A SURVEY OF CONVOLUTIVE BLIND SOURCE SEPARATION METHODS

Michael Syskind Pedersen[1], Jan Larsen[2], Ulrik Kjems[1], and Lucas C. Parra[3]

[1] Oticon A/S, 2765, Smørum, Denmark, {msp, uk}@oticon.dk
[2] Technical University of Denmark, Informatics and Mathematical
Modelling, 2800 Kgs. Lyngby, Denmark, jl@imm.dtu.dk
[3] The City College of New York, Biomedical Engineering,
New York, NY 10031, parra@ccny.cuny.edu

ABSTRACT

In this chapter, we provide an overview of existing algorithms for blind source separation of convolutive audio mixtures. We provide a taxonomy, wherein many of the existing algorithms can be organized, and we present published results from those algorithms that have been applied to real-world audio separation tasks.

1. INTRODUCTION

During the past decades, much attention has been given to the separation of mixed sources, in particular for the *blind* case where both the sources and the mixing process are unknown and only recordings of the mixtures are available. In several situations it is desirable to recover all sources from the recorded mixtures, or at least to segregate a particular source. Furthermore, it may be useful to identify the mixing process itself to reveal information about the physical mixing system.

In some simple mixing models each recording consists of a sum of differently weighted source signals. However, in many real-world applications, such as in acoustics, the mixing process is more complex. In such systems, the mixtures are weighted and delayed, and each source contributes to the sum with multiple delays corresponding to the multiple paths by which an acoustic signal propagates to a microphone. Such filtered sums of different sources are called convolutive mixtures. Depending on the situation, the filters may consist of a few delay elements, as in radio communications, or up to several thou-

sand delay elements as in acoustics. In these situations the sources are the desired signals, yet only the recordings of the mixed sources are available and the mixing process is unknown.

There are multiple potential applications of convolutive blind source separation. In acoustics different sound sources are recorded simultaneously with possibly multiple microphones. These sources may be speech or music, or underwater signals recorded in passive sonar [1]. In radio communications, antenna arrays receive mixtures of different communication signals [2, 3]. Source separation has also been applied to astronomical data or satellite images [4]. Finally, convolutive models have been used to interpret functional brain imaging data and bio-potentials [5, 6, 7, 8].

This chapter considers the problem of separating linear convolutive mixtures focusing in particular on acoustic mixtures. The *cocktail-party problem* has come to characterize the task of recovering speech in a room of simultaneous and independent speakers [9, 10]. Convolutive blind source separation (BSS) has often been proposed as a possible solution to this problem as it carries the promise to recover the sources exactly. The theory on linear noise-free systems establishes that a system with multiple inputs (sources) and multiple output (sensors) can be inverted under some reasonable assumptions with appropriately chosen multi-dimensional filters [11]. The challenge lies in finding these convolution filters.

There are already a number of partial reviews available on this topic [12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22]. The purpose of this chapter is to pro-

vide a complete survey of convolutive BSS and identify a taxonomy that can organize the large number of available algorithms. This may help practitioners and researchers new to the area of convolutive source separation obtain a complete overview of the field. Hopefully those with more experience in the field can identify useful tools, or find inspiration for new algorithms. Figure 1 provides an overview of the different topics within convolutive BSS and in which section they are covered. An overview of published results is given in Section 8.

2. THE MIXING MODEL

First we introduce the basic model of convolutive mixtures. At the discrete time index t, a mixture of N source signals $s(t) = (s_1(t), \ldots, s_N(t))$ are received at an array of M sensors. The received signals are denoted $x(t) = (x_1(t), \ldots, x_M(t))$. In many real-world applications the sources are said to be *convolutively* (or dynamically) mixed. The convolutive model introduces the following relation between the m'th mixed signal, the original source signals, and some additive sensor noise $v_m(t)$:

$$x_m(t) = \sum_{n=1}^{N} \sum_{k=0}^{K-1} a_{mnk} s_n(t-k) + v_m(t) \quad (1)$$

The mixed signal is a linear mixture of filtered versions of each of the source signals, and a_{mnk} represents the corresponding mixing filter coefficients. In practice, these coefficients may also change in time, but for simplicity the mixing model is often assumed stationary. In theory the filters may be of infinite length (which may be implemented as IIR systems), however, again, in practice it is sufficient to assume $K < \infty$. In matrix form, the convolutive model can be written as:

$$x(t) = \sum_{k=0}^{K-1} A_k s(t-k) + v(t), \quad (2)$$

where A_k is an $M \times N$ matrix which contains the k'th filter coefficients. $v(t)$ is the $M \times 1$ noise vector. In the z-domain the convolutive mixture (2) can be written as:

$$X(z) = A(z)S(z) + V(z), \quad (3)$$

where $A(z)$ is a matrix with FIR polynomials in each entry [23].

2.1. Special cases

There are some special cases of the convolutive mixture which simplify Eq. (2):

Instantaneous Mixing Model: Assuming that all the signals arrive at the sensors at the same time without being filtered, the convolutive mixture model (2) simplifies to

$$x(t) = As(t) + v(t). \quad (4)$$

This model is known as the *instantaneous* or delay-less (linear) mixture model. Here, $A = A_0$, is an $M \times N$ matrix containing the mixing coefficients. Many algorithms have been developed to solve the instantaneous mixture problem, see e.g. [17, 24].

Delayed Sources: Assuming a reverberation-free environment with propagation delays the mixing model can be simplified to

$$x_m(t) = \sum_{n=1}^{N} a_{mn} s_n(t - k_{mn}) + v_m(t) \quad (5)$$

where k_{mn} is the propagation delay between source n and sensor m.

Noise Free: In the derivation of many algorithms, the convolutive model (2) is assumed to be noise-free, i.e.:

$$x(t) = \sum_{k=0}^{K-1} A_k s(t-k). \quad (6)$$

Over and Under-determined Sources: Often it is assumed that the number of sensors equals (or exceeds) the number of sources in which case linear methods may suffice to invert the linear mixing. However, if the number of sources exceeds the number of sensors the problem is under-determined, and even under perfect knowledge of the mixing system linear methods will not be able to recover the sources.

2.2. Convolutive model in the frequency domain

The convolutive mixing process (2) can be simplified by transforming the mixtures into the frequency domain. The linear convolution in the time domain can
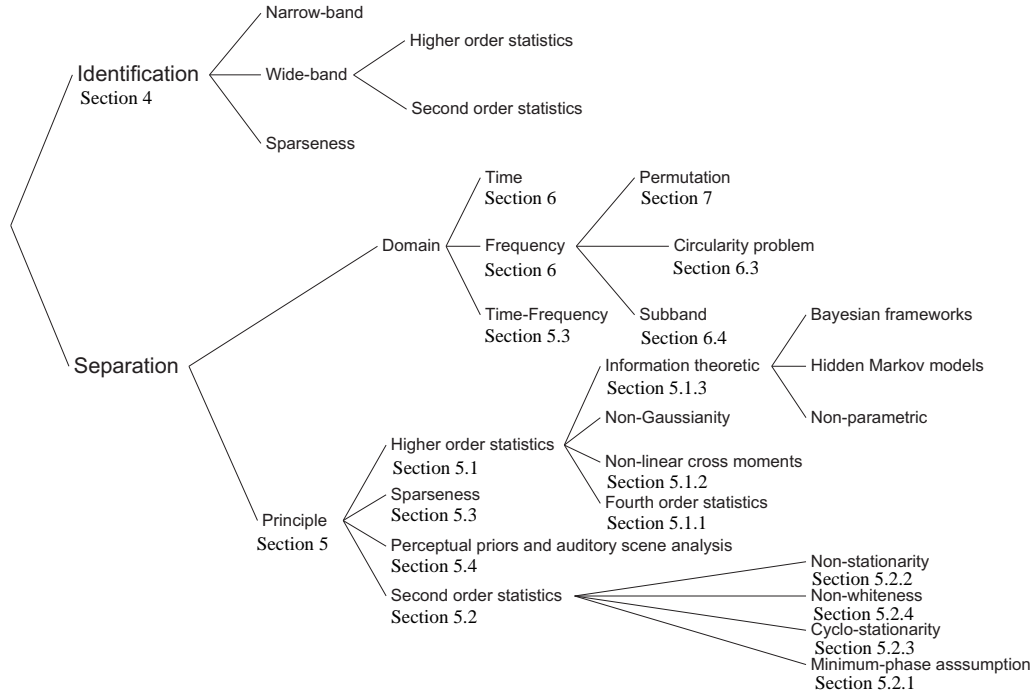
Figure 1: Overview of important areas within blind separation of convolutive sources.

be written in the frequency domain as separate multiplications for each frequency:

$$\boldsymbol{X}(\omega) = \boldsymbol{A}(\omega)\boldsymbol{S}(\omega) + \boldsymbol{V}(\omega). \qquad (7)$$

At each frequency, $\omega = 2\pi f$, $\boldsymbol{A}(\omega)$ is a complex $M \times N$ matrix, $\boldsymbol{X}(\omega)$ and $\boldsymbol{V}(\omega)$ are complex $M \times 1$ vectors, and similarly $\boldsymbol{S}(\omega)$ is a complex $N \times 1$ vector. The frequency transformation is typically computed using a discrete Fourier transform (DFT) within a time frame of size T starting at time t:

$$\boldsymbol{X}(\omega, t) = \mathrm{DFT}([\boldsymbol{x}(t), \cdots, \boldsymbol{x}(t + T - 1)]), \quad (8)$$

and correspondingly for $\boldsymbol{S}(\omega, t)$ and $\boldsymbol{V}(\omega, t)$. Often a windowed discrete Fourier transform is used:

$$\boldsymbol{X}(\omega, t) = \sum_{\tau=0}^{T-1} w(\tau)\boldsymbol{x}(t + \tau)e^{-j\omega\tau/T}, \quad (9)$$

where the window function $w(\tau)$ is chosen to minimize band-overlap due to the limited temporal aperture. By using the fast Fourier transform (FFT) convolutions can be implemented efficiently in the discrete Fourier domain, which is important in acoustics as it often requires long time-domain filters.

2.3. Block-based Model

Instead of modeling individual samples at time t one can also consider a block consisting of T samples. The equations for such a block can be written as follows:

$$
\begin{aligned}
\boldsymbol{x}(t) &= \boldsymbol{A}_0 \boldsymbol{s}(t) + \cdots + \boldsymbol{A}_{K-1}\boldsymbol{s}(t - K + 1) \\
\boldsymbol{x}(t-1) &= \boldsymbol{A}_0 \boldsymbol{s}(t-1) + \cdots + \boldsymbol{A}_{K-1}\boldsymbol{s}(t - K) \\
\boldsymbol{x}(t-2) &= \boldsymbol{A}_0 \boldsymbol{s}(t-2) + \cdots + \boldsymbol{A}_{K-1}\boldsymbol{s}(t - K - 1) \\
&\vdots
\end{aligned}
$$

The M-dimensional output sequence can be written as an $MT \times 1$ vector:

$$\widehat{\boldsymbol{x}}(t) = \left[\boldsymbol{x}^T(t), \boldsymbol{x}^T(t-1), \cdots, \boldsymbol{x}^T(t-T+1) \right]^T, \tag{10}$$

where $\boldsymbol{x}^T(t) = [x_1(t), \cdots, x_N(t)]$. Similarly, the N-dimensional input sequence can be written as an $N(T + K - 1) \times 1$ vector:

$$\widehat{\boldsymbol{s}}(t) = \left[\boldsymbol{s}^T(t), \boldsymbol{s}^T(t-1), \cdots, \boldsymbol{s}^T(t-T-K+2) \right]^T \tag{11}$$

From this the convolutive mixture can be expressed formally as:

$$\widehat{\boldsymbol{x}}(t) = \widehat{\boldsymbol{A}}\widehat{\boldsymbol{s}}(t) + \widehat{\boldsymbol{v}}(t), \tag{12}$$

where $\widehat{\boldsymbol{A}}$ has the following form:

$$\widehat{\boldsymbol{A}} = \begin{bmatrix} \boldsymbol{A}_0 & \cdots & \boldsymbol{A}_{K-1} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \ddots & \ddots & \ddots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{A}_0 & \cdots & \boldsymbol{A}_{K-1} \end{bmatrix}. \tag{13}$$

The block-Toeplitz matrix $\widehat{\boldsymbol{A}}$ has dimensions $MT \times N(T + K - 1)$. On the surface, Eq. (12) has the same structure as an instantaneous mixture given in Eq. (4), and the dimensionality has increased by a factor T. However, the models differ considerably as the elements within $\widehat{\boldsymbol{A}}$ and $\widehat{\boldsymbol{s}}(t)$ are now coupled in a rather specific way.

The majority of the work in convolutive source separation assumes a mixing model with a finite impulse response (FIR) as in Eq. (2). A notable exception is the work by Cichocki which considers also an auto-regressive (AR) component as part of the mixing model [18]. The ARMA mixing system proposed there is equivalent to a first-order Kalman filter with an infinite impulse response (IIR).

3. THE SEPARATION MODEL

The objective of blind source separation is to find an estimate, $\boldsymbol{y}(t)$, which is a model of the original source signals $\boldsymbol{s}(t)$. For this, it may not be necessary to identify the mixing filters $\boldsymbol{A}_k$ explicitly. Instead, it is often sufficient to estimate separation filters $\boldsymbol{W}_l$ that remove the cross-talk introduced by the mixing process. These separation filters may have a feed-back structure with an infinite impulse response (IIR), or may have a finite impulse response (FIR) expressed as feed-forward structure.

3.1. Feed-forward Structure

An FIR separation system is given by

$$y_n(t) = \sum_{m=1}^{M} \sum_{l=0}^{L-1} w_{nml} x_m(t-l) \tag{14}$$

or in matrix form

$$\boldsymbol{y}(t) = \sum_{l=0}^{L-1} \boldsymbol{W}_l \boldsymbol{x}(t-l). \tag{15}$$

As with the mixing process, the separation system can be expressed in the z-domain as

$$\boldsymbol{Y}(z) = \boldsymbol{W}(z)\boldsymbol{X}(z), \tag{16}$$

and it can also be expressed in block Toeplitz form with the corresponding definitions for $\widehat{\boldsymbol{y}}(t)$ and $\widehat{\boldsymbol{W}}$ [25]:

$$\widehat{\boldsymbol{y}}(t) = \widehat{\boldsymbol{W}}\widehat{\boldsymbol{x}}(t). \tag{17}$$

Table 1 summarizes the mixing and separation equations in the different domains.

3.2. Relation between source and separated signals

The goal in source separation is not necessarily to recover identical copies of the original sources. Instead, the aim is to recover model sources without interferences from other sources, i.e., each separated signal $y_n(t)$ should contain signals originating from a single source only (see Figure 3). Therefore, each model source signal can be a filtered version of the original source signals, i.e.:

$$\boldsymbol{Y}(z) = \boldsymbol{W}(z)\boldsymbol{A}(z)\boldsymbol{S}(z) = \boldsymbol{G}(z)\boldsymbol{S}(z). \tag{18}$$

This is illustrated in Figure 2. The criterion for separation, i.e., interference-free signals, is satisfied if the recovered signals are permuted, and possibly scaled and filtered versions of the original signals, i.e.:

$$\boldsymbol{G}(z) = \boldsymbol{P}\boldsymbol{\Lambda}(z), \tag{19}$$

where $\boldsymbol{P}$ is a permutation matrix, and $\boldsymbol{\Lambda}(z)$ is a diagonal matrix with scaling filters on its diagonal. If one can identify $\boldsymbol{A}(z)$ exactly, and choose $\boldsymbol{W}(z)$ to be its (stable) inverse, then $\boldsymbol{\Lambda}(z)$ is an identity matrix, and one recovers the sources exactly. In source separation, instead, one is satisfied with convolved versions of the sources, i.e. arbitrary diagonal $\boldsymbol{\Lambda}(z)$.

Table 1: The convolutive mixing equation and its corresponding separation equation are shown for different domains in which blind source separation algorithms have been derived.

