14th European Signal Processing Conference (EUSIPCO 2006), Florence, Italy, September 4-8, 2006, copyright by EURASIP

A DOA ESTIMATION METHOD FOR 3D MULTIPLE SOURCE SIGNALS USING INDEPENDENT COMPONENT ANALYSIS

Hao Yuan, Makoto Yamada, and Hisashi Kawai

KDDI R&D Laboratories, Inc.
2-1-15 Ohara, Fujimino City, Saitama 356-8502, Japan
Phone: +81-49-278-7588, Fax: +81-49-278-7510
E-mail: {yuan, ma-yamada, Hisashi.Kawai}@kddilabs.jp
web: http://www.kddilabs.jp/eng/index.html

ABSTRACT

A new DOA (direction of arrival) estimation method is proposed for 3D (three-dimensional) multiple source signals using independent component analysis (ICA). The multiple source signals travel and mix in a reverberant environment and are observed at a sensor array. These observed signals are separated based on independent component analysis and the DOAs of source signals are estimated. This method can deal with signals up to the number of sensors while the conventional method based on subspace analysis, such as the well-known MUSIC algorithm, can merely be applied to those cases where the number of source signals is less than that of the sensors. A two step estimation method is also proposed to improve the estimation accuracy and the dispersion of the estimated DOAs. Experimental results reveal that the proposed method is better than the MUSIC algorithm from the perspective of small dispersion.

1. INTRODUCTIONN

DOA estimation is a very important technique both in wireless telecommunication systems and audio/speech processing systems [1, 2]. The estimated DOAs of incoming signals can be used to suppress the interference and enhance the desired signal in an adaptive array antenna system. In a TV conference system, the estimated DOAs of speech signals can be applied to a video camera and drive the camera in the direction of the speaker. Another applicable field of DOA estimation is robotics.

To estimate the DOA, a variety of approaches have been proposed in previous works such as [3, 4]. In recent years, BSS (blind source separation) based on ICA has been a significant area of focus, which reconstructs the original source signals from their mixtures without training signals. If the signals are mixed instantaneously, the separating matrix can be simply obtained from the ICA solution. In a reverberant environment, however, the mixture is convolutive and a matrix whose elements are filters is necessary to perform the separation. An alternative approach for the separation is the frequency domain BSS, where the convolutive mixture is converted into a set of instantaneous mixtures, that is, into the instantaneous mixtures in each frequency bin, and the elements of the separating matrices are simply complex values, but not filters. The cost of using the frequency domain BSS is the permutation ambiguity. If we know the directions of the source signals, we can align the signal order and reconstruct the separated source in the time domain. This is the motivation for estimating DOAs using ICA. The basic consideration of estimating DOA by ICA is using the direction information contained in the separating matrix of ICA.

To solve the permutation inconsistency problem, a DOA estimation method was proposed based on directivity patterns [5]. This method uses the separating matrix of ICA to plot directivity patterns and searches the nulls as the incident directions. Two sensors were spaced close enough to ensure no grating nulls appeared over the entire voice spectrum. It was later extended to arbitrary sensor spacing and the null searching was performed at lower frequencies when large spacing was used [6]. This method is effective in cases where there are only two signals but difficult to apply when there are 3 signals. To handle estimating DOAs for more than two signals, Sawada *et al*. proposed a method directly calculating the incident angles, using the inverse of the ICA separating matrix with a linear sensor array [7]. This method can deal with signals up to the number of sensors, and because it is unnecessary to search the nulls of directivity patterns, the calculation cost is reduced. Due to the symmetry of a linear array, the angles from a single broadside can merely be estimated. By orienting pairs of sensors in different directions, this method was extended to estimate angles from arbitrary azimuths around the array [8]. Since pairs of sensors inherit the concept of the linear array, this extension can only handle signals from the horizontal plane and no information on elevation can be obtained.

