STABLE AND FAST UPDATE RULES FOR INDEPENDENT VECTOR ANALYSIS BASED ON AUXILIARY FUNCTION TECHNIQUE

Nobutaka Ono

National Institute of Informatics
2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo, 101-8430, Japan
onono@nii.ac.jp

ABSTRACT

This paper presents stable and fast update rules for independent vector analysis (IVA) based on auxiliary function technique. The algorithm consists of two alternative updates: 1) weighted covariance matrix updates and 2) demixing matrix updates, which include no tuning parameters such as step size. The monotonic decrease of the objective function at each update is guaranteed. The experimental evaluation shows that the derived update rules yield faster convergence and better results than natural gradient updates.

***Index Terms*—** independent vector analysis, blind source separation, natural gradient, auxiliary function

1. INTRODUCTION

Blind source separation (BSS) has still been one of the most interest topics in signal processing field. In recent years, multivariate-type independent component analysis (ICA), which can be referred as independent vector analysis (IVA), has been developed [1, 2, 3] and applied for frequency-domain approach for convolutive mixtures [4]. In conventional frequency-domain ICA, sources are separated at each frequency bin, individually. Hence, the permutation problem has to be solved as a post processing [5]. While, in IVA, the whole frequency components are modeled as a stochastic vector variable and simultaneously processed. Due to the model including dependencies over frequency components, IVA is theoretically not affected by the permutation ambiguity, which is its remarkable advantage.

One of the standard solutions of IVA is the natural gradient update [1, 2, 3]. However, there is a tradeoff between the convergence speed and the stability. While, recently, the author has derive new effective update rules for ICA based on auxiliary function technique [6]. In this paper, we show that the similar technique as [6] is also applicable for the objective function of IVA, and derive efficient update rules for IVA in the similar manner.

2. INDEPENDENT VECTOR ANALYSIS

2.1. BSS in Frequency Domain

Assume here that K sources are observed by K microphones and their Short-time Fourier Transform (STFT) representations are obtained. Let $s_k(\omega)$ and $x_k(\omega)$ be the kth source signal and the kth observation signal at ωth frequency bin, respectively. We here omit the index of time frame and regard the observation of each frame as a realization of stochastic process.

In frequency-domain approach for convolutive mixture [4], the linear mixing model in frequency domain:

$$\boldsymbol{x}(\omega) = A(\omega)\boldsymbol{s}(\omega) \tag{1}$$

is assumed and the sources are estimated by a linear demixing process:

$$\boldsymbol{y}(\omega) = W(\omega)\boldsymbol{x}(\omega), \tag{2}$$

where $A(\omega)$ is a mixing matrix,

$$W(\omega) = (\boldsymbol{w}_1(\omega) \ \cdots \ \boldsymbol{w}_K(\omega))^h \tag{3}$$

is a demixing matrix where ^{h} denotes Hermitian transpose, $\boldsymbol{s}(\omega)$, $\boldsymbol{x}(\omega)$, and $\boldsymbol{y}(\omega)$ are the frequency-wise vector representation of the sources, the observations, and the estimated sources, respectively, which are defined as

$$\boldsymbol{s}(\omega) = (s_1(\omega) \ \cdots \ s_K(\omega))^t, \tag{4}$$

$$\boldsymbol{x}(\omega) = (x_1(\omega) \ \cdots \ x_K(\omega))^t, \tag{5}$$

$$\boldsymbol{y}(\omega) = (y_1(\omega) \ \cdots \ y_K(\omega))^t, \tag{6}$$

where ^{t} denotes vector transpose.

2.2. Objective Function of IVA

In IVA, assuming a multivariate p.d.f. for sources to exploit the dependencies over frequency components, the demixing matrices are estimated by minimizing the following objective function.

$$J(\boldsymbol{W}) = \sum_{k=1}^{K} E[G(\boldsymbol{y}_k)] - \sum_{\omega=1}^{N_\omega} \log|\det W(\omega)| \tag{7}$$

where $\boldsymbol{W}$ denotes a set of $W(\omega)$ $(1 \le \omega \le N_\omega)$, $E[\cdot]$ the expectation operation, $\boldsymbol{y}_k$ is the source-wise vector representation defined as

$$\boldsymbol{y}_k = (y_k(1) \ \cdots \ y_k(N_\omega))^t, \tag{8}$$

and $G(\boldsymbol{y}_k)$ is called contrast function and has a relationship $G(\boldsymbol{y}_k) = -\log p(\boldsymbol{y}_k)$ where $p(\boldsymbol{y}_k)$ represents a multivariate p.d.f. of sources.

