# BLIND SEPARATION OF MORE SOURCES THAN SENSORS IN CONVOLUTIVE MIXTURES

*Rasmus Kongsgaard Olsson and Lars Kai Hansen*

Informatics and Mathematical Modelling, Technical University of Denmark
2800 Lyngby, Denmark, rko,lkh@imm.dtu.dk

**ABSTRACT**

We demonstrate that blind separation of more sources than sensors can be performed based solely on the second order statistics of the observed mixtures. This a generalization of well-known robust algorithms that are suited for equal number of sources and sensors. It is assumed that the sources are non-stationary and sparsely distributed in the time-frequency plane. The mixture model is convolutive, i.e. acoustic setups such as the cocktail party problem are contained. The limits of identifiability are determined in the framework of the PARAFAC model. In the experimental section, it is demonstrated that real room recordings of 3 speakers by 2 microphones can be separated using the method.

## 1. INTRODUCTION

The human auditory system solves the so-called cocktail party problem, i.e it separates out a single speech signal from a composition of speech signals and possibly other interfering noises. This is an instance of blind source separation (BSS). Machines capable of emulating this function have potential applications in e.g. hearing aids and audio communication. The *convolutive* mixture model accounts for the various delays and attenuations of e.g. an acoustic mixture:

$$\mathbf{x}[t] = \sum_{k=0}^{L-1} \mathbf{A}[k]\mathbf{s}[t-k] \tag{1}$$

where $\mathbf{x}[t]$ and $\mathbf{s}[t]$ are an $N$ dimensional sensor vector and an $R$ dimensional source vector, respectively, sampled at discrete time $t$. The matrices $\mathbf{A}[k]$ contain the impulse responses of the signal channels. The sources can be recovered blindly, i.e. $\mathbf{A}[k]$ unknown, up to arbitrary filtering and permutation of source index.

In many cases, such as the cocktail party situation where the speakers are independent on the timescale of interest, the problem can to some extend be solved by algorithms that are based on independent component analysis (ICA), [1]. In particular, the instantaneous mixture model, which arises as a results of $L = 1$, is a well-solved problem, see, e.g., [2]. However, this mixing model ($L = 1$) is inappropriate and insufficient for the separation of acoustically mixed audio signals for the reasons already mentioned. ICA algorithms determine $\mathbf{s}_t$ by assuming statistical independency of the sources and certain properties of the distribution of $\mathbf{s}_t$, where non-Gaussianity, non-stationarity and non-whiteness are the most important. Convolutive ICA algorithms, i.e. $L > 1$, have been devised by e.g. [3] and [4]. The most efficient methods use transformation to the discrete Fourier domain, where convolution approximately translates to multiplication, yielding a separate instantaneous ICA problem for each evaluated frequency. As a result, the resulting arbitrary permutations across frequency problem must be resolved. Algorithms that function in the time-domain, such as [5], can benefit more directly from domain-oriented source modelling, but typically at a higher computational cost.

Common to the algorithms mentioned above are that they assume *quadratic* mixtures, that is, the number of sensors equals the number of sources, or $N = R$. The class of mixtures, where $R > N$, are termed *underdetermined*.[1] Instantaneous ICA algorithms have been devised to solve the underdetermined problem, i.e. [6], [7] and [8]. These methods assume a sparse distribution of the sources, either directly in the time-domain or in a transformed domain.

An alternative approach to blind source separation is the use of binary masks in the spectrogram, i.e. assigning each point in the time-frequency plane to a source. The masks are often constructed using segmentation cues inspired by the human auditory system, such as interaural intensity and time differences (IID/ITD), [9]. Efforts to combine ICA with binary masks have been undertaken by [10]. A problem introduced by binary masks is that artifacts, or unnatural sounds, may appear in the reconstructed signals.