In a real environment, however, a source signal is located in a three-dimensional space and arrives at a receiver in not only an azimuth, but also in an elevation. To determine the characteristics of source signals, such as DOAs or locations, azimuth only is insufficient. If another parameter describing 3D DOA can also be obtained, signals with the same azimuth but different elevations are expected to be distinguishable. Therefore, there is significant potential for the enhancement of signal separation when the information of 3D DOA is applied to solve the permutation inconsistency problem in BSS.

In this paper, we propose a DOA estimation method with ICA for 3D signals in Section 2. After giving a definition of triangle linearity, which is used to adaptively select a better sensor combination, we propose an accuracy enhancement method in Section 3. Experimental results in Section 4 reveal that the proposed method is better than the MUSIC algorithm from the perspective of small dispersion and finally we summarize the paper in Section 5.

2. 3D DOA ESTIMATION METHOD WITH ICA

In this paper, we assume that the sensors are arbitrarily located in the X-Y plane and the number of sensors M is greater than or equal to that of signals L. The source signals, propagating through the space, are mixed and received by the sensors. The sensor outputs can be expressed as

$$x_i(t) = \sum_{l=1}^{L}\sum_{k=0}^{\infty} h_{il}(k) s_l(t-k) + n_i(t) \quad (i = 1, \cdots, M) \quad (1)$$

where s_l is the l-th source signal, h_{il} the impulse response from the l-th source to the i-th sensor and n_i the Gaussian sensor noise.

Applying the Fourier transformation to (1), we get the observed signals in the frequency domain,

$$X(f,t) = H(f)S(f,t) + N(f) \quad (2)$$

where $X(f,t) = [x_1(f,t), \cdots, x_M(f,t)]^T$ is the observed signal vector, $S(f,t) = [s_1(f,t), \cdots, s_L(f,t)]^T$ the source signal vector and $N(f)$ the noise vector in frequency f, respectively. $H(f)$ is the $M\times L$ complex mixing matrix in which the l-th column represents the transfer function from the l-th source to the sensors. The superscript ^{T} denotes transposition. The purpose of ICA in the frequency domain is to find a separating matrix $W(f)$ in each frequency bin so that the observed signals can be separated into individual signals, denoted as a vector $Y(f,t)$.

$$Y(f,t) = W(f)X(f,t) \quad (3)$$

The separated signal vector $Y(f,t)$ is an estimation of the source signal vector $S(f,t)$. If the separating matrix has converged and the noise can be ignored, the following formula holds:

$$Y(f,t) = D(f)P(f)W(f)H(f)S(f,t) \quad (4)$$

where $D(f)$ and $P(f)$ are the scaling matrix and the permutation matrix, respectively. Then the pseudo inverse (or Moore-Penrose inverse) of the separating matrix can be obtained and expressed as follows:

$$W(f)^+ = H(f)D(f)P(f) \quad (5)$$

If the number of sensors is equal to that of source signals (M=L), the pseudo inverse matrix regresses to the normal inverse. In a real environment, however, separation errors and noise exit, and the inverse of the separating matrix becomes an approximation of the mixing matrix.

Although the signals are mixed in a reverberant condition, we can approximate the elements of the mixing matrix in (2), which is the frequency response of the impulse response, as follows:

$$H_{ml}(f) = a_{ml} \exp\left\{ j\frac{2\pi f}{v} \vec{r}_m^{\,T} \cdot \vec{k}(\theta_l, \phi_l) \right\} \quad (6)$$

where a_{ml} is the gain, $\vec{r}_m = (x_m, y_m, z_m)^T$ the coordinate vector of the m-th sensor, v the velocity of the signal, and $\vec{k}(\theta_l, \phi_l) = \left(\sin\theta_l \cos\phi_l, \sin\theta_l \sin\phi_l, \cos\theta_l\right)^T$ is the look direction of the l-th signal. We select 3 sensors that are not aligned, say sensor 1, 2 and 3, to calculate the azimuth and zenith angles. The arbitrary scaling can be removed by the ratio of two elements in the same column of the mixing matrix $H(f)$. (7) and (8) provide the element ratios corresponding to sensor 1 and 2, and sensor 1 and 3, respectively.