In the literature [1, 2, 3], spherical contrast functions:

$$G(\boldsymbol{y}_k) = G_R(r_k) \tag{9}$$

$$r_k = ||\boldsymbol{y}_k||_2 = \sqrt{\sum_{\omega=1}^{N_\omega} |y_k(\omega)|^2} \tag{10}$$

are often used where $||\cdot||_2$ denotes L_2 norm of a vector. This paper also focuses on this type of contrast functions.

Minimizing Eq. (7) is a nonlinear optimization problem for which there is generally no closed-form solutions. The most standard approach to solve it is iteratively applying the following update rule based on natural gradient [1, 2, 3]:

$$W(\omega) \leftarrow W(\omega) + \mu(I - E[\boldsymbol{\phi}_\omega(\boldsymbol{y})\boldsymbol{y}^h(\omega)])W(\omega), \tag{11}$$

$$\boldsymbol{\phi}_\omega(\boldsymbol{y}) = (\phi_{1\omega}(\boldsymbol{y}_1) \cdots \phi_{K\omega}(\boldsymbol{y}_K))^t, \tag{12}$$

$$\phi_{k\omega}(\boldsymbol{y}_k) = \frac{\partial G(\boldsymbol{y}_k)}{\partial y_k^*(\omega)}, \tag{13}$$

where * denotes complex conjugate and μ denotes a step size parameter. As well known, the step size determines the tradeoff between the convergence speed and the stability. A larger step size would leads faster convergence but may cause the divergence.

3. AUXILIARY FUNCTION OF CONTRAST FUNCTION

3.1. Auxiliary Function Technique

To avoid a step size tuning and derive an effective iterative update rules, we here introduce auxiliary function technique, which is an extension of EM algorithm and has been recently applied to solve various kinds of optimization problems in the signal processing field [7, 8, 6].

In order to introduce the auxiliary function technique, let us consider a general optimization problem to find a parameter vector $\boldsymbol{\theta} = \boldsymbol{\theta}^\dagger$ satisfying

$$\boldsymbol{\theta}^\dagger = \mathrm{argmin}_{\boldsymbol{\theta}}\ J(\boldsymbol{\theta}), \tag{14}$$

where $J(\boldsymbol{\theta})$ is an objective function.

In the auxiliary function technique, a function $Q(\boldsymbol{\theta}, \tilde{\boldsymbol{\theta}})$ is designed such that it satisfies

$$J(\boldsymbol{\theta}) = \min_{\tilde{\boldsymbol{\theta}}} Q(\boldsymbol{\theta}, \tilde{\boldsymbol{\theta}}). \tag{15}$$

$Q(\boldsymbol{\theta}, \tilde{\boldsymbol{\theta}})$ is called an auxiliary function for $J(\boldsymbol{\theta})$, and $\tilde{\boldsymbol{\theta}}$ are called auxiliary variables. Then, instead of directly minimizing the objective function $J(\boldsymbol{\theta})$, the auxiliary function $Q(\boldsymbol{\theta}, \tilde{\boldsymbol{\theta}})$ is minimized in terms of $\boldsymbol{\theta}$ and $\tilde{\boldsymbol{\theta}}$, alternatively, the variables being iteratively updated as

$$\tilde{\boldsymbol{\theta}}^{(i+1)} = \mathrm{argmin}_{\tilde{\boldsymbol{\theta}}}\ Q(\boldsymbol{\theta}^{(i)}, \tilde{\boldsymbol{\theta}}), \tag{16}$$

$$\boldsymbol{\theta}^{(i+1)} = \mathrm{argmin}_{\boldsymbol{\theta}}\ Q(\boldsymbol{\theta}, \tilde{\boldsymbol{\theta}}^{(i+1)}), \tag{17}$$

where i denotes the iteration index. The monotonic decrease of $J(\boldsymbol{\theta})$ under the above updates is guaranteed. When both of Eq. (16) and Eq. (17) can be written in closed forms, the auxiliary function technique gives us efficient iterative update rules. However, how to find appropriate $Q(\boldsymbol{\theta}, \tilde{\boldsymbol{\theta}})$ is problem-dependent.

3.2. Auxiliary Function of IVA Contrast Function

Focusing on the resemblance of the IVA objective function to the standard ICA objective function, let us apply the similar technique used in deriving an efficient auxiliary function in the ICA case [6] to the IVA case. We first begin with a definition.

