MAXIMUM LIKELIHOOD APPROACH FOR BLIND AUDIO SOURCE SEPARATION USING TIME-FREQUENCY GAUSSIAN SOURCE MODELS

*Cédric Févotte**

Signal Processing Lab
Cambridge University Engineering Dept
Cambridge CB2 1PZ, UK
cf269@cam.ac.uk

Jean-François Cardoso

Lab. de Trait. et Comm. de l'Information
CNRS/LTCI & ENST/TSI
46, rue Barrault, 75634 Paris Cedex 13, France
cardoso@tsi.enst.fr

ABSTRACT

In this paper we propose a simple time-frequency Gaussian model of audio signals that allows for separation of possibly underdetermined and noisy linear instantaneous mixtures. An efficient EM algorithm is proposed to estimate the mixing matrix, the noise covariance and covariances of the source t-f coefficients over a chosen frame/subband tiling of the time-frequency domain. Results are given on 4×4 and 3×4 noisy mixtures of audio sources.

1. INTRODUCTION

Blind Source Separation (BSS) consists in estimating n signals (the sources) from the sole observation of m mixtures of them (the observations). In this paper we consider linear instantaneous mixtures of time series: at each time index, the observations are a linear combination of the sources at the same time index.

Determined ($m = n$) noise-free linear instantaneous mixtures have been widely studied, within the field of Independent Component Analysis, assuming independent and identically distributed (i.i.d) sources and using higher order statistics (see [1, 2], for a survey), or using correlation (*e.g*, [3], [4]), non-stationarity (*e.g*, [5], [6]), or both (*e.g*, [7],[8]), leading to second order statistics based methods. These methods have proved to perform well for noise-free determined mixtures. They might still be able to estimate the mixing matrix reasonably well in noisy conditions provided the noise level is not too unfavorable, but even in this case, they usually do not provide *explicit* denoising of the sources estimates. In this paper, we focus on a more difficult case: noisy, possibly underdetermined mixtures. The underdetermined case in particular is very challenging because contrary to (over)determined mixtures, estimating the mixing system is not sufficient for reconstruction of the sources, since for $m < n$ the mixing matrix is not invertible. In this context, prior information about the sources plays a key role for their reconstruction. Similarly, prior information is also required for optimal source reconstruction from noisy mixtures.

In this paper we propose a simple time-frequency ($t-f$) model of the sources: in the $t - f$ plane, the coefficients of the source signals are modeled as Gaussian random variables with constant variances over time frames and frequency subbands. As usual, the sources are furthermore assumed mutually independent. This frame/subband Gaussian model allows for easy maximum likelihood estimation of all parameters of interest using the Expectation-Maximization EM algorithm. The sources are reconstructed by Wiener filtering of the observations in each frame/subband block.

The paper is organized as follows: Section 2 introduces notations and the source models. Our maximum likelihood approach is presented in Section 3 and results over 4×4 and 3×4 noise-free and noisy mixtures of audio sources are presented in Section 4.

2. MODEL

We consider the following standard linear instantaneous model, $\forall t = 1, \ldots, N$:

$$\mathbf{x}(t) = \mathbf{A}\,\mathbf{s}(t) + \mathbf{n}(t) \tag{1}$$

where $\mathbf{x}(t) = [x_1(t), \ldots, x_m(t)]^T$ is a vector of size m containing the observations, $\mathbf{s}(t) = [s_1(t), \ldots, s_n(t)]^T$ is a vector of size n containing the sources and $\mathbf{n}(t) = [n_1(t), \ldots, n_m(t)]^T$ is a vector of size m containing additive noise. Variables without time index t denote whole sequences of samples, *e.g*, $\mathbf{x} = [\mathbf{x}(1), \ldots, \mathbf{x}(N)]$ and $x_1 = [x_1(1), \ldots, x_1(N)]$.

The aim of the following work is to estimate the sources $\mathbf{s}$ and the mixing matrix $\mathbf{A}$ up to the standard BSS indeterminacies on gain and order, that is, compute $\hat{\mathbf{s}}$ and $\hat{\mathbf{A}}$ such that

$$\hat{\mathbf{A}} = \mathbf{A}\,\mathbf{D}\,\mathbf{P} \tag{2}$$
$$\hat{\mathbf{s}} = \mathbf{P}^T\,\mathbf{D}^{-1}\,\mathbf{s} \tag{3}$$

where $\mathbf{D}$ is a diagonal matrix and $\mathbf{P}$ is a permutation matrix.