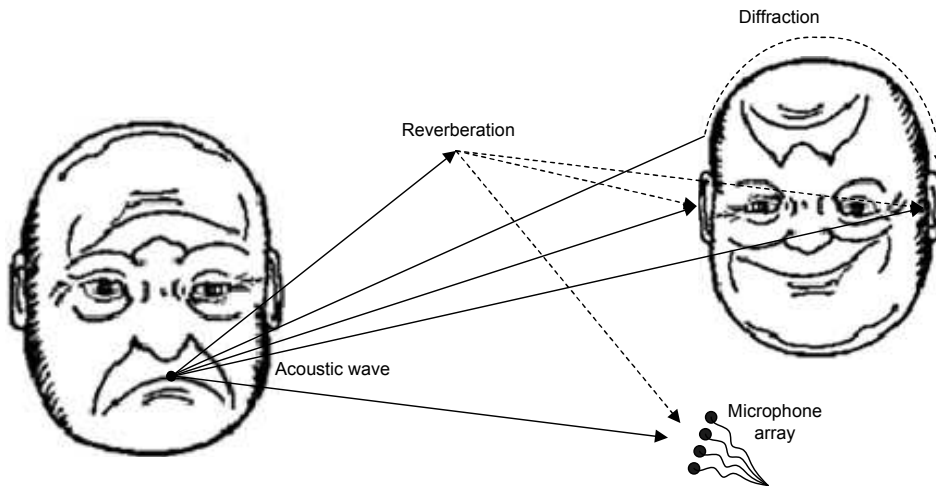|	**Mixing Process**	**Separation Model**
Time	$x_m(t) = \sum_{n=1}^{N} \sum_{k=0}^{K-1} a_{mnk} s_n(t-k) + v_m(t)$	$y_n(t) = \sum_{m=1}^{M} \sum_{l=0}^{L-1} w_{nml} x_m(t-l)$
	$\boldsymbol{x}(t) = \sum_{k=0}^{K-1} \boldsymbol{A}_k \boldsymbol{s}(t-k) + \boldsymbol{v}(t)$	$\boldsymbol{y}(t) = \sum_{l=0}^{L-1} \boldsymbol{W}_l \boldsymbol{x}(t-l)$
z-domain	$\boldsymbol{X}(z) = \boldsymbol{A}(z)\boldsymbol{S}(z) + \boldsymbol{V}(z),$	$\boldsymbol{Y}(z) = \boldsymbol{W}(z)\boldsymbol{X}(z)$
Frequency domain	$\boldsymbol{X}(\omega) = \boldsymbol{A}(\omega)\boldsymbol{S}(\omega) + \boldsymbol{V}(\omega)$	$\boldsymbol{Y}(\omega) = \boldsymbol{W}(\omega)\boldsymbol{X}(\omega)$
Block Toeplitz Form	$\widehat{\boldsymbol{x}}(t) = \widehat{\boldsymbol{A}}\widehat{\boldsymbol{s}}(t)$	$\widehat{\boldsymbol{y}}(t) = \widehat{\boldsymbol{W}}\widehat{\boldsymbol{x}}(t)$

Figure 3: Illustration of a speech source. It is not always clear what the desired acoustic source should be. It could be the acoustic wave as emitted from the mouth. This corresponds to the signal as it would have been recorded in an anechoic chamber in the absence of reverberations. It could be the individual source as it is picked up by a microphone array. Or it could be the speech signal as it is recorded on microphones close to the two eardrums of a person. Due to reverberations and diffraction, the recorded speech signal is most likely a filtered version of the signal at the mouth. NOTE TO PUBLISHER: THIS FIGURE IS A PLACE HOLDER ONLY. IT WILL REQUIRE MODIFICATION BY YOUR PRODUCTION DEPARTMENT. THE FACES ARE TO BE REPLACED WITH ANY REASONABLE REPRESENTATION OF A "SOURCE" AND "RECEIVER" OF A SPEECH SIGNAL.

3.3. Feedback Structure

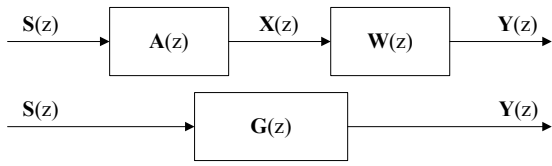The mixing system given by (2) is called a feedforward system. Often such FIR filters are inverted by a feedback structure using IIR filters. The esti-

Figure 2: The source signals $\boldsymbol{Y}(z)$ are mixed with the mixing filter $\boldsymbol{A}(z)$. An estimate of the source signals is obtained through an unmixing process, where the received signals $\boldsymbol{X}(z)$ are unmixed with the filter $\boldsymbol{W}(z)$. Each estimated source signal is then a filtered version of the original source, i.e., $\boldsymbol{G}(z) = \boldsymbol{W}(z)\boldsymbol{A}(z)$. Note that the mixing and the unmixing filters do not necessarily have to be of the same order.
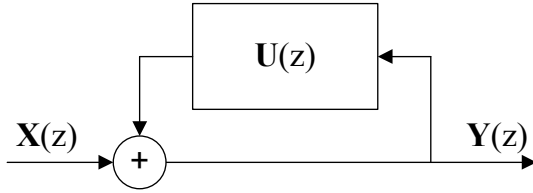

Figure 4: Recurrent unmixing (feedback) network given by equation (21). The received signals are separated by a IIR filter to achieve an estimate of the source signals.

mated sources are then given by the following equation, where the number of sources equals the number of receivers:

$$y_n(t) = x_n(t) + \sum_{l=0}^{L-1} \sum_{m=1}^{M} u_{nml} y_m(t-l), \quad (20)$$

and u_{nml} are the IIR filter coefficients. This can also be written in matrix form

$$\boldsymbol{y}(t) = \boldsymbol{x}(t) + \sum_{l=0}^{L-1} \boldsymbol{U}(l)\boldsymbol{y}(t-l). \quad (21)$$

The architecture of such a network is shown in Figure 4. In the z-domain, (21) can be written as [26]

$$\boldsymbol{Y}(z) = (\boldsymbol{I} + \boldsymbol{U}(z))^{-1}\boldsymbol{X}(z), \quad (22)$$

provided $(\boldsymbol{I}+\boldsymbol{U}(z))^{-1}$ exists and all poles are within the unit circle. Therefore,

$$\boldsymbol{W}(z) = (\boldsymbol{I} + \boldsymbol{U}(z))^{-1}. \quad (23)$$

The feed-forward and the feedback network can be combined to a so-called hybrid network, where a feed-forward structure is followed by a feedback network [27, 28].

3.4. Example: The TITO system

A special case, which is often used in source separation work is the two-input-two-output (TITO) system [29]. It can be used to illustrate the relationship between the mixing and unmixing system, feed-forward and feed-back structures, and the difference between recovering sources versus generating separated signals.

Figure 5 shows a diagram of a TITO mixing and unmixing system. The signals recorded at the two microphones are described by the following equations:

$$x_1(z) = s_1(z) + a_{12}(z)s_2(z) \quad (24)$$
$$x_2(z) = s_2(z) + a_{21}(z)s_1(z). \quad (25)$$

The mixing system is thus given by

$$\mathbf{A}(z) = \begin{bmatrix} 1 & a_{12}(z) \\ a_{21}(z) & 1 \end{bmatrix}, \quad (26)$$

which has the following inverse

$$[\mathbf{A}(z)]^{-1} = \frac{1}{1 - a_{12}(z)a_{21}(z)} \begin{bmatrix} 1 & -a_{12}(z) \\ -a_{21}(z) & 1 \end{bmatrix}. \quad (27)$$

If the two mixing filters $a_{12}(z)$ and $a_{21}(z)$ can be identified or estimated as $\bar{a}_{12}(z)$ and $\bar{a}_{21}(z)$, the separation system can be implemented as

$$y_1(z) = x_1(z) - \bar{a}_{12}(z)x_2(z) \quad (28)$$
$$y_2(z) = x_2(z) - \bar{a}_{21}(z)x_1(z). \quad (29)$$

A sufficient FIR separating filter is

$$\mathbf{W}(z) = \begin{bmatrix} 1 & -a_{12}(z) \\ -a_{21}(z) & 1 \end{bmatrix} \quad (30)$$

However, the exact sources are not recovered until this model sources $\boldsymbol{y}(t)$ are filtered with the IIR filter
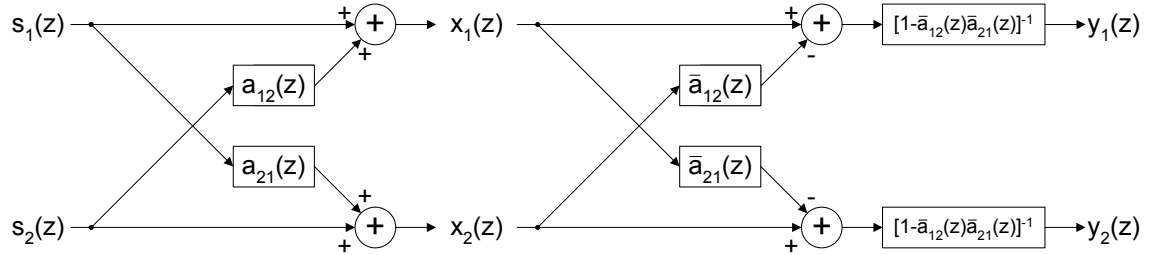
Figure 5: The two mixed sources s_1 and s_2 are mixed by a FIR mixing system. The system can be inverted by an alternative system, if the estimates $\bar{a}_{12}(z)$ and $\bar{a}_{21}(z)$ of the mixing filters $a_{12}(z)$ and $a_{12}(z)$ are known. Further, if the filter $[1 - \bar{a}_{12}(z)\bar{a}_{21}(z)]^{-1}$ is stable, the sources can be perfectly reconstructed as they were recorded at the microphones.

$[1 - \bar{a}_{12}(z)\bar{a}_{21}(z)]^{-1}$. Thus, the mixing process is invertible, provided this inverse IIR filter is stable. If a filtered version of the separated signals is acceptable, we may disregard the potentially unstable recursive filter in (27) and limit separation to the FIR inversion of the mixing system with (30).

4. IDENTIFICATION

Blind identification deals with the problem of estimating the coefficients in the mixing process $\boldsymbol{A}_k$. In general, this is an ill-posed problem, and no unique solution exists. In order to determine the conditions under which the system is blindly identifiable, assumptions about the mixing process and the input data are necessary. Even though the mixing parameters are known, it does not imply that the sources can be recovered. Blind identification of the sources refers to the exact recovery of sources. Therefore one should distinguish between the conditions required to identify the mixing system and the conditions necessary to identify the sources. The limitations for the exact recovery of sources when the mixing filters are known are discussed in [30, 11, 31]. For a recent review on identification of acoustic systems see [32]. This review considers single and multiple input-output systems for the case of completely known sources as well as blind identification, where both the sources and the mixing channels are unknown.

5. SEPARATION PRINCIPLE

Blind source separation algorithms are based on different assumptions on the sources and the mixing system. In general, the sources are assumed to be *independent* or at least decorrelated. The separation criteria can be divided into methods based on higher order statistics (HOS), and methods based on second order statistics (SOS). In convolutive separation it is also assumed that sensors receive N linearly independent versions of the sources. This means that the sources should originate from different locations in space (or at least emit signals into different orientations) and that there are at least as many sources as sensors for separation, i.e., $M \geq N$.

Instead of spatial diversity a series of algorithms make strong assumptions on the statistics of the sources. For instance, they may require that sources do not overlap in the time-frequency domain, utilizing therefore a form of *sparseness* in the data. Similarly, some algorithms for acoustic mixtures exploit regularity in the sources such as common onset, harmonic structure, etc. These methods are motivated by the present understanding on the grouping principles of auditory perception commonly referred to as "Auditory Scene Analysis". In radio communications a reasonable assumption on the sources is cyclo-stationarity (see Section 5.2.3) or the fact that source signals take on only discrete values. By using such strong assumptions on the source statistics it is sometimes possible to relax the conditions on the number of sensors, e.g. $M < N$. The different

Table 2: Assumptions made for separation

$N < M$	$N = M$	$N > M$
• Subspace methods [25].	• Asymmetric sources by 2nd and 3rd order cumulants [33]	• Non-stationary, column-wise co-prime sources [34]
• Reduction of problem to instantaneous mixture [35, 36, 37, 25, 38, 39, 40]	• Separation criteria based on SOS and HOS for 2×2 system [41]	• Cross-cumulants [42, 43]
	• Uncorrelated sources with distinct power spectra [44].	• Sparseness in time and frequency [45, 46, 47]
	• 2×2, temporally colored sources [48] • Cumulants of order > 2, ML principle [49]. • Known cross filters [41] • 2×2, each with different correlation [50, 51], extended to $M \times M$ in [52] • Non-linear odd functions [53, 26, 54, 55, 56, 57, 58] • Non-linearity approximating the cdf see e.g. [59]	

criteria for separation are summarized in Table 5.

5.1. Higher Order Statistics

Source separation based on higher order statistics is based on the assumption that the sources are statistically independent. Many algorithms are based on minimizing second and fourth order dependence between the model signals. A way to express independence is that all the cross-moments between the model sources are zero, i.e.:

$$E[y_n(t)^\alpha, y_{n'}(t - \tau)^\beta] = 0 \quad , \\ n \neq n', \alpha, \beta = \{1, 2, \ldots\}, \forall \tau,$$

where $E[\cdot]$ denotes the statistical expectation. Successful separation using higher order moments requires that the underlying sources are non-Gaussian (with the exception of at most one), since Gaussian sources have zero higher cumulants [60] and therefore equations (31) are trivially satisfied without providing useful conditions.

5.1.1. 4th-order statistic

It is not necessary to minimize all cross-moments in order to achieve separation. Many algorithms are based on minimization of second and fourth order dependence between the model source signals. This minimization can either be based on second and

fourth order cross-moments or second and fourth order cross-cumulants. Whereas off-diagonal elements of cross-cumulants vanish for independent signals the same is not true for all cross-moments [61]. Source separation based on cumulants has been used by several authors. Separation of convolutive mixtures by means of fourth order cumulants has been addressed by [62, 63, 41, 64, 65, 66, 67, 68, 61, 69, 70, 71]. In [72, 73, 74], the JADE algorithm for complex-valued signals [75] was applied in the frequency domain in order to separate convolved source signals. Other cumulant-based algorithms in the frequency domain are given in [76, 77]. Second and third order cumulants have been used by Ye et al. (2003) [33] for separation of asymmetric signals. Other algorithms based on higher order cumulants can be found in [78, 79]. For separation of more sources than sensors, cumulant-based approaches have been proposed in [80, 70]. Another popular 4th-order measure of non-Gaussianity is *kurtosis*. Separation of convolutive sources based on kurtosis has been addressed in [81, 82, 83].

5.1.2. Non-linear cross-moments

Some algorithms apply higher order statistics for separation of convolutive sources indirectly using non-linear functions by requiring:

$$E[f(y_n(t)), g(y_{n'}(t - \tau))] = 0. \qquad (31)$$

Here $f(\cdot)$ and $g(\cdot)$ are odd non-linear functions. The Taylor expansion of these functions captures higher order moments and this is found sufficient for separation of convolutive mixtures. This approach was among of the first for separation of convolutive mixtures [53] extending an instantaneous blind separation algorithm by Herault and Jutten (H-J) [84]. In Back and Tsoi (1994) [85], the H-J algorithm was applied in the frequency domain, and this approach was further developed in [86]. In the time domain, the approach of using non-linear odd functions has been used by Nguyen Thi and Jutten (1995) [26]. They present a group of TITO (2×2) algorithms based on 4th order cumulants, non-linear odd functions, and second and fourth order cross-moments. This algorithm has been further examined by Serviere (1996) [54], and it has also been used by Ypma et al. (2002) [55]. In Cruces and Castedo (1998) [87] a separation algorithm can be found, which can be regarded as a generalization of previous results from [26, 88]. In Li and Sejnowski (1995) [89], the H-J algorithm has been used to determine the delays in a beamformer. The H-J algorithm has been investigated further by Charkani and Deville (1997,1999) [90, 57, 58]. They extended the algorithm further to colored sources [56, 91]. Depending on the distribution of the source signals, also optimal choices of non-linear functions were found. For these algorithms, the mixing process is assumed to be minimum-phase, since the H-J algorithm is implemented as a feedback network. A natural gradient algorithm based on the H-J network has been applied in Choi et al. (2002) [92]. A discussion of the H-J algorithm for convolutive mixtures can be found in Berthommier and Choi (2003) [93]. For separation of two speech signals with two microphones, the H-J model fails if the two speakers are located on the same side, as the appropriate separating filters can not be implemented without delaying one of the sources and the FIR filters are constrained to be causal. HOS independence obtained by applying antisymmetric non-linear functions has also been used in [94, 95].

5.1.3. *Information Theoretic*

Statistical independence between the source signals can also be expressed in terms of the probability density functions (PDF). If the model sources $\boldsymbol{y}$ are independent, the joint probability density function can be written as

$$p(\boldsymbol{y}) = \prod_n p(y_n). \qquad (32)$$

This is equivalent to stating that model sources y_n do not carry mutual information. Information theoretic methods for source separation are based on maximizing the entropy in each variable. Maximum entropy is obtained when the sum of the entropy of each variable y_n equals the total joint-entropy in $\boldsymbol{y}$. In this limit variables do not carry any mutual information and are hence mutually independent [96]. A well-known algorithm based on this idea is the Infomax algorithm by Bell and Sejnowski (1995) [97] which was significantly improved in convergence speed by the natural gradient method of Amari [98]. The Infomax algorithm can also be derived directly from model equation (32) using Maximum Likelihood [99], or equivalently, using the Kullback-Leibler divergence between the empirical distribution and the independence model [100].

In all instances it is necessary to assume or model the probability density function $p_s(s_n)$ of the underlying sources s_n. In doing so, one captures higher order statistics of the data. In fact, most information theoretic algorithms contain expressions rather similar to the non-linear cross-statistics in (31) with $f(y_n) = \partial \ln p_s(y_n)/\partial y_n$, and $g(y_n) = y_n$. The PDF is either assumed to have a specific form or it is estimated directly from the recorded data, leading to *parametric* and *non-parametric* methods respectively [16]. In non-parametric methods the PDF is captured implicitly through the available data. Such methods have been addressed in [101, 102, 103]. However, the vast majority of convolutive algorithms have been derived based on explicit parametric representations of the PDF.

Infomax, the most common parametric method, was extended to the case of convolutive mixtures by Torkkola (1996) [59] and later by Xi and Reilly (1997,1999) [104, 105]. Both feed-forward and feedback networks were shown. In the frequency domain it is necessary to define the PDF for complex variables. The resulting analytic non-linear functions can be derived with [106, 107]

$$\boldsymbol{f}(Y) = -\frac{\partial \ln p(|Y|)}{\partial |Y|} e^{j \arg(Y)}, \qquad (33)$$

where $p(Y)$ is the probability density of the model source $Y \in \mathbb{C}$. Some algorithms assume circular sources in the complex domain, while other algorithms have been proposed that specifically assume non-circular sources [108, 109].

