MAXIMUM A POSTERIORI ICA: APPLYING PRIOR KNOWLEDGE TO THE SEPARATION OF ACOUSTIC SOURCES

Graham W. Taylor[*]

University of Toronto
Dept. of Computer Science
Toronto, ON M5S 2Z9 Canada

Michael L. Seltzer, Alex Acero

Speech Technology Group
Microsoft Research
Redmond, WA 98052 USA

ABSTRACT

Independent component analysis (ICA) for convolutive mixtures is often applied in the frequency domain due to the desirable decoupling into independent instantaneous mixtures per frequency bin. This approach suffers from a well-known scaling and permutation ambiguity. Existing methods perform a computation-heavy and sometimes unreliable phase of post-processing which typically makes use of knowledge regarding the geometry of the sensors post-ICA. In this paper, we propose a natural way to incorporate *a priori* knowledge of the unmixing matrix in the form of a prior distribution. This softly constrains ICA in a manner that avoids the permutation problem, and also allows us to integrate information about the environment, such as likely user configurations, into ICA using a unified statistical framework. Maximum *a priori* ICA easily follows from the maximum likelihood derivation of ICA. Its effectiveness is demonstrated through a series of experiments on convolutive mixtures of speech signals.

Index Terms**—** Unsupervised learning, Acoustic signal processing, Array signal processing

1. INTRODUCTION

Blind source separation (BSS) has been an active area of research for many years. It aims to recover original source signals using only the information from a set of mixed observation signals. One approach to BSS is independent component analysis (ICA), which assumes that the original sources are statistically independent. ICA can be derived from several standpoints, including information maximization [1], maximizing non-Gaussianity [2], and maximum likelihood [3]. Typically ICA is derived for the case of instantaneous mixing, where sources are mixed via a "mixing matrix". Realistic room environments cause reverberation which generate convolutive mixtures. Instantaneous mixing is a poor approximation to such environments.

Performing convolutive ICA in the time domain is computationally demanding. A more desirable approach is to convert the observations to the frequency domain, and perform ICA independently in each frequency bin. Unfortunately, frequency-domain ICA suffers from a well-known scaling and permutation ambiguity which must be resolved post-ICA. There is no clear solution to the permutation problem, but many approaches exploit knowledge regarding the room and sensor geometry in this stage (e.g. [4]). These approaches are computationally complex, especially when scaling to a large number of sources, and can be unreliable. One approach [5] constrains ICA to a specific "look" direction, but in our opinion such a constraint is too rigid. It may be unnecessary for frequencies in which there is little or no energy, and if there is an error in the DOA estimation, the hard constraint may limit performance. Others have used a sophisticated prior model of speech, such as a mixture of Gaussians [6].

In this paper, we propose that knowledge of the unmixing matrix or room environment can be integrated into ICA *a priori*, addressing the permutation ambiguity, and avoiding an expensive post-ICA repair phase. This approach also allows us to naturally incorporate knowledge we have regarding the sources or environment. Unlike a hard constraint, a prior encourages the solution toward a direction of interest, but the data has the final say. Maximum *a priori* (MAP) ICA can be derived in a manner that is similar to maximum likelihood ICA. The greatest challenge is in choosing the form of prior for the unmixing matrix. We propose using a beamformer-based prior model as beamforming and ICA essentially solve for the same parameters but with different objective functions. With a suitable prior, we can unify the two approaches in a principled probabilistic model.

2. MAXIMUM LIKELIHOOD ICA

We first review the maximum likelihood approach to deriving ICA. More details can be found in [2]. We assume that a source vector, $\mathbf{x} \in \Re^N$ is sampled according to a joint density $p(x_1, \ldots, x_N) = p(x_1)p(x_2), \ldots, p(x_N)$, where N represents the number of independent sources.