2.1. Time domain / Transform domain

We propose to solve the problem defined by Eq. (1) using a time-frequency model of the sources. Let x be a real valued sequence of length N. Let

$$\mathbb{R}^{1\times N} \rightarrow \mathbb{R}^{l_{\mathrm{frame}}\times n_{\mathrm{frame}}} \tag{4}$$
$$x \rightarrow \tilde{x} \tag{5}$$

denote a bijective linear time-frequency transform, preferably orthonormal, with time resolution $l_{\mathrm{frame}}.\frac{1}{f_s}$ and frequency resolution $\frac{1}{l_{\mathrm{frame}}}.\frac{f_s}{2}$, where f_s is the sampling frequency and $n_{\mathrm{frame}} = \frac{N}{l_{\mathrm{frame}}}$. Such transforms includes the Modulated Lapped Transforms family to which the Modified Discrete Cosine Transform (MDCT) belongs (see [9] for a survey). In the $t - f$ domain, by linearity of the $t - f$ transform, model (1) reads:

$$\tilde{\mathbf{x}}(q, k) = \mathbf{A}\,\tilde{\mathbf{s}}(q, k) + \tilde{\mathbf{n}}(q, k) \tag{6}$$

*Cédric Févotte acknowledges support of the European Commission funded Research Training Network HASSIP (HPRN-CT-2002-00285)

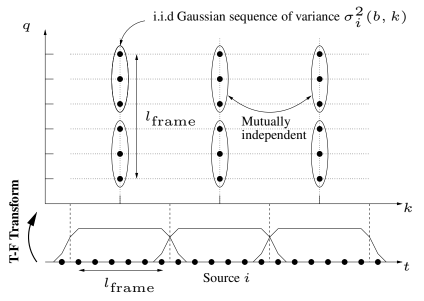

Figure 1: Time-frequency model of the sources

with $\tilde{\mathbf{x}}(q,k) = [\tilde{x}_1(q,k), \ldots, \tilde{x}_m(q,k)]^T$, for all $(q,k) \in [\![1, l_{\text{frame}}]\!] \times [\![1, n_{\text{frame}}]\!]$ and similarly for $\tilde{\mathbf{n}}(q,k)$ and $\tilde{\mathbf{s}}(q,k)$.

2.2. Source model

Let $[B^{(1)}, \ldots, B^{(n_{\text{block}})}]$ be an arbitrary partition of the frequency axis $[\![1, l_{\text{frame}}]\!]$, where each block $B^{(b)}$ corresponds to a frequency subband.

The statistical model for the source signals is (see Fig. 1):

Model for each source:

1) Each subset of $t-f$ coefficients $\{\tilde{s}_i(q,k), q \in B^{(b)}\}$ of a source i in a frequency block $B^{(b)}$ of a frame k is modelled by a iid Gaussian sequence of variance $\sigma_i^2(b,k)$,

2) The subsets of $t-f$ coefficients of a source i over all the frequency blocks and all the frames are mutually independent.

Model of independent components:

3) The sets of $t-f$ coefficients of the sources are mutually independent.

By assumptions 1) and 3), the covariance matrix of the source coefficients is diagonal: $\forall(b,k) \in [\![1, n_{\text{block}}]\!] \times [\![1, n_{\text{frame}}]\!]$,

$$\mathbf{S}_{bk} \overset{\text{def}}{=} \text{Cov}(\tilde{\mathbf{s}}(q,k)) = \text{diag}\left(\sigma_1^2(b,k), \ldots, \sigma_n^2(b,k)\right) \tag{7}$$

2.3. Noise model

For simplicity, we model the noise coefficients $\tilde{\mathbf{n}}(q,k)$ as zero-mean Gaussian vectors with covariance given by

$$\text{E}\{\tilde{\mathbf{n}}(q,k)\tilde{\mathbf{n}}(q',k')^T\} = \delta(q,q')\delta(k,k')\mathbf{J}$$

where $\mathbf{J}$ is an arbitrary diagonal matrix. Hence, m parameters are devoted to model the noise structure. Our method, however, does not rely heavily on this assumption and could easily be extended to deal, for instance, with a stationary colored noise model, that is, $\text{E}\{\tilde{\mathbf{n}}(q,k)\tilde{\mathbf{n}}(q',k')^T\} = \delta(q,q')\delta(k,k')g(b)\mathbf{J}$ when $q \in B^{(b)}$ for some function $g(b)$.