The major contribution of this work is to generalize to the overcomplete case the robust algorithms of [3] and [11], which handle quadratic mixtures relying solely on robust time-varying second order statistics in the Fourier domain. The essentially time-varying Gaussian signal model is an instance of the trilinear PARAFAC[2] model, [12], the results from this field are employed to construct the algorithm and certify the identifiability of the model under various assumptions. One observation derived from the PARAFAC formulation is that the source power spectra are identifiable for ($N = 2$, $R = 3$) provided that the mixing process parameters ($\mathbf{A}[k]$) are available at that stage. As a consequence, the maximum posteriori estimates of the sources can be computed as opposed to the usual binary mask reconstructions. A key component in determining $\mathbf{A}[k]$ is the sparsity of the sources in the time-frequency plane, which allows for estimation of this part of the model through k-means clustering, [8]. As evidence of the usefulness of the approach, it is demonstrated that speech mixtures ($N = 2$, $R = 3$) recorded in a *real* office environment can be handled, see
`www.imm.dtu.dk/~rko/underdetermined`.

In sections 2 and 3, the PARAFAC formulation of the blind source separation problem is motivated. In sections 4 and 6, the estimation of the parameters and the source inference, respectively, are covered. The limits of identifiability are discussed in section 5. Implementation issues are summarized in section 7. The perfor-

[1] One-sensor separation is a topic in its own right and is not discussed here.

[2] Parallel Factor Analysis. Also known as Canonical Decomposition (CANDECOMP).

Oticon Fonden financially supported this work.

mance of the algorithm is gauged on benchmark audio data in the experimental section.

## 2. PARALLEL FACTOR ANALYSIS

A thorough review of the following theory can be found in [13]. Consider the 3-way array $x_{ijk}$ indexed by $i \in [1, .., I]$, $j \in [1, .., J]$, $k \in [1, .., K]$. The trilinear PARAFAC decomposition is defined:

$$x_{ijk} = \sum_{f=1}^{F} a_{if} b_{jf} c_{kf}$$

with loading matrices $(\mathbf{A})_{if} = a_{if}$, $(\mathbf{B})_{jf} = b_{jf}$ and $(\mathbf{C})_{kf} = c_{kf}$. The PARAFAC model can equivalently be expressed in terms of its matrices or 'slabs':

$$\mathbf{X}_k = \mathbf{A}\mathrm{diag}_k[\mathbf{C}]\mathbf{B}^\top \tag{2}$$

where $\mathrm{diag}_k[\cdot]$ operating on a matrix constructs a diagonal matrix with the $k$th row of the matrix as diagonal elements. The PARAFAC model could equivalently be expressed along $\mathbf{A}$ or $\mathbf{B}$. The *matrization* of the PARAFAC model is yet another representation: $\mathbf{X}_{IJ\times K} = (\mathbf{B} \odot \mathbf{A})\,\mathbf{C}^\top$, where the Khatri-Rao product is defined

$$(\mathbf{B} \odot \mathbf{A}) \equiv \begin{bmatrix} \mathbf{A}\mathrm{diag}_1[\mathbf{B}] \\ \vdots \\ \mathbf{A}\mathrm{diag}_J[\mathbf{B}] \end{bmatrix}$$

The indices of $\mathbf{X}_{IJ\times K}$ indicate the direction and hence the dimensions of the matrization. The former index varies more rapidly. A sufficient condition for uniqueness was provided by Kruskal in [12]:

$$k[\mathbf{A}] + k[\mathbf{B}] + k[\mathbf{C}] \geq 2(F-1) \tag{3}$$

where the k-rank, denoted $k[\mathbf{A}]$, of a matrix $\mathbf{A}$, is defined as the maximal integer $m$ such that any $m$ columns of $\mathbf{A}$ form a linearly independent set. Clearly, $k[\mathbf{A}] \leq r[\mathbf{A}]$, where $r[\mathbf{A}]$ is the rank of $\mathbf{A}$.

## 3. MODEL

The convolutive mixture of equation (1) forms the basis of the model. Only second-order statistics in the Fourier domain, i.e. auto/cross spectra are included in the analysis, which is similar to imposing a multivariate Gaussian model on the data. The assumptions can be summarized as:

A1: the sources, $\mathbf{s}[t]$, are zero mean, uncorrelated, i.e. $\langle \mathbf{s}[t]\mathbf{s}^\top[t-\tau]\rangle$ is a diagonal matrix for all $\tau$.