The performance of the algorithm depends to a certain degree on the selected PDF. It is important to determine if the data has super-Gaussian or sub-Gaussian distributions. For speech commonly a Laplace distribution is used. The non-linearity is also known as the Bussgang non-linearity [110]. A connection between the Bussgang blind equalization algorithms and the Infomax algorithm is given in Lambert and Bell (1997) [111]. Multichannel blind deconvolution algorithms derived from the Bussgang approach can be found in [112, 23, 111]. These learning rules are similar to those derived in Lee et al. (1997) [113].

Choi et al. (1999) [114] have proposed a *non-holonomic* constraint for multichannel blind deconvolution. Non-holonomic means that there are some restrictions related to the direction of the update. The non-holonomic constraint has been applied for both a feed-forward and a feedback network. The non-holonomic constraint was applied to allow the natural gradient algorithm by Amari et al. (1997) [98] to cope with over-determined mixtures. The non-holonomic constraint has also been used in [115, 116, 117, 118, 119, 120, 121, 122]. Some drawbacks in terms of stability and convergence in particular when there are large power fluctuations within each signal (e.g. for speech) have been addressed in [115].

Many algorithms have been derived from (32) directly using Maximum Likelihood (ML) [123]. The ML approach has been applied in [124, 125, 126, 127, 128, 129, 99, 130, 131, 132]. A method closely related to the ML is the Maximum a Posteriori (MAP) methods. In MAP methods, prior information about the parameters of the model are taken into account. MAP has been used in [23, 133, 134, 135, 136, 137, 138, 139, 140, 141].

The convolutive blind source separation problem has also been expressed in a Bayesian formulation [142]. The advantage of a Bayesian formulation is that one can derive an optimal, possibly non-linear estimator of the sources enabling the estimation of more sources than the number of available sensors. The Bayesian framework has also been applied in

[143, 144, 145, 135, 137].

A strong prior on the signal can also be realized via Hidden Markov Models (HMMs). HMMs can incorporate state transition probabilities of different sounds [136]. A disadvantage of HMMs is that they require prior training and they carry a high computational cost [146]. HMMs have also been used in [147, 148].

5.2. Second Order Statistics

In some cases, separation can be based on second order statistics (SOS) by requiring only non-correlated sources rather then the stronger condition of independence. Instead of assumptions on higher order statistics these methods make alternate assumptions such as the non-stationarity of the sources [149], or a minimum phase mixing system [50]. By itself, however, second order conditions are not sufficient for separation. Sufficient conditions for separation are given in [150, 15]. The main advantage of SOS is that they are less sensitive to noise and outliers [13], and hence require less data for their estimation [50, 150, 151, 34, 152]. The resulting algorithms are often also easier to implement and computationally efficient.

5.2.1. Minimum-phase mixing

Early work by Gerven and Compernolle [88] had shown that two source signals can be separated by decorrelation if the mixing system is minimum phase. The FIR coupling filters have to be strictly causal and their inverses stable. The condition for stability is given as $|a_{12}(z)a_{21}(z)| < 1$, where $a_{12}(z)$ and $a_{21}(z)$ are the two coupling filters (see Figure 5). These conditions are not met if the mixing process is non-minimum phase [153]. Algorithms based on second order statistic assuming minimum-phase mixing can be found in [154, 38, 39, 51, 50, 155, 156, 52, 157, 158].

5.2.2. Non-stationarity

The fact that many signals are non-stationary has been successfully used for source separation. Speech signals in particular can be considered non-stationary on time scales beyond 10 ms [159, 160]).

The temporally varying statistics of non-stationarity sources provides additional information for separation. Changing locations of the sources, on the other hand, generally complicate source separation as the mixing channel changes in time. Separation based on decorrelation of non-stationary signals was proposed by Weinstein et al. (1993) [29] who suggested that minimizing cross-powers estimated during different stationarity times should give sufficient conditions for separation. Wu and Principe (1999) proposed a corresponding joint diagonalization algorithm [103, 161] extending an earlier method developed for instantaneous mixtures [162]. Kawamoto et al. (1998) extend an earlier method [163] for instantaneous mixtures to the case of convolutive mixtures in the time domain [164, 153] and frequency domain [165]. This approach has also been employed in [166, 167, 168, 169] and an adaptive algorithm was suggested by Aichner et al. (2003) [170]. By combining this approach with a constraint based on whiteness, the performance can be further improved [171].

Note that not all of these papers have used simultaneous decorrelation, yet, to provide sufficient second-order constraints it is necessary to minimize multiple cross-correlations simultaneously. An effective frequency domain algorithm for simultaneous diagonalization was proposed by Parra and Spence (2000) [149]. Second-order statistics in the frequency domain is captured by the cross-power spectrum,

$$\boldsymbol{R}_{yy}(\omega,t) = E\left[\boldsymbol{Y}(\omega,t)\boldsymbol{Y}^H(\omega,t)\right] \quad (34)$$

$$= \boldsymbol{W}(\omega)\boldsymbol{R}_{xx}(\omega,t)\boldsymbol{W}^H(\omega), \quad (35)$$

where the expectations are estimated around some time t. The goal is to minimize the cross-powers on the off-diagonal of this matrix, e.g. by minimizing:

$$J = \sum_{t,\omega} \|\boldsymbol{R}_{yy}(\omega,t) - \boldsymbol{\Lambda}_y(\omega,t)\|^2, \quad (36)$$

where $\boldsymbol{\Lambda}_y(\omega,t)$ is an estimate of the cross-power spectrum of the model sources and is assumed to be diagonal. This cost function simultaneously captures multiple times and multiple frequencies, and has to be minimized with respect to $\boldsymbol{W}(\omega)$ and $\boldsymbol{\Lambda}_y(\omega,t)$ subject to some normalization constraint. If the source signals are non-stationary the cross-powers

estimated at different times t differ and provide independent conditions on the filters $\boldsymbol{W}(\omega)$. This algorithm has been successfully used on speech signals [172, 173] and investigated further by Ikram and Morgan (2000, 2001, 2002, 2005) [174, 175, 176] to determine the trade-offs between filter length, estimation accuracy, and stationarity times. Long filters are required to cope with long reverberation times of typical room acoustics, and increasing filter length also reduces the error of using the circular convolution in (35) (see Section 6.3). However, long filters increase the number of parameters to be estimated and extend the effective window of time required for estimating cross-powers thereby potentially loosing the benefit of non-stationarity of speech signals. A number of variations of this algorithm have been proposed subsequently, including time domain implementations [177, 178, 179], and other method that incorporate additional assumptions [180, 174, 181, 182, 183, 184, 185, 186, 187]. A recursive version of the algorithm was given in Ding et al. (2003) [188]. In Robeldo-Arnuncio and Juang (2005) [189], a version with non-causal separation filters was suggested. Based on a different way to express (35), Wang et al. (2003, 2004, 2005) [190, 191, 148, 192] propose a slightly different separation criterion, that leads to a faster convergence than the original algorithm by Parra and Spence (2000) [149].

Other methods that exploit non-stationarity have been derived by extending the algorithm of Molgedey and Schuster (1994) [193] to the convolutive case [194, 195] including a common two step approach of 'sphering' and rotation [159, 196, 197, 198, 199]. (Any matrix, for instance matrix $\boldsymbol{W}$, can be represented as a concatenation of a rotation with subsequent scaling (which can be used to remove second-order moments, i.e. sphering) and an additional rotation).

In Yin and Sommen (1999) [160] a source separation algorithm was presented based on non-stationarity and a model of the direct path. The reverberant signal paths are considered as noise. A time domain decorrelation algorithm based on different cross-correlations at different time lags is given in Ahmed et al. (1999) [200]. In Yin and Sommen (2000) [201] the cost function is based on minimization of the power spectral density between the

source estimates. The model is simplified by assuming that the acoustic transfer function between the source and closely spaced microphones is similar. The simplified model requires fewer computations. An algorithm based on joint diagonalization is suggested in Rahbar and Reilly (2003, 2005) [152, 152]. This approach exploits the spectral correlation between the adjacent frequency bins in addition to non-stationarity. Also in [202, 203] a diagonalization criterion based on non-stationarity has been used.

In Olsson and Hansen (2004) [139, 138] the non-stationary assumption has been included in a state-space Kalman filter model.

In Buchner et al. (2003) [204], an algorithm that uses a combination of non-stationarity, non-Gaussianity and non-whiteness has been suggested. This has also been applied in [205, 206, 207]. In the case of more source signals than sensors, an algorithm based on non-stationarity has also been suggested [70]. In this approach, it is possible to separate three signals: a mixture of two non-stationary source signals with short-time stationarity and one signal which is long-term stationary. Other algorithms based on the non-stationary assumptions can be found in [208, 209, 210, 211, 212, 213, 214].

5.2.3. Cyclo-stationarity

If a signal is assumed to be cyclo-stationary, the signals' cumulative distribution is invariant with respect to time shifts of some period T or any integer multiples of T. Further, a signal is said to be wide-sense cyclo-stationary if the signals mean and auto-correlation is invariant to shifts of some period T or any integer multiples of T [215], i.e.:

$$E[\boldsymbol{s}(t)] = E[\boldsymbol{s}(t + \alpha T)] \tag{37}$$
$$E[\boldsymbol{s}(t_1), \boldsymbol{s}(t_2)] = E[\boldsymbol{s}(t_1 + \alpha T), \boldsymbol{s}(t_2 + \alpha T)] \tag{38}$$

An example of a cyclo-stationary signal is a random amplitude sinusoidal signal. Many communication signals have the property of cyclo-stationarity, and voiced speech is sometimes considered approximately cyclo-stationary [216]. This property has been used explicitly to recover mixed source in e.g. [216, 217, 218, 55, 219, 220, 34, 118, 221, 222]. In [220] cyclo-stationarity is used to solve the frequency permutation problem (see Section 6.1) and in [118] it

is used as additional criteria to improve separation performance.

5.2.4. Non-whiteness

Many natural signals, in particular acoustic signals, are temporally correlated. Capturing this property can be beneficial for separation. For instance, capturing temporal correlations of the signals can be used to reduce a convolutive problem to an instantaneous mixture problem, which is then solved using additional properties of the signal [35, 25, 36, 37, 38, 39, 40]. In contrast to instantaneous separation where decorrelation may suffice for non-white signals, for convolutive separation additional conditions on the system or the sources are required. For instance, Mei and Yin (2004) [223] suggest that decorrelation is sufficient provided the sources are an ARMA process.

5.3. Sparseness in the Time/Frequency domain

Numerous source separation applications are limited by the number of available microphones. It is in not always guaranteed that the number of sources is less than or equal to the number of sensors. With linear filters it is in general not possible to remove more than $M - 1$ sources from the signal. By using non-linear techniques, in contrast, it may be possible to extract a larger number of source signals. One technique to separate more sources than sensors is based on sparseness. If the source signals do not overlap in the time-frequency (T-F) domain it is possible to separate them. A mask can be applied in the T-F domain to attenuate interfering signal energy while preserving T-F bins where the signal of interest is dominant. Often a binary mask is used giving perceptually satisfactory results even for partially overlapping sources [224, 225]. These methods work well for anechoic mixtures (delay-only) [226]. However, under reverberant conditions, the T-F representation of the signals is less sparse. In a mildly reverberant environment ($T_{60} \leq 200$ ms) under-determined sources have been separated with a combination of independent component analysis (ICA) and T-F masking [47]. The first $N - M$ signals are removed from the mixtures by applying a T-F mask estimated from the direction of arrival of the signal (cf. Section 7.1). The

remaining M sources are separated by conventional BSS techniques. When a binary mask is applied to a signal, artifacts (musical noise) are often introduced. In order to reduce the musical noise, smooth masks have been proposed [227, 47].

Sparseness has also been used as a post processing step. In [77], a binary mask has been applied as post-processing to a standard BSS algorithm. The mask is determined by comparison of the magnitude of the outputs of the BSS algorithm. Hereby a higher signal to interference ratio is obtained. This method was further developed by Pedersen et al. (2005, 2006) in order to segregate under-determined mixtures [228, 229]. Because the T-F mask can be applied to a single microphone signal, the segregated signals can be maintained as e.g. stereo signals.

Most of the T-F masking methods do not effectively utilize information from more than two microphones because the T-F masks are applied to a single microphone signal. However, some methods have been proposed that utilize information from more than two microphones [225, 230].

Clustering has also been used for sparse source separation [231, 232, 233, 234, 140, 141, 235, 236, 230]. If the sources are projected into a space where each source groups together, the source separation problem can be solved with clustering algorithms. In [46, 45] the mask is determined by clustering with respect to amplitude and delay differences.

In particular when extracting sources from single channels sparseness becomes an essential criterion. Pearlmutter and Zador (2004) [237] use strong prior information on the source statistic in addition to knowledge of the head-related transfer functions (HRTF). An *a priori* dictionary of the source signals as perceived through a HRTF makes it possible to separate source signals with only a single microphone. In [238], *a priori* knowledge is used to construct basis functions for each source signals to segregate different musical signals from their mixture. Similarly, in [239, 240] sparseness has been assumed in order to extract different music instruments.

Techniques based on sparseness are further discussed in the survey by O'Grady et al. (2005) [21].

5.4. Priors from Auditory Scene Analysis and Psycho-Acoustics

Some methods rely on insights gained from studies of the auditory system. The work by Bergman [241] on auditory scene analysis characterized the cues used by humans to segregate sound sources. This has motivated computational algorithms that are referred to as computational auditory scene analysis (CASA). For instance, the phenomenon of auditory masking, i.e., the dominant perception of the signal with largest signal power has motivated the use of T-F masking for many years [242]. In addition to the direct T-F masking methods outlined above, separated sources have been enhanced by filtering based on perceptual masking and auditory hearing thresholds [191, 243].

Another important perceptual cue that has been used in source separation is pitch frequency, which typically differs for simultaneous speakers [135, 244, 245, 137, 138, 147]. In Tordini and Piazza (2000) [135] pitch is extracted from the signals and used in a Bayesian framework. During unvoiced speech, which lacks a well-defined pitch they use an ordinary blind algorithm. In order to separate two signals with one microphone, Gandhi and Hasegawa-Johnson (2004) [137] have proposed a state-space separation approach with strong *a priori* information. Both pitch and Mel-frequency cepstral coefficients (MFCC) were used in their method. A pitch codebook as well as an MFCC codebook have to be known in advance. Olsson and Hansen [138] have used a Hidden-Markov Model, where the sequence of possible states is limited by the pitch frequency that is extracted in the process. As a pre-processing step to source separation, Furukawa et al. (2003) [245] use pitch in order to determine the number of source signals.

A method for separation of more sources than sensors is given in Barros et al. (2002) [244]. They combined ICA with CASA techniques such as pitch tracking and auditory filtering. Auditory filter banks are used in order to model the cochlea. In [244] wavelet filtering has been used for auditory filtering. Another commonly used auditory filter bank is the Gammatone filter-bank (see e.g. Patterson (1994) [246] or [247, 248]). In Roman et al. (2003) [248] binaural cues have been used to segregate sound sources, whereby inter-aural time and inter-aural intensity differences (ITD, IID) have been used to

group the source signals.

6. TIME VERSUS FREQUENCY DOMAIN

The blind source separation problem can either be expressed in the time domain

$$\boldsymbol{y}(t) = \sum_{l=0}^{L-1} \boldsymbol{W}_l \boldsymbol{x}(t-l) \qquad (39)$$

or in the frequency domain

$$\boldsymbol{Y}(\omega, t) = \boldsymbol{W}(\omega)\boldsymbol{X}(\omega, t). \qquad (40)$$

A survey of frequency-domain BSS is provided in [22]. In Nishikawa et al. (2003) [249] the advantages and disadvantages of the time and frequency domain approaches have been compared. This is summarized in Table 3.

An advantage of blind source separation in the frequency domain is that the separation problem can be decomposed into smaller problems for each frequency bin in addition to the significant gains in computational efficiency. The convolutive mixture problem is reduced to "instantaneous" mixtures for each frequency. Although this simplifies the task of convolutive separation a set of new problems arise: The frequency domain signals obtained from the DFT are complex-valued. Not all instantaneous separation algorithms are designed for complex-valued signals. Consequently, it is necessary to modify existing algorithms correspondingly [250, 251, 252, 5]. Another problem that may arise in the frequency domain is that there are no longer enough data points available to evaluate statistical independence [131]. For some algorithms [149] it is necessary that the frame size T of the DFT is much longer than the length of the room impulse response K (see Section 6.3). Long frames result in fewer data samples per frequency [131], which complicates the estimation of the independence criteria. A method that copes with this issue has been proposed by Servière (2004) [253].

6.1. Frequency Permutations

Another problem that arises in the frequency domain is the permutation and scaling ambiguity. If the convolutive problem is treated for each frequency as a separate problem, the source signals in each frequency bin may be estimated with an arbitrary permutation and scaling, i.e.:

$$\boldsymbol{Y}(\omega, t) = \boldsymbol{P}(\omega)\boldsymbol{\Lambda}(\omega)\boldsymbol{S}(\omega, t). \qquad (41)$$

If the permutation $\boldsymbol{P}(\omega)$ is not consistent across frequency then converting the signal back to the time domain will combine contributions from different sources into a single channel, and thus annihilate the separation achieved in the frequency domain. An overview of the solutions to this permutation problem is given in Section 7. The scaling indeterminacy at each frequency – arbitrary solution for $\boldsymbol{\Lambda}(\omega)$ – will result in an overall filtering of the sources. Hence, even for perfect separation the separated sources may have a different frequency spectrum than the original sources.