Definition 1 *A set of real-valued functions of a vector random variable $\boldsymbol{z}$, S_G, is defined as*

$$S_G = \{G(\boldsymbol{z})|\, G(\boldsymbol{z}) = G_R(||\boldsymbol{z}||_2)\} \tag{18}$$

where $G_R(r)$ is a continuous and differentiable function of a real variable r satisfying that $G'_R(r)/r$ is continuous everywhere and it is monotonically decreasing in $r \geq 0$.

The condition of $G_R(r)$ is derived from the super-Gaussianity of the assumed source p.d.f. [6]. Note that most of the IVA contrast functions used in the literature [1, 2, 3], belong to S_G such as

$$G_1(\boldsymbol{z}) = Cr, \tag{19}$$

$$G_2(\boldsymbol{z}) = m \log \cosh(Cr), \tag{20}$$

where $r = ||\boldsymbol{z}||_2$ and m and C are positive constants.

Based on this definition of S_G, an explicit auxiliary function for the IVA objective function is obtained by the following two theorems.

Thorem 1 *For any $G(\boldsymbol{z}) = G_R(||\boldsymbol{z}||_2) \in S_G$,*

$$G(\boldsymbol{z}) \leq \frac{G'_R(r_0)}{2r_0}||\boldsymbol{z}||_2^2 + \left(G_R(r_0) - \frac{r_0 G'_R(r_0)}{2}\right) \tag{21}$$

holds for any $\boldsymbol{z}$ and r_0. The equality sign is satisfied if and only if $r_0 = ||\boldsymbol{z}||_2$.

The proof of theorem 1 can be obtained by the same manner as written in [6]. Theorem 1 indicates that the right side in Eq. (21) can be an auxiliary function for $G(z)$ with auxiliary variable r_0. Note that we use $V_k(\omega)$ as auxiliary variables instead of r_k in Theorem 2, which is a different notation from [6].

Thorem 2 *For any $G(z) = G_R(||z||_2) \in S_G$, let*

$$Q(\boldsymbol{W}, \boldsymbol{V}) = \sum_{\omega=1}^{N_\omega} Q_\omega(W(\omega), \boldsymbol{V}(\omega)), \tag{22}$$

$$Q_\omega(W(\omega), \boldsymbol{V}(\omega)) = \frac{1}{2}\sum_{k=1}^{K} \boldsymbol{w}_k^h(\omega) V_k(\omega) \boldsymbol{w}_k(\omega) - \log|\det W(\omega)| + R, \tag{23}$$

where

$$V_k(\omega) = E\left[\frac{G'_R(r_k)}{r_k}\boldsymbol{x}(\omega)\boldsymbol{x}^h(\omega)\right], \tag{24}$$

and r_k is a positive random variable, $\boldsymbol{V}(\omega)$ represents a set of $V_k(\omega)$ for any k, $\boldsymbol{V}$ represents a set of $V_k(\omega)$ for any k and ω, and R is a constant term independent of W. Then,

$$J(\boldsymbol{W}) \leq Q(\boldsymbol{W}, \boldsymbol{V}) \tag{25}$$

holds for any $\boldsymbol{W}$ and any $\boldsymbol{V}$ defined as Eq. (24). The equality sign holds if and only if

$$r_k = ||\boldsymbol{y}_k||_2 = \sqrt{\sum_{\omega=1}^{N_\omega} |\boldsymbol{w}_k^h(\omega)\boldsymbol{x}(\omega)|^2}. \tag{26}$$

Proof: Applying theorem 1 to $E[G(\boldsymbol{y}_k)]$, we have

$$\begin{aligned}
&E[G(\boldsymbol{y}_k)] \\
\leq\ & E\left[\frac{G'_R(r_k)}{2r_k} \cdot \sum_{\omega=1}^{N_\omega} |y_k(\omega)|^2\right] + R_k \\
=\ & E\left[\frac{G'_R(r_k)}{2r_k} \cdot \sum_{\omega=1}^{N_\omega} \boldsymbol{w}_k(\omega)^h \boldsymbol{x}(\omega)\boldsymbol{x}(\omega)^h \boldsymbol{w}_k(\omega)\right] + R_k \\
=\ & \sum_{\omega=1}^{N_\omega} \boldsymbol{w}_k(\omega)^h E\left[\frac{G'_R(r_k)}{2r_k} \cdot \boldsymbol{x}(\omega)\boldsymbol{x}(\omega)^h\right] \boldsymbol{w}_k(\omega) + R_k \\
=\ & \sum_{\omega=1}^{N_\omega} \boldsymbol{w}_k(\omega)^h V_k(\omega) \boldsymbol{w}_k(\omega) + R_k
\end{aligned} \tag{27}$$

where $V_k(\omega)$ is defined as Eq. (24) and R_k is a constant term independent of $\boldsymbol{w}_k(\omega)$ for any ω. The equality sign holds if and only if $r_k = ||\boldsymbol{y}_k||_2$. Summing up Eq. (27) over all k and rearraning it, we have Eq. (25). ■