A2: the signal channels of $\mathbf{A}[k]$ are constant on the time-scale of analysis. No columns are collinear, as this would effectively constitute a reduction of $N$.

A3: the autocorrelation functions $\langle \mathbf{s}[t]\mathbf{s}^\top[t-\tau]\rangle$ are time-varying as in [3]. The variation patterns are independent for each source.

A4: At most one source is non-zero in any time-frequency block $(\{n, \ldots, n+M-1\}, k)$, where $M$ is the block length not to be confused with the frame length, $K$. This effectively is an assumption of sparsity as in, e.g., [8]. In figure 1, it is demonstrated that for certain quasi-periodic signals such as speech, the number of $(n, k)$ bins required to represent the signal is indeed small.

The discrete Fourier transform (DFT) is applied to windowed frames of $\mathbf{x}_t$, length $K$, obtaining:

$$\mathbf{x}_k^{(n)} = \mathbf{A}_k \mathbf{s}_k^{(n)} + \mathbf{e}_k^{(n)} \tag{4}$$

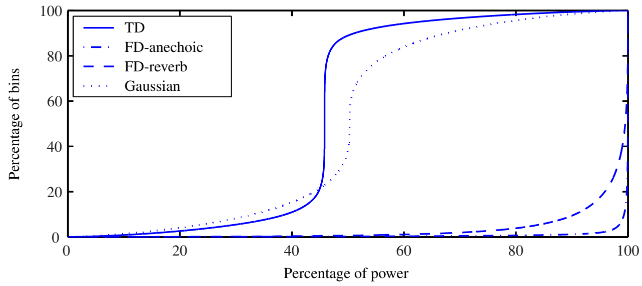


**Fig. 1**. The sparsity of speech ($F_s = 8kHz$) under various conditions presented as the minimal percentage of frequency bins required to represent a given percentage of power. The raw time-domain speech signal (TD) is shown as well as its time-frequency transformation (FD-anechoic) and a version with simulated large-room reverberation (FD-reverb). DFT lenght: $K = 512$, no overlap.

where $\mathbf{x}_k^{(n)}$, $\mathbf{A}_k$ and $\mathbf{s}_k^{(n)}$ are the DFT's of the corresponding time-domain signals at discrete frequencies $k \in [0, 1, \ldots, K-1]$ and frame $n$. The residual term, $\mathbf{e}_k^{(n)}$, is due to equation (1) being a linear convolution rather than a circular one. When $L \ll K$, the mismatch vanishes, that is $\langle \frac{|\mathbf{e}_k|}{|\mathbf{x}_k|} \rangle \rightarrow 0$. Furthermore, the auto/cross power spectra of $\mathbf{x}_t$ as a function of frame index $n$, $\mathbf{C}_k^{(n)}$, are related as follows to the power time-spectra of $\mathbf{s}_t$:

$$\mathbf{C}_k^{(n)} = \mathbf{A}_k \mathbf{D}_k^{(n)} \mathbf{A}_k^H + \mathbf{E}_k^{(n)} \tag{5}$$

where $\mathbf{D}_k^{(n)}$ is a diagonal matrix (due to A1) with the power of the sources in time-frequency bin $(n, k)$ as diagonal elements. The channels, $\mathbf{A}_k$, are independent of $n$ due to A2. The power spectrum residual, $\mathbf{E}_k^{(n)}$ can be neglected when $\mathbf{e}_k^{(n)}$ is small. As was also noted in [11], any linear channel that exhibits a sufficiently rapidly decaying autocorrelation function can be treated by our approach, not just a convolutive channel.

By comparing with equation (2), it is seen that equation (5) is approximately a PARAFAC model. The following reformulation of (5) is convenient:

$$\mathbf{C}_{NN\times K}[k] = (\mathbf{A}_k \odot \mathbf{A}_k^*)\,\mathbf{\Lambda}_k^\top \tag{6}$$

where $\mathbf{C}_{NN\times K}[k]$ is the matricized auto/cross power at frequency $k$ and $(\mathbf{\Lambda}_k)_{nj}$ is the power of source $j$ at time-frequency bin $(n, k)$.