6.2. Time-Frequency Algorithms

Algorithms that define a separation criteria in the time domain do typically not exhibit frequency permutation problems, even when computations are executed in the frequency domain. A number of authors have therefore used time-domain criteria combined with frequency domain implementations that speed up computations. [254, 113, 255, 256, 121, 101, 257, 179, 171]. However, note that second-order criteria may be susceptible to the permutation problem even if they are formulated in the time domain [184].

6.3. Circularity Problem

When the convolutive mixture in the time domain is expressed in the frequency domain by the DFT, the convolution becomes separate multiplications, i.e.:

$$\boldsymbol{x}(t) = \boldsymbol{A} * \boldsymbol{s}(t) \longleftrightarrow \boldsymbol{X}(\omega, t) \approx \boldsymbol{A}(\omega)\boldsymbol{S}(\omega, t). \qquad (42)$$

However, this is only an approximation which is exact only for periodic $\boldsymbol{s}(t)$ with period T, or equivalently, if the time convolution is *circular*:

$$\boldsymbol{x}(t) = \boldsymbol{A} \circledast \boldsymbol{s}(t) \longleftrightarrow \boldsymbol{X}(\omega) = \boldsymbol{A}(\omega)\boldsymbol{S}(\omega). \qquad (43)$$

For a *linear convolution* errors occur at the frame boundary, which are conventionally corrected with

Table 3: Advantages and disadvantages for separation in the time domain or separation in the frequency domain.

Time Domain		Frequency Domain	
Advantages	Disadvantages	Advantages	Disadvantages
• The independence assumption holds better for full-band signals	• Degradation of convergence in strong reverberant environment	• The convolutive mixture can be transformed into instantaneous mixture problems for each frequency bin	• For each frequency band, there is a permutation and a scaling ambiguity which needs to be solved
• Possible high convergence near the optimal point	• Many parameters need to be adjusted for each iteration step	• Due to the FFT, computations are saved compared to an implementation in the time domain	• Problem with too few samples in each frequency band may cause the independence assumption to fail
		• Convergence is faster	• Circular convolution deteriorates the separation performance.
			• Inversion of **W** is not guaranteed

the overlap-save method. However, a correct overlap-save algorithm is difficult to implement when computing cross-powers such as in (35) and typically the approximate expression (42) is assumed.

The problem of linear/circular convolution has been addressed by several authors [62, 149, 258, 171, 121]. Parra and Spence (2000) [149] note that the frequency domain approximation is satisfactory provided that the DFT length T is significantly larger than the length of the mixing channels. In order to reduce the errors due to the circular convolution, the filters should be at least two times the length of the mixing filters [131, 176].

To handle long impulse responses in the frequency domain, a frequency model which is equivalent to the time domain linear convolution has been proposed in [253]. When the time domain filter extends beyond the analysis window the frequency response is under-sampled [258, 22]. These errors can be mitigated by spectral smoothing or equivalently by windowing in the time domain. According to [259] the circularity problem becomes more severe when the number of sources increases.

Time domain algorithms are often derived using Toeplitz matrices. In order to decrease the complexity and improve computational speed, some calculations involving Toeplitz matrices are performed using the fast-Fourier transform. For that purpose, it is

necessary to express the Toeplitz matrices in circulant Toeplitz form [23, 260, 261, 195, 121, 171]. A method that avoids the circularity effects but maintains the computational efficiency of the FFT has been presented in [262]. Further discussion on the circularity problem can be found in [189].

6.4. Subband filtering

Instead of the conventional linear Fourier domain some authors have used subband processing. In [142] a long time-domain filter is replaced by a set of short independent subband-filters, which results in faster convergence as compared to the full-band methods [214]. Different filter lengths for each subband filter have also been proposed motivated by the varying reverberation time of different frequencies (typically low-frequencies have a longer reverberation time) [263].

7. THE PERMUTATION AMBIGUITY

The majority of algorithms operate in the frequency domain due to the gains in computational efficiency, which are important in particular for acoustic mixtures that require long filters. However, in frequency domain algorithms the challenge is to solve the permutation ambiguity, i.e., to make the permutation

matrix $\boldsymbol{P}(\omega)$ independent of frequency. Especially when the number of sources and sensors is large, recovering consistent permutations is a severe problem. With N model sources there are $N!$ possible permutations in each frequency bin. Many frequency domain algorithms provide *ad hoc* solutions, which solve the permutation ambiguity only partially, thus requiring a combination of different methods. Table 4 summarizes different approaches. They can be grouped into two categories

1. Consistency of the filter coefficients

2. Consistency of the spectrum of the recovered signals

The first exploits prior knowledge about the mixing filters, and the second uses prior knowledge about the sources. Within each group the methods differ in the way consistency across frequency is established, varying sometimes in the metric they use to measure *distance* between solutions at different frequencies.

7.1. Consistency of the Filter Coefficients

Different methods have been used to establish consistency of filter coefficients across frequency, such as constraints on the length of the filters, geometric information, or consistent initialization of the filter weights.

Consistency across frequency can be achieved by requiring continuity of filter values in the frequency domain. One may do this directly by comparing the filter values of neighboring frequencies after adaptation, and pick the permutation that minimize the Euclidean distance between neighboring frequencies [269, 74]. Continuity (in a discrete frequency domain) is also expressed as smoothness, which is equivalent with a limited temporal support of the filters in the time domain. The simplest way to implement such a smoothness constraint is by zero-padding the time domain filters prior to performing the frequency transformation [264]. Equivalently, one can restrict the frequency domain updates to have a limited support in the time domain. This method is explained in Parra et al. [149] and has been used extensively [283, 161, 269, 174, 190, 188, 201, 119, 122, 192]. Ikram and Morgan [174, 176] evaluated this constraint and point out that there is a trade-off

between the permutation alignment and the spectral resolution of the filters. Moreover, restricting the filter length may be problematic in reverberant environments where long separation filters are required. As a solution they have suggest to relax the constraint on filter length after the algorithm converges to satisfactory solutions [176].

Another suggestion is to assess continuity after accounting for the arbitrary scaling ambiguity. To do so, the separation matrix can be normalized as proposed in [265]:

$$\boldsymbol{W}(\omega) = \widetilde{\boldsymbol{W}}(\omega)\boldsymbol{\Lambda}(\omega), \qquad (44)$$

where $\boldsymbol{\Lambda}(\omega)$ is a diagonal matrix and $\widetilde{\boldsymbol{W}}(\omega)$ is a matrix with unit diagonal. The elements of $\widetilde{\boldsymbol{W}}(\omega)$, $\widetilde{W}_{mn}(\omega)$ are the ratios between the filters and these are used to assess continuity across frequencies [48, 220].

Instead of restricting the *unmixing* filters, Pham et al. (2003) [202] have suggested to require continuity in the *mixing* filters, which is reasonable as the mixing process will typically have a shorter time constant. A specific distance measure has been proposed by Asano et al. (2003) [284, 267]. They suggest to use the cosine between the filter coefficients of different frequencies ω_1 and ω_2:

$$\cos \alpha_n = \frac{\boldsymbol{a}_n^H(\omega_1)\boldsymbol{a}_n(\omega_2)}{\|\boldsymbol{a}_n^H(\omega_1)\|\|\boldsymbol{a}_n(\omega_2)\|}, \qquad (45)$$

where $\boldsymbol{a}_n(\omega)$ is the n'th column vector of $\boldsymbol{A}(\omega)$, which is estimated as the pseudo-inverse of $\boldsymbol{W}(\omega)$. Measuring distance in the space of separation filters rather than mixing filters was also suggested because these may better reflect the spacial configuration of the sources [285].

In fact, continuity across frequencies may also be assessed in terms of the estimated spatial locations of sources. Recall that the mixing filters are impulse responses between the source locations and the microphone locations. Therefore, the parameters of the separation filters should account for the position of the source in space. Hence, if information about the sensor location is available it can be used to address the permutation problem.

To understand this, consider the signal that arrives at an array of sensors. Assuming a distant

Table 4: Categorization of approaches to solve the permutation problem in the frequency domain.

Class	Metric	Reference
Consistency of the filter coefficients	Smooth spectrum	[264, 149]
	Source locations	[265]
	Directivity pattern	[266, 175, 73]
	Location vectors	[267]
	DOA	[184, 268, 72]
	Adjacent matrix distance	[269]
	Invariances	[48]
	Split spectrum	[270]
	Frequency link in update process	[127]
	Initialization	[250, 271]
	Moving sources	[167]
	Vision	[148]
Consistency of the spectrum of the recovered signals	Amplitude modulation	[159, 197, 272, 126, 203]
	Pitch	[135, 147]
	Psychoacoustics	[243, 243]
	Fundamental frequency	[244]
	Cyclo-stationarity	[273]
	Periodic signals	[221]
	Cross-correlation	[62, 274, 209]
	Cross-cumulants	[275]
	Kurtosis	[86]
	Source distribution	[276, 134]
	Multidimensional prior	[277, 278]
	Clustering	[230, 279]
Time-frequency information	FIR polynomial	[23, 254, 113, 255]
	TD cost function	[178]
	Apply ICA to whole spectrogram	[280]
Combined approaches		[106, 258, 281, 282]

source in an reverberation-free environment the signal approximates a plane-wave. If the plane-waves arrives at an angle to the microphone array it will impinge on each microphone with a certain delay (see Figure 6). This delay τ is given by the microphone distance d, the velocity of the wave c, and the direction-of-arrival (DOA) angle θ:

$$\tau = \frac{d}{c} \sin \theta, \qquad (46)$$

Filters that compensate for this delay can add the microphone signals constructively (or destructively) to produce a maximum (or minimum) response in the DOA. Hence, the precise delay in filters (which in the frequency domain correspond to precise phase relationships) establishes a relationship between different frequencies that can be used to identify correct permutations. This was first considered by Soon et al. (1993) [286].

To be specific, each row in the separation matrix $\boldsymbol{W}(\omega)$ defines a *directivity pattern*, and therefore each row can be thought of as a separate beamformer. This directivity pattern is determined by the transfer function between the source and the filter output. The magnitude response of the n-th output is given by

$$\boldsymbol{r}_n(\omega, \boldsymbol{\theta}) = |\boldsymbol{w}_n^H(\omega)\boldsymbol{a}(\omega, \boldsymbol{\theta})|^2, \qquad (47)$$

where $\boldsymbol{a}(\omega)$ is an $M \times 1$ vector representing the propagation of a distant source with DOA θ to the sensor array. When M sensors are available, it is possible to place $M - 1$ nulls in each of the M *directivity patterns*, i.e., directions from which the arriving signal is canceled out. In an ideal, reverberation-free environment separation is achieved if these nulls point to the directions of the interfering sources. The lo-
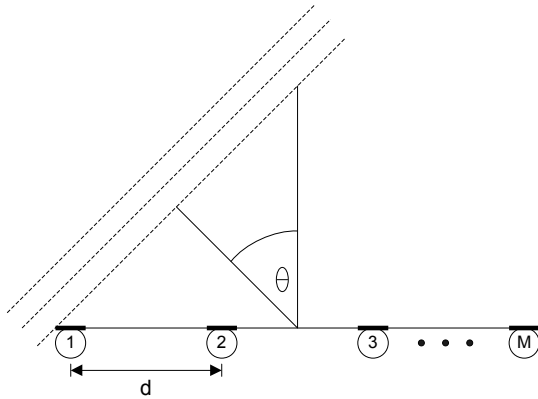
Figure 6: A sensor array consisting of M sensors linearly distributed with the distance d to the adjacent sensor. The sensors are placed in a free field. A source signal is considered coming from a point source of a distance r away from the sensor array. The source signal is placed in the far-field, i.e., $r \gg d$. Therefore the incident wave is planar and the arrival angle θ is the same for all the sensors.

cations of these nulls, as they may be identified by the separation algorithm, can be used to resolve the permutation ambiguity [266, 287, 288, 81, 77, 131, 289, 290]. These techniques draw strong parallels between source separation solutions and *beamforming*. The DOA's do not have to be known in advance and can instead be extracted from the resulting separation filters. Note, however, that the ability to identify source locations is limited by the physics of wave propagation and sampling: distant microphones will lead to grading lobes which will confuse the source locations, while small aperture limits spatial resolution at low frequencies. Ikram and Morgan (2002) [175] extend the idea of Kurita et al. (2000) [266] to the case where the sensor space is wider than one half of the wavelength. Source locations are estimated at lower frequencies, which do not exhibit grating lobes. These estimates are then used to determine the correct nulls for the higher frequencies and hereby the correct permutations. In order to resolve permutations when sources arrive from the same direction, Mukai et al. (2004) [291] use a near-field model. Mitianoudis and Davies (2004) [268] suggested frequency alignment based on DOA esti-

mated with the MuSIC algorithm [292]. A subspace method has been used in order to avoid constraints on the number of sensors. Knaak et al. (2003) [222] include DOA information as a part of the BSS algorithm in order to avoid the permutation. Although all these methods assume a reverberation-free environment they give reasonable results in reverberant environments as long as the source has a strong direct path to the sensors.

Two other methods also utilize geometry. In the case of moving sources, where only one source is moving, the permutation can be resolved by noting that only one of the parameters in the separation matrix changes [167]. If visual cues are available, they may also be used to solve the permutation ambiguity [148].

Instead of using geometric information as a separate step to solve the permutation problem Parra and Alvino (2002) include geometric information directly into the cost function [184, 185]. This approach has been applied to microphone arrays under reverberant conditions [187]. Baumann et al. (2003) [72] have also suggested a cost function, which includes the DOA estimation. The arrival angles of the signals are found iteratively and included in the separation criterion. Baumann et al. [73] also suggest a maximum likelihood approach to solve the permutation problem. Given the probability of a permuted or non-permuted solution as function of the estimated zero directions, the most likely permutation is found.

Gotanda et al. (2003) [270] have proposed a method to reduce the permutation problem based on the split spectral difference, and the assumption that each source is closer to one microphone. The split spectrum is obtained when each of the separated signals are filtered by the estimated mixing channels.

Finally, for iterative update algorithms a proper initialization of the separation filters can result in consistent permutations across frequencies. Smaragdis [250] proposed to estimate filter values sequentially starting with low frequencies and initializing filter values with the results of the previous lower frequency. This will tend to select solutions with filters that are smooth in the frequency domain, or equivalently, filters that are short in the time domain. Filter values may also be initialized to simple beamforming filters that point to estimated source locations. The separation algorithm will then tend

to converge to solutions with the same target source across all frequencies [184, 271].

7.2. Consistency of the Spectrum of the Recovered Signals

Some solutions to the permutation ambiguity are based on the properties of speech. Speech signals have strong correlations across frequency due to a common amplitude modulation.

At the coarsest level the power envelope of the speech signal changes depending on whether there is speech or silence, and within speech segments the power of the carrier signal induces correlations among the amplitude of different frequencies. A similar argument can be made for other natural sounds. Thus, it is fair to assumed that natural acoustic signals originating from the same source have a correlated amplitude envelope for neighboring frequencies. A method based on this co-modulation property was proposed by Murata et al. (1998) [159, 196]. The permutations are sorted to maximize the correlation between different envelopes. This is illustrated in Figure 7. This method has also been used in [293, 198, 199, 287, 263, 203]. Rahbar and Reilly (2001, 2005) [209, 152] suggest efficient methods for finding the correct permutations based on cross-frequency correlations.

Asano and Ikeda (2000) [294] report that the method sometimes fails if the envelopes of the different source signals are similar. They propose the following function to be maximized in order to estimate the permutation matrix:

$$\widehat{\boldsymbol{P}}(\omega) = \arg \max_{\boldsymbol{P}(\omega)} \sum_{t=1}^{T} \sum_{j=1}^{\omega-1} [\boldsymbol{P}(\omega)\bar{\boldsymbol{y}}(\omega,t)]^H \bar{\boldsymbol{y}}(j,t),$$

(48)

where $\bar{\boldsymbol{y}}$ is the power envelope of $\boldsymbol{y}$ and $\boldsymbol{P}(\omega)$ is the permutation matrix. This approach has also been adopted by Peterson and Kadambe (2003) [232]. Kamata et al. (2004) [282] report that the correlation between envelopes of different frequency channels may be small, if the frequencies are too far from each other. Anemüller and Gramms (1999) [127] avoid the permutations since the different frequencies are linked in the update process. This is done by serially switching from low to high frequency components while updating.
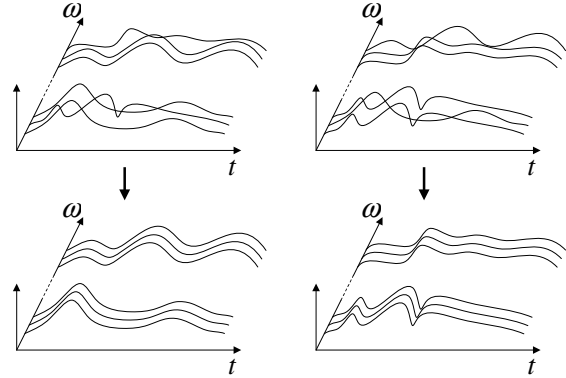

Figure 7: For speech signals, it is possible to estimate the permutation matrix by using information on the envelope of the speech signal (amplitude modulation). Each speech signal has a particular envelope. Therefore, by comparison with the envelopes of the nearby frequencies, it is possible to order the permuted signals.