2.4. Observation model

Using model (1) and the above assumptions, the covariance matrix for the transformed data vector $\tilde{\mathbf{x}}(q,k)$ depends on the block b to which q belongs:

$$\mathbf{R}_{bk} \overset{\text{def}}{=} \text{Cov}(\tilde{\mathbf{x}}(q,k)) \quad \text{for} \quad q \in B^{(b)}$$

and is given by

$$\mathbf{R}_{bk} = \mathbf{A}\,\mathbf{S}_{bk}\,\mathbf{A}^T + \mathbf{J} \tag{8}$$

These covariance matrices can be estimated by sample averaging:

$$\widehat{\mathbf{R}}_{bk} = \frac{1}{w_b} \sum_{q \in B^{(b)}} \tilde{\mathbf{x}}(q,k)\tilde{\mathbf{x}}(q,k)^T \tag{9}$$

where w_b denotes the number of data points in the b-th block.

3. METHOD

Our approach to blind separation of noisy mixtures is in two steps. In a first step, all unknown parameters, collectively denoted by $\boldsymbol{\theta}$:

$$\boldsymbol{\theta} = \{\mathbf{A}, \mathbf{J}, \sigma_i^2(b,k), i \in [\![1,n]\!], b \in [\![1, n_{\text{block}}]\!], k \in [\![1, n_{\text{frame}}]\!]\}$$

are estimated by maximizing the likelihood of the model described in section 2. In second step, the source coefficients in a frame k at a frequency q belonging to the block $B^{(b)}$ are obtained as the conditional expectation of $\tilde{\mathbf{s}}(q,k)$ given the observation $\tilde{\mathbf{x}}(q,k)$ and the (estimated) parameters. If true parameters were used, this technique would yield the best estimates in a least-square sense. It is easily implemented in our Gaussian framework where it reduces to linear Wiener filtering:

$$\text{E}\{\tilde{\mathbf{s}}(q,k)|\tilde{\mathbf{x}}(q,k), \boldsymbol{\theta}\} = \mathbf{W}_{bk}(\boldsymbol{\theta})\tilde{\mathbf{x}}(q,k)$$

where the $n \times m$ matrix $\mathbf{W}_{bk}(\boldsymbol{\theta})$ depends on $\boldsymbol{\theta}$ according to

$$\mathbf{W}_{bk}(\boldsymbol{\theta}) = (\mathbf{A}^T\mathbf{J}^{-1}\mathbf{A} + \mathbf{S}_{bk}^{-1})^{-1}\mathbf{A}^T\mathbf{J}^{-1}$$

In practice, the source coefficients are obtained by $\mathbf{W}_{bk}(\hat{\boldsymbol{\theta}})\tilde{\mathbf{x}}(q,k)$ (where $\hat{\boldsymbol{\theta}}$ is the maximum likelihood estimate of $\boldsymbol{\theta}$, see below) and the source signals are then reconstructed in the time domain via inverse transform of the estimated coefficients in each frame.

Maximum likelihood estimation of the unknown parameters is obtained as follows. The distribution of the $t-f$ coefficients of the observations for a frame k at a frequency q belonging to the block $B^{(b)}$ is multivariate Gaussian with zero mean and covariance matrix $\mathbf{R}_{bk}$. In the Gaussian model described at section 2.2, the log-likelihood of the transformed data $\tilde{\mathbf{x}}$ is:

$$\log p(\tilde{\mathbf{x}}|\boldsymbol{\theta}) = \sum_{k=1}^{n_{\text{frame}}} \sum_{b=1}^{n_{\text{block}}} \sum_{q \in B^{(b)}} \log \mathcal{N}(\tilde{\mathbf{x}}(q,k)|0, \mathbf{R}_{bk}(\boldsymbol{\theta}))$$

Simple computations shows that $-2\log p(\tilde{\mathbf{x}}|\boldsymbol{\theta}) = \phi(\boldsymbol{\theta}) + \text{cst}$ with

$$\phi(\boldsymbol{\theta}) = \sum_{k=1}^{n_{\text{frame}}} \sum_{b=1}^{n_{\text{block}}} w_q\{\text{trace}(\widehat{\mathbf{R}}_{bk}\mathbf{R}_{bk}(\boldsymbol{\theta})^{-1}) + \log\det\mathbf{R}_{bk}(\boldsymbol{\theta})\}$$