## 4. PARAMETER ESTIMATION

A standard approach to estimating the parameters of equation (6) is the alternating least squares algorithm (ALS), which alternatingly minimizes the Frobenius norm of $\mathbf{R}_{NN\times K}[k] - \mathbf{C}_{NN\times K}[k]$ with respect to the matrices $\mathbf{A}_k$ and $\mathbf{\Lambda}_k$, where $\mathbf{R}_{NN\times K}[k]$ is the estimated auto/cross power spectra of $\mathbf{x}_t$, e.g.:

$$\hat{\mathbf{\Lambda}}_k^\top = \arg\min_{\mathbf{\Lambda}} \left\| \mathbf{R}_{NN\times K}[k] - (\mathbf{A}_k \odot \mathbf{A}_k^*)\mathbf{\Lambda}_k^\top \right\|^2 \tag{7}$$

The arbitrary scaling of the model is fixed by normalizing the columns of $\mathbf{A}_k$ so that $\|\mathbf{a}_{i,k}\|^2 = 1$ and zeroing the phases of the 1st row: $\angle(\mathbf{A})_{1i,k} = 0\ \forall i$. The solution to equation (7) is just a least-squares fit:

$$\hat{\mathbf{\Lambda}}_k^\top = (\mathbf{A}_k \odot \mathbf{A}_k^*)^\dagger\, \mathbf{R}_{NN\times K}[k] \tag{8}$$

| | assumptions | bound |
|---|---|---|
| I | A1, A2, A3 | $2N-2 \geq R$ |
| II | A1, A2, A3, A4 | $\frac{1}{2}N(N+1) \geq R$ |

**Table 1**. Bounds of identifiability of the source spectrograms depending on the set of model assumptions.

where † is the pseudoinverse operator. An alternative means of estimating $\mathbf{A}_k$ and $\mathbf{D}[k,n]$ from $\mathbf{C}[k,n]$ is the application of a joint diagonalization algorithm such as in [14]. However, the ALS is conceptually simple and has good convergence properties under certain circumstances as was demonstrated in [11]. In the case of a 1-component model, i.e. when $\mathbf{A}_k$ and $\mathbf{\Lambda}_k$ consist of a single column ($\mathbf{a}_k$ and $\lambda_k$), a particularly simple solution exists:

$$\left[\sum_n \mathbf{R}_k^{(n)}\right] \mathbf{a}_k = \left[\sum_n \left(\lambda_k^{(n)}\right)^2\right] \mathbf{a}_k \quad (9)$$

where $\mathbf{R}_k^{(n)}$ and $\lambda_k^{(n)}$ are the measured auto/cross power of $\mathbf{x}_t$ and the power of $\mathbf{s}_t$ in time-frequency bin $(n,k)$, respectively. The $\mathbf{a}_k$ corresponding to the maximal eigenvalue is the least-squares estimate.

## 5. IDENTIFIABILITY

In the following will be discussed mainly the limits of recovering the time-varying source power spectra, $\mathbf{\Lambda}_k$, i.e. the spectrograms. The MAP inference of $\mathbf{s}_k^{(n)}$ is treated in section 6. It was mentioned in section 1 that $\mathbf{\Lambda}_k$ can only be blindly recovered up to an unknown scaling and ordering, i.e. it is only possible to estimate $\mathbf{P}_k\mathbf{H}_k\mathbf{\Lambda}_k$, where $\mathbf{H}_k$ and $\mathbf{P}_k$ are scaling and permutation matrices, respectively. The frequency permutation problem, i.e. estimating $\mathbf{P}_k$ for all $k$, can be remedied by defining a similarity measure across frequencies as in e.g. [11]. The scaling matrix was fixed in the parameter estimation process due to certain assumptions about the scale of $\mathbf{A}_k$.