Another solution based on amplitude correlation is the so-called Amplitude Modulation Decorrelation (AMDecor)–algorithm presented by Anemüller and Kollmeier (2000, 2001) [272, 126]. They propose to solve, the source separation problem and the permutation problems simultaneously. An amplitude modulation correlation is defined, where the correlation between the frequency channels ω_k and ω_l of the two spectrograms $\boldsymbol{Y}_a(\omega,t)$ and $\boldsymbol{Y}_b(\omega,t)$ is calculated as

$$c(\boldsymbol{Y}_a(\omega,t), \boldsymbol{Y}_b(\omega,t)) = \\ E[|\boldsymbol{Y}_a(\omega,t)||\boldsymbol{Y}_b(\omega,t)|] \\ -E[|\boldsymbol{Y}_a(\omega,t)|]E[|\boldsymbol{Y}_b(\omega,t)|].$$

(49)

This correlation can be computed for all combinations of frequencies. This results in a square matrix $\boldsymbol{C}(\boldsymbol{Y}_a, \boldsymbol{Y}_b)$ with sizes equal to the number of frequencies in the spectrogram, whose k, lth element is given by (49). Since the unmixed signals $\boldsymbol{y}(t)$ have to be independent, the following decorrelation property must be fulfilled

$$C_{kl}(\boldsymbol{Y}_a, \boldsymbol{Y}_b) = 0 \quad \forall a \neq b, \forall k, l. \qquad (50)$$

This principle also solves the permutation ambiguity. The source separation algorithm is then based on the

minimization of a cost function given by the Frobenius norm of the amplitude modulation correlation matrix.

A priori knowledge about the source distributions has also been used to determine the correct permutations. Based on assumptions of Laplacian distributed sources, Mitianopudis and Davies (2001, 2002) [251, 276, 134] propose a likelihood ratio test to test which permutation is most likely. A time-dependent function that imposes frequency coupling between frequency bins is also introduced. Based on the same principle, the method has been extended to more than two sources by Rahbar and Reilly (2003) [152]. A hierarchical sorting is used in order to avoid errors introduced at a single frequency. This approach has been adopted in Mertins and Russel (2003) [212].

Finally, one of the most effective convolutive BSS methods to-date (see Table 5) uses this statistical relationship of signal powers across frequencies. Rather than solving separate "instantaneous" source separation problems in each frequency band Kim et al. (2006) [295, 278, 277] propose a multidimensional version of the density estimation algorithms described in Section 5.1.3. The density function captures the power of the entire model source rather than the power at individual frequencies. As a result, the joint-statistics across frequencies are effectively captured and the algorithm converges to satisfactory permutations in each frequency.

Other properties of speech have also been suggested in order to solve the permutation indeterminacy. A *pitch*-based method has been suggested by Tordini and Piazza (2002) [135]. Also Sanei et al. (2004) [147] use the property of different pitch frequency for each speaker. The pitch and formants are modeled by a coupled hidden Markov model (HMM). The model is trained based on previous time frames.

Motivated by psycho-acoustics, Guddeti and Mulgrew (2005) [243] suggest to disregard frequency bands that are perceptually masked by other frequency bands. This simplifies the permutation problem as the number of frequency bins that have to be considered is reduced. In Barros et al. (2002) [244], the permutation ambiguity is avoided due to *a priori* information of the phase associated with the fundamental frequency of the desired speech signal.

Non-speech signals typically also have properties which can be exploited. Two proposals for solving the permutation in the case of cyclo-stationary signals can be found in Antoni et al. (2005) [273]. For machine acoustics, the permutations can be solved easily since machine signals are (quasi) periodic. This can be employed to find the right component in the output vector [221].

Continuity of the frequency spectra has been used by Capdevielle et al. (1995) [62] to solve the permutation ambiguity. The idea is to consider the sliding Fourier transform with a delay of one point. The cross correlation between different sources are zero due to the independence assumption. Hence, when the cross correlation is maximized, the output belongs to the same source. This method has also been used by Servière (2004) [253]. A disadvantage of this method is that it is computationally very expensive since the frequency spectrum has to be calculated with a window shift of one. A computationally less expensive method based on this principle has been suggested by Dapena and Servière (2001) [274]. The permutation is determined from the solution that maximizes the correlation between only two frequencies. If the sources have been whitened as part of separation, the approach by Capdevielle et al.(1995) [62] does not work. Instead, Kopriva et al. (2001) [86] suggest that the permutation can be solved by independence tests based on kurtosis. For the same reason, Mejuto et al. (2000) [275] consider fourth order cross-cumulants of the outputs at all frequencies. If the extracted sources belong to the same sources, the cross-cumulants will be non-zero. Otherwise, if the sources belong to different sources, the cross-cumulants will be zero.

Finally, Hoya et al. (2003) [296] use pattern recognition to identify speech pauses that are common across frequencies, and in the case of over-complete source separation, K-means clustering has been suggested. The clusters with the smallest variance are assumed to correspond to the desired sources [230]. Dubnov et al. (2004) [279] also address the case of more sources than sensors. Clustering is used at each frequency and Kalman tracking is performed in order to link the frequencies together.

7.3. Global permutations

In many applications only one of the source signals is desired and the other sources are only considered as interfering noise. Even though the local (frequency) permutations are solved, the global (external) permutation problem still exists. Only few algorithms address the problem of selecting the desired source signal from the available outputs. In some situations, it can be assumed that the desired signal arrives from a certain direction (e.g. the speaker of interest is in front of the array). Geometric information can determine which of the signals is the target [184, 171]. In other situations, the desired speaker is selected as the most dominant speaker. In Low et al. (2004) [289], the most dominant speaker is determined on a criterion based on kurtosis. The speaker with the highest kurtosis is assumed to be the dominant. In separation techniques based on clustering, the desired source is assumed to be the cluster with the smallest variance [230]. If the sources are moving it is necessary to maintain the global permutation by tracking each source. For block-based algorithm the global permutation might change at block-boundaries. This problem can often be solved by initializing the filter with the estimated filter from the previous block [186].

8. RESULTS

The overwhelming majority of convolutive source separation algorithms have been evaluated on simulated data. In the process, a variety of simulated room responses have been used. Unfortunately, it is not clear if any of these results transfer to real data. The main concerns are the sensitivity to microphone noise (often not better than -25 dB), non-linearity in the sensors, and strong reverberations with a possibly weak direct path. It is suggestive that only a small subset of research teams evaluate their algorithms on actual recordings. We have considered more than 400 references and found results on real room recordings in only 10% of the papers. Table 5 shows a complete list of those papers. The results are reported as signal-to-interference ratio (SIR), which is typically averaged over multiple output channels. The resulting SIR are not directly comparable as the results for a given algorithm are very likely to dependent on the recording equipment, the room that was used, and the

SIR in the recorded mixtures. A state-of-the art algorithm can be expected to improve the SIR by 10-20 dB for two stationary sources. Typically a few seconds of data (2 s-10 s) will be sufficient to generate these results. However, from this survey nothing can be said about moving sources. Note that only 8 (of over 400) papers reported separation of more than 2 sources indicating that this remains a challenging problem.

9. CONCLUSION

We have presented a taxonomy for blind separation of convolutive mixtures with the purpose of providing a survey and discussion of existing methods. Further we hope that this might stimulate the development of new models and algorithms which more efficiently incorporate specific domain knowledge and useful prior information.

In the title of the BSS review by Torkkola (1999) [13], it was asked: *Are we there yet?* Since then numerous algorithms have been proposed for blind separation of convolutive mixtures. Many convolutive algorithms have shown good performance when the mixing process is stationary, but still only few methods work in real-world, time-varying environments. In real-time-varying environments, there are too many parameters to update in the separation filters, and too little data available in order to estimate the parameters reliably, while the less complicated methods such as null-beamformers may perform just as well. This may indicate that the long de-mixing filters are not the solution for real-world, time-varying environments such as the cocktail-party party situation.

Acknowledgments

Table 5: An overview of algorithms applied in real rooms, where the SIR improvement has been reported.

Room size (approx.) [m]	T_{60} [ms]	N	M	SIR [dB]	Reference
$6 \times 3 \times 3$	300	2	2	13	[169, 170][1]
$6 \times 3 \times 3$	300	2	2	8–10	[271][1]
$6 \times 3 \times 3$	300	2	2	12	[249]
$6 \times 3 \times 3$	300	2	2	5.7	[290]
$6 \times 3 \times 3$	300	2	2	18–20	[297, 132][1]
	50	2	2	10	[207]
	250	2	2	16	[262]
$6 \times 6 \times 3$	200	2	2	< 16	[205][2]
$6 \times 6 \times 3$	150	2	2	< 15	[206]
$6 \times 6 \times 3$	150	2	2	< 20	[171]
	500	2	2	6	[262]
$4 \times 4 \times 3$	130	3	2	4–12	[298]
$4 \times 4 \times 3$	130	3	2	14.3	[227]
$4 \times 4 \times 3$	130	3	2	< 12	[47]
$4 \times 4 \times 3$	130	2	2	7–15	[130]
$4 \times 4 \times 3$	130	2	2	4–15	[22, 299][2]
$4 \times 4 \times 3$	130	2	2	12	[291]
$4 \times 4 \times 3$	130	6	8	18	[300]
$4 \times 4 \times 3$	130	4	4	12	[259]
	130	3	2	10	[140, 141]
Office		2	2	5.5–7.6	[142]
6×5	130	2	8	1.6–7.0	[269]
8×7	300	2	2	4.2–6.0	[73]
15×10	300	2	2	5–8.0	[72]
		2	2	< 10	[57, 91]
Office		2	2	6	[122]
Many rooms		2	2	3.1–27.4	[115]
Small room		2	2	4.7–9.5	[252]
$4 \times 3 \times 2$		2	2	< 10	[181]
$4 \times 3 \times 2$		2	2	14.4	[183]
$4 \times 3 \times 2$		2	2	4.6	[182]
		2	2	< 15	[245]
6×7	580	2	3	< 73	[31][3]
	810	2	2	< 10	[167][2]
Conf. room		4	4	14	[278]
	150	3	3	10	[222]
$15 \times 10 \times 4$	300	2	2	10	[77]
	360	2	2	5	[266]
5×5	200	2	2	6–21	[301]
	300	2	2–12	8–12	[302]
3×6		3	8	10	[184]
$4 \times 3 \times 2$		2	2	15	[149]
$5 \times 5 \times 3$		2	2	5	[187]
8×4	700	2	4	16	[152]
$7 \times 4 \times 3$	250	2	2	9.3	[253][1]
4×4	200	2	2	15	[303]
Office	500	3	2	4.3–10.6	[45]
	300	2	6	< 15	[213]

[1] Sources convolved with real impulse responses.

[2] Moving sources.

[3] This method is not really blind. It requires that sources are on one at a time.

10. REFERENCES

[1] A. Mansour, N. Benchekroun, and C. Gervaise, "Blind separation of underwater acoustic signals," in *ICA'06*, 2006, pp. 181–188.

[2] S. Cruces-Alvarez, A. Cichocki, and L. Castedo-Ribas, "An iterative inversion approach to blind source separation," *IEEE Trans. Neural Networks*, vol. 11, no. 6, pp. 1423–1437, Nov 2000.

[3] K. I. Diamantaras and T. Papadimitriou, "MIMO blind deconvoluition using subspace-based filter deflation," in *ICASSP'04*, vol. IV, 2004, pp. 433–436.

[4] D. Nuzillard and A. Bijaoui, "Blind source separation and analysis of multispectral astronomical images," *Astron. Astrophys. Suppl. Ser.*, vol. 147, pp. 129–138, Nov. 2000.

[5] J. Anemüller, T. J. Sejnowski, and S. Makeig, "Complex independent component analysis of frequency-domain electroencephalographic data," *Neural Networks*, vol. 16, no. 9, pp. 1311–1323, Nov 2003.

[6] M. Dyrholm, S. Makeig, and L. K. Hansen, "Model structure selection in convolutive mixtures," in *ICA'06*, 2006, pp. 74–81.

[7] C. Vayá, J. J. Rieta, C. Sánchez, and D. Moratal, "Performance study of convolutive BSS algorithms applied to the electrocardiogram of atrial fibrillation," in *ICA'06*, 2006, pp. 495–502.

[8] L. K. Hansen, "ICA of fMRI based on a convolutive mixture model," in *Ninth Annual Meeting of the Organization for Human Brain Mapping (HBM 2003)*, 2003.

[9] E. C. Cherry, "Some experiments on the recognition of speech, with one and two ears," *J. Acoust. Soc. Am.*, vol. 25, no. 5, pp. 975–979, Sep 1953.

[10] S. Haykin and Z. Chen, "The cocktail party problem," *Neural Computation*, vol. 17, pp. 1875–1902, Sep 2005.

[11] M. Miyoshi and Y. Kaneda, "Inverse filtering of room acoustics," *IEEE Trans. Acout. Speech. Sig. Proc.*, vol. 36, no. 2, pp. 145–152, Feb 1988.

[12] K. J. Pope and R. E. Bogner, "Blind Signal Separation II. Linear Convolutive Combinations," *Digital Signal Processing*, vol. 6, pp. 17–28, 1996.

[13] K. Torkkola, "Blind separation for audio signals – are we there yet?" in *ICA'99*, 1999, pp. 239–244.

[14] ——, "Blind separation of delayed and convolved sources," in *Unsupervised Adaptive Filtering, Blind Source Separation*, S. Haykin, Ed. Wiley, John and Sons, Incorporated, Jan 2000, vol. 1, ch. 8, pp. 321–375.

[15] R. Liu, Y. Inouye, and H. Luo, "A system-theoretic foundation for blind signal separation of MIMO-FIR convolutive mixtures - a review," in *ICA'00*, 2000, pp. 205–210.

[16] K. E. Hild, "Blind separation of convolutive mixtures using renyi's divergence," Ph.D. dissertation, University of Florida, 2003.

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

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

[19] S. C. Douglas, "Blind separation of acoustic signals," in *Microphone Arrays*, M. S. Brandstein and D. B. Ward, Eds. Springer, 2001, ch. 16, pp. 355–380.

[20] ——, "Blind signal separation and blind deconvolution," in *Handbook of neural network signal processing*, ser. Electrical engineering and applied signal processing, Y. H. Hu and J.-N. Hwang, Eds. CRC Press LLC, 2002, ch. 7.

[21] P. D. O'Grady, B. A. Pearlmutter, and S. T. Rickard, "Survey of sparse and non-sparse methods in source separation," *IJIST*, vol. 15, pp. 18–33, 2005.

[22] S. Makino, H. Sawada, R. Mukai, and S. Araki, "Blind source separation of convolutive mixtures of speech in frequency domain," *IEICE Trans. Fundamentals*, vol. E88-A, no. 7, pp. 1640–1655, Jul 2005.

[23] R. Lambert, "Multichannel blind deconvolution: FIR matrix algebra and separation of multipath mixtures," Ph.D. dissertation, University of Southern California, Department of Electrical Engineering, May 1996.

[24] S. Roberts and R. Everson, *Independent Components Analysis: Principles and Practice*. Cambridge University Press, 2001.

[25] A. Gorokhov and P. Loubaton, "Subspace based techniques for second order blind separation of convolutive mixtures with temporally correlated sources," *IEEE Trans. Circ. Syst.*, vol. 44, no. 9, pp. 813–820, Sep 1997.

[26] H.-L. N. Thi and C. Jutten, "Blind source separation for convolutive mixtures," *Signal Processing, Elsevier*, vol. 45, no. 2, pp. 209–229, 1995.

[27] S. Choi and A. Cichocki, "Adaptive blind separation of speech signals: Cocktail party problem," in *ICSP'97*, 1997, pp. 617–622.

[28] ——, "A hybrid learning approach to blind deconvolution of linear MIMO systems," *Electronics Letters*, vol. 35, no. 17, pp. 1429–1430, Aug 1999.

[29] E. Weinstein, M. Feder, and A. Oppenheim., "Multichannel signal separation by decorrelation." *IEEE Trans. Speech Audio Proc.*, vol. 1, no. 4, pp. 405–413, Oct 1993.

[30] S. T. Neely and J. B. Allen, "Invertibility of a room impulse response," *J. Acoust. Soc. Am.*, vol. 66, no. 1, pp. 165–169, Jul 1979.

[31] Y. A. Huang, J. Benesty, and J. Chen, "Blind channel identification-based two-stage approach to separation and dereverberation of speech signals in a reverberant environment," *IEEE Trans. Speech Audio Proc.*, vol. 13, no. 5, pp. 882–895, Sep 2005.

[32] Y. Huang, J. Benesty, and J. Chen, "Identification of acoustic MIMO systems: Challenges and opportunities," *Signal Processing, Elsevier*, vol. 86, pp. 1278–1295, 2006.

[33] Z. Ye, C. Chang, C. Wang, J. Zhao, and F. H. Y. Chan, "Blind separation of convolutive mixtures based on second order and third order statistics," in *ICASSP'03*, vol. 5, 2003, pp. V–305–308.

[34] K. Rahbar, J. P. Reilly, and J. H. Manton, "Blind identification of MIMO FIR systems driven by quasistationary sources using second-order statistics: A frequency domain approach," *IEEE Trans. Sig. Proc.*, vol. 52, no. 2, pp. 406–417, Feb 2004.