$$\frac{W_{1l}^+}{W_{2l}^+} = \left|\frac{W_{1l}^+}{W_{2l}^+}\right| e^{j\frac{2\pi f}{v}\left[(x_1-x_2)\sin\theta_l\cos\phi_l + (y_1-y_2)\sin\theta_l\sin\phi_l\right]} \quad (7)$$

$$\frac{W_{1l}^+}{W_{3l}^+} = \left|\frac{W_{1l}^+}{W_{3l}^+}\right| e^{j\frac{2\pi f}{v}\left[(x_1-x_3)\sin\theta_l\cos\phi_l + (y_1-y_3)\sin\theta_l\sin\phi_l\right]} \quad (8)$$

Following the argument yields the phase difference between two sensors and solving the simultaneous equations on phase difference, we obtain

$$\sin\theta_l \cos\phi_l = \frac{(y_1-y_3)A_1 - (y_1-y_2)B_1}{(x_1-x_2)(y_1-y_3)-(x_1-x_3)(y_1-y_2)} \quad (9)$$

$$\sin\theta_l \sin\phi_l = \frac{(x_1-x_2)B_1 - (x_1-x_3)A_1}{(x_1-x_2)(y_1-y_3)-(x_1-x_3)(y_1-y_2)} \quad (10)$$

where

$$A_1 = arg\left(W_{1l}^+ / W_{2l}^+\right) \big/ \left(2\pi f / v\right) \quad (11)$$

$$B_1 = arg\left(W_{1l}^+ / W_{3l}^+\right) \big/ \left(2\pi f / v\right) \quad (12)$$

We define the right hand of (9) and (10) as A and B, respectively, then the azimuth and zenith angles can be obtained from the following formulas, respectively:

$$\phi_l = arg(A + jB) \quad (13)$$

$$\theta_l = \sin^{-1}\sqrt{A^2 + B^2} \quad (14)$$

where j is the imaginary unit. Since the selected three sensors are out of alignment, the denominators in (9) and (10) will not become zero. Because the sensor array locates on the X-Y plane, it cannot discriminate two signals from the upper and lower half space respectively, and we confine the estimation to the upper half space, meaning $0 \le \theta \le 90°$.

3. IMPROVING ESTIMATION ACCURACY WITH THE TWO STEP METHOD

The previous section revealed that 3 sensors that are not aligned are necessary to estimate the 3D DOAs. If the sensor number is more than 3, we have many choices to select a combination of the 3 sensors. Therefore, we may ask whether there is an optimal or at least a better combination than a fixed one.

In this section, we firstly give a definition of triangle linearity to describe the degree of approximation of a triangle to a line. Then, we give a two step estimation method to enhance the estimation accuracy.

3.1 The definition of the linearity of a triangle

From our experience, we can judge which of two triangles is more similar to a line in configuration. When the third apex goes toward the line connecting the other apexes, this triangle converts toward a line. Two triangles are plotted in Fig. 1 and we can easily state that the triangle in Fig. 1 (b) resembles a line more approximately than that in (a), even though we just simply shifted sensor 2 to the left. Here we give a definition of linearity concerning a triangle.

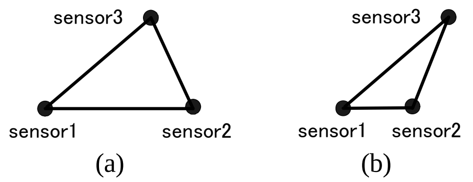

Fig. 1. Illustration of the linearity of triangles.

Definition: The linearity of a triangle is defined as the difference between the maximum interior angle and the minimum interior angle of the triangle as follows:

$$\mathbb{L} = \max_i \{\alpha_i\} - \min_i \{\alpha_i\} \qquad (15)$$

here α_i denotes the i-th interior angle. From this definition, we know that the equilateral triangle has the most non-linearity.