The identifiability of the source spectrograms, $\mathbf{\Lambda}_k$, is determined in the general blind case ($\mathbf{A}_k$ is unknown) under assumptions A1, A2 and A3. The uniqueness theorem (3) then yields a lower bound for identifiability:

$$2N \geq R+2$$

where the full rank of $\mathbf{A}_k$ and $\mathbf{\Lambda}_k$ was assumed, consequences of A2 and A3, respectively. This means that $\mathbf{\Lambda}_k$ can be estimated in many cases where $R > N$, however notably excluding the $N=2$, $R=3$ case. In [15], it was shown that the bound is tight for $R=3$, but not necessarily for $R>3$.

In order to retrieve the source power spectrograms in the $N=2$, $R=3$ case, the sparsity assumption, A4 is required. The rationale is that if only a single source is active in $M$ consecutive frames, the local dimensionality of the PARAFAC model will be $R=1$, and the corresponding column of $\mathbf{A}_k$ is available through equation (9). When estimating $\mathbf{a}_k$ across time, the estimates should ideally occupy $R$ discrete points in space pertaining to the $R$ columns of $\mathbf{A}_k$. In a realistic setting, model bias and slight violations of A4 will cause an amount of dispersion of the estimated $\mathbf{a}_k$ around the true values. Provided this effect is not too severe, $\mathbf{A}_k$ can still be estimated by means of a clustering algorithm. The dispersion around the cluster centers can be quantified by computing the within-class variance, $Q$.

After $\mathbf{A}_k$ has been acquired through the clustering, $\mathbf{\Lambda}_k$ can be estimated via equation (8) from all frames under less strict conditions on $R$ and $N$. The typical rank of $(\mathbf{A}\odot\mathbf{A}^*)$ is $\min\left\{\frac{1}{2}N(N+1), R\right\}$.

- discrete Fourier transform Hann-windowed data
- prewhiten
- for all frequencies $k$:
  - for all blocks of frames $\{n, n+1, \ldots, n+M-1\}$:
    - fit 1-component PARAFAC model
  - estimate $\mathbf{A}_k$ via k-means clustering of $\mathbf{a}_k$ estimates
  - compute $\mathbf{\Lambda}_k$ from $\hat{\mathbf{A}}_k$
  - compute MAP estimate of $\mathbf{s}_k$ from $\mathbf{A}_k$ and $\mathbf{\Lambda}_k$
- solve permutation problem using power corr. across freq.
- inverse prewhiting
- reconstruct time-domain signal by IDFT and overlap-add

**Table 2**. The blind source separation algorithm for underdetermined convolutive mixtures of sources that are sparse in the time-frequency plane.

Therefore, $\mathbf{\Lambda}_k$ generally has a unique solution if $\frac{1}{2}N(N+1) \geq R$. As a consequence, the source spectrograms can be recovered when $N=2$, $R=3$, provided $\mathbf{A}_k$ is known or has been estimated sufficiently accurately. The identifiability of $\mathbf{\Lambda}_k$ including or excluding the sparsity assumption is summarized in table 1.

## 6. SOURCE RECONSTRUCTION

Once $\mathbf{A}_k$ and $\mathbf{D}_k^n$ have been estimated, the sources can be inferred from the data and the parameters of the model. A maximum posteriori (MAP) scheme, which builds on the joint Gaussianity of $\mathbf{s}_t$ and $\mathbf{x}_t$ and the assumptions A1-A3, is employed as suggested in [3]:

$$\hat{\mathbf{s}}_k = \mathbf{D}_k^{(n)}\mathbf{A}_k\left(\mathbf{A}_k\mathbf{D}_k^{(n)}\mathbf{A}_k^H\right)^{-1}\mathbf{x}_k$$

In order to cancel the effects of the arbitrary scaling of $\mathbf{A}_k$ and $\mathbf{D}_k^n$, the sources are computed as they appear at the sensors, e.g. the $j$'th source at the $i$'th sensor: $(\mathbf{A}_k)_{ij}\hat{s}_{j,k}$. Where the number of sources and sensors are locally equal in the time-frequency plane, the above simply reduces to $\mathbf{s}_k = \mathbf{A}_k^{-1}\mathbf{x}_k$. In case of underdetermined mixtures, the degree of success of the MAP estimation depends on the sparsity assumption, A4. Reconstruction in the time domain is performed by inverse DFT and overlap-add.

