frac{a_{21}a_{22}}{a_{11}} s_2 + \frac{a_{21}^2}{a_{11}} s_1 + \frac{a_{21}a_{22}}{a_{11}} s_2 \end{bmatrix} \tag{21}$$

$$= \frac{a_{11}a_{12}}{a_{12}a_{21} - a_{11}a_{22}} \begin{bmatrix} \left(-\frac{a_{11}^2 a_{22}}{a_{11}a_{12}} + \frac{a_{11}a_{12}a_{21}}{a_{11}a_{12}}\right) s_1 \\ \left(-\frac{a_{11}a_{21}a_{22}}{a_{11}a_{12}} + \frac{a_{12}a_{21}^2}{a_{11}a_{12}}\right) s_1 \end{bmatrix} \tag{22}$$

$$= \frac{a_{11}a_{12}}{a_{12}a_{21} - a_{11}a_{22}} \begin{bmatrix} \left(\frac{a_{12}a_{21}-a_{11}a_{22}}{a_{11}a_{12}}\right) a_{11}s_1 \\ \left(\frac{a_{12}a_{21}-a_{11}a_{22}}{a_{11}a_{12}}\right) a_{21}s_1 \end{bmatrix} \tag{23}$$

$$= \begin{bmatrix} a_{11}s_1 \\ a_{21}s_1 \end{bmatrix} \tag{24}$$

$$\begin{bmatrix} v_{21} \\ v_{22} \end{bmatrix} = \frac{1}{r_2 - r_1} \begin{bmatrix} r_2 & -1 \\ r_1 r_2 & -r_1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

$$= \frac{1}{\frac{a_{21}}{a_{11}} - \frac{a_{22}}{a_{12}}} \begin{bmatrix} \frac{a_{21}}{a_{11}} & -1 \\ \frac{a_{21}}{a_{11}}\frac{a_{22}}{a_{12}} & -\frac{a_{22}}{a_{12}} \end{bmatrix} \begin{bmatrix} a_{11}s_1 + a_{12}s_2 \\ a_{21}s_1 + a_{22}s_2 \end{bmatrix} \tag{25}$$

$$= \frac{a_{11}a_{12}}{a_{12}a_{21} - a_{11}a_{22}} \begin{bmatrix} a_{21}s_1 + \frac{a_{12}a_{21}}{a_{11}} s_2 - a_{21}s_1 - a_{22}s_2 \\ \frac{a_{21}a_{22}}{a_{12}} s_1 + \frac{a_{21}a_{22}}{a_{11}} s_2 - \frac{a_{21}a_{22}}{a_{12}} s_1 - \frac{a_{22}^2}{a_{12}} s_2 \end{bmatrix} \tag{26}$$

$$= \frac{a_{11}a_{12}}{a_{12}a_{21} - a_{11}a_{22}} \begin{bmatrix} \left(\frac{a_{12}^2 a_{21}}{a_{11}a_{12}} - \frac{a_{11}a_{12}a_{22}}{a_{11}a_{12}}\right) s_2 \\ \left(\frac{a_{12}a_{21}a_{22}}{a_{11}a_{12}} - \frac{a_{11}a_{22}^2}{a_{11}a_{12}}\right) s_2 \end{bmatrix} \tag{27}$$

$$= \frac{a_{11}a_{12}}{a_{12}a_{21} - a_{11}a_{22}} \begin{bmatrix} \left(\frac{a_{12}a_{21}-a_{11}a_{22}}{a_{11}a_{12}}\right) a_{12}s_2 \\ \left(\frac{a_{12}a_{21}-a_{11}a_{22}}{a_{11}a_{12}}\right) a_{22}s_2 \end{bmatrix} \tag{28}$$

$$= \begin{bmatrix} a_{12}s_2 \\ a_{22}s_2 \end{bmatrix} \tag{29}$$

The estimated signals v_{kl} are derived to be expressed as a product of the source signal s_n and the transfer function a_{mn}. Therefore, v_{nm} has no indeterminacy of scaling.