Therefore, the most likely parameters are obtained as $\hat{\boldsymbol{\theta}}_{ML} = \arg\min \phi(\theta)$. The important points regarding maximum likelihood estimation in this context are:

- The likelihood (or, equivalently, function $\phi(\boldsymbol{\theta})$) depends on the data only through the collection $\{\mathbf{R}_{bk}\}$ of sample covariance matrices.
- In the determined, noise-free case ($\mathbf{J} = 0$), criterion $\phi(\boldsymbol{\theta})$ can be understood as a joint diagonalization criterion. A very efficient algorithm due to Pham [4] can be used to minimize it.

- In the noisy case, possibly underdetermined, a simple Gaussian EM algorithm can be used to minimize $\phi(\boldsymbol{\theta})$. The EM algorithm for likelihood maximization is completely determined once a set of 'latent' or 'unobserved' variables are defined. In our case, these latent variables are taken to be the source coefficients $\tilde{\mathbf{s}}(q,k)$. The EM algorithm is derived in [10] in the case when the noise covariance matrix does not depend on (b,k). It can be extended to more complex noise scenarios, like arbitrary dependence on (b,k) or, more realistically, a dependence on the frequency band b only.

4. RESULTS

We present results over mixtures of $n=4$ audio sources (speech, piano, rhythmic guitar, solo guitar). The signals are sampled at 8kHz with $N=65536$ ($\approx 8s$). The time-frequency transform used is the Modified Discrete Cosine Transform, an overlap/add transform [9] which has proved to give good sparse approximations of audio signals, with many coding applications [11, 12]. The MDCT was used with a sine bell window analysis with $l_{\text{frame}} = 512$, which corresponds to a time resolution of $64ms$ and a frequency resolution of $7.8Hz$. We used a linearly spaced partition of the frequency axis to model the sources, with equal subband length $l_{\text{block}} = 16$ ($125Hz$), yielding $n_{\text{block}} = 32$.

The proposed method was applied to 3 mixtures. The first one is 4×4 noise-free mixture, the second mixture is obtained by adding noise to the first one, and the third mixture is obtained by discarding one observation of the second one, thus yielding a 3×4 noisy mixture.

All sound samples, including mixtures, original sources and estimates can be listened to at `http://www-sigproc.eng.cam.ac.uk/~cf269/waspaa05_1/sound_files`.

The separation evaluation criteria used in the following are described in [13], but basically, the SDR (Source to Distortion Ratio) provides an overall separation performance criterion, the SIR (Source to Interferences Ratio) measures the level of interferences from the other sources in each source estimate, SNR (Source to Noise Ratio) measures the error due to the additive noise on the sensors and the SAR (Source to Artifacts Ratio) measures the level of artifacts in the source estimates. We point out that the performance criteria are invariant to a change of basis, so that figures can be computed either on the time sequences ($\hat{\mathbf{s}}$ compared to $\mathbf{s}$) or the MDCT coefficients ($\hat{\tilde{\mathbf{s}}}$ compared to $\tilde{\mathbf{s}}$).

4.1. Noise-free determined mixture

The determined mixture is obtained by mixing the 4 sources by the following mixing matrix:

$$\mathbf{A} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0.8 & 1.3 & -0.9 & 1 \\ 1.2 & -0.7 & 1.1 & 0.6 \\ 0.6 & -0.8 & 0.5 & 1.2 \end{bmatrix} \quad (10)$$

Table 1 shows the SIRs for the source estimates obtained by JADE [14], BGML [5], and the proposed maximum likelihood approach (referred to as TFGML). In the noise-free determined case, TFGML amounts to the joint diagonalization of the set of matrices $\widehat{\mathbf{R}}_{bk}$ and is, except for the use of the MDCT, identical to [7]. As these three methods yield source estimates via application of a separating matrix to the observations, they do not produce artifacts and the SARs are thus infinite. The mixture being noise-free, SNRs are also infinite. Table 1 illustrates the gain in performance provided by TFGML. Note however, that, from an audio point of view, no interference can be heard above $30dB$ SNRs, so that the 3 methods perform reasonably well.

	$\hat{s}_1$	$\hat{s}_2$	$\hat{s}_3$	$\hat{s}_4$
TFGML	75.2	88.5	49.5	45.9
BGML	63.0	37.7	39.6	46.3
JADE	35.6	30.6	29.7	31.9

Table 1: Performance criteria in the determined noise-free case.