## 7. ALGORITHM

The full set of assumptions, A1-A4, form the basis of the presented algorithm, yet a number of implementational issues remains. A preprocessing step is included to better condition the data for clustering and separation. In a band of frequencies a whitening matrix is applied to $\mathbf{x}_k$, i.e. $\tilde{\mathbf{x}}_k = \mathbf{W}\mathbf{x}_k$, so that $\langle\tilde{\mathbf{x}}_k\tilde{\mathbf{x}}_k^H\rangle = \mathbf{I}$, where averaging is over a suitable bandwidth and the signal length. The clustering is carried out in the polar coordinates of $\mathbf{a}_k$ using k-means. This requires the distance measure to be able to handle circularity. To be more resilient against outliers and violations of the sparsity assumption, the *median* rather than the mean was taken as the cluster center.

The problem of permutation across frequencies was solved by iteratively growing a set in which the permutations are corrected. The measure of similarity is correlation of amplitude as in e.g. [11]. Starting from the highest frequency index, $k=K-1$, to the lowest, $k=0$, the permutation matrix $\mathbf{P}_k$ was chosen so that it (greedily) maximizes the correlation coefficients of

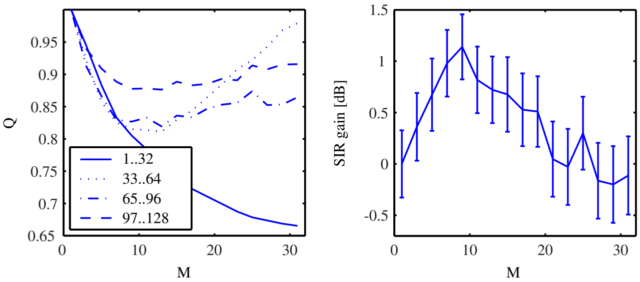


**Fig. 2**. Effects of the block length $M$ on the clustering of $\mathbf{a}_k$. Left: Within-cluster variance, $Q$, in 4 frequency bands. Right: The signal to interference (SIR) gain as a function of the segment length $M$. The results were obtained by averaging across the sources of 3 underdetermined mixtures, $K = 256$.

| | $K = 256$ | $K = 512$ | Araki et al. |
|---|---|---|---|
| `mmf` | $12, 13, 11$ | $16, 15, 10$ | $14, 14, 7$ |
| `mmm` | $12, 7, 15$ | $15, 8, 16$ | $11, 4, 14$ |
| `fff` | $10, 11, 16$ | $8, 9, 15$ | $5, 18, 18$ |

**Table 3**. Estimated signal to interference (SIR) ratios for $R = 3$, $N = 2$ mixtures in a simulated reverberant room. f's and m's represent male and female speakers in the mixture.

$|\mathbf{P}_k \hat{\mathbf{s}}_k|$ to $\sum_{k'=k+1}^{K-1} |\mathbf{P}_{k'} \hat{\mathbf{s}}_{k'}|$. This simple method proved fairly robust, and better results were not obtained with the approach of [11]. The algorithm is summarized in table 2.

## 8. RESULTS AND DISCUSSION

For the initial simulations, experimental audio data generated by researchers Araki et al., see e.g. [10], was used.[3] In each of the mixtures, $R = 3$ speech signals, sample rate $F_s = 8kHz$, were convolved with real room impulse functions of length $\tau_r = 130ms$ to construct $N = 2$ mixtures. The microphones were situated $d_m = 4cm$ apart, the distance to the speakers was $d_s = 110cm$ and the angles to the microphones were $\theta_1 = 50°$, $\theta_2 = 100°$ and $\theta_3 = 135°$. The room dimensions were $\approx 4m \times 4m \times 3m$. Mixture fragments of length $T = 7s$ were used. To measure the degree of separation, the signal-to-interference ratio (SIR) quality index was computed. The SIR's were estimated from the original and estimated sources by means of the `BSS_EVAL` toolbox, see [16] for definitions, and provided a reasonable correlate with human subjective evaluation.[4] Using the sparse algorithm, results on mixtures of same-sex and mixed-sex speech were obtained, see table 3. For reference, the quality measures were also computed for the audio files of [10]. The new method appears to exhibit similar performance to the reference. These results were replicated in a real office environment - 3 male speakers reading aloud from Hans Christian Andersen's fairy tales were separated using the algorithm. The room measures $4.25m \times 5.82m \times 3.28m$. The values of $d_m$, $d_s$ were left unchanged from the generated mixtures. The estimated source signals can be appreciated at www.imm.dtu.dk/∼rko/underdetermined/index.htm.