4. DERIVATION OF UPDATE RULES

4.1. Derivative of Auxiliary Function

Based on the principle of the auxiliary function technique, update rules should be obtained by minimizing $Q(\boldsymbol{W}, \boldsymbol{V})$ in terms of $\boldsymbol{W}$ and $\boldsymbol{V}$, alternatively. From Theorem 2, the minimization of Q in terms of $\boldsymbol{V}$ is easily obtained by just applying Eq. (26) to Eq. (24). Then, let us focus on minimizing Q in terms of $\boldsymbol{W}$.

Since the auxiliary function defined as Eq. (22) is written by a sum of the contribution from each frequency, $\partial Q(\boldsymbol{W}, \boldsymbol{V})/\partial \boldsymbol{w}_k^*(\omega) = 0$ can be written in the same way as the ICA case [6] like

$$\frac{1}{2} V_k(\omega)\boldsymbol{w}_k(\omega) - \frac{\partial}{\partial \boldsymbol{w}_k^*(\omega)} \log|\det W(\omega)| = 0 \tag{28}$$

Rearranging Eq. (28) using a matrix formula $(\partial/\partial W)\det W = W^{-t}\det W$ yields the following simultaneous vector equations for $1 \le k \le K,\ 1 \le l \le K$.

$$\boldsymbol{w}_l^h(\omega) V_k(\omega)\boldsymbol{w}_k(\omega) = \delta_{lk} \tag{29}$$

Actually, it is exactly the same problem as Hybrid Exact-Approximate Joint Diagonalization (HEAD) problem [11] and a closed-form solution for updating all of $\boldsymbol{w}_k(\omega)$ simultaneously is still an open problem.

4.2. Sequential Update Rules

Instead of simultaneously updating all of $\boldsymbol{w}_k(\omega)$, let us consider an update of only a $\boldsymbol{w}_k(\omega)$ with keeping other $\boldsymbol{w}_l$s $(l \ne k)$ fixed. In this case, the problem to be solved can be written as follows.

$$\boldsymbol{w}_k^h(\omega) V_k(\omega)\boldsymbol{w}_k(\omega) = 1, \tag{30}$$

$$\boldsymbol{w}_l^h(\omega) V_k(\omega)\boldsymbol{w}_k(\omega) = 0 \quad (l \ne k). \tag{31}$$

Here we introduce a simpler solution than the one presented in [6]. Eq. (30) and eqs. (31) determine the scale and the direction of $w_k(\omega)$, respectively. Adding a dummy equation $\boldsymbol{a}^h V_k(\omega)\boldsymbol{w}_k(\omega) = 1$ where $\boldsymbol{a}$ is an arbitrary vector to eqs. (31), the direction of $\boldsymbol{w}_k(\omega)$ can be obtained from

$$\begin{pmatrix} \boldsymbol{w}_1^h(\omega) \\ \vdots \\ \boldsymbol{w}_{k-1}^h(\omega) \\ \boldsymbol{a}^h \\ \boldsymbol{w}_{k+1}^h(\omega) \\ \vdots \\ \boldsymbol{w}_K^h(\omega) \end{pmatrix} V_k(\omega)\boldsymbol{w}_k(\omega) = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}. \tag{32}$$

Putting $\boldsymbol{w}_k(\omega)$ obtained in the previous iteration into $\boldsymbol{a}$, the update of the direction of $\boldsymbol{w}_k(\omega)$ can be simply written by

$$\boldsymbol{w}_k(\omega) \leftarrow (W(\omega)V_k(\omega))^{-1}\boldsymbol{e}_k \tag{33}$$

where $\boldsymbol{e}_k$ denotes the unit vector with the kth element unity. Finally, the normalization should be performed to satisfy Eq. (30). They can be applied sequentially and iteratively for all of k. Consequently, the algorithm is summarized as the following alternative updates for all k, which are applied in order until convergence. We refer it AuxIVA.