Furthermore, we assume the system undergoes linear mixing with no noise, $\mathbf{y} = \mathbf{H}\mathbf{x}$, and we observe only the mixture, $\mathbf{y}$. In ICA, we seek the the unmixing matrix $\mathbf{W} = \mathbf{H}^{-1}$ such that $\hat{\mathbf{x}} = \mathbf{W}\mathbf{y}$. If we assume that $\mathbf{x}$ is distributed according to $p(\mathbf{x})$, then the likelihood of the observed vector can be written as

$$p_Y(\mathbf{y}) = |\mathbf{H}^{-1}|p(\mathbf{H}^{-1}\mathbf{y}) = |\mathbf{W}|p(\mathbf{W}\mathbf{y}) \tag{1}$$

where we have applied a change of variable inside the distribution. Given a sequence of T independent observations $\mathcal{Y} = \{\mathbf{y}_1, \ldots, \mathbf{y}_T\}$, and treating $\mathbf{W}$ as a parameter, we can compute its ML estimate as:

$$\mathbf{W}_{ML} = \operatorname*{argmax}_{\mathbf{W}} p(\mathcal{Y}; \mathbf{W}) \tag{2}$$

where $p(\mathcal{Y}; \mathbf{W})$ is

$$p(\mathcal{Y}; \mathbf{W}) = \prod_t |\mathbf{W}|p(\mathbf{W}\mathbf{y}_t). \tag{3}$$

This likelihood expression can be maximized using gradient descent. The gradient of the log-likelihood can be expressed as

$$\frac{\partial \log p(\mathcal{Y}; \mathbf{W})}{\partial \mathbf{W}} = T(\mathbf{W}^{-1})^T + \sum_t g(\mathbf{W}\mathbf{y}_t)\mathbf{y}_t^T \tag{4}$$

[*]The author performed the work while at Microsoft Research.

1-4244-1484-9/08/$25.00 ©2008 IEEE

where $g(\mathbf{q}) = \frac{p'(\mathbf{q})}{p(\mathbf{q})}$. With the appropriate choice of density[1], $p(\mathbf{x})$, this result is the well-known Bell-Sejnowski InfoMax gradient-based update rule [2]. Convergence using this update rule can be improved by using the natural gradient, which is obtained by multiplying through by $\mathbf{W}^T\mathbf{W}$ to obtain

$$\Delta\mathbf{W} \propto \mathbf{W} + \frac{1}{T}\sum_t g(\hat{\mathbf{x}})\hat{\mathbf{x}}_t{}^T\mathbf{W} = \left(I + \frac{1}{T}\sum_t g(\hat{\mathbf{x}})\hat{\mathbf{x}}_t{}^T\right)\mathbf{W}.$$

2.1. Frequency domain convolutive ICA

The above derivation is valid for the simple case of linear, instantaneous mixing. In most common acoustic environments, such as rooms, the mixing process is not instantaneous, but convolutive, as the source signals reflect off of the room's surfaces and arrive at the sensors as delayed and attenuated copies. This situation is considerably more complicated as the elements of the mixing and unmixing matrices are no longer scalar elements but rather FIR filters, often of considerable length.

To avoid such complications, separation of convolutive mixtures is typically started by transforming the observed signals into the frequency domain. This converts the convolutive time-domain mixture into a series of independent instantaneous mixtures, expressed as

$$p(\mathcal{Y};\mathbf{W}) = \prod_\tau \prod_{\omega=1}^{K} |\mathbf{W}(\omega)| p(\mathbf{W}(\omega)\mathbf{y}_\tau(\omega)) \tag{5}$$

where $\mathbf{y}_\tau(\omega)$ is the vector of observed signals at frame τ and frequency ω and K is the total number of frequency components. Despite this computationally attractive decoupling, the ICA solution has permutation and scaling ambiguities: permuting the rows of $\mathbf{W}(\omega)$ or multiplying a row by a constant still produces a valid ICA solution. Because the ICA solutions at each frequency can have a different scaling and permutation, re-assembling the signal from its parts is not possible without a subsequent post-processing step.