3.2 The two step estimation method

At first, we roughly estimate the DOAs with a fixed combination of 3 sensors. The configuration of the fixed 3 sensors should have a high non-linearity so that it can deal with signals from arbitrary directions. The sensors should be spaced close enough to ensure no spatial aliasing appears over the entire signal spectrum. With this fixed choice, we use (9)-(14) to roughly estimate DOAs as the first step.

Then, we use the DOA information obtained from the first step to select an optimal or a better sensor combination for each incoming signal. With a roughly estimated direction, we select two sensors which have the largest aperture to the direction. We then select the third sensor so that the triangle formed from the three sensors has the lowest linearity. With this kind of selected sensor combination, the phase difference between the sensors has been enlarged and the angle reservation can be expected to be improved, especially at low frequencies.

To avoid the spatial aliasing, the distances between the sensors for all roughly estimated DOAs are calculated and the maximum of them is converted to a half-wave length at a certain frequency, referred to as a threshold frequency. If a frequency bin does not exceed this threshold frequency, the DOAs are estimated again with the selected sensor combination. If the frequency bin exceeds the threshold, the DOAs to be estimated are obtained from the fixed sensor combination, or estimated from other closer sensor combinations.

4. EXPERIMENTAL RESULTS

Experiments were conducted to show the effectiveness of the proposed methods and for a comparison with the MUSIC algorithm for 3D signals. We briefly describe the MUSIC algorithm in the following subsection.

4.1 The MUSIC algorithm for 3D DOA estimation

The MUSIC algorithm proposed in [3] is a well known method for estimating the DOAs of signals. The correlation matrix of the observed signal vector

$$R_{XX}(f) = E\{X(f,t)X^H(f,t)\}$$

is eigenvalue decomposed and the M-dimensional space is divided into a signal subspace and a noise subspace according to the eigenvalues and the eigenvectors. Here E and ^{H} denote expectation and conjugate transpose. The decomposed correlation matrix can be expressed as

$$R_{XX}(f) = (E_L E_N) diag(\lambda_1, \cdots \lambda_L, \lambda_{L+1}, \cdots, \lambda_M)(E_L E_N)^H$$

where $\lambda_1 \geq, \cdots \geq \lambda_L > \lambda_{L+1} =, \cdots, = \lambda_M$ are eigenvalues corresponding to L signals and the noise, and E_L and E_N are the matrices composed of the corresponding eigenvectors. Since the signal subspace is orthogonal to the noise subspace, $a^H(\theta,\phi)E_N E_N^H a(\theta,\phi)$ will become zero when the steering vector $a(\theta,\phi)$ coincides with one of the signal directions $a(\theta_l,\phi_l)$, $l = 1, 2, \cdots, L$. Therefore, we can use the following MUSIC spectrum to search for L peaks that correspond to L signal directions.

$$P_{MU}(\theta,\phi) = \frac{a^H(\theta,\phi)a(\theta,\phi)}{a^H(\theta,\phi)E_N E_N^H a(\theta,\phi)} \qquad (16)$$

To utilize the MUSIC algorithm to estimate DOAs, the noise subspace is necessary. This means that the number of signals L must be less than that of sensors M. The MUSIC algorithm cannot handle signals in the same number as sensors.

4.2 The experimental results of the proposed method and the MUSIC algorithm

A planar array of 6 omnidirectional microphones, installed in a plastic box of L11.2W7.2H3.2 cm, was placed in the middle of a soundproof chamber shown in Fig. 2. Around the microphone array, two speakers were set up as sound sources.

Two streams of signals were played simultaneously from the speakers. The sound signals were received by the 6 microphones, and digitized at a sampling rate of 8 kHz. The speech signals were ATR phonetically-balanced sentences in Japanese [9] in 15 seconds. MIC1, 2 and 4 were chosen as the fixed combination to estimate the azimuth and zenith angles at the first step.