[35] A. Mansour, C. Jutten, and P. Loubaton, "Subspace method for blind separation of sources and for a convolutive mixture model," in *Signal Processing VIII, Theories and Applications*. Elsevier, Sep 1996, pp. 2081–2084.

[36] W. Hachem, F. Desbouvries, and P. Loubaton, "On the identification of certain noisy FIR convolutive mixtures," in *ICA'99*, 1999.

[37] A. Mansour, C. Jutten, and P. Loubaton, "Adaptive subspace algorithm for blind separation of independent sources in convolutive mixture," *IEEE Trans. Sig. Proc.*, vol. 48, no. 2, pp. 583–586, Feb 2000.

[38] N. Delfosse and P. Loubaton, "Adaptive blind separation of convolutive mixtures," in *ICASSP'96*, 1996, pp. 2940–2943.

[39] ——, "Adaptive blind separation of independent sources: A second-order stable algorithm for the general case," *IEEE Trans. Circ. Syst.–I: Fundamental Theory and Applications*, vol. 47, no. 7, pp. 1056–1071, Jul 2000.

[40] L. K. Hansen and M. Dyrholm, "A prediction matrix approach to convolutive ICA," in *NNSP'03*, 2003, pp. 249–258.

[41] D. Yellin and E. Weinstein, "Multichannel signal separation: Methods and analysis," *IEEE Trans. Sig. Proc.*, vol. 44, no. 1, pp. 106–118, Jan 1996.

[42] B. Chen and A. P. Petropulu, "Frequency domain blind MIMO system identification based on second- and higher order statistics," *IEEE Trans. Sig. Proc.*, vol. 49, no. 8, pp. 1677–1688, Aug 2001.

[43] B. Chen, A. P. Petropulu, and L. D. Lathauwer, "Blind identification of complex convolutive MIMO systems with 3 sources and 2 sensors," in *ICASSP'02*, vol. II, 2002, pp. 1669–1672.

[44] Y. Hua and J. K. Tugnait, "Blind identifiability of FIR-MIMO systems with colored input using second order statistics," *IEEE Sig. Proc. Lett.*, vol. 7, no. 12, pp. 348–350, Dec 2000.

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

[46] N. Roman, "Auditory-based algorithms for sound segregatoion in multisource and reverberant envirrontments," Ph.D. dissertation, The Ohio State University, Columbus, OH, 2005.

[47] A. Blin, S. Araki, and S. Makino, "Underdetermined blind separation og convolutive mixtures of speech using time-frequency mask and mixing matrix estimation," *IEICE Trans. Fundamentals*, vol. E88-A, no. 7, pp. 1693–1700, Jul 2005.

[48] K. I. Diamantaras, A. P. Petropulu, and B. Chen, "Blind Two-Input-TwoOutput FIR Channel Identification Based on Frequency Domain Second- Order Statistics," *IEEE Trans. Sig. Proc.*, vol. 48, no. 2, pp. 534–542, Feburary 2000.

[49] E. Moulines, J.-F. Cardoso, and E. Cassiat, "Maximum likelihood for blind separation and deconvolution of noisy signals using mixture models," in *ICASSP'97*, vol. 5, 1997, pp. 3617–3620.

[50] U. A. Lindgren and H. Broman, "Source separation using a criterion based on second-order statistics," *IEEE Trans. Sig. Proc.*, vol. 46, no. 7, pp. 1837–1850, Jul 1998.

[51] H. Broman, U. Lindgren, H. Sahlin, and P. Stoica, "Source separation: A TITO system identification approach," *Signal Processing, Elsevier*, vol. 73, no. 1, pp. 169–183, 1999.

[52] H. Sahlin and H. Broman, "MIMO signal separation for FIR channels: A criterion and performance analysis," *IEEE Trans. Sig. Proc.*, vol. 48, no. 3, pp. 642–649, Mar 2000.

[53] C. Jutten, L. Nguyen Thi, E. Dijkstra, E. Vittoz, and J. Caelen, "Blind separation of sources: An algorithm for separation of convolutive mixtures," in *Higher Order Statistics. Proceedings of the International Signal Processing Workshop*, J. Lacoume, Ed.

Elsevier, 1992, pp. 275–278.

[54] C. Servière, "Blind source separation of convolutive mixtures," in *SSAP'96*, 1996, pp. 316–319.

[55] A. Ypma, A. Leshem, and R. P. Duina, "Blind separation of rotating machine sources: Bilinear forms and convolutive mixtures," *Neurocomp.*, vol. 49, no. 1–4, pp. 349–368, 2002.

[56] N. Charkani, Y. Deville, and J. Herault, "Stability analysis and optimization of time-domain convolutive source separation algorithms," in *SPAWC'97*, 1997, pp. 73–76.

[57] N. Charkani and Y. Deville, "A convolutive source separation method with self-optimizing nonlinearities," in *ICASSP'99*, vol. 5, 1999, pp. 2909–2912.

[58] ——, "Self-adaptive separation of convolutively mixed signals with a recursive structure. part I: Stability analysis and optimization of asymptotic behaviour," *Signal Processing, Elsevier*, vol. 73, no. 3, pp. 225–254, 1999.

[59] K. Torkkola, "Blind separation of convolved sources based on information maximization," in *NNSP'96*, 1996, pp. 423–432.

[60] P. Comon, "Independent Component Analysis, a new concept ?" *Signal Processing, Elsevier*, vol. 36, no. 3, pp. 287–314, Apr 1994, special issue on Higher-Order Statistics.

[61] P. Comon and L. Rota, "Blind separation of independent sources from convolutive mixtures," *IEICE Trans. on Fundamentals*, vol. E86-A, no. 3, pp. 542–549, Mar 2003.

[62] V. Capdevielle, C. Servire, and J. L. Lacoume, "Blind separation of wide-band sources in the frequency domain," in *ICASSP95*, vol. III, Detroit, MI, USA, May 9–12 1995, pp. 2080–2083.

[63] S. Icart and R. Gautier, "Blind separation of convolutive mixtures using second and fourth order moments," in *ICASSP'96*, vol. 5, 1996, pp. 3018–3021.

[64] M. Girolami and C. Fyfe, "A temporal model of linear anti-hebbian learning," *Neural Processing Letters*, vol. 4, no. 3, pp. 139–148, 1996.

[65] J. K. Tugnait, "On blind separation of convolutive mixtures of independent linear signals in unknown additive noise," *IEEE Trans. on Sig. Proc.*, vol. 46, no. 11, pp. 3117–3123, Nov 1998.

[66] C. Simon, P. Loubaton, C. Vignat, C. Jutten, and G. d'Urso, "Separation of a class of convolutive mixtures: A contrast function approach," in *ICASSP'99*, 1999, pp. 1429–1432.

[67] Y. Su, L. He, and R. Yang, "An improved

cumulant-based blind speech separation method," in *ICASSP'00*, 2000, pp. 1867–1870.

[68] P. Baxter and J. McWhirter, "Blind signal separation of convolutive mixtures," in *AsilomarSSC*, vol. 1, 2003, pp. 124–128.

[69] S. Hornillo-Mellado, C. G. Puntonet, R. Martin-Clemente, and M. Rodríguez-Álvarez, "Characterization of the sources in convolutive mixtures: A cumulant-based approach," in *ICA'04*, 2004, pp. 586–593.

[70] Y. Deville, M. Benali, and F. Abrard, "Differential source separation for underdetermined instantaneous or convolutive mixtures: Concept and algorithms," *Signal Processing*, vol. 84, no. 10, pp. 1759–1776, Oct 2004.

[71] M. Ito, M. Kawamoto, N. Ohnishi, and Y. Inouye, "Eigenvector algorithms with reference signals for frequency domain BSS," in *ICA'06*, 2006, pp. 123–131.

[72] W. Baumann, D. Kolossa, and R. Orglmeister, "Beamforming-based convolutive source separation," in *ICASSP'03*, vol. V, 2003, pp. 357–360.

[73] ——, "Maximum likelihood permutation correction for convolutive source separation," in *ICA'03*, 2003, pp. 373–378.

[74] M. S. Pedersen and C. M. Nielsen, "Gradient flow convolutive blind source separation," in *MLSP'04*, 2004, pp. 335–344.

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

[76] D. Yellin and E. Weinstein, "Criteria for multi-channel signal separation," *IEEE Trans. Sig. Proc.*, vol. 42, no. 8, pp. 2158–2168, Aug 1994.

[77] D. Kolossa and R. Orglmeister, "Nonlinear post-processing for blind speech separation," in *ICA'04*, 2004, pp. 832–839.

[78] P. Comon, E. Moreau, and L. Rota, "Blind separation of convolutive mixtures: A contrast-based joint diagonalization approach," in *ICA'01*, 2001, pp. 686–691.

[79] E. Moreau and J. Pesquet, "Generalized contrasts for multichannel blind deconvolution of linear systems," *IEEE Sig. Proc. Lett.*, vol. 4, no. 6, pp. 182–183, Jun 1997.

[80] Y. Li, J. Wang, and A. Cichocki, "Blind source extraction from convolutive mixtures in ill-conditioned multi-input multi-output channels," *IEEE Trans. Circ. Syst. I: Regular Papers*, vol. 51, no. 9, pp. 1814–1822, Sep 2004.

[81] R. K. Prasad, H. Saruwatari, and K. Shikano, "Problems in blind separation of convolutive speech mixtures by negentropy maximization," in *IWAENC'03*, 2003, pp. 287–290.

[82] X. Sun and S. Douglas, "Adaptive paraunitary filter banks for contrast-based multichannel blind deconvolution," in *ICASSP'01*, vol. 5, 2001, pp. 2753–2756.

[83] J. Thomas, Y. Deville, and S. Hosseini, "Time-domain fast fixed-point algorithms for convolutive ICA," *IEEE Sig. Proc. Lett.*, vol. 13, no. 4, pp. 228–231, Apr 2006.

[84] C. Jutten and J. Herault, "Blind separation of sources, part I: An adaptive algorithm based on neuromimetic architecture," *Signal Processing, Elsevier*, vol. 24, no. 1, pp. 1–10, 1991.

[85] A. D. Back and A. C. Tsoi, "Blind deconvolution of signals using a complex recurrent network," in *NNSP'94*, 1994, pp. 565–574.

[86] I. Kopriva, Željko Devčić, and H. Szu, "An adaptive short-time frequency domain algorithm for blind separation of nonstationary convolved mixtures," in *IJCNN'01*, 2001, pp. 424–429.

[87] S. Cruces and L. Castedo, "A Gauss-Newton method for blind source separation of convolutive mixtures," in *ICASSP'98*, vol. IV, 1998, pp. 2093–2096.

[88] S. V. Gerven and D. V. Compernolle, "Signal separation by symmetric adaptive decorrelation: Stability, convergence, and uniqueness," *IEEE Trans. Sig. Proc.*, vol. 43, no. 7, pp. 1602–1612, Jul 1995.

[89] S. Li and T. J. Sejnowski, "Adaptive separation of mixed broadband sound sources with delays by a beamforming Herault-Jutten network," *IEEE J. Ocean. Eng.*, vol. 20, no. 1, pp. 73–79, Jan 1995.

[90] N. Charkani and Y. Deville, "Optimization of the asymppotic preformance of time-domain convolutive source separation algorithms," in *ESANN'97*, 1997, pp. 273–278.

[91] ——, "Self-adaptive separation of convolutively mixed signals with a recursive structure. part II: Theoretical extensions and application to synthetic and real signals," *Signal Processing, Elsevier*, vol. 75, no. 2, pp. 117–140, 1999.

[92] S. Choi, H. Hong, H. Glotin, and F. Berthommier, "Multichannel signal separation for cocktail party speech recognition: A dynamic recurrent network," *Neurocomp.*, vol. 49, no. 1–4, pp. 299–314, Dec 2002.

[93] F. Berthommier and S. Choi, "Several improvements of the Hérault-Jutten model for speech segregation," in *ICA'03*, 2003, pp. 1089–1094.

[94] M. Cohen and G. Cauwenbergh, "Blind separation of linear convolutive mixtures through parallel stochastic optimization," in *ISCAS'98*, vol. 3, 1998, pp. 17–20.

[95] M. Stanacevic, M. Cohen, and G. Cauwenberghs, "Blind separation of linear convolutive mixtures using orthogonal filter banks," in *ICA'01*, 2001, pp. 260–265.

[96] G. Deco and D. Obradovic, *An Information-Theoretic Approach to Neural Computing*. New York: Springer Verlag, 1996.

[97] A. J. Bell and T. J. Sejnowski, "An information-maximization approach to blind separation and blind deconvolution," *Neural Computation*, vol. 7, no. 6, pp. 1129–1159, 1995.

[98] S. Amari, S. Douglas, A. Cichocki, and H.H.Yang, "Multichannel blind deconvolution and equalization using the natural gradient," in *IEEE International Workshop on Wireless Communication*, 1997, pp. 101–104.

[99] B. A. Pearlmutter and L. C. Parra, "Maximum likelihood blind source separation: A context-sensitive generalization of ICA," *NIPS'97*, pp. 613–619, 1997.

[100] J.-F. Cardoso, "Blind signal separation: Statistical principles," *Proc. IEEE*, vol. 9, no. 10, pp. 2009–2025, Oct 1998.

[101] K. Kokkanikis and A. K. Nandi, "Optimal blind separation of convolutive audio mixtures without temporal constraints," in *ICASSP'04*, vol. I, 2004, pp. 217–220.

[102] K. Kokkinakis and A. K. Nandi, "Multichannel speech separation using adaptive parameterization of source pdfs," in *ICA'04*, 2004, pp. 486–493.

[103] H.-C. Wu and J. C. Principe, "Generalized anti-hebbian learning for source separation," in *ICASSP'99*, vol. 2, 1999, pp. 1073–1076.

[104] J. Xi and J. P. Reilly, "Blind separation and restoration of signals mixed in convolutive environment," in *ICASSP'97*, 1997, pp. 1327 – 1330.

[105] J. P. Reilly and L. C. Mendoza, "Blind signal separation for convolutive mixing environments using spatial-temporal processing," in *ICASSP'99*, 1999, pp. 1437–1440.

[106] H. Sawada, R. Mukai, S. Araki, and S. Makino, "A robust and precise method for solving the permutation problem of frequency-domain blind source separation," *IEEE Trans. Speech Audio Proc.*, vol. 12, no. 5, pp. 530–538, Sep 2004.

[107] ——, "Polar coordinate based nonlinear function for frequency domain blind source separation," in *ICASSP'02*, 2002, pp. 1001–1004.

[108] J.-F. Cardoso and T. Adali, "The maximum likelihood approach to complex ICA," in *ICASSP*, vol. V, 2006, pp. 673–676.

[109] M. Novey and T. Adali, "Adaptable nonlinearity for complex maximization of nongaussianity and a fixed-point algorithm," in *MLSP*, 2006.

[110] M. Joho, H. Mathis, and G. S. Moschytz, "An FFT-based algorithm for multichannel blind deconvolution," in *ISCAS'99*, vol. 3, 1999, pp. 203–206.

[111] R. H. Lambert and A. J. Bell, "Blind separation of multiple speakers in a multipath environment," in *ICASSP'97*, vol. 1, 1997, pp. 423–426.

[112] R. H. Lambert, "A new method for source separation," in *ICASSP'95*, vol. 3, 1995, pp. 2116–2119.

[113] T.-W. Lee, A. J. Bell, and R. Orglmeister, "Blind source separation of real world signals," in *ICNN'97*, 1997, pp. 2129–2135.

[114] S. Choi, S. Amari, A. Cichocki, and R. wen Liu, "Natural gradient learning with a nonholonomic constraint for blind deconvolution of multiple channels," in *ICA'99*, 1999, pp. 371–376.

[115] S. C. Douglas and X. Sun, "Convolutive blind separation of speech mixtures using the natural gradient," *Speech Communication, Elsevier*, vol. 39, pp. 65–78, 2003.

[116] K. Matsuoka, Y. Ohba, Y. Toyota, and S. Nakashima, "Blind separation for convolutive mixture of many voices," in *IWAENC'03*, 2003, pp. 279–282.

[117] K. Matsuoka and S. Nakashima, "Minimal distortion principle for blind source separation," in *ICA'01*, 2001, pp. 722–727.

[118] W. Wang, M. G. Jafari, S. Sanei, and J. A. Chambers, "Blind separartion of convolutive mixtures of cyclostationary signals," *International Journal of Adaptive Control and Signal Processing*, vol. 18, pp. 279–298, Mar 2004.

[119] G.-J. Jang, C. Choi, Y. Lee, and Y.-H. Oh, "Adaptive cross-channel interference cancellation on blind signal separation outputs using source absence/presence detection and spectral subtraction," in *ICLSP'04*, vol. IV, 2004, pp. 2865–2868.

[120] S. H. Nam and S. Beack, "A frequency-domain normalized multichannel blind deconvolution algorithm for acoustical signals," in *ICA'04*, 2004, pp. 524–531.

[121] M. Joho and P. Schniter, "Frequency domain realization of a multichannel blind deconvolution algorithm based on the natural gradient," in *ICA'03*,

2003, pp. 543–548.

[122] C. Choi, G. Jang, Y. Lee, and S. R. Kim, "Adaptive cross-channel interference cancellation on blind source separation outputs," in *ICA'04*, 2004, pp. 857–864.

[123] L. Parra, C. Spence, and B. de Vries, "Convolutive source separation and signal modeling with ML," in *ISIS'97*, 1997.