The scaling ambiguity is easily handled by choosing a reference sensor for each estimated source, to which the estimated signals are normalized [4]. The permutation problem has been more challenging, though several methods have been proposed for its solution. Most use either spatial information such as the direction-of-arrival estimation, or spectral information such as the correlations of output samples across frequency and time. While these methods have been shown to work reasonably well, they are complex, computationally intensive, and do not scale well to a high number of sources.

This "unmix-then-repair" approach to frequency-domain ICA is somewhat strange in that the unmixing occurs in a truly blind manner using very minimal assumptions about the independence of the sources. Then some knowledge about the mixing process, such as sensor geometry, is introduced in a downstream repair stage. In the following section, we introduce a principled and natural way to incorporate such prior knowledge directly into the ICA algorithm in order to prevent the permutation problem from arising during unmixing, and as a result, eliminate the need for permutation fixing.

3. INCORPORATING PRIOR KNOWLEDGE INTO ICA

Recall from Sec. 2 that conventional ICA seeks the maximum likelihood estimate of the unmixing parameters $\mathbf{W}$ based on the observed data $\mathcal{Y}$. Alternatively, we can treat $\mathbf{W}$ as a random variable that is generated according to some prior distribution $p(\mathbf{W})$. Given this prior distribution, we can reformulate ICA as a MAP estimation problem:

$$\mathbf{W}_{MAP} = \underset{\mathbf{W}}{\text{argmax}}\ p(\mathbf{W}|\mathcal{Y}) = \underset{\mathbf{W}}{\text{argmax}}\ p(\mathcal{Y}|\mathbf{W})p(\mathbf{W}). \tag{6}$$

Substituting (3) into (6), the expression to be maximized can be rewritten as

$$p(\mathcal{Y}|\mathbf{W})p(\mathbf{W}) = \left(\prod_t |\mathbf{W}| p(\mathbf{W}\mathbf{y}_t)\right) p(\mathbf{W}). \tag{7}$$

As before, taking the log and differentiating leads to

$$\frac{\partial \log(p(\mathbf{W}|\mathcal{Y}))}{\partial \mathbf{W}} = T(\mathbf{W}^{-1})^T + \sum_t g(\mathbf{W}\mathbf{y}_t)\mathbf{y}_t^H + h(\mathbf{W}) \tag{8}$$

where $h(\mathbf{W}) = \frac{p'(\mathbf{W})}{p(\mathbf{W})}$. Dividing by T, the gradient becomes

$$\Delta\mathbf{W} \propto (\mathbf{W}^{-1})^T + \frac{1}{T}\sum_t g(\mathbf{W}\mathbf{y}_t)\mathbf{y}_t^H + \frac{1}{T}h(\mathbf{W}). \tag{9}$$

This expression shows that the role of the prior in the gradient update changes as a function of the amount of data T. As T grows larger, the prior plays a decreasingly important role. Of course, to compute $h(\mathbf{W})$, the form of $p(\mathbf{W})$ must be known. Furthermore, if the form of the prior distribution is not chosen properly, the frequency domain version of MAP ICA will suffer from the same permutation problems as conventional ICA. In the next section, we will describe one method for constructing an useful prior distribution for $\mathbf{W}$.

3.1. A prior model for frequency domain ICA

In frequency domain ICA, $p(\mathbf{W})$ really represents the prior distributions over the unmixing matrices in all frequency bins, i.e. $p(\mathbf{W}) = p(\mathbf{W}(1), \ldots, \mathbf{W}(K))$. If we assume the frequency bins are all independent as usual, this can be factorized simply as the product of a series of independent distributions $p(\mathbf{W}(\omega))$. However, we would like the prior distribution to somehow connect the unmixing matrices across all frequencies in such a manner so as to prevent the occurrence of the permutation problem. To do so, we will assume that the prior model for $\mathbf{W}$ is generated by a hidden variable, $\Theta = \{\theta_1, \ldots, \theta_N\}$, a vector that represents the direction of arrival (DOA) of each of the N sources. Since each θ is a continuous variable, we quantize θ into directional "bins" to avoid integration. Under this model, the prior distribution can be expressed as

$$p(\mathbf{W}) = \sum_\Theta p(\mathbf{W}|\Theta)p(\Theta) = \sum_\Theta \prod_\omega p(\mathbf{W}(\omega)|\Theta)p(\Theta). \tag{10}$$