We adopted frequency domain ICA to separate the convolutive mixed sound signals. We can utilize any ICA algorithm such as the information maximization approach [10], but here we chose the FastICA algorithm [11] to converge the separating matrix $W(f)$ in each frequency bin, with a FFT length of 1024 points and frame shift of 512 points. Then we calculated the inverse matrix $W(f)^+$ and estimated the azimuth and zenith angles using formulas (9) to (14). Figure 3

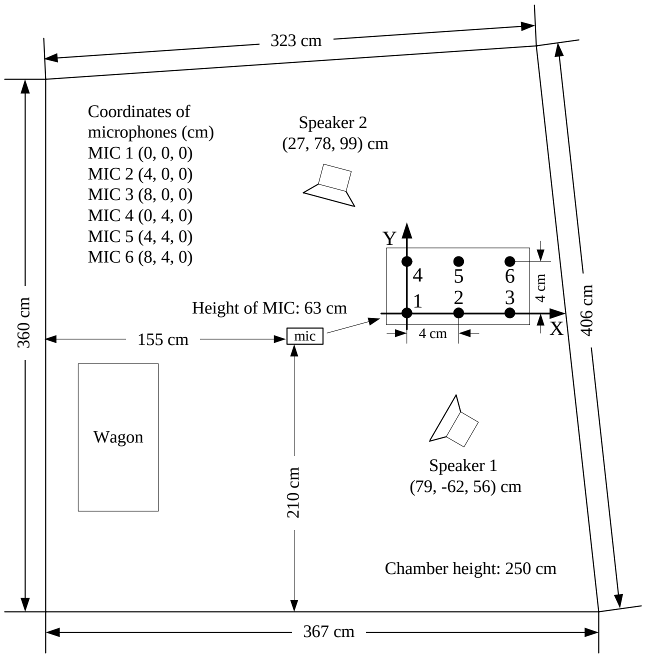

Fig. 2. Layout of soundproof chamber.

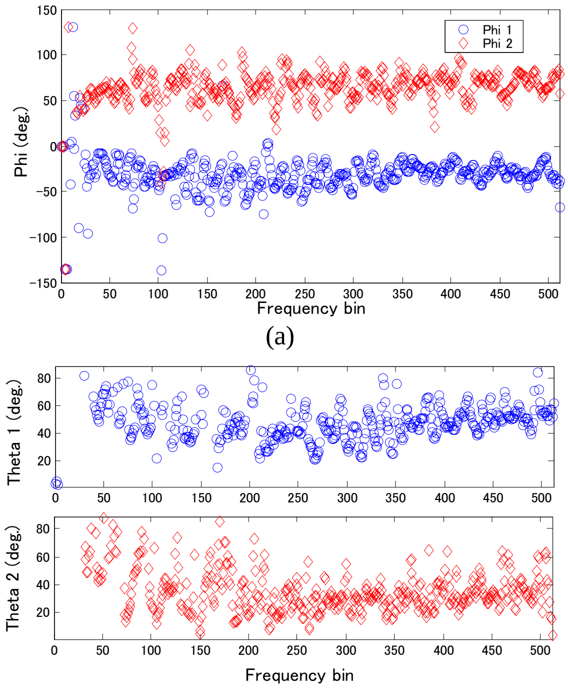

Fig. 3. The estimated azimuth angles (a) and zenith angles (b) by the proposed method.

and Fig. 4 show the estimated azimuth and zenith angles in each frequency bin for the proposed method and the MUSIC algorithm, respectively. Both the proposed method and the MUSIC algorithm can estimate the 3D DOAs, but the MUSIC algorithm obtained a large dispersion around its averaged value, especially at low frequencies. Table 1 shows the averaged values and the standard deviation of the estimated DOA for the proposed method and the MUSIC algorithm, together with the measured data for reference. It is