The improved SIRs obtained with TFGML are to be linked with the sparsity of the source $t-f$ coefficients: $t-f$ areas where one source is "silent" allows for *super-efficiency* of the criterion and perfect reconstruction of the corresponding row of $\mathbf{A}^{-1}$ [5].

4.2. Noisy determined mixture

We now add iid Gaussian noise to the previous mixture, yielding $\approx 18dB$ input SNR on each observation. Table 2 shows the separation criteria for the source estimates obtained by JADE and TFGML (with its EM implementation), as well as oracle results. The oracle provides upper bounds corresponding to the best results that can be expected from the chosen model of the sources. Oracle source estimates are obtained by reconstructing the sources via Wiener filtering of the observations in each frame/subband with the *true* mixing matrix, input noise variance, and batch covariances matrices $\widehat{\mathbf{S}}_{bk}$ computed on the *original* sources. In Table 2 "Oracle diag" gives upper bounds when the mutual independence of the source coefficients is enforced, that is when $\widehat{\mathbf{S}}_{bk}$ is constrained to be diagonal. The figures obtained by both oracles are very similar, which thus justifies the reliability of the mutual independence assumption.

When TFGML is initialized with random values of the parameter $\boldsymbol{\theta}$, satisfactory convergence is observed after 1000 iterations, taking approximately 2 minutes on a Mac G4 clocked at 1.25 GHz. Table 2 shows that TFGML nearly reaches the oracle results and thus illustrates the accuracy of the EM estimation in the determined case (and similar results were observed over several runs of the algorithm with different initializations). As expected, Table 2 also shows the benefits of the $t-f$ Gaussian source model in terms of denoising but also interference rejection of TFMGL with comparison to JADE.

4.3. Noisy underdetermined mixture

We now consider the difficult case consisting of discarding one observation of the previous mixture, thus yielding an underdetermined problem. As in previous section, Table 3 shows the separation criteria for the source estimates obtained by TFGML with comparison to the oracle results. TFGML was again initialized with random values of the parameters and convergence was observed after 3000 iterations. However, in the underdetermined case the algorithm appears to be more sensitive to initialization, and convergence happened to be slower on other runs. Preliminary results did not however indicate local maxima problems in the 3×4, but this is to be verified on more simulations.

Because the mixture is now underdetermined, the upper bounds given by the oracles are much lower than in the determined case.

	$\hat{s}_1$ (speech)			
	SDR	SIR	SAR	SNR
Oracle	14.2	35.1	15.0	21.8
Oracle diag	13.9	34.7	14.8	21.4
TFGML	12.7	29.1	14.9	17.1
JADE	6.3	18.8	x	6.6
	$\hat{s}_2$ (piano)			
	SDR	SIR	SAR	SNR
Oracle	20.9	45.7	22.4	26.6
Oracle diag	20.7	45.2	22.2	26.2
TFGML	20.0	37.3	22.8	23.3
JADE	14.9	28.2	x	15.2
	$\hat{s}_3$ (rhythmic guitar)			
	SDR	SIR	SAR	SNR
Oracle	17.7	40.0	19.3	22.8
Oracle diag	17.5	39.5	19.2	22.4
TFGML	16.7	28.8	20.0	19.9
JADE	11.6	26.2	x	11.7
	$\hat{s}_4$ (solo guitar)			
	SDR	SIR	SAR	SNR
Oracle	18.8	36.8	21.7	22.1
Oracle diag	18.6	36.3	21.6	21.9
TFGML	17.9	30.7	21.7	20.8
JADE	12.8	29.7	x	12.9

Table 2: Performance criteria in the determined noisy case.

The audio results are however still reasonable. TFGML still provides good source estimates, though it fails to reach the oracle results.

5. CONCLUSIONS

We have presented a maximum likelihood approach to blind audio source separation under a simple but powerful time-frequency Gaussian model of the sources. Even though the model is rather crude, it still yields accurate source estimates: for determined mixtures and also, to some extent, for underdetermined mixtures, the denoising and interference rejection properties are almost as good in blind separation as with an oracle.

An important feature of our Gaussian is that its likelihood can be simply and reasonably quickly maximized by the EM algorithm. This is in contrast to non Gaussian models which may offer better sound quality but at the cost of much more computationally intensive optimization techniques [15].