Thus, according to (10), the distributions of the different unmixing matrices are not independent, but rather are *conditionally independent*, given the hidden variable Θ. This is the key to preventing the permutation problem: Θ ties $\mathbf{W}(\omega)$ across frequencies.

To compute the MAP estimate of $\mathbf{W}$ under this model, we can rewrite the posterior distribution in (6) as

$$p(\mathbf{W}|\mathcal{Y}) = \sum_\Theta p(\mathbf{W},\Theta|\mathcal{Y}) \propto p(\mathcal{Y}|\mathbf{W}) \sum_\Theta p(\mathbf{W},\Theta) \tag{11}$$

where we have applied Bayes' rule, dropped the normalizing term that is independent of $\mathbf{W}$, and pulled $p(\mathcal{Y}|\mathbf{W})$ out of the summation. The gradient of the logarithm of (11) can be expressed as

$$\frac{\partial}{\partial\mathbf{W}}\log p(\mathbf{W}|\mathcal{Y}) = \frac{\partial}{\partial\mathbf{W}}\log p(\mathcal{Y}|\mathbf{W}) + \frac{\partial}{\partial\mathbf{W}}\log\sum_\Theta p(\mathbf{W},\Theta). \tag{12}$$

[1]We assume here that we can use one of the standard formulations of $p(\mathbf{x})$ used in ICA. In this work, we use the Laplacian distribution described in [4].

The first term on the right side of (12) is identical to the maximum likelihood gradient in (4). Noting that $x = \exp(\log(x))$ and applying the chain rule, the second term can be written as

$$\begin{aligned}
&\frac{\partial}{\partial \mathbf{W}} \log \sum_{\Theta} p(\mathbf{W}, \Theta) \\
&= \frac{1}{\sum_{\Theta} p(\mathbf{W}, \Theta)} \sum_{\Theta} \exp(\log(p(\mathbf{W}, \Theta))) \frac{\partial}{\partial \mathbf{W}} \log p(\mathbf{W}, \Theta) \\
&= \frac{1}{p(\mathbf{W})} \sum_{\Theta} p(\mathbf{W}, \Theta) \frac{\partial}{\partial \mathbf{W}} \log \left(\prod_{\omega} p(\mathbf{W}(\omega), \Theta) \right) \\
&= \sum_{\Theta} p(\Theta|\mathbf{W}) \sum_{\omega} \frac{\partial}{\partial \mathbf{W}(\omega)} \log(p(\mathbf{W}(\omega)|\Theta)) . \qquad (13)
\end{aligned}$$

The posterior probability $p(\Theta|\mathbf{W})$ is computed from the prior distribution of $\mathbf{W}$ over all frequency bins using Bayes rule:

$$p(\Theta|\mathbf{W}) = \frac{\prod_{\omega} p(\mathbf{W}(\omega)|\Theta) p(\Theta)}{\sum_{\Theta'} \prod_{\omega} p(\mathbf{W}(\omega)|\Theta) p(\Theta)}. \qquad (14)$$

In summary, MAP ICA is a gradient descent algorithm with an update rule involving two terms: one that is the same as the conventional ICA update rule, and another that depends on the prior model chosen for $\mathbf{W}$. The posterior term, $p(\Theta|\mathbf{W})$, depends on $\mathbf{W}$, which is updated with each iteration of ICA. So each iteration of our algorithm consists of two steps, much like the EM algorithm:

Step 1: Update the posterior, $p(\Theta|\mathbf{W})$, given $\mathbf{W}$

. **Step 2:** Update $\mathbf{W}$, given $p(\Theta|\mathbf{W})$ (by gradient descent).