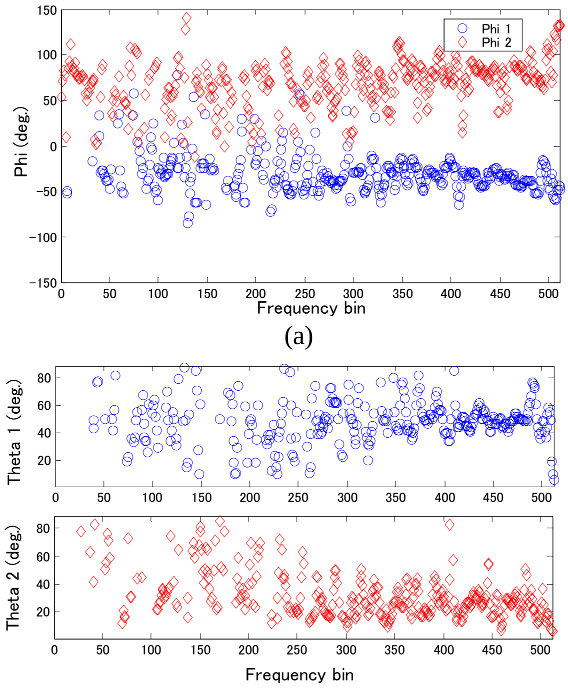

Fig. 4. The estimated azimuth angles (a) and zenith angles (b) by the MUSIC algorithm.

Table 1. The estimated azimuth and zenith angles for the proposed method and the MUSIC algorithm (degree) and their standard deviation (SD):

	ϕ_1	θ_1	ϕ_2	θ_2
Measured	-38	60	70	40
Proposal	-33	48	67	35
MUSIC	-31	39	69	32
SD (Proposal)	15	12	16	15
SD (MUSIC)	21	18	26	16

seen that the estimated angles of the proposed method were in accordance with the reference values, and the differences were within 12 degrees, while the maximum difference of the MUSIC algorithm exceeded 20 degrees. Another advantage of the proposed method is that it has a smaller standard deviation than the MUSIC algorithm does.

4.3 The two step estimation method

Another experiment was also conducted in the same chamber to deal with the three signals and show the effectiveness of the two step estimation method. The coordinates of the sensors and sources are given in Table 2. We used all observed signals from the 6 sensors to separate the mixed signals and chose MIC 1, 2 and 4 as the initial sensor combination to estimate DOAs. In this experiment, source 1 and 3 have high elevations and the estimated azimuth angles tend to go to zero at low frequencies, because of the phase differences between sensors being small for high elevation signals. To deal with this problem, we use the two step estimation method described in Section 3.2. Figure 5 shows the estimated azimuths in each frequency bin with the initial

Table 2. The coordinates of microphones and speakers. (cm):

	m1	m2	m3	m4	m5	m6	sp1	sp2	sp3
X	0	4	8	0	4	8	-28	51	-33
Y	0	0	0	4	4	4	-50	79	41
Z	0	0	0	0	0	0	104	109	105

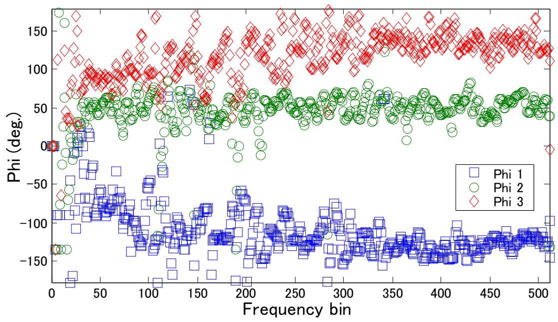

Fig. 5 The azimuth angles in the initial estimation with MIC 1, 2 and 4.