[124] L. C. Parra, "Temporal models in blind source separation," *Lecture Notes in Computer Science*, vol. 1387, pp. 229–247, 1998.

[125] K. Yamamoto, F. Asano, W. van Rooijen, E. Ling, T. Yamada, and N. Kitawaki, "Estimation of the number of sound sources using support vector machines and its application to sound source separation," in *ICASSP'03*, vol. 5, 2003, pp. 485–488.

[126] J. Anemüller, "Across-frequency processing in convolutive blind source separation," Ph.D. dissertation, Oldenburg, Univ., Jul 30st 2001.

[127] J. Anemüller and T. Gramms, "On-line blind separation of moving sound sources," in *ICA'99*, 1999, pp. 331–334.

[128] S. Deligne and R. Gopinath, "An EM algorithm for convolutive independent component analysis," *Neurocomp.*, vol. 49, pp. 187–211, 2002.

[129] J. Rosca, C. Borss, and R. Balan, "Generalized sparse signal mixing model and application to noisy blind source separation," in *ICASSP'04*, vol. III, 2004, pp. 877–880.

[130] S. C. Douglas, H. Sawada, and S. Makino, "Natural gradient multichannel blind deconvolution and speech separation using causal FIR filters," *IEEE Trans. Speech Audio Proc.*, vol. 13, no. 1, pp. 92–104, Jan 2005.

[131] 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. Speech Audio Proc.*, vol. 11, no. 2, pp. 109–116, Mar 2003.

[132] S. Ukai, T. Takatani, T. Nishikawa, and H. Saruwatari, "Blind source separation combining SIMO-model-based ICA and adaptive beamforming," in *ICASSP'05*, vol. III, 2005, pp. 85–88.

[133] R. H. Lambert and C. L. Nikias, "Polynomial matrix whitening and application to the multichannel blind deconvolution problem," in *MILCOM'95*, vol. 3, 1995, pp. 988–992.

[134] N. Mitianoudis and M. Davies, "Audio source separation of convolutive mixtures," 2002, iEEE Trans. Speech Audio Proc.

[135] F. Tordini and F. Piazza, "A semi-blind approach to the separation of real world speech mixtures," in *IJCNN'02*, vol. 2, 2002, pp. 1293–1298.

[136] H. Attias, "Source separation with a sensor array using graphical models and subband filtering," in *NIPS'02*, vol. 15, 2002, pp. 1229–1236.

[137] M. A. Gandhi and M. A. Hasegawa-Johnson, "Source separation using particle filters," in *ICLSP'04*, vol. III, 2004, pp. 2449–2452.

[138] R. K. Olsson and L. K. Hansen, "A harmonic excitation state-space approach to blind separation of speech," in *NIPS*, Dec 2004.

[139] ——, "Probabilistic blind deconvolution of nonstationary sources," in *EUSIPCO'04*, 2004, pp. 1697–1700.

[140] S. Winter, H. Sawada, S. Araki, and S. Makino, "Hierarchical clustering applied to overcomplete BSS for convolutive mixtures," in *SAPA'04*, 2004.

[141] ——, "Overcomplete BSS for convolutive mixtures based on hierarchical clustering," in *ICA'04*, 2004, pp. 652–660.

[142] H. Attias, "New EM algorithms for source separation and deconvolution," in *ICASSP'03*, vol. V, 2003, pp. 297–300.

[143] C. Andrieu and S. Godsill, "A particle filter for model based audio source separation," in *ICA'00*, 2000, pp. 381–386.

[144] J. R. Hopgood, "Bayesian blind MIMO deconvolution of nonstationary subband autoregressive sources mixed through subband all-pole channels," in *SSP'03*, 2003, pp. 422–425.

[145] S. J. Godsill and C. Andrieu, "Bayesian separation and recovery of convolutively mixed autoregressive sources," in *ICASSP'99*, vol. III, 1999, pp. 1733–1736.

[146] K. Abed-Meraim, W. Qiu, and Y. Hua, "Blind system identification," *Proc. IEEE*, vol. 85, no. 8, pp. 1310–1322, Aug 1997.

[147] S. Sanei, W. Wang, and J. A. Chambers, "A coupled HMM for solving the permutation problem in frequency domain BSS," in *ICASSP'04*, vol. V, 2004, pp. 565–568.

[148] W. Wang, D. Cosker, Y. Hicks, S. Sanei, and J. Chambers, "Video assisted speech source seperation," in *ICASSP'05*, vol. V, 2005, pp. 425–428.

[149] L. Parra and C. Spence, "Convolutive blind separation of non-stationary sources," *IEEE Trans. Speech Audio Proc.*, vol. 8, no. 3, pp. 320–327, May 2000.

[150] D. W. E. Schobben and P. C. W. Sommen, "On the indeterminacies of convolutive blind signal separa-

tion based on second-order statistics," in *ISSPA'99*, 1999, pp. 215–218.

[151] J. Liang and Z. Ding, "Blind MIMO system identification based on cumulant subspace decomposition," *IEEE Trans. Sig. Proc.*, vol. 51, no. 6, pp. 1457–1468, Jun 2003.

[152] K. Rahbar and J. P. Reilly, "A frequency domain method for blind source separation of convolutive audio mixtures," *IEEE Trans. Speech Audio Proc.*, vol. 13, no. 5, pp. 832–844, Sep 2005.

[153] M. Kawamoto, K. Matsuoka, and N. Ohnishi, "A method of blind separation for convolved non-stationary signals," *Neurocomp.*, vol. 22, no. 1–3, pp. 157–171, Nov 1998.

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

[155] A. G. Lindgren, T. P. von Hoff, and A. N. Kaelin, "Stability and performance of adaptive algorithms for multichannel blind source separation and deconvolution," in *EUSIPCO'00*, vol. 2, 2000, pp. 861–864.

[156] H. Sahlin and H. Broman, "Separation of real-world signals," *Signal Processing, Elsevier*, vol. 64, pp. 103–113, 1998.

[157] D. C. B. Chan, P. J. W. Rayner, and S. J. Godsill, "Multi-channel blind signal separation," in *ICASSP'96*, 1996, pp. 649–652.

[158] C. Simon, C. Vignat, P. Loubaton, C. Jutten, and G. d'Urso, "On the convolutive mixture - source separation by the decorrelation approach," in *ICASSP'98*, vol. 4, 1998, pp. 2109–2112.

[159] N. Murata, S. Ikeda, and A. Ziehe, "An approach to blind source separation based on temporal structure of speech signals," RIKEN Brain Science Institute, Japan, BSIS Technical Reports 98-2, Apr 1998.

[160] B. Yin and P. Sommen, "Adaptive blind signal separation using a new simplified mixing model," in *ProRISC'99*, J. Veen, Ed., 1999, pp. 601–606.

[161] H.-C. Wu and J. C. Principe, "Simultaneous diagonalization in the frequency domain (SDIF) for source separation," in *ICA'99*, 1999, pp. 245–250.

[162] A. Souloumiac, "Blind source detection and separation using second order non-stationarity," in *ICASSP'95*, vol. III, 1995, pp. 1912–1915.

[163] K. Matsuoka, M. Ohya, and M. Kawamoto, "A neural net for blind separation of nonstationary signals," *Neural Networks*, vol. 8, no. 3, pp. 411–419, 1995.

[164] M. Kawamoto, A. K. Barros, A. Mansour, K. Matsuoka, and N. Ohnishi, "Blind separation for convolutive mixtures of non-stationary signals," in *Int. Conf. Neural Inf. Proc.*, 1998, pp. 743–746.

[165] M. Kawamoto, A. K. Barros, K. Matsuoka, and N. Ohnishi, "A method of real-world blind separation implemented in frequency domain," in *ICA'00*, 2000, pp. 267–272.

[166] M. Ito, M. Maruyoshi, M. Kawamoto, T. Mukai, and N. Ohneshi, "Effectiveness of directional microphones and utilization of source arriving directions in source separation," in *ICONIP'02*, 2002, pp. 523–526.

[167] M. Ito, Y. Takeuchi, T. Matsumoto, H. Kudo, M. Kawamoto, T. Mukai, and N. Ohnishi, "Moving-source separation using directional microphones," in *ISSPIT'02*, 2002, pp. 523–526.

[168] Y. Katayama, M. Ito, Y. Takeuchi, T. Matsumoto, H. Kudo, N. Ohnishi, and T. Mukai, "Reduction of source separation time by placing microphones close together," in *ISSPIT'02*, 2002, pp. 540–544.

[169] R. Aichner, S. Araki, S. Makino, T. Nishikawa, and H. Saruwatari, "Time domain blind source separation of non-stationary convolved signals by utilizing geometric beamforming," in *NNSP'02*, 2002, pp. 445–454.

[170] R. Aichner, H. Buchner, S. Araki, and S. Makino, "On-line time-domain blind source separation of nonstationary convolved signals," in *ICA'03*, 2003, pp. 987–992.

[171] H. Buchner, R. Aichner, and W. Kellermann, "A generalization of blind source separation algorithms for convolutive mixtures based on second-order statistics," *IEEE Trans. on Speech Audio Proc.*, vol. 13, no. 1, pp. 120–134, Jan 2005.

[172] E. Visser and T.-W. Lee, "Speech enhancement using blind source separation and two-channel energy based speaker detection," in *ICASSP'03*, vol. 1, 2003, pp. 884–887.

[173] E. Visser, K. Chan, S. Kim, and T.-W. Lee, "A comparison of simultaneous 3-channel blind source separation to selective separation on channel pairs using 2-channel BSS," in *ICLSP'04*, vol. IV, 2004, pp. 2869–2872.

[174] M. Z. Ikram and D. R. Morgan, "A multiresolution approach to blind separation of speech signals in a reverberant environment," in *ICASSP'01*, vol. V, 2001.

[175] M. Ikram and D. Morgan, "A beamforming approach to permutation alignment for multichannel frequency-domain blind speech separation," in *ICASSP'02*, 2002, pp. 881–884.

[176] M. Z. Ikram and D. R. Morgan, "Permutation inconsistency in blind speech separation: Investigation and solutions," *IEEE Trans. Speech Audio Proc.*, vol. 13, no. 1, pp. 1–13, Jan 2005.

[177] L. Parra and C. Spence, "On-line convolutive source separation of non-stationary signals," *IEEE J. VLSI Sig. Proc.*, vol. 26, no. 1/2, pp. 39–46, Aug 2000.

[178] M. Joho, "Blind signal separation of convolutive mixtures: A time-domain joint-diagonalization approach," in *ICA'04*, 2004, pp. 578–585.

[179] ——, "Convolutive blind signal separation in acoustics by joint approximate diagonalization of spatiotemporal correlation matrices," in *Asilomar SSC*, 2004.

[180] S. Araki, S. Makino, R. Mukai, and H. Saruwatari, "Equivalence between frequency domain blind source separation and frequency domain adaptive beamformers," in *CRAC'01*, 2001.

[181] C. L. Fancourt and L. Parra, "The coherence function in blind source separation of convolutive mixtures of non-stationary signals," in *NNSP*, 2001, pp. 303–312.

[182] ——, "The generalized sidelobe decorrelator," in *WASPAA'01*, 2001.

[183] C. Fancourt and L. Parra, "A comparison of decorrelation criteria for the blind source separation of non-stationary signals," in *SAM'02*, 2002.

[184] L. Parra and C. Alvino, "Geometric source separation: Merging convolutive source separation with geometric beamforming," *IEEE Trans. Speech Audio Proc.*, vol. 10, no. 6, pp. 352–362, Sep 2002.

[185] L. Parra and C. Fancourt, *An Adaptive Beamforming Perspective on Convolutive Blind Source Separation.* CRC Press LLC, 2002, book chapter in: Noise Reduction in Speech Applications, Ed. Gillian Davis.

[186] E. Visser and T.-W. Lee, "Blind source separation in mobile environments usinbg a priori knowledge," in *ICASSP'04*, vol. III, 2004, pp. 893–896.

[187] M. S. Pedersen, L. K. Hansen, U. Kjems, and K. B. Rasmussen, "Semi-blind source separation using head-related transfere functions," in *ICASSP'04*, vol. V, 2004, pp. 713–716.

[188] S. Ding, T. Hikichi, T. Niitsuma, M. Hamatsu, and K. Sugai, "Recursive method for blind source separation and its applications to real-time separations of acoustic signals," in *ICA'03*, 2003, pp. 517–522.

[189] E. Robledo-Arnuncio and B. F. Juang, "Issues in frequency domain blind source separation - a critical revisit," in *ICASSP'05*, vol. V, 2005, pp. 281–284.

[190] W. Wang, J. A. Chambers, and S. Sanei, "A joint diagonalization method for convolutive blind separation of nonstationary sources in the frequency domain," in *ICA'03*, 2003, pp. 939–944.

[191] ——, "A novel hybrid approach to the permutation problem of frequency domain blind source separation," in *ICA'04*, 2004, pp. 532–539.

[192] W. Wang, S. Sanei, and J. A. Chambers, "Penalty function-based joint diagonalization approach for convolutive blind separation of nonstationary sources," *IEEE Trans. Sig. Proc.*, vol. 53, no. 5, pp. 1654–1669, May 2005.

[193] L. Molgedey and H. Schuster, "Separation of independent signals using time-delayed correlations," *Physical Review Letters*, vol. 72, no. 23, pp. 3634–3637, 1994.

[194] T. J. Ngo and N. Bhadkamkar, "Adaptive blind separation of audio sources by a physically compact device using second order statistics," in *ICA'99*, 1999, pp. 257–260.

[195] D. W. Schobben and P. C. W. Sommen, "A frequency domain blind signal separation method based on decorrelation," *IEEE Trans. Sig. Proc.*, vol. 8, no. 50, pp. 1855–1865, Aug 2002.

[196] S. Ikeda and N. Murata, "An approach to blind source separation of speech signals," in *ICANN'98*, vol. 2, 1998, pp. 761–766.

[197] ——, "A method of blind separation on temporal structre of signals," in *ICONIP'98*, vol. 2, 1998, pp. 737–742.

[198] ——, "A method of ICA in time-frequency domain," in *ICA'99*, 1999, pp. 365–371.

[199] N. Murata, S. Ikeda, and A. Ziehe, "An approach to blind source separation based on temporal structure of speech signals," *Neurocomp.*, vol. 41, no. 1–4, pp. 1–24, 2001.

[200] A. Ahmed, P. J. W. Rayner, and S. J. Godsill, "Considering non-stationarity for blind signal separation," in *WASPAA'99*, 1999, pp. 111–114.

[201] B. Yin and P. Sommen, "A new convolutive blind signal separation algorithm based on second order statistics using a simplified mixing model," in *EUSIPCO'00*, vol. 4, 2000, pp. 2049–2053.

[202] D.-T. Pham, C. Servière, and H. Boumaraf, "Blind separation of convolutive audio mixtures using non-stationarity," in *ICA'03*, 2003, pp. 981–986.

[203] C. Servière and D. Pham, "A novel method for permutation correction in frequency-domain in blind separation of speech mixtures," in *ICA'04*, 2004, pp. 807–815.

[204] H. Buchner, R. Aichner, and W. Kellermann, "Blind source separation for convolutive mixtures exploiting nongaussianity, nonwhiteness, and nonstationarity," in *IWAENC'03*, 2003, pp. 275–278.

[205] R. Aichner, H. Buchner, F. Yan, and W. Kellermann, "Real-time convolutive blind source separation based on a broadband approach," in *ICA'04*, 2004, pp. 840–848.

[206] H. Buchner, R. Aichner, and W. Kellermann, "Trinicon: A versatile framework for multichannel blind signal processing," in *ICASSP'04*, vol. III, 2004, pp. 889–892.

[207] R. Aichner, H. Buchner, and W. Kellermann, "On the causality problem in time-domain blind source separation and deconvolution algorithms," in *ICASSP'05*, vol. V, 2005, pp. 181–184.

[208] B. S. Krongold and D. L. Jones, "Blind source separation of nonstationary convolutively mixed signals," in *SSAP'00*, 2000, pp. 53–57.

[209] K. Rahbar and J. P. Reilly, "Blind source separation algorithm for MIMO convolutive mixtures," in *ICA'01*, 2001.

[210] A. Holobar and D. Zazula, "A novel approach to convolutive blind separation of close-to-orthogonal pulse sources using second-order statistics," in *EU-SIPCO'04*, 2004, pp. 381–384.

[211] K.-C. Yen and Y. Zhao, "Adaptive co-channel speech separation and recognition," *IEEE Trans Speech Audio Proc.*, vol. 7, no. 2, pp. 138–151, Mar 1999.

[212] A. Mertins and I. Russel, "An extended ACDC algorithm for the blind estimation of convolutive mixing systems," in *ISSPA'03*, vol. 2, 2003, pp. 527–530.

[213] Y. Zhao and R. Hu, "Fast convergence speech source separation in reverberant acoustic environment," in *ICASSP'04*, vol. III, 2004, pp. 897–900.

[214] I. Russell, J. Xi, A. Mertins, and J. Chicharo, "Blind source separation of nonstationary convolutively mixed signals in the subband domain," in *ICASSP'04*, vol. V, 2004, pp. 484–484.

[215] A. Leon-Garcia, *Probability and Random Processes for Electrical Engineering*, 2nd ed. Addison Wesley, May 1994.

[216] S. Shamsunder and G. B. Giannakis, "Multichannel blind signal separation and reconstruction," *IEEE Trans. Speech, Audio Proc.*, vol. 5, no. 6, pp. 515–528, Nov 1997.

[217] L. Deneire and D. T. Slock, "A Schur method for multiuser multichannel blind identification," in *ICASSP'99*, 1999, pp. 2905–2908.

[218] C. T. Ma, Z. Ding, and S. F. Yau, "A two-stage algorithm for MIMO blind deconvolution of nonstationary colored signals," *IEEE Trans. Sig. Proc.*, vol. 48, no. 4, pp. 1187–1192, Apr 2000.

[219] I. Bradaric, A. P. Petropulu, and K. I. Diamantaras, "On blind identifiability of FIR-MIMO systems with cyclostationary inputs using second order statistics," in *ICASSP'02*, vol. II, 2002, pp. 1745–1748.

[220] ——, "On blind identifiability of FIR-MIMO systems with cyclostationary inputs using second order statistics," *IEEE Trans. Sig. Proc.*, vol. 51, no. 2, pp. 434–441, Feburary 2003.

[221] M. Knaak, M. Kunter, and D. Filbert, "Blind source separation for acoustical machine diagnosis," in *DSP'02*, 2002.

[222] M. Knaak, S. Araki, and S. Makino, "Geometrically constrained ICA for robust separation of sound mixtures," in *ICA'03*, 2003, pp. 951–956.

[223] T. Mei and F. Yin, "Blind separation of convolutive mixtures by decorrelation," *Signal Processing, Elsevier*, vol. 84, no. 12, pp. 2297–2213, Nov 2004.

[224] S. Rickard and O. Yilmaz, "On the approximate W-disjoint orthogonality of speech," in *ICASSP'02*, vol. I, 2002, pp. 529–532.

[225] S. Rickard, T. Melia, and C. Fearon, "DESPRIT - histogram based blind source separation of more sources than sensors using subspace methods," in *WASPAA'05*, 2005, pp. 5–8.

[226] A. Jourjine, S. Rickard, and O. Yilmaz, "Blind separation of disjoint orthogonal signals: Demixing N sources from 2 mixtures," in *ICASSP'00*, vol. V, 2000, pp. 2985–2988.

[227] S. Araki, S. Makino, H. Sawada, and R. Mukai, "Reducing musical noise by a fine-shift overlap-add method applied to source separation using a time-frequency mask," in *ICASSP'05*, vol. III, 2005, pp. 81–84.

[228] M. S. Pedersen, D. Wang, J. Larsen, and U. Kjems, "Overcomplete blind source separation by combining ICA and binary time-frequency masking," in *MLSP'05*, 2005.

[229] M. S. Pedersen, D. L. Wang, J. Larsen, and U. Kjems, "Separating underdetermined convolutive speech mixtures," in *ICA'06*, 2006, pp. 674–681.

[230] H. Sawada, S. Araki, R. Mukai, and S. Makino, "Blind extraction of a dominant source signal from mixtures of many sources," in *ICASSP'05*, vol. III, 2005, pp. 61–64.

[231] H.-C. Wu, J. C. Principe, and D. Xu, "Exploring the time-frequency microstructure of speech for blind

source separation," in *ICASSP'98*, vol. 2, 1998, pp. 1145–1148.

[232] J. M. Peterson and S. Kadambe, "A probabilistic approach for blind source separation of underdetermined convolutive mixtures," in *ICASSP'03*, vol. 6, 2003, pp. VI–581–584.

[233] D. Luengo, I. Santamaria, L. Vielva, and C. Pantaleon, "Underdetermined blind separation of sparse sources with instantaneous and convolutive mixtures," in *NNSP'03*, 2003, pp. 279–288.

[234] S. A. Abdallah and M. D. Plumbley, "Application of geometric dependency analysis to the seperation of convolved mixtures," in *ICA'04*, 2004, pp. 540–547.

[235] M. Babaie-Zadeh, A. Mansour, C. Jutten, and F. Marvasti, "A geometric approach for separating several speech signals," in *ICA'04*, 2004, pp. 798–806.

[236] Y. Li, A. Cichocki, and L. Zhang, "Blind source estimation of FIR channels for binary sources: A grouping decision approach," *Signal Processing, Elsevier*, vol. 84, no. 12, pp. 2245–2263, Nov 2004.

[237] B. A. Pearlmutter and A. M. Zador, "Monaural source separation using spectral cues," in *ICA'04*, 2004, pp. 478–485.

[238] P. Smaragdis, "Non negative matrix factor deconvolution, extraction of multiple sound sources from monophonic inputs," in *ICA'04*, 2004, pp. 494–499.

[239] T. Virtanen, "Separation of sound sources by convolutive sparse coding," in *SAPA'04*, 2004.

[240] M. S. Pedersen, T. Lehn-Schiøler, and J. Larsen, "BLUES from music: BLind Underdetermined Extraction of Sources from Music," in *ICA'06*, 2006, pp. 392–399.

[241] A. S. Bregman, *Auditory Scene Analysis*, 2nd ed. MIT Press, 1990.

[242] M. Weintraub, "The GRASP sound separation system," in *ICASSP'84*, 1984, pp. 69–72.

[243] R. R. Guddeti and B. Mulgrew, "Perceptually motivated blind source separation of convolutive mixtures," in *ICASSP'05*, vol. V, 2005, pp. 273–276.

[244] A. K. Barros, T. Rutkowski, F. Itakura, and N. Ohnishi, "Estimation of speech embedded in a reverberant and noisy environment by independent component analysis and wavelets," *IEEE Trans. Neural Networks*, vol. 13, no. 4, pp. 888–893, Jul 2002.

[245] M. Furukawa, Y. Hioka, T. Ema, and N. Hamada, "Introducing new mechanism in the learning process of FDICA-based speech separation," in *IWAENC'03*, 2003, pp. 291–294.

[246] R. D. Patterson, "The sound of a sinusoid: Spectral models," *J. Acoust. Soc. Am.*, vol. 96, pp. 1409–1418, May 1994.

[247] T. Rutkowski, A. Cichocki, and A. K. Barros, "Speech enhancement from interfering sounds using CASA techniques and blind source separation," in *ICA'01*, 2001, pp. 728–733.

[248] N. Roman, D. Wang, and G. J. Brown, "Speech segregation based on sound localization," *J. Acoust. Soc. Am.*, vol. 114, no. 4, pp. 2236–2252, Oct 2003.

[249] T. Nishikawa, H. Saruwatari, and K. Shikano, "Blind source separation of acoustic signals based on multistage ICA combining frequency-domain ICA and time-domain ICA," *IEICE Trans. Fundamentals*, vol. E86-A, no. 4, pp. 846–858, Sep 2003.

[250] P. Smaragdis, "Efficient blind separation of convolved sound mixtures," in *WASPAA'97*, Oct 19–22 1997.

[251] M. Davies, "Audio source separation," in *Mathematics in Signal Processing V*. Oxford University Press, 2001.

[252] F. Duplessis-Beaulieu and B. Champagne, "Fast convolutive blind speech separation via subband adaptation," in *ICASSP'03*, vol. 5, 2003, pp. V–513–516.

[253] C. Servière, "Separation of speech signals under reverberant conditions," in *EUSIPCO'04*, 2004, pp. 1693–1696.

[254] T.-W. Lee, A. J. Bell, and R. H. Lambert, "Blind separation of delayed and convolved sources," in *NIPS*, vol. 9, 1997, pp. 758–764.

[255] T.-W. Lee, A. Ziehe, R. Orglmeister, and T. J. Sejnowski, "Combining time-delayed decorrelation and ICA: towards solving the cocktail party problem," in *ICASSP'98*, vol. 2, 1998, pp. 1249–1252.

[256] A. Westner and V. M. Bove Jr., "Blind separation of real world audio signals using overdetermined mixtures," in *ICA'99*, 1999.

[257] K. Kokkinakis and A. K. Nandi, "Multichannel blind deconvolution for source separation in convolutive mixtures of speech," *IEEE Trans. Audio, Speech, Lang. Proc.*, vol. 14, no. 1, pp. 200–212, Jan. 2006.

[258] H. Sawada, R. Mukai, S. F. G. M. de la Kethulle de Ryhove, S. Araki, and S. Makino, "Spectral smoothing for frequency-domain blind source separation," in *IWAENC'03*, 2003, pp. 311–314.

[259] H. Sawada, R. Mukai, S. Araki, and S. Makino, "Convolutive blind source separation for more than two sources in the frequency domain," in *ICASSP'04*, vol. III, 2004, pp. 885–888.

[260] D. W. E. Schobben and P. C. W. Sommen, "A new blind signal separation algorithm based on second-order statistics," in *IASTED SIP'06*, 1998, pp. 564–569.

[261] H. Attias, J. C. Platt, A. Acero, and L. Deng, "Speech denoising and dereverberation using probabilistic models," *NIPS'01*, vol. 13, 2001.

[262] R. Aichner, H. Buchner, and W. Kellermann, "A novel normalization and regularization scheme for broadband convolutive blind source separation," in *ICA'06*, 2006, pp. 527–535.

[263] H. Sawada, S. Araki, R. Mukai, and S. Makino, "Blind source separation with different sensor spacing and filter length for each frequency range," in *NNSP'02*, 2002, pp. 465–474.

[264] P. Smaragdis, "Blind separation of convolved mixtures in the frequency domain," in *Neurocomp.*, ser. 1–3, vol. 22, Nov 1998, pp. 21–34.

[265] V. C. Soon, L. Tong, Y. F. Huang, and R. Liu, "A wideband blind identification approach to speech acquisition using a microphone array," in *ICASSP'92*, vol. 1, 1992, pp. 293–296.

[266] S. Kurita, H. Saruwatari, S. Kajita, K. Takeda, and F. Itakura, "Evaluation of frequency-domain blind signal separation using directivity pattern under reverberant conditions," in *ICASSP'00*, 2000, pp. 3140–3143.

[267] F. Asano, S. Ikeda, M. Ogawa, H. Asoh, and N. Kitawaki, "Combined approach of array processing and independent component analysis for blind separation of acoustic signals," *IEEE Trans. Speech Audio Proc.*, vol. 11, no. 3, pp. 204–215, May 2003.

[268] N. Mitianoudis and M. E. Davies, "Permutation alignment for frequency domain ICA using subspace beamforming methods," in *ICA'04*, 2004, pp. 669–676.

[269] W. Baumann, B.-U. Köhler, D. Kolossa, and R. Orglmeister, "Real time separation of convolutive mixtures," in *ICA'01*, 2001, pp. 65–69.

[270] H. Gotanda, K. Nobu, T. Koya, K. ichi Kaneda, T. aki Ishibashi, and N. Haratani, "Permutation correction and speech extraction based on split spectrum through fastICA," in *ICA'03*, 2003, pp. 379–384.

[271] S. Araki, S. Makino, R. Aichner, T. Nishikawa, and H. Saruwatari, "Subband based blind source separation with appropriate processing for each frequency band," in *ICA'03*, 2003, pp. 499–504.

[272] J. Anemüller and B. Kollmeier, "Amplitude modulation decorrelation for convolutive blind sourc separation," in *ICA'00*, 2000, pp. 215–220.

[273] J. Antoni, F. Guillet, M. El Badaoui, and F. Bonnardot, "Blind separation of convolved cyclostationary processes," *Signal Processing, Elsevier*, vol. 85, no. 1, pp. 51–66, Jan 2005.

[274] A. Dapena and C. Serviere, "A simplified frequency-domain approach for blind separation of convolutive mixtures," in *ICA'01*, 2001, pp. 569–574.

[275] C. Mejuto, A. Dapena, and L. Castedo, "Frequency-domain infomax for blind separation of convolutive mixtures," in *ICA'00*, 2000, pp. 315–320.

[276] N. Mitianoudis and M. Davies, "New fixed-point ICA algorithms for convolved mixtures," in *ICA'01*, 2001, pp. 633–638.

[277] I. Lee, T. Kim, and T.-W. Lee, "Complex fastiva: A robust maximum likelihood approach of mica for convolutive bss," in *ICA'06*, 2006, pp. 625–632.

[278] T. Kim, H. Attias, S.-Y. Lee, and T.-W. Lee, "Blind source separation exploiting higher-order frequency dependencies," *IEEE Trans. Audio, Speech, Lang. Proc.*, vol. 15, no. 1, Jan. 2007.

[279] S. Dubnov, J. Tabrikain, and M. Arnon-Targan, "A method for directionally-disjoint source separation in convolutive environment," in *ICASSP'04*, vol. V, 2004, pp. 489–492.

[280] A. Hiroe, "Solution of permutation problem in frequency domain ica, using multivariate probability density functions," in *ICA'06*, 2006, pp. 601–608.

[281] D. Kolossa, B. uwe Khler, M. Conrath, and R. Orgelmeister, "Optimal permutation correlation by multiobjective genetic algorithms," in *ICA'01*, 2001, pp. 373–378.

[282] K. Kamata, X. Hu, and H. Kobataka, "A new approach to the permutation problem in frequency domain blind source separation," in *ICA'04*, 2004, pp. 849–856.

[283] H. Attias and C. E. Schreiner, "Blind source separation and deconvolution: The dynamic component analysis algorithm," *Neural Computation*, vol. 11, pp. 803–852, 1998.

[284] F. Asano, S. Ikeda, M. Ogawa, H. Asoh, and N. Kitawaki, "Blind source separation in reflective sound fields," in *HSC'01*, Apr 9–11 2001.

[285] H. Sawada, S. Araki, R. Mukai, and S. Makino, "On calculating the inverse of separation matrix in frequency-domain blind source separation," in *ICA'06*, 2006, pp. 691–699.

[286] V. C. Soon, L. Tong, Y. F. Huang, and R. Liu, "A robust method for wideband signal separation," in *ISCAS'93*, 1993, pp. 703–706.

[287] R. Mukai, S. Araki, H. Sawada, and S. Makino, "Re-

moval of residual cross-talk components in blind source separation using LMS filters," in *NNSP'02*, 2002, pp. 435–444.

[288] H. Saruwatari, S. Kurita, K. Takeda, F. Itakura, and K. Shikano, "Blind source separation based on sub-band ICA and beamforming," in *ICSLP'00*, vol. III, 2000, pp. 94–97.

[289] S. Y. Low, S. Nordholm, and R. Togneri, "Convolutive blind signal separation with post-processing," *IEEE Trans. Speech, Audio Proc.*, vol. 12, no. 5, pp. 539–548, Sep 2004.

[290] H. Saruwatari, T. Kawamura, T. Nishikawa, A. Lee, and K. Shikano, "Blind source separation based on a fast-convergence algorithm combining ICA and beamforming," *IEEE Trans. Audio, Speech, Lang. Proc.*, vol. 14, no. 2, pp. 666–678, Mar 2006.

[291] R. Mukai, H. Sawada, S. Araki, and S. Makino, "Near-field frequency domain blind source separation for convolutive mixtures," in *ICASSP'04*, vol. IV, 2004, pp. 49–52.

[292] R. O. Schmidt and R. E. Franks, "Multiple Source DF Signal Processing: An Experimental System," *IEEE Trans. Ant. Prop.*, vol. AP-34, no. 3, pp. 281–290, Mar 1986.

[293] N. Murata and S. Ikeda, "An on-line algorightm for blind source separation on speech signals," in *International Symposium on Theory and its Applications*, vol. 3, 1998, pp. 923–926.

[294] F. Asano and S. Ikeda, "Evaluation and real-time implementation of blind source separation system using time-delayed decorrelation," in *ICA'00*, 2000, pp. 411–416.

[295] T. Kim, T. Eltoft, and T.-W. Lee, "Independent vector analysis: An extension of ICA to multivariate components," in *ICA'06*, 2006, pp. 165–172.

[296] T. Hoya, A. K. Barros, T. Rutkowski, and A. Cichocki, "Speech extraction based upon a combined subband independent component analysis and neural memory," in *ICA'03*, 2003, pp. 355–360.

[297] S. Ukai, H. Saruwatari, T. Takatani, and R. Mukai, "Multistage SIMO-model-based blind source separation combining frequent-domain ICA and time-domain ICA," in *ICASSP'04*, vol. IV, 2004, pp. 109–112.

[298] S. Araki, S. Makino, A. Blin, R. Mukai, and H. Sawada, "Blind separation of more speech than sensors with less distortion by combining sparseness and ICA," in *IWAENC'03*, 2003, pp. 271–274.

[299] R. Mukai, H. Sawada, S. Arakiand, and S. Makino, "Blind source separation for moving speech signals using blockwise ICA and residual crosstalk sub-traction," *IEICE Trans. Fundamentals*, vol. E87-A, no. 8, pp. 1941–1948, Aug 2004.

[300] R. Mukai, H. Sawada, S. Araki, and S. Makino, "Frequency domain blind source separation for many speech signals," in *ICA'04*, 2004, pp. 461–469.

[301] Y. Mori, H. Saruwatari, T. Takatani, K. Shikano, T. Hiekata, and T. Morita, "ICA and binary-mask-based blind source separation with small directional microphones," in *ICA'06*, 2006, pp. 649–657.

[302] T. Nishikawa, H. Saruwatari, and K. Shikano, "Stable learning algorithm for blind separation of temporally correlated signals combining multistage ICA and linear prediction," in *ICA'03*, 2003, pp. 337–342.

[303] T. Takatani, T. Nishikawa, H. Saruwatari, and K. Shikano, "Blind separation of binaural sound mixtures using SIMO-model-based independent component analysis," in *ICASSP'04*, vol. IV, 2004, pp. 113–116.

Index