Of course, as (13) indicates, the exact gradient expression depends on the exact form of $p(\mathbf{W}(\omega)|\Theta)$. This will be discussed in the following section.

3.2. A beamformer-based prior model

ICA and traditional beamforming both optimize the same set of parameters in order to achieve a similar high level goal: enhancement of the target signal and attenuation of interference, typically by spatial filtering. However, their objective functions are quite different as are the assumptions made in each class of algorithms. Nevertheless, researchers have studied the equivalence between solutions of the two classes of algorithms under certain conditions [7]. That is, in certain cases, the rows of the learned unmixing matrix in ICA are much like a series of beamformers or interference cancellers, and vice versa.

Based on this observation, we believe a beamformer, which is analagous to a row of $\mathbf{W}$, is an attractive source of a prior model $p(\mathbf{W})$. For example, a superdirective or delay-and-sum beamformer for the prior on $\mathbf{W}$ is advantageous because it can be computed in closed form and is dependent only on the direction of arrival of the target source. As a result, the N rows of $\mathbf{W}$, corresponding to each of the sources, will be independent of each other. Because our hidden variable Θ operates on quantized DOA regions, we require that the model allows for uncertainty in DOA, while maintaining a computationally tractable update rule. Therefore the unmixing matrix in each frequency bin is modeled as the joint probability of N independent multivariate Gaussians:

$$p(\mathbf{W}(\omega)|\Theta) = \prod_{i=1}^{N} p(\mathbf{w}_i(\omega)|\theta_i) = \prod_{i=1}^{N} \mathcal{N}(\mathbf{w}_i(\omega); \boldsymbol{\mu}_{\theta_i}(\omega), \boldsymbol{\Sigma}_{\theta_i}(\omega)) \quad (15)$$

Recall that the MAP ICA update rule involves a term from the ML update (4) and a term involving the prior (13). When $P(\mathbf{W}(\omega)|\Theta)$ is Gaussian as in (15), the gradient with respect to $\mathbf{w}_i(\omega)$ is simply

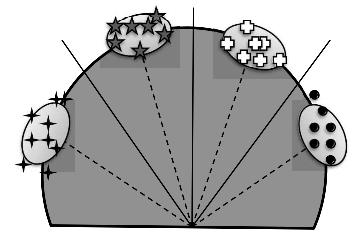


Fig. 1. For each directional bin, several source locations are sampled. Beamformers are directly computed from source locations, and used to estimate the mean and covariance for $p(\mathbf{w}_i(\omega)|\theta_i)$.

$$\sum_{\Theta} p(\Theta|\mathbf{W}) \left(-\boldsymbol{\Sigma}_{\theta_i}(\omega)^{-1} (\mathbf{w}_i(\omega) - \boldsymbol{\mu}_{\theta_i}(\omega)) \right) \qquad (16)$$

where $\boldsymbol{\Sigma}_{\theta_i}(\omega)$ and $\boldsymbol{\mu}_{\theta_i}(\omega)$ are estimated in an offline training phase from an ensemble of beamformers. This process is described in more detail in Section 4.

The last component to $p(\mathbf{W})$ is the choice of the prior distribution over the hidden DOA variable Θ. Any knowledge about the likelihood of particular source configurations can be reflected in the setting of $p(\Theta)$. For example, if nothing is known about the source locations, then it can be uniform, making all configurations of the N sources equally likely. On the other hand, an appropriate choice of $p(\Theta)$ can reflect *a priori* knowledge of the source locations.