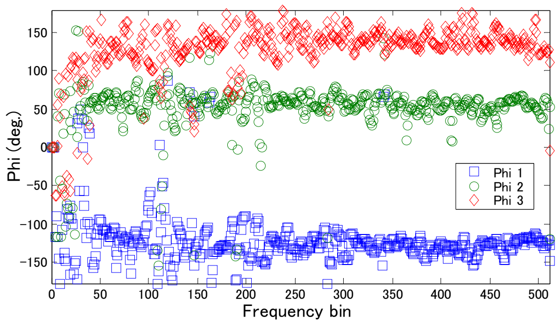

Fig. 6 The azimuth angles in the second step estimation with the adaptively selected sensors.

sensor combination and Fig. 6 shows the results of the second step estimation. Comparing Fig. 5 and Fig. 6, it is clear that Fig. 6 has a smaller deviation than Fig. 5 does, especially at low frequencies. The averaged results and the standard deviations are listed in Table 3. It is seen that the second step increases the angle reservation and reduces the standard deviations. This result may help to easily solve the permutation inconsistency problem.

5. CONCLUSIONS

We proposed a new DOA estimation method for 3D signals by making use of the separating matrix of ICA in the frequency domain. The method can estimate azimuth and zenith angles simultaneously with at least 3 sensors, which are not laid in a straight line. It is superior to the MUSIC algorithm because it can estimate source signals up to the number of sensors. Experimental results showed a smaller spread of the estimated angles than when using the MUSIC algorithm.

Table 3. The estimated azimuth and zenith angles (degree) and their standard deviation (SD) for two step estimation:

	ϕ_1	θ_1	ϕ_2	θ_2	ϕ_3	θ_3
Measured	-119	29	57	41	128	27
MIC1,2,4	-113	23	46	32	119	23
Two step	-122	24	51	30	128	22
SD(MIC1,2,4)	36	10	25	13	27	10
SD(Two step)	34	9	23	12	24	9

This paper also proposed a two step estimation method to enhance the estimation accuracy. Experimental results revealed that the improvement of angle reservation was achieved and this may help to easily solve the permutation inconsistency problem.

REFERENCES

[1] R.T. Compton, Jr., *Adaptive Antennas – Concepts and Performance*, Prentice-Hall, 1988.

[2] M. Brandstein, and D. Ward, Eds., *Microphone Arrays*, Springer, 2001.

[3] R.O. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antennas Propagat.*, vol. 34, no. 3, pp. 276-280, Mar. 1986.

[4] R. Roy and T. Kailath, "ESPRIT—Estimation of signal parameters via rotational invariance techniques," *IEEE Trans. Acoust. Speech, Signal Processing*, vol. 37, pp. 984-995, July 1989.

[5] S. Kurita, H. Saruwatari, S. Kajita, K. Takeda, and F. Itakura, "Evaluation of blind signal separation method using directivity pattern under reverberant conditions," *Proc. ICASSP 2000*, pp. 3140-3143, June 2000.

[6] M. Z. Ikram and D. R. Morgan, "A beamforming approach to permutation alignment for multichannel frequency-domain blind speech separation," *Proc. ICASSP 2002*, pp. 881-884, May 2002.

[7] H. Sawada, R. Mukai and S. Makino, "Direction of arrival estimation for multiple source signals using independent component analysis," *Proc. ISSPA 2003*, vol. 2, pp. 411-414, 2003.

[8] R. Mukai, H. Sawada, S. Araki and S. Makino, "Frequency domain blind source separation using small and large spacing sensor pairs," *Proc. ISCAS 2004*, vol. 5, pp. 1-4, 2004.

[9] Abe, M., Sagisaka, Y., Umeda, T., and Kuwabara, H. "Speech Database User's Manual," *ATR Technical Report*, TR-I-0166 (*in Japanese*), 1990.

[10] A. Bell and T. Sejnowski, "An information-maximization approach to blind separation and blind deconvolution," *Neural Computation*, vol. 7, no. 6, pp. 1129-1159, 1995.

[11] E. Bingham and A. Hyvärinen, "A fast fixed-point algorithm for independent component analysis of complex-valued signals," *International Journal of Neural Systems*, vol. 10, no. 1, pp. 1-8, Feb. 2000.