[3]The audio wave files are available at www.kecl.ntt.co.jp/icl/signal/araki/nbficademo.html.

[4]The toolbox is downloadable at www.irisa.fr/metiss/bss_eval/. The script `bss_decomp_filt.m` was used with $L = 100$.

Furthermore, the effect of the block length $M$ was assessed, using the same convolutive mixtures, $K = 256$. It is seen in figure 2 that the within-class variance $Q$ at most frequencies approximately peaks at $M \approx 10$. The SIR index has a similar optimum, suggesting that $Q$ can be used to determine the stationarity properties of the signal and select an optimal model order.

## 9. REFERENCES


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

[2] J. Cardoso, "Blind signal separation: statistical principles," in *IEEE, Blind Identification and Estimation*, 1998, vol. 90.

[3] L. Parra and C. Spence, "Convolutive blind separation of non-stationary sources," *IEEE Transactions, Speech and Audio Processing*, pp. 320–7, 5 2000.

[4] T.W. Lee, A. J. Bell, and R. H. Lambert, "Blind separation of delayed and convolved sources," in *Advances in NIPS*. 1997, vol. 9, p. 758, The MIT Press.

[5] R. K. Olsson and L. K. Hansen, "A harmonic excitation state-space approach to blind separation of speech," in *Advances in NIPS 17*, 2005, pp. 993–1000.

[6] T. Lee, M. Lewicki, M. Girolami, and T. Sejnowski, "Blind source separation of more sources than mixtures using overcomplete representations," *IEEE Signal Processing Letters*, vol. 6, no. 4, 1999.

[7] P. Comon, "Blind identification and source separation in 2x3 under-determined mixtures," *IEEE Transactions on Signal Processing*, vol. 52, pp. 11–22, 2004.

[8] P. Bofill and M. Zibulevsky, "Underdetermined blind source separation using sparse representations," *Signal Processing*, vol. 81, pp. 2353–2362, 2001.

[9] O. Yilmaz and S. Rickard, "Blind separation of speech mixtures via time-frequency masking," *IEEE Transactions on Signal Processing*, vol. 52, pp. 1830–1847, 2004.

[10] S. Araki, S. Makino, A. Blin, R. Mukai, and H. Sawada, "Blind separation of more speakers than sensors with less distortion by combining sparseness and ICA," in *IWAENC2003*, 2003, pp. 271–274.

[11] K. Rahbar and J. P. Reilly, "A new frequency domain method for blind source separation of convolutive audio mixtures," *IEEE Transactions on Speech and Audio Processing*, to appear.

[12] J. Kruskal, "Three-way arrays: Rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics," *Linear Algebra and its Applications*, vol. 18, pp. 95–138, 1977.

[13] A. Smilde, R Bro, and P. Geladi, *Multi-way Analysis: Applications in the Chemical Sciences*, Wiley, 2004.

[14] A. Belouchrani, K. Abed-Meraim, J.F. Cardoso, and E. Moulines, "A blind source separation technique using second order statistics," *IEEE Trans. on Signal Processing*, , no. 2, pp. 320–7, 1997.

[15] J. M. F. ten Berge and N. D. Sidiropoulos, "On uniqueness in CANDECOMP / PARAFAC," *Psychometrika*, vol. 67, no. 3, pp. 399–409, 2002.

[16] C. Févotte, R. Gribonval, and E. Vincent, "BSS_EVAL Toolbox User Guide," Tech. Rep. 1706, IRISA, Rennes, France, 2005.