Regardless of source, each θ_i will be discretized into B bins, and all N sources use the same model of $p(\mathbf{w}_i(\omega)|\theta_i)$. Thus, the prior model for $\mathbf{W}$ consists of $B \times K$ Gaussians, one for each frequency and DOA bin. Since a superdirective beamformer can be computed in closed form given a source direction [8], the parameters of $p(\mathbf{w}_i(\omega)|\theta_i)$ can be fit by sampling several source DOA from a bin, computing the respective vectors representing the beamformers (at each frequency), and then taking the sample mean and covariance of the beamformer vector. This process is shown in Figure 1.

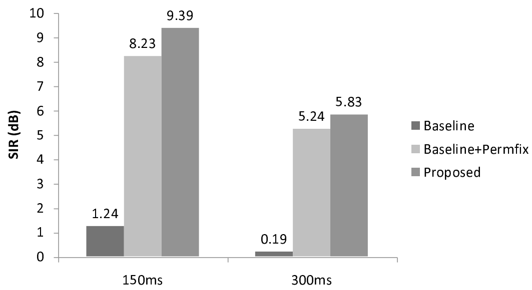
4. EXPERIMENTS

We carried out a series of experiments to separate convolutive mixtures of speech signals captured by a microphone array in a reverberant environment. We restrict ourselves to the case where the number of sources N equals the number of sensors, M, mainly for convenience. This allows us to more accurately compare to the baseline frequency-domain ICA algorithm that was described in Sec. 2.1.

Our experiments are carried out for $N{=}2$ and $N{=}3$ scenarios, using speech utterances convolved with impulse responses created via the image method [9] simulating a 5.73 m$\times$3.12 m$\times$2.70 m room with reverberation times of 150ms and 300ms. The experimental setup for the $N{=}2$ case matches that described in [7], with sources 1.15 m from the array, one at -30° and one at 40°. For the $N{=}3$ experiments, a third microphone was added to the array as well as a third source at -5°. All experiments used a microphone spacing of 4 cm, a sampling rate of 8 kHz and a frame size of 2048 samples.

4.1. Training the prior distribution

The components needed for the prior distribution are the conditional distribution $p(\mathbf{W}|\Theta)$ and discrete prior distribution $p(\Theta)$. To create $p(\mathbf{W}|\Theta)$, we first quantize the array's working area and then train the Gaussian distributions for the beamformers in each directional bin. In these experiments, we used 9 sectors, each 20^o wide,



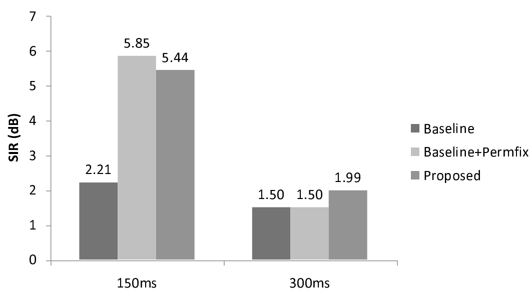


Fig. 2. SIR results for the traditional ICA followed by permutation fixing and the proposed MAP ICA algorithm. Top: 2 sources, 2 microphones. Bottom: 3 sources, 3 microphones.

spanning -90^o to $+90^o$ where 0^o is directly in front of the array. For each directional bin i, 2000 source location samples were drawn from a Gaussian centered at a point between the boundaries of that bin, at a radius of 1.15m. For every location, a superdirective beamformer was computed and this ensemble of beamformers was used to estimate the mean and covariance of $p(\mathbf{w}_i(\omega)|\theta_i)$. This was repeated for every directional bin and every frequency to create $p(\mathbf{W}|\Theta)$.

The prior distribution over source configurations, $p(\Theta)$, was constructed by hand to favor configurations where users were between -70^o to -10^o or $+10^o$ to $+50^o$ (additionally -30^o to $+10^o$ for the $N = 3$ case), reflecting knowledge of approximate user location. Remaining configurations were considered unlikely; those with multiple sources in the same sector were considered extremely unlikely.