Table 1: Experimental conditions

microphone spacing	2.83cm
source-microphone distance	2m
source direction	10° to 170° by 20°
reverberation time	300ms
signal length	10s
sampling frequency	16kHz
frame length	2048
frame shift	1024
window function	hamming

Auxiliary variable updates: Update the weighted covariance matrices $V_k(\omega)$ for all ω as follows.

$$r_k = \sqrt{\sum_{\omega=1}^{N_\omega} |\boldsymbol{w}_k^h(\omega)\boldsymbol{x}(\omega)|^2}, \tag{34}$$

$$V_k(\omega) = E\left[\frac{G'(r_k)}{r_k}\boldsymbol{x}(\omega)\boldsymbol{x}^h(\omega)\right]. \tag{35}$$

Note that r_k is common for all ω.

Demixing matrix updates: Apply the following updates in order for all ω.

$$\boldsymbol{w}_k(\omega) \leftarrow (W(\omega)V_k(\omega))^{-1}\boldsymbol{e}_k, \tag{36}$$

$$\boldsymbol{w}_k(\omega) \leftarrow \boldsymbol{w}_k(\omega)/\sqrt{\boldsymbol{w}_k^h(\omega)V_k(\omega)\boldsymbol{w}_k(\omega)}. \tag{37}$$

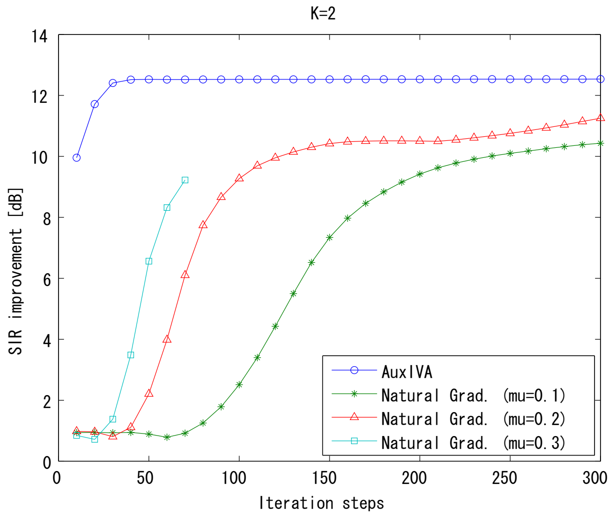
5. EXPERIMENTAL EVALUATIONS

In order to evaluate the performance of the proposed algorithm, AuxIVA, it was applied for synthesized convolutive mixtures of speech. The source signals were randomly selected from ATR Japanese speech database (Set B). While, the impulse responses recorded in a variable reverberation room (E2A) from RWCP Sound Scene Database in Real Acoustical Environments [10] were used. We prepared 20 mixtures by convoluting both of the speech source and the impulse response after downsampling to 16kHz for each of $K = 2$ or $K = 3$ (K: number of sources), where the source directions were randomly selected from 10° to 170° by 20°. Other experimental conditions are summarized in Table 1. The proposed algorithm was compared with the natural gradient update in Eq. (11) with different step sizes ($\mu = 0.1, 0.2, 0.3$). In all of the algorithms, $G(\boldsymbol{y}_k) = G_R(r_k) = r_k$ was used as a contrast function. The initial value of the demixing matrix was given by the identity matrix for simplicity. The estimated sources were calculated by applying estimated demixing matrix with Projection back [9]. The performance was evaluated by the average of SIR improvement over all sources and trials calculated by BSS toolbox [12] at every 10 iterations.

Fig. 1 shows resultant SIR improvements. AuxIVA showed much faster convergence and better results than natural gradient updates. While, the natural gradient updates suffered from the tradeoff between the convergence speed and the stability. The step size

Table 2: Averaged calculation time per one iteration [s]

	AuxIVA	Natural gradient
$K = 2$	0.15	0.10
$K = 3$	0.34	0.16



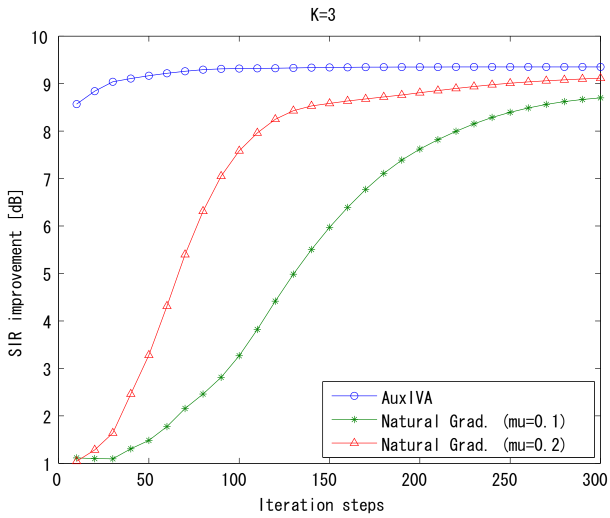


Figure 1: Averaged SIR improvements at every ten iterations by AuxIVA updates and the natural gradient updates with different step sizes for $K = 2$ case (top) and $K = 3$ case (bottom)