Future work will consider acceleration techniques for maximizing the likelihood and investigate the influence of the frame/subband tiling: "wavelet-like" tilings could be used to improve frequency resolution of the model in low and medium frequency bands (where audio signals gather most of their energy) while keeping a low number of parameters to estimate.

	$\hat{s}_1$ (speech)			
	SDR	SIR	SAR	SNR
Oracle	7.4	17.2	8.1	24.3
Oracle diag	6.6	15.5	7.5	23.7
TFGML	2.0	8.1	4.2	16.5
	$\hat{s}_2$ (piano)			
	SDR	SIR	SAR	SNR
Oracle	14.8	24.0	15.9	25.2
Oracle diag	14.3	22.4	15.5	24.7
TFGML	10.8	18.6	12.3	20.9
	$\hat{s}_3$ (rhythmic guitar)			
	SDR	SIR	SAR	SNR
Oracle	17.3	38.0	19.0	22.5
Oracle diag	17.2	37.3	18.9	22.2
TFGML	16.3	26.7	19.5	20.0
	$\hat{s}_4$ (solo guitar)			
	SDR	SIR	SAR	SNR
Oracle	9.9	17.9	11.1	21.7
Oracle diag	9.1	16.1	10.5	21.5
TFGML	5.0	11.0	6.8	21.7

Table 3: Performance criteria in the underdetermined noisy case.

6. REFERENCES

[1] J.-F. Cardoso, "Blind signal separation: statistical principles," *Proceedings of the IEEE. Special issue on blind identification and estimation*, vol. 9, no. 10, pp. 2009–2025, Oct. 1998.

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

[3] A. Belouchrani, K. Abed-Meraim, J.-F. Cardoso, and E. Moulines, "A blind source separation technique based on second order statistics," *IEEE Trans. Signal Processing*, vol. 45, no. 2, pp. 434–444, Feb. 1997.

[4] D. Pham, "Blind separation of instantaneous mixture of sources via the Gaussian mutual information criterion," *Signal Processing*, no. 4, pp. 855–870, 2001.

[5] D.-T. Pham and J.-F. Cardoso, "Blind separation of instantaneous mixtures of non stationary sources," *IEEE Trans. Signal Processing*, vol. 49, no. 9, pp. 1837–1848, Sep. 2001.

[6] L. Parra and C. Spence, "Convolutive blind source separation of non-stationary sources," *IEEE Trans. on Speech and Audio Processing*, pp. 320–327, may 2000.

[7] D.-T. Pham, "Exploiting source non stationary and coloration in blind source separation," in *Proceeding of DSP 2002 Conference*, Santorini, Greek, Jul. 2002.

[8] C. Févotte and C. Doncarli, "Two contributions to blind source separation using time-frequency distributions," *IEEE Signal Processing Letters*, vol. 11, no. 3, Mar. 2004.

[9] S. Mallat, *A wavelet tour of signal processing*. Academic Press, 1998.

[10] J.-F. Cardoso, H. Snoussi, J. Delabrouille, and G. Patanchon, "Blind separation of noisy Gaussian stationary sources. Application to cosmic microwave background imaging," in *Proc. EUSIPCO*, vol. 1, 2002, pp. 561–564.

[11] K. Brandenburg, "MP3 and AAC explained," in *Proc. AES 17th Int. Conf. High Quality Audio Coding*, Florence, Italy, Sep 1999.

[12] L. Daudet and M. Sandler, "MDCT analysis of sinusoids: exact results and applications to coding artifacts reduction," *IEEE Trans. Speech and Audio Processing*, vol. 12, no. 3, pp. 302–312, May 2004.

[13] R. Gribonval, L. Benaroya, E. Vincent, and C. Févotte, "Proposals for performance measurement in source separation," in *Proc. 4th Symposium on Independent Component Analysis and Blind Source Separation (ICA'03)*, Nara, Japan, Apr. 2003.

[14] J.-F. Cardoso and A. Souloumiac, "Blind beamforming for non Gaussian signals," *IEE Proceedings-F*, vol. 140, no. 6, pp. 362–370, 1993.

[15] C. Févotte and S. J. Godsill, "A bayesian approach for blind separation of sparse sources," *Technical Report of Cambridge University Engineering Dept. CUED/F-INFENG/TR.511*, Jan. 2005, http://www-sigproc.eng.cam.ac.uk/~cf269/TechRep/techrep05.pdf.