4.2. Source separation evaluation

To evaluate the proposed MAP ICA approach, we compared the separation performance of the baseline frequency domain ICA algorithm described in Section 2.1, both with and without a post-ICA permutation fixing stage, to the proposed MAP ICA algorithm. In the baseline method (with or without permutation fixing), the unmixing matrices were initialized to the identity matrix for each frequency bin. In our proposed MAP ICA approach, the unmixing matrices were initialized to the mean of the prior $p(\mathbf{W})$.

In all experiments, ICA is performed for 2000 iterations at a conservative learning rate (1×10^{-5}). After ICA is completed, we perform scaling alignment and then perform an integrated permutation alignment which involves source localization and inter-frequency correlation of separated signals. This approach is described in [4]. No permutation fixing is applied to the MAP ICA algorithm.

The *signal-to-interference ratio* (SIR) of the ICA output of the three methods is shown in Figure 2. For each mixture, the SIR scores were computed for each source in the mixture and averaged across all sources. In all cases, permutation fixing matches or improves the baseline SIR scores as expected. We see that typically, incorporating the prior allows us to achieve results comparable to or slightly better than to the baseline algorithm with permutation fixing. However, it is interesting to note that while the permutation fixing approach uses both spatial information (DOA) and spectral information (envelope correlation), our MAP ICA model uses a prior that only captures spatial information. Thus, it is likely that incorporating an appropriate prior of the source distributions, such as the GMM used in [6] would capture the spectral correlation and result in further improvements. Nevertheless, the fact that the proposed MAP ICA algorithm achieves performance similar to permutation fixing, while using only half of the knowledge, demonstrates that it is a promising approach.

5. CONCLUSIONS

We have proposed a framework for maximum *a posteriori* ICA, and experimented with a beamformer-based prior model of the unmixing matrix. Incorporating a prior model of $\mathbf{W}$ can allow us to avoid the well-known permutation problem in convolutive, frequency-domain ICA, while achieving or outperforming the traditional approach of unmixing following by permutation fixing.

In future work, we hope to extend the beamformer prior to incorporate knowledge of both the source direction and the directions of the competing sources. In addition, we hope to experiment with other prior models for the unmixing matrix, potentially data-driven, which could be modeled by more representationally powerful distributions than a mixture of Gaussians.

6. REFERENCES

[1] A. Cichocki and S.-I. Amari, *Adaptive Blind Signal and Image Processing*, Wiley, 2002.

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

[3] B. Pearlmutter and L. C. Parra, "Maximum likelihood blind source separation: A context-sensitive generalization of ICA," in *Proc. NIPS 9*, 1997, pp. 613–619.

[4] H. Sawada, R. Mukai, S. Araki, and S. Makino, "Frequency-domain blind source separation," in *Speech Enhancement*, J. Benesty, S. Makino, and J. Chen, Eds., pp. 299–327. Springer, 2005.

[5] L. Parra and C. Alvino, "Geometric source separation: merging convolutive source separation with geometric beamforming," in *Proc. IEEE Signal Proc. Soc. Workshop*, 2001, pp. 273 –282.

[6] Hagai Attias, "Source separation with a sensor array using graphical models and subband filtering," in *Proc. NIPS 15*, pp. 1205–1212. MIT Press, Cambridge, MA, 2003.

[7] S. Araki, R. Mukai, S. Makino, T. Nishikawa, and H. Saruwatari, "The fundamental limitation of frequency domain blind source separation for convolutive mixtures of speech," *IEEE Trans. SAP*, vol. 11, no. 2, pp. 109–116, 2003.

[8] J. Bitzer and K. Simmer, "Superdirective microphone arrays," in *Microphone Arrays*, M. Brandstein and D. Ward, Eds. Springer, 2001.

[9] J.B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *J. Acoust. Soc. Amer.*, vol. 65, no. 4, pp. 943–950, 1979.