$\mu = 0.2$ showed faster convergence than $\mu = 0.1$. However, $\mu = 0.3$ causes the divergence in 70-80 iterations for some of 20 trials when $K = 2$, and in the first 10 iterations when $K = 3$.

The experiments were performed in Matlab ver. 7.12 (R2011a). on a laptop PC with 2.66GHz CPU. The comparison of actual computational time in this environment is shown in Table 2. Since AuxIVA includes the matrix inversion in the update, it needs more time for larger K. However, AuxIVA is still more effective than the natural gradient update. The comparison of the separation performance with other kinds of BSS techniques like [13] is one of the future work.

6. CONCLUSION

This paper presents stable and fast update rules for IVA based on auxiliary function technique. The derived update rules show faster convergence and better results than natural gradient updates in experimental evaluation. They will facilitate real-time implementation of BSS in real environment.

7. ACKNOWLEDGMENT

This research was partially supported by Grand-in-Aid for Scientific Research (A) (23240023) from the Ministry of Education, Culture, Sports, Science and Technology of Japan.

8. REFERENCES

[1] A. Hiroe, "Solution of Permutation Problem in Frequency Domain ICA Using Multivariate Probability Density Functions," *Proc. ICA*, pp. 601–608, 2006.

[2] T. Kim, T. Eltoft, and T.-W. Lee, "Independent Vector Analysis: An Extension of ICA to Multivariate Components," *Proc. ICA*, pp. 165–172, 2006.

[3] T. Kim, H. T. Attias, S.-Y. Lee, and T.-W. Lee, "Blind Source Separation Exploiting Higher-order Frequency Dependencies," *IEEE Trans. ASLP*, vol. 15, no. 1, pp. 70–79, 2007.

[4] P. Smaragdis, "Blind Separation of Convolved Mixtures in the Frequency Domain," *Neurocomputing*, vol. 22, no. 1-3, pp. 21–34, 1998.

[5] H. Sawada, R. Mukai, S. Araki, S. Makino, "A Robust and Precise Method for Solving the Permutation Problem of Frequency-Domain Blind Source Separation," *IEEE Trans. SAP*, vol.12, no.5, pp.530–538, Sept. 2004.

[6] N. Ono and S. Miyabe, "Auxiliary-function-based Independent Component Analysis for Super-Gaussian Sources," *Proc. LVA/ICA*, pp.165-172, 2010.

[7] D. D. Lee and H. S. Seung, "Algorithms for Non-Negative Matrix Factorization" *Proc. NIPS*, pp. 556–562, 2000.

[8] N. Ono and S. Sagayama, "R-Means Localization: A Simple Iterative Algorithm for Source Localization Based on Time Difference of Arrival," *Proc. ICASSP*, pp. 2718–2721, Mar. 2010.

[9] N. Murata, S. Ikeda, and A. Ziehe, "An Approach to Blind Source Separation Based on Temporal Structure of Speech Signals," *Neurocomputing*, vol. 41, no. 1, pp. 1–24, 2001.

[10] S. Nakamura, K. Hiyane, F. Asano, T. Nishiura, and T. Yamada, "Acoustical Sound Database in Real Environments for Sound Scene Understanding and Hands-Free Speech Recognition," *Proc. LREC*, pp. 965-968, 2000.

[11] A. Yeredor, "On Hybrid Exact-Approximate Joint Diagonalization," *Proc. CAMSAP*, pp. 312–315, 2009.

[12] E. Vincent, C. Fevotte, and R. Gribonval, "Performance Measurement in Blind Audio Source Separation," *IEEE Trans. ASLP*, vol. 14, no. 4, pp. 1462–1469, 2006.

[13] F. Nesta, T. S. Wada, B. H. Juang, "Coherent spectral estimation for a robust solution of the permutation problem," *Proc. WASPAA*, pp. 105–108, Oct. 2009.