

BLIND SEPARATION FOR AUDIO SIGNALS - ARE WE THERE YET?

Kari Torkkola

Motorola, Phoenix Corporate Research Labs
2100 East Elliot Road, MD EL508, Tempe, AZ 85284, USA
a540aa@email.mot.com <http://members.home.net/torkkola>

ABSTRACT

We attempt to give an overview of current research in blind separation of convolutive mixing of signals, concentrating on audio signals, and methods applicable thereof. We briefly enumerate some application areas, we present two possible taxonomies of separation methods, one based on system parametrizations, the other on different criteria to solve the problem. We wade through the literature following these taxonomies. We also discuss what might yet be missing in the current research.

1. INTRODUCTION

We presuppose familiarity with the concepts of Independent Component Analysis (ICA) and Blind Signal Separation (BSS). If this is not the case the reader is advised to refer to [1, 8, 15, 20, 37].

BSS seemingly has a large number of potential applications in the audio realm. The generic application is, of course, separation of simultaneous audio sources in reverberating or echoing environment, that is, in a natural environment, for example, inside a room. We will enumerate here only a few actual applications, the reader is free to use her imagination to develop more.

A very desirable application area would be signal enhancement by removing noise or other unwanted signal components using blind separation methods as in [31], for example. In this area only one signal is of interest, the rest is considered as nuisance. Enhancement of voice quality in mobile phones, would be one important application, especially in car environments. Since voice coders used in cell phones are optimized for coding speech alone, the combination of excessive noise with a speech signal results in a poor sound quality. Some initial experiments in this area can be found in [78].

Making voice dialling or speech recognition in general more viable in noisy environments would fall into the same category [115, 48]. Spying, intelligence, or forensic applications fall also under the same category whereby the interest might be in picking up one important signal amongst others [69].

In audio communications transparency refers to reproduced audio being ideally free from reverberation, noise, acoustical echoes, and mixed other speakers [82]. Teleconferencing and speakerphones are two areas where speech signal acquisition with transparency is desirable. Combining existing multi-channel acoustical echo cancellation technology with BSS has been shown to be useful in a teleconferencing setup [82].

Hearing aids are also another lucrative application area for speech enhancement through BSS.

Whether some of the above can yet be a viable and profitable application area, is an open question and will be touched upon in the concluding section. Current limitations in the methods might render some applications if not impossible, at least impractical.

Besides audio, an extremely fruitful application arena is digital communications. While the basic concept remains the same (mul-

tiplex transmitters at same frequency, multiple antennas receiving multiple mixtures, for example), there are a few important differences. Signals in this context are man-made and thus their properties are completely known in advance. This can (and should be) be exploited in devising separation methods. Another difference is that signals could be transmitted in short bursts, which might call for block-based algebraic methods rather than adaptive methods [102]. The paper at hand, however, concentrates in methods applicable to audio.

The main contribution of the paper is an attempt to collect a major part of the literature pertinent to BSS and audio, and to present that literature in light of two possible taxonomies of separation methods. These are based on how to parametrize the separating structure, and on the criteria used for separation. Due to the breadth of the literature and page limitations, the presentation cannot be but superficial. We also discuss what the technology yet might lack to produce successful applications.

2. PARAMETRIZATION

Any method for separation of convolutive mixtures can be roughly divided in three essential components: 1) parametrization of the separation system (filters or matrices), 2) the separation criterion, and 3) the method to optimize the chosen criterion. We concentrate in looking at the first two components, and mention only briefly that the optimization methods can be coarsely divided into adaptive and algebraic approaches, and the former category can further be subdivided into stochastic gradient type algorithms (with or without 2nd order information, i.e. Newton's method), and function zero search algorithms or fixed point methods. The latter category mainly consists of methods to jointly and/or approximately diagonalize a number of matrices. In this section we discuss what alternatives exist for the parametrization of the separating system and possibly for the parametrization of the source signals if the method at hand requires this.

2.1. Feedforward, Feedback?

Most real convolutive mixing scenarios with audio can be modeled as a feedforward mixing network having FIR filters in its branches. A room with multiple simultaneous sound sources and multiple microphones is an example, where the mixing filters are room impulse responses between each source and each microphone.

The separation system, ideally inverting the effect of the mixing system, can also be modeled as a feedforward network of FIR-filters that approximate the required inverse filters. Consider a simple 2x2 mixing case in the z-domain:

$$\begin{aligned} X_1(z) &= A_{11}(z)S_1(z) + A_{12}(z)S_2(z) \\ X_2(z) &= A_{21}(z)S_1(z) + A_{22}(z)S_2(z), \end{aligned} \quad (1)$$

Solving for the sources gives the ideal solution:

$$\begin{aligned} S_1(z) &= (A_{22}(z)X_1(z) - A_{21}(z)X_2(z))/G(z) \\ S_2(z) &= (-A_{12}(z)X_1(z) + A_{11}(z)X_2(z))/G(z) \end{aligned} \quad (2)$$

where $G(z) = A_{12}(z)A_{21}(z) - A_{11}(z)A_{22}(z)$. In a feedforward filter separation system with W_{ij} operating on X_j , for example, $W_{11}(z) = A_{22}(z)/G(z)$ to reproduce the exact source. However, the convolutive separation problem has the indeterminacy upto arbitrary filtering of the sources. Depending on the learning algorithm, the sources will be distorted by filtering. Some separation methods tend to produce temporally whitened outputs as they aim at redundancy reduction. In the feedforward case it would be actually very desirable to learn filters without the $G(z)^{-1}$ factor as the filters to be learned are the same mixing filters and not their inverses (the adjunct of the mixing, that is). There is not, though, an effective way of achieving this.

Another possible arrangement is the feedback architecture, which in the 2x2 case is

$$\begin{aligned} U_1(z) &= X_1(z) - W_{12}(z)U_2(z) \\ U_2(z) &= X_2(z) - W_{21}(z)U_1(z), \end{aligned} \quad (3)$$

W_{12} and W_{21} have the following ideal separation solution:

$$\begin{aligned} W_{12}(z) &= A_{12}(z)A_{22}(z)^{-1} \\ W_{21}(z) &= A_{21}(z)A_{11}(z)^{-1}. \end{aligned} \quad (4)$$

Note that the inverses of the *direct* mixing filters are required. If these are minimum-phase filters they have stable causal inverses. That is, if the direct paths are “good” a feedback network with causal filters is able to invert the mixing. Otherwise one must use the feedforward network with acausal filters or implement acausal filters within the feedback network as described in [35].

2.2. Frequency domain

The parameters to be learned above are the time-domain coefficients of the filters. However, in the audio case the filters may need to be thousands of taps long to properly invert the mixing. Computationally it may be lighter to move to the frequency domain as convolutions with long filters in the time domain become efficient multiplications in the frequency domain under certain conditions [65, 72]. Now there are two avenues to take. In the first one everything including the actual separation could be done in the frequency domain. This has the great advantage of decomposing a convolutive mixing problem into multiple instantaneous mixing problems (i.e., ICA), that can be solved using any desired method. The downside is that now the standard ICA indeterminacy of scaling and permutation appears at each output frequency bin! Reconstruction of the time-domain output signal requires all frequency components of the same source. Various methods to overcome the scaling and permutation problem using different continuity criteria are presented in [92] and [13, 86, 85, 91, 65, 70, 72, 110].

The second avenue is that the actual separation is not done in the frequency domain but only one or some aspects of the separation algorithm. The rest is done in the time domain. Filters may be easier to learn in the frequency domain as components are now orthogonal and not dependent on each other like the time domain coefficients [7, 81]. Examples of methods that apply their separation criterion (independence, HOS, nonlinearities) in the time domain but do the rest in the frequency domain are reported in [7, 42, 48]. A frequency domain representation of the filters is learned, and they are also applied in the frequency domain. The final time-domain result is reconstructed using standard, e.g., overlap-save signal processing techniques. Thus, the permutation and scaling

problem does not exist. An example of learning a filter of matrices in time domain, when the criterion is in frequency domain is presented in [58]. Back et al. also present an example of the H-J-algorithm in the frequency domain, whereas the nonlinear functions are applied in the time domain [7].

Another type of parametrization is presented in [93]. The source location is parametrized, and frequency bin information is clustered to produce consistent source location estimates. Alternative discrete time operators are considered in [6].

2.3. Decomposition

Rather than trying to learn these possibly huge filters all at once, it is possible to decompose the problem [113, 68]. At this point it is useful to discuss the relation between independent component analysis (ICA) and blind separation of convolutive mixtures (BSCM). ICA (in the separation context) makes use of *spatial* statistics of the mixture signals to learn a *spatial* separation system. For stationary sources, ICA needs to use higher than 2nd order spatial statistics to succeed. However, if the sources can be assumed to be nonstationary 2nd order spatio-temporal statistics is sufficient as shown in [97]. In contrast, BSCM needs to make use of *spatio-temporal* statistics of the mixture signals to learn a *spatio-temporal* separation system. Stationarity of the sources is decisive for BSCM, too. If the sources are not stationary, only 2nd order spatio-temporal statistics is enough as briefly discussed in [105] and later, for example, in [70, 65]. Stationary sources require again higher than 2nd order statistics, but in the following fashion. Spatio-temporal 2nd order statistics can be made use of to decorrelate the mixtures. This step returns the problem to that of conventional ICA, which again requires higher-order spatial statistics. Examples of these approaches are given in [39, 23, 34] with linear prediction based methods, in [18] with an adaptive approach, and in [52] with a beamforming approach.

Alternatively, for sources that cannot be assumed non-stationary one can resort to higher-order spatio-temporal statistics from the beginning as done in a lot of papers.

Another way to decompose the problem is presented in [67]. In this case the microphone arrangement is rather peculiar, a compact microphone array: two omnidirectional microphones 1cm apart in a reverberant environment. Now, for each source, the transfer functions differ predominantly by a small delay. Mixtures are modeled as $x = DRs$, where D is a matrix of bandlimited approximations of delta functions at some delays, and R represents all other acoustic effects. First stage estimates delays \hat{D}

$$y = \text{adj}(\hat{D})x = \text{adj}(\hat{D})DRs = Hs \quad (5)$$

Second stage cancels off-diagonal elements of H (small relative to diagonal elements) using a feedback configuration

$$z = y - Mz = (M + I)^{-1}Hs. \quad (6)$$

Now H is easier to estimate than DR directly since its off-diagonal elements are small: Initial guess $M = 0$ is likely to lie near a global optimum. Simple decorrelation adaptation is then sufficient to learn M .

2.4. Sources parametrized, too

Some separation methods require assumptions about the sources (see Sec. 3.3) and parametrize some aspects of them, such as the densities, some parameters thereof [64, 31, 5], or some parameters related to temporal statistics of the sources, such as AR parameters [73, 71, 49].

2.5. System identification approach

With room acoustics the inverse system generally contains more parameters than the mixing system. This would suggest rather learning the mixing system than the separation system.

3. SEPARATION CRITERIA

The actual separation criterion can be quite independent of the chosen parametrization. In this section we go through a major part of them that have appeared in the literature in the audio context.

3.1. Minimizing mutual information

Independence criterion can be based on factoring the marginal output density.

$$p_u(\mathbf{u}) = \prod_i p_{u_i}(u_i). \quad (7)$$

Divergence between the two sides of (7) reflects thus the independence of the u_i . For the two sides of (7) the divergence is

$$\begin{aligned} D(p_u(\mathbf{u}), \prod_i p_{u_i}(u_i)) &= \int p_u(\mathbf{u}) \log \frac{p_u(\mathbf{u})}{\prod_i p_{u_i}(u_i)} du \quad (8) \\ &= -H(U; W) + \sum_i H(U_i; W), \end{aligned}$$

the joint entropy of output minus the sum of entropies of individual components, which is also the mutual information between the output components. Minimizing this expression results in statistical independence of the output components. If W is just a matrix, and $u = Wx$ then

$$H(U; W) = H(X) + \log \det(W) \quad (9)$$

Minimizing this mutual information requires entropies of individual components, which are not available, but can be estimated by means of statistical expansions [20, 1, 36] or other methods [38]. In the convolutive mixture context MMI has been used in [58, 19, 17].

3.2. InfoMax, or Entropy maximization

Consider now passing the outputs through bounded nonlinear functions, that approximate the cumulative densities of the sources $y = g(Wx)$. If outputs u are the separated true sources, the y have density close to a uniform density, the density that has the largest entropy among bounded distributions. Maximizing the entropy of y is thus equivalent of producing the true sources. This criterion makes use of the source densities implicitly.

This criterion was applied to delayed and convolutive mixing in [99, 98] and further developed in [19, 47, 112, 2, 3, 27, 35]. Fully or partially frequency domain approaches were developed in [49, 48, 41, 90, 89, 91].

3.3. Latent source models and ML

Source densities can be made use of explicitly, too. If they are known, or if their shapes are known less perhaps some parameters that need to be estimated while W is learned, the divergence between the marginal output density and the known source densities can be used as the criterion to be minimized, that is, find W that drives the outputs towards known distributions. This approach lends itself naturally to the maximum likelihood estimation [75, 9, 14]. With convolutive mixing this approach has been taken in [71, 73, 64, 28, 31].

Even if the densities are unknown we can estimate them as sums of some simple, for example Gaussian or logistic distributions while the separation matrix is being estimated. Either gradient ascent [74] or expectation-maximization in some cases [10, 64] can then be used. This situation lends itself also to the following interpretation. The sources can be interpreted as latent signals whose convolutive mixture is the only observable [5, 4]. General literature on latent models and their estimation is then applicable.

3.4. Bussgang approach

Bussgang methods have been used as tools for blind deconvolution, where the true source is estimated from a signal corrupted by convolutional noise by a nonlinear function which optimally equals $g(s) = (\partial p_s(s)/\partial s)/p_s(s)$. Equalization algorithms are derived by finding a filter that minimizes the difference between the output and the true source estimated through g . LMS is typically the criterion. Extensions for multichannel blind deconvolution and separation were presented by Lambert [42, 43, 45]. Coupled with FIR-matrix algebra [46, 43] efficient separation methods seem to result. See also [106, 107] for application of these methods for the overdetermined mixing case. It is notable that the non-linearity has the same exact form as in the entropy maximization, and maximum likelihood approaches, leading to similar separation algorithms [50].

3.5. Central Limit Theorem

Assumption of non-Gaussian sources enables us to apply the central limit theorem: The sum of a “small” number of non-Gaussian signals has a distribution that is “closer” to a Gaussian than the densities of the sources. The distance of the marginal density of the output p_u to a Gaussian can be expressed by using the divergence, which is equivalent to the differential entropy of a Gaussian (which has the same mean and covariance as p_u) minus the differential entropy of p_u . This measure is also called negentropy. From this measure it is straightforward to derive a stochastic gradient ascent adaptation rule to find a W that maximizes this distance and thus separates the sources, as shown by Girolami in [33].

Deviation from Gaussianity can also be measured directly by higher order statistics since these are zero for Gaussians. For example, kurtosis, which is written as $E\{u^4\} - 3(E\{u^2\})^2$ for a zero mean variable, could be used. Both kurtosis and negentropy require us to know to which direction we are driving the output densities from a Gaussian. For example, densities that have negative kurtosis (flat densities like uniform distribution) require the kurtosis of the output signal to be minimized whereas for positively kurtotic (sharply peaked) signals kurtosis needs to be maximized. Either we need to know the signs of kurtoses of our sources, or we need to estimate them while learning the W as discussed in [33].

3.6. Minimization of cross-HOS

As mentioned in Section 2.3 it is possible use the traditional ICA-methods by constructing the criteria for separation from high-order statistics, but by doing it spatio-temporally. One can construct an objective function, usually from fourth order cross-cumulants, and minimize this using stochastic gradient descent on the filters [30, 66, 114, 22, 100], or a contrast function as in [63]. Alternatively, one can construct a simple algorithm that aims just at canceling the criteria [94]. HOS can also be used implicitly through nonlinear functions. Examples and analysis are given in [95, 40, 94, 25, 7, 57, 24]. Platt and Faggin first derived similar rules employing nonlinear functions, but from the minimum output power principle [76]. Approaches that utilize HOS in the frequency domain are

described in [13, 84, 86, 87]. An algebraic approach is presented in [96].

In general, estimation of higher-order statistics is more sensitive to noise and outliers than that of 2nd order statistics. Thus, based on this argument it would seem to be more robust to work on 2nd order statistics as much as possible.

3.7. Spatio-Temporal decorrelation

Let us look at spatial covariance in instantaneous separation

$$R_x(t) = AR_s(t)A^T + R_{noise}(t), \quad x(t) = As(t) + n(t) \quad (10)$$

If s is non-stationary, that is, $R_s(t) \neq R_s(t + \tau)$ we get multiple conditions for different choices of τ to solve for A , $R_s(t)$, and $R_{noise}(t)$, where the covariances are diagonal matrices.

For convolutive mixtures we may look at cross covariances over time $R_x(t, t + \tau) = E\{x(t)x(t + \tau)^T\}$. This approach was mentioned in [105] and utilized in [61] and [59].

In frequency domain for sample averages we can write

$$\bar{R}_x(\omega, t) = \mathbf{A}(\omega)\mathbf{R}_s(\omega, t)\mathbf{A}^H(\omega) + \mathbf{R}_{noise}(\omega, t) \quad (11)$$

Again, if s is non-stationary we can write multiple linearly independent equations for different time lags and solve for unknowns or find LMS estimates of them by diagonalizing a number of matrices in frequency domain [29, 108, 65, 70, 72, 60, 110].

For minimum-phase mixing decorrelation only can provide a unique solution without having to make use of the non-stationarity [12, 55, 54, 88]. There is a multitude of adaptive approaches [101, 103, 104, 105, 57, 55, 16, 115, 26, 78, 79, 81, 82], a few algebraic [56], and many that are derived from anti-Hebbian learning rule considerations [59, 77, 32, 18, 31].

4. DISCUSSION

Despite the seeming scope of research good results in realistic scenarios are hard to come by. What makes it difficult to move BSS into real world applications? Are there some inherent limitations in the audio setup that make it an extremely hard (if indeed solvable) problem? We pose some questions and try to present some directions in this section.

Are all assumptions really true? Some methods assume stationarity of the sources, some do not. Speech is certainly nonstationary in larger time scale, but quasi-stationary in a short scale. Derivation of some methods might expect doubly-infinite filters, which can only be approximated in reality. This approximation might cause local minima in optimizing the chosen criterion [88].

Are there always (much) more sources than sensors? An intriguing application would be to separate speech from car noise in mobile communications. However, one quickly realizes that there are multiple noise sources, in fact an infinite number, since the whole car interior is vibrating and acts as a delocalized noise source. A 2x2 speech-noise separation system would not work in this case.

Is the reality too dynamic? In general, the reported performance figures for static sources (loudspeakers) are higher than figures for real people, even in seemingly static positions. The effect of a speaker turning her head 10-20 degrees, or leaning backwards a couple of inches can have a drastic effect to the impulse response between the speaker and the microphones. These effects have been studied in acoustics, and further cross-fertilization between the fields would seem to be necessary to establish bounds to what performance could be expected from BSS in situations involving live speakers. Bradstein simulates these effects in [11] and concludes that "Any system which attempts to estimate the reverberation effects and apply some means of inverse filtering would

have to be adaptable on almost a frame-by-frame basis to be effective." This gives quite a pessimistic view for applications that involve on-line adaptation to dynamic situations.

Related to the previous question is how much data do we really need to converge? Are there too many parameters? Filters with thousands of taps in the time domain need tens (or hundreds) of thousands of samples to converge. That might be too much for the application.

What else can be done? It would be reasonable to combine different approaches to make use of every bit of available knowledge. For example, we should combine the nonstationarity-exploiting decorrelation approaches with approaches that make use of the source densities. Initial attempts are reported in [51]. A promising approach seems to be decomposing the problem into smaller and independent subproblems instead of trying to solve all at once with a huge number of parameters.

The sensitivity to (unavoidable) noise is much examined in communications but has not been studied enough in the audio context. Some studies or methods taking noise into account are presented in [21, 62, 64, 72].

Using more microphones than there are sources to separate should in theory be able to improve the noise tolerance, in addition to separation quality. Westner, though, reports rather pessimistic results in his initial experiments [106].

An avenue that has not been looked at enough in BSS is to make use of the fact that the target signal is a speech signal. In [11] Brandstein advocates explicitly incorporating the nature of the speech signal including non-stationarity, model of production, pitch, voicing, formant structure, and a model of source radiation, into a beamforming context. He feels that this is essential to realize the goal, high-quality speech signal acquisition from an unconstrained talker in a hands-free environment surrounded by interfering sources. This goal equals the goal of many BSS applications. Some recent BSS work to these directions is presented in [111, 109].

The last but not the least question is *What is the best method for separating audio signals mixed convolutively?* Given the breadth of the field and the lack of common criteria and databases this question remains unanswerable. It is possible to derive some theoretical bounds for some algorithms [80, 53, 113], but not for all of them. Empirical comparisons using agreed upon databases and measurements may be the only way to find partial answers [83, 44].

We conclude with these open questions, and state that as coping with artificial signals seems to be more straightforward than with natural signals, it might be that BSS with convolutive mixtures finds its first real applications in the area of communications.

5. REFERENCES

- [1] S. Amari, A. Cichocki, and H. H. Yang. A new learning algorithm for blind signal separation. In *Advances in Neural Information Processing Systems 8*, pages 757–763. MIT Press, 1996.
- [2] S. Amari, S. Douglas, A. Cichocki, and H. H. Yang. Multichannel blind deconvolution and equalization using the natural gradient. In *Proc. 1st IEEE Signal Processing Workshop on Signal Processing Advances in Wireless Communications*, pages 101–104, Paris, France, April 16-18 1997.
- [3] S. Amari, S. Douglas, A. Cichocki, and H. H. Yang. Novel on-line adaptive learning algorithms for blind deconvolution using the natural gradient algorithm. In *Proc. 11th IFAC Symposium on System Identification*, pages 1057–1062, Kitakyushu City, Japan, July 1997.
- [4] H. Attias and C. Schreiner. Blind source separation and deconvolution by dynamic component analysis. In *IEEE Workshop on Neural Networks for Signal Processing*, pages 456–465, Amelia Island, FL, USA, September 1997.
- [5] H. Attias and C. Schreiner. Blind source separation and deconvolution: The dynamic component analysis algorithm. *Neural Computation*, 10:1373–1424, 1998.
- [6] A. D. Back and A. Cichocki. Blind source separation and deconvolution of fast sampled signals. In *Proc. ICONIP*, 1997.

- [7] A. D. Back and A. C. Tsoi. Blind deconvolution of signals using a complex recurrent network. In *IEEE Workshop on Neural Networks for Signal Processing*, pages 565–574, 1994.
- [8] A. Bell and T. Sejnowski. An information-maximisation approach to blind separation and blind deconvolution. *Neural Computation*, 7(6):1129–1159, 1995.
- [9] A. Belouchrani and J.-F. Cardoso. Maximum likelihood source separation for discrete sources. In *Signal Processing VII: Theories and Applications (Proc. of the EUSIPCO-94