BLIND SPARSE SOURCE SEPARATION FOR UNKNOWN NUMBER OF SOURCES USING GAUSSIAN MIXTURE MODEL FITTING WITH DIRICHLET PRIOR

Shoko Araki, Tomohiro Nakatani, Hiroshi Sawada, Shoji Makino

NTT Communication Science Laboratories, NTT Corporation
2-4 Hikaridai, Seika-cho, Soraku-gun, Kyoto 619-0237, Japan

ABSTRACT

In this paper, we propose a novel sparse source separation method that can be applied even if the number of sources is unknown. Recently, many sparse source separation approaches with time-frequency masks have been proposed. However, most of these approaches require information on the number of sources in advance. In our proposed method, we model the histogram of the estimated direction of arrival (DOA) with a Gaussian mixture model (GMM) with a Dirichlet prior. Then we estimate the model parameters by using the maximum a posteriori estimation based on the EM algorithm. In order to avoid one cluster being modeled by two or more Gaussians, we utilize a sparse distribution modeled by the Dirichlet distributions as the prior of the GMM mixture weight. By using this prior, without any specific model selection process, our proposed method can estimate the number of sources and time-frequency masks simultaneously. Experimental results show the performance of our proposed method.

Index Terms— Blind source separation, Dirichlet distribution, prior, number of sources, sparse

1. INTRODUCTION

Blind source separation (BSS) is an approach for estimating source signals that uses only the mixed signal information observed at each sensor. The BSS technique for speech dealt with in this paper has many applications, including hands-free teleconference systems and preprocessing for an automatic speech recognizer.

Let us formulate the task. Suppose that $N_s \geq 2$ speech sources $\mathrm{s}_1, \ldots, \mathrm{s}_{N_s}$ are convolutively mixed and observed at N_m sensors,

$$\mathrm{x}_j(t) = \sum_{i=1}^{N_s} \sum_l \mathrm{h}_{ji}(l)\, \mathrm{s}_i(t-l),\ j=1,\ldots,N_m, \qquad (1)$$

where $\mathrm{h}_{ji}(l)$ represents the impulse response from source i to sensor j. Our goal is to obtain estimates y_i of each source signal s_i from the sensor observations x_j without information about the number of sources N_s, the speech sources s_i or the mixing process h_{ji}.

Two approaches have been widely studied and employed to solve the BSS problem: one is based on independent component analysis (ICA) (e.g., [1]) and the other relies on the sparseness of source signals (e.g., [2]). In this paper we focus on the latter approach, more specifically, the time-frequency mask approach [2–4].

With the time-frequency mask approach, we assume that signals are sufficiently sparse and, therefore, that at most one source is dominant at each time-frequency slot. If these assumptions hold, a histogram of the phase differences between sensor observations (or direction of arrival (DOA) estimated from the phase differences) has N_s clusters. Because an individual cluster in the histogram corresponds to an individual source, we can separate each signal by collecting the observation signal at time-frequency points in each cluster. This is the time-frequency mask approach.

In the previous work [2–4], to automatically estimate each cluster, the number of sources N_s is assumed to be known. For example, in our previous work [4], we employed the k-means algorithm for the clustering by assuming N_s is given. However, in real situations (e.g., a meeting situation [5]) we usually cannot obtain information on the number of sources N_s in advance. Moreover, especially for an underdetermined case ($N_s > N_m$), the source number estimation is difficult, and few papers have dealt with this problem.

In this paper, we propose a novel sparse source separation method that can estimate the number of sources and time-frequency masks simultaneously. We model the histogram of the DOA with a Gaussian mixture model (GMM) with a Dirichlet prior [6], and we estimate the model parameters by using the maximum a posteriori estimation based on the EM algorithm. Here, it is expected that an individual cluster in the histogram corresponds to one Gaussian. In order to avoid one cluster being modeled by two or more Gaussians, thus making it possible to estimate the number of sources correctly, we propose utilizing a sparse distribution modeled by the Dirichlet distribution as the prior of the GMM mixture weight. The authors of [7, 8] also modeled the histogram with a GMM and derived the EM algorithm. However, they still needed to know the number of sources N_s in advance. On the other hand, our proposed algorithm in this paper does not need information on the source number, thanks to the weight prior. Moreover, our method can estimate the source number without any specific model selection process [6], which is usually computationally expensive.

The experimental results show that our proposed method can estimate the number of sources from sensor observations and can separate signals by time-frequency masks obtained by the posterior probability for each cluster.

2. MIXING AND SEPARATION PROCESSES

This paper employs a time-frequency domain approach. With an F-point short-time Fourier transform (STFT), (1) is converted into:

$$x_j(f,\tau) = \sum_{i=1}^{N_s} h_{ji}(f) s_i(f,\tau), \qquad (2)$$

where $h_{ji}(f)$ is the frequency response from source i to sensor j, $s_i(f,\tau)$ is the STFT of a source s_i. $f \in \{0,\ \frac{1}{F}f_s,\ \ldots,\ \frac{F-1}{F}f_s\}$ is a frequency (f_s is the sampling frequency) and $\tau(=0,\cdots,T-1)$ is a time-frame index.

In this paper, we assume the sparseness of the sources [2]:

$$x_j(f,\tau) \approx h_{ji}(f) s_i(f,\tau), \qquad (3)$$

where $s_i(f,\tau)$ is a dominant source at the time-frequency slot (f,τ). This is approximately true for speech signals in the time-frequency domain (see [2, 4] and their references).

2.1. Separation method

First, by assuming the source sparseness, we estimate the DOA value at each time-frequency slot by using the time difference of arrival (TDOA) between sensors [9]:

$$\begin{bmatrix} \cos\psi(f,\tau)\cos\varphi(f,\tau) \\ \sin\psi(f,\tau)\cos\varphi(f,\tau) \\ \sin\varphi(f,\tau) \end{bmatrix} = v\mathbf{D}^{+}\mathbf{q}(f,\tau). \quad (4)$$

where ψ and φ are the source azimuth and elevation, respectively, v is the sound velocity, $\mathbf{D} = [\cdots, \mathbf{d}_j - \mathbf{d}_{j'}, \cdots]^T$ is the microphone coordinate where $\mathbf{d}_j$ is the three-dimensional vector representing the location of the sensor j, $\mathbf{q}(f,\tau)$ is a TDOA vector consisting of the $q_{jj'}(f,\tau) = \frac{1}{2\pi f}\arg[x_j(f,\tau)\ x^*_{j'}(f,\tau)]$ of all microphone pairs j-j', and $^+$ denotes the Moore-Penrose pseudo-inverse [9]. In this paper, we utilize just the azimuth $\psi(f,\tau)$ information as DOA. We assume that the microphone spacing is sufficiently small to avoid the spatial aliasing problem.

Then we classify the estimated DOA values in some way. For example, the k-means clustering is utilized in [4], and a GMM fitting with the EM algorithm is employed in [7] and in this paper. Each estimated cluster corresponds to an individual source.

Finally, we estimate the separated signals $y_i(f,\tau)$ with time-frequency masks $M_i(f,\tau)$, which extract time-frequency points of members in the i-th cluster:

$$y_i(f,\tau) = x_1(f,\tau)M_i(f,\tau). \quad (5)$$

3. PROPOSED METHOD

3.1. Problem

In this subsection, we explain a typical problem that occurs when we apply a GMM fitting method to a DOA histogram. Figure 1 shows an example. Here we have four sources, and thus four clusters in the DOA histogram (Fig. 1(a)). Figure 1(c) shows the fitting result of GMM of eight Gaussians, and Fig. 1(b) plots each Gaussian. From Fig. 1(b), we can see that two Gaussians are fit to the cluster around -115 degrees. However, we expect just one Gaussian for this peak. In this paper, in order to avoid the case where one cluster is modeled by two or more Gaussians, we propose utilizing a sparse distribution for the prior of the GMM mixture weight parameter.

3.2. Probabilistic model

If we observe one source from one direction, the DOA $d(f,\tau) = \psi(f,\tau)$ at time-frequency slot (f,τ) is an observation, and the power $a(f,\tau) = |x(f,\tau)|^2$ is considered as a weight. Moreover, we assume that the DOA observation follows a Gaussian distribution, whose mean value gives us the DOA of the source. Now, let $n = \tau F + f$, where F is the number of frequency bins, and let the observations be the DOA values $d = \{d_1, d_2, \cdots, d_n, \cdots, d_N\}$ and power values $a = \{a_1, a_2, \cdots, a_n, \cdots, a_N\}$. Here, N is the number of observations, i.e., when we have T time frames and F frequency bins, $N = T \times F$.

In our observed mixture, we assume that there are a sufficient number of source signals from different directions, where some are dominant and the others are much less dominant. We also assume that observed data d_n follows a Gaussian mixture model (GMM),

$$\sum_{m=1}^{M} \alpha_m \mathcal{N}(d_n; \mu_m, \sigma_m), \quad (6)$$

where one Gaussian is assigned to one direction (= one source). We prepare a sufficient number M of Gaussians for our GMM model and estimate the mean μ_m, variance σ_m^2 and weight α_m for each Gaussian m.

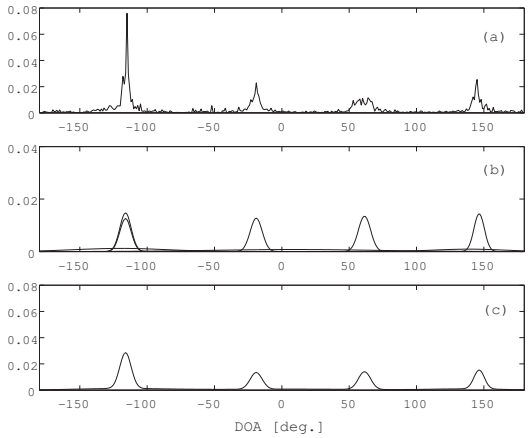

Fig. 1. An example GMM fitting result without prior ($\phi = 1.0$) when $N_s = 4$. (a) DOA histogram, (b) each Gaussian of GMM, (c) GMM plot. Two Gaussians are fit to the cluster around -115 degrees. Their mean values are -116.01 and -116.07 degrees, and their mixture weights are 0.13 and 0.15.

In this paper, in order to deal with the distribution of DOA values, we utilize a modified Gaussian distribution. When the DOA value is observed, it is folded into a range between $-\pi$ and π. However, if the mean of the given data is close to $\pm\pi$, the distribution wraps and becomes bimodal [10]. In order to handle such a distribution, we model the DOA data with a wrapped phase model [10]. More specifically, instead of (6), as the GMM we use $\sum_{m=1}^{M}\alpha_m \sum_{k=-\infty}^{\infty}\mathcal{N}(d_n + 2\pi k; \mu_m, \sigma_m)$, where $-\pi \le d_n < \pi$ and k is an integer that is uniquely determined from the observed DOA.

In order to solve the problem mentioned in Section 3.1, that is, in order to model the observed DOA data by allocating one Gaussian to each source, we assume *the sparseness of the source directions*. For this purpose, as the prior of the mixture weight, we employ the Dirichlet distribution:

$$p(\alpha) = \frac{1}{B(\phi)}\prod_{m=1}^{M}\alpha_m^{\phi-1}, \quad (7)$$

where $\alpha = \{\alpha_1, \cdots, \alpha_m, \cdots, \alpha_M\}$, $\sum_{m=1}^{M}\alpha_m = 1$, $0 \le \alpha_m \le 1$, and $B(\phi)$ is the beta distribution (regularization term). When we make the hyper parameter ϕ small ($\phi < 1$), the prior takes a larger value as the number of mixture weights whose values are close to zero increases, which is desirable for representing the sparseness of the source direction [6]. In addition, the Dirichlet distribution is known to be a conjugate prior of the mixture weight [6], and it can be incorporated into the GMM fitting approach in a computationally efficient manner.

3.3. Cost function based on GMM

Let $\theta = \{\alpha_m, \mu_m, \sigma_m, \cdots\}$ be a model parameter set. The observations are the DOA values $d = \{d_1, d_2, \cdots, d_n, \cdots, d_N\}$ and power values $a = \{a_1, a_2, \cdots, a_n, \cdots, a_N\}$ (see Sec. 3.2). In the following, Gaussian indices m and k of the wrapped phase model are assumed not to be observed, and therefore dealt with as hidden variables.

The cost function of the maximum a posteriori estimation is defined based on a log of a joint probability density function as

$$\mathcal{L}(\theta) = \log p(d,\theta) = \log p(d|\theta) + \log p(\alpha) + \text{const.} \quad (8)$$

$$= \sum_{n=1}^{N} f(a_n) \log p(d_n|\theta) + \log p(\alpha) + \text{const.} \quad (9)$$

$$= \sum_{n=1}^{N} f(a_n) \log \left(\sum_{m=1}^{M} \sum_{k=-\infty}^{\infty} p(m,k,d_n|\theta) \right) + \log p(\alpha) + \text{const.}, \quad (10)$$

where

$$p(m,k,d_n|\theta) = \frac{\alpha_m}{\sqrt{2\pi\sigma_m^2}} \exp\left(\frac{-(d_n + 2\pi k - \mu_m)^2}{2\sigma_m^2} \right), \quad (11)$$

and $-\pi \leq d_n < \pi$. We disregarded the priors of the model parameters except for α in (8), and

$$f(a_n) = c a_n / \sum_{n=1}^{N} a_n. \quad (12)$$

gives a power weight and controls the importance of the observation relative to the prior term (2nd term of (10)), where c is a control parameter.

In (10), $p(\alpha)$ follows the Dirichlet distribution (7). As mentioned in Section 3.2, for the sparse representation of the GMM, $\phi < 1$ is preferred for the Dirichlet distribution (7). Note that $\phi = 1$ is equivalent to the case without a prior for the mixture weight.

3.4. EM algorithm

Here we derive an algorithm for estimating parameter θ by the EM algorithm.

The auxiliary function Q is given as

$$Q(\theta|\theta^t) = \mathrm{E}\left[\log p(d,\theta)|d_n;\theta^t\right] \quad (13)$$

$$= \sum_n \sum_m \sum_k \left[p(m,k|d_n,\theta^t) f(a_n) \log p(m,k,d_n|\theta) \right] + \log(\alpha), \quad (14)$$

where θ^t is the estimate of the parameters after the t-th iteration, and

$$p(m,k|d_n,\theta^t) = \frac{p(m,k,d_n|\theta^t)}{\sum_m \sum_k p(m,k,d_n|\theta^t)}. \quad (15)$$

By setting $\frac{\partial Q(\theta|\theta^t)}{\partial \mu_m} = 0$ and $\frac{\partial Q(\theta|\theta^t)}{\partial \sigma_m^2} = 0$, we obtain

$$\mu_m^{t+1} = \frac{\sum_n \sum_k p(m,k|d_n,\theta^t) f(a_n)(d_n + 2\pi k)}{\sum_n \sum_k p(m,k|d_n,\theta^t) f(a_n)} \quad (16)$$

$$(\sigma_m^2)^{t+1} = \frac{\sum_n \sum_k p(m,k|d_n,\theta^t) f(a_n)(d_n + 2\pi k)^2}{\sum_n \sum_k p(m,k|d_n,\theta^t) f(a_n)} - (\mu_m^{t+1})^2. \quad (17)$$

Moreover, by using the Lagrange multiplier method, $\sum_m^M \alpha_m = 1$ and (12), the mixture weight is obtained as follows:

$$\alpha_m^{t+1} = \frac{1}{c + M(\phi - 1)} \left\{ \sum_n \sum_k p(m,k|d_n,\theta^t) f(a_n) + (\phi - 1) \right\} \quad (18)$$

Since $\alpha_m \geq 0$, $c > M(1-\phi)$ must hold from (18).

In the E-step we calculate (15), then in the M-step the parameters θ are calculated by using (16), (17) and (18). Sometimes $\alpha_m < 0$ occurs. In such a case, we can factor out the corresponding Gaussian (by setting $\alpha_m = \epsilon$, where ϵ is a very small value) and re-caluculate the parameters.

Table 1. Example estimated parameters θ for (a) 40 and (b) 100 iterations for an $N_s = 4$ case (see Sec. 4 for experimental conditions).

(a) 40 iterations:								
m	1	2	3	4	5	6	7	8
μ_m	64.2	63.8	79.8	161.2	252.9	241.1	352.5	326.4
σ_m	4.3	4.5	**46.9**	4.7	**59.8**	4.1	2.6	5.5
α_m	**0.00**	0.33	0.13	0.15	0.12	0.15	**0.00**	0.17
(b) 100 iterations:								
m	1	2	3	4	5	6	7	8
μ_m	6.8	61.9	8.1	158.3	7.5	242.4	278.8	325.9
σ_m	2.5	14.6	2.0	16.1	2.2	12.7	27.4	10.8
α_m	**0.00**	0.44	**0.00**	0.23	**0.00**	0.22	**0.00**	0.21

3.5. Source number estimation

Table 1 shows example estimated parameters θ after (a) 40 and (b) 100 iterations when the number of sources $N_s = 4$. After a sufficient number of iterations (Table 1 (b)), four mixture weights α_m become 0.00 and four α_m indicate meaningful values. That is, the source number $N_s = 4$ is estimated by thresholding the mixture weight α_m. However, such a large number of iterations requires heavy computational time. On the other hand, when we use fewer iterations (Table 1 (a)), six mixture weights α_m still have significant values. However, we can see that two of them have very large variance σ_m. Obviously, such Gaussians with large variance do not represent one cluster.

Therefore, to save computational time, we determine the number of sources N_s by counting the number of Gaussians whose parameters meet conditions $\alpha_m \geq \epsilon$ and $\sigma_m \leq th$, where ϵ is a sufficiently small threshod value and th is an appropriate threshold value ($\epsilon = 10^{-10}$ and $th = 20$ degrees are used in this paper).

3.6. Source Separation

Let x_n and y_{nm} be the observation vector at sensor 1 and the estimated m-th separated signal, respectively. The time-frequency mask for the m-th separated source is obtained by marginalizing the estimated pdf (15) with respect to k,

$$p(m|d_n,\theta) = \sum_{k=-\infty}^{\infty} p(m,k|d_n,\theta). \quad (19)$$

For implementation, we summed up $-1 \leq k \leq 1$. Now, the separated signal is obtained by

$$y_{nm} = x_n p(m|d_n,\theta), \quad (20)$$

or, with the Section 2.1 notation,

$$y_m(f,\tau) = x_1(f,\tau) p(m|\psi(f,\tau),\theta), \quad (21)$$

where $n = \tau F + f$.

4. EXPERIMENTS

4.1. Experimental setup

We performed experiments with measured impulse responses h_{ji} in a room, whose reverberation time was 130 ms (see Fig. 4 of [4]). We utilized three microphones arranged at the apexes of an equilateral triangle, 4 cm on a side. Mixtures were made by convolving the measured room impulse responses and 5-second English speech signals. The sampling rate was 8 kHz. The frame size F for STFT was 512 (64 ms), and the frame shift was 128 (16 ms).

In the EM algorithm, we utilized $M = 8$ Gaussians, whose initial values were μ_m=[25, 75, 115, 160, 205, 250, 295, 340] in de-

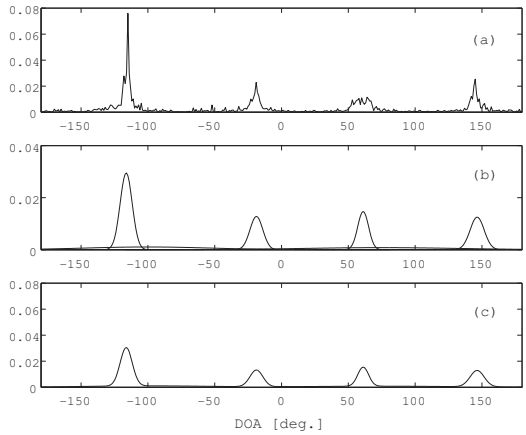

Fig. 2. GMM fitting result with prior ($\phi = 0.9$) for the same example as Fig. 1. (a) DOA histogram, (b) each Gaussian of GMM, (c) GMM plot. We have just four ($= N_s$) dominant Gaussians.

gree, $\sigma_m = 40$ in degree, and $\alpha_m = 1/M = 0.125$ for all m. For the wrapped phase model, we summed up $-1 \leq k \leq 1$. The number of iterations was 40. As the hyper parameter for (7), we utilized $\phi = 0.9$ for our proposed method and $\phi = 1.0$ for a conventional EM algorithm that corresponds to the case without any prior for the mixture weights. The control parameter c for (12) was 5.

We evaluated the signal-to-interference ratio (SIR) as a separation performance measure, and the signal-to-distortion ratio (SDR) as a sound quality measure:

$$\mathsf{InputSIR}_i = 10\log_{10}\frac{\sum_t |\mathrm{x}_{Ji}(t)|^2}{\sum_t |\sum_{k\neq i}\mathrm{x}_{Jk}(t)|^2} \quad [\mathrm{dB}], \tag{22}$$

$$\mathsf{OutputSIR}_i = 10\log_{10}\frac{\sum_t |\mathrm{y}_{ii}(t)|^2}{\sum_t |\sum_{k\neq i}\mathrm{y}_{ik}(t)|^2} \quad [\mathrm{dB}], \tag{23}$$

$$\mathsf{SDR}_i = 10\log_{10}\frac{\sum_t |\mathrm{x}_{Ji}(t)|^2}{\sum_t |\mathrm{x}_{Ji}(t) - \beta\mathrm{y}_{ii}(t-\Delta)|^2} \quad [\mathrm{dB}], \tag{24}$$

where $\mathrm{y}_{ik}(t)$ is the s_k component that appears at output $\mathrm{y}_i(t)$: $\mathrm{y}_i(t) = \sum_{k=1}^{N_s}\mathrm{y}_{ik}(t)$, $\mathrm{x}_{Ji}(t) = \sum_l \mathrm{h}_{Ji}(l)\,\mathrm{s}_i(t-l)$, and β and Δ are parameters used to compensate for the amplitude and phase difference, respectively, between x_{Ji} and y_{ii}.

We calculated SIR and SDR values for the separated sources that are counted as the sources by the method in Section 3.5. We tested 20 trials with different speech source combinations and location combinations, and then averaged the results.

4.2. Results

Figure 2 shows a GMM fitting result with prior ($\phi = 0.9$) for the same DOA distribution as that of Fig. 1. Although we had two Gaussians around -115 degrees in Fig. 1, we can see that just four ($= N_s$) Gaussians are dominant in this case. That is, using the prior ((7) with $\phi = 0.9$), more correct GMM fitting was performed.

Table 2 reports the experimental results. In the table, $\phi = 0.9$ means the results with sparse prior (7) and $\phi = 1.0$ indicates the results without a prior. The percentage values are shown where the method estimates the number of sources as $\hat{N}_s$ within 20 trials. The average separation performance results, SIR and SDR in dB, are also reported. The results with the k-means method [4], where the number of sources was given, are also shown.

Table 2. Experimental results. $\phi = 0.9$: with prior; $\phi = 1.0$: without prior, K: with the k-means. InputSIR was 0.0 [dB] ($N_s = 2$), -3.1 [dB] ($N_s = 3$), and -4.9 [dB] ($N_s = 4$).

N_s	ϕ	Accuracy of $\hat{N}_s$ estimation [%]								Performance [dB]	
		$\hat{N}_s$:1	2	3	4	5	6	7	8	OutputSIR	SDR
2	0.9		100							19.6	11.3
	1.0		0		10	25	35	20	10	11.6	4.2
	K	given								15.9	14.5
3	0.9		5	95						16.0	8.9
	1.0			35	15	30	20			14.5	7.5
	K	given								12.2	11.4
4	0.9				100					13.7	7.8
	1.0				20	75	5			12.8	6.7
	K	given								10.7	9.7

From Table 2, we can see that with the prior ($\phi = 0.9$) the number of sources is almost perfectly estimated. On the other hand, without the prior, the number of sources is overestimated, and the accuracy rate was quite low. This overestimation tendency is stronger for a small number of sources N_s.

As for the separation performance, we obtained better performance by using the prior ($\phi = 0.9$) than without the prior ($\phi = 1.0$).

5. CONCLUSION

In this paper, we proposed a sparse source separation method that can be applied even if the number of sources is unknown. We modeled the DOA histogram with a GMM. We proposed employing the Dirichlet distribution as the prior of the GMM mixture weight to avoid the case where one cluster is modeled by two or more Gaussians. Our experimental results show that the proposed method can estimate the number of sources correctly without using any specific model selection process. We also confirmed that the proposed method gives good separation performance.

6. REFERENCES

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

[2] O. Yilmaz and S. Rickard, "Blind separation of speech mixtures via time-frequency masking," *IEEE Trans. Signal Processing*, vol. 52, no. 7, pp. 1830–1847, July 2004.

[3] N. Roman and D. Wang, "Binaural sound segregation for multisource reverberant environments," in *Proc. ICASSP 2004*, May 2004, vol. II, pp. 373–376.

[4] S. Araki, H. Sawada, R. Mukai, and S. Makino, "Underdetermined blind sparse source separation for arbitrarily arranged multiple sensors," *Signal Processing*, vol. 77, no. 8, pp. 1833–1847, Aug. 2007.

[5] S. Araki, M. Fujimoto, K. Ishizuka, H. Sawada, and S. Makino, "Speaker indexing and speech enhancement in real meeting / conversations," in *Proc. ICASSP'08*, Apr. 2008, vol. I, pp. 93–96.

[6] C. M. Bishop, *Pattern recognition and machine learning*, Springer, 2008.

[7] M. Mandel, D. Ellis, and T. Jebara, "An EM algorithm for localizing multiple sound sources in reverberant environments," in *Proc. Neural Info. Proc. Sys.* , 2006.

[8] P. O'Grady and B.Pearlmutter, "Soft-LOST: EM on a mixture of oriented lines," in *Proc. ICA 2004 (LNCS 3195)*, Sept. 2004, pp. 430–436.

[9] S. Araki, H. Sawada, R. Mukai, and S. Makino, "DOA estimation for multiple sparse sources with normalized observation vector clustering," in *Proc. ICASSP'06*, May 2006, vol. 5, pp. 33–36.

[10] P. Smaragdis and P. Boufounos, "Learning source trajectories using wrapped-phase hidden Markov models," in *Proc. of WASPAA'05*, Oct. 2005, pp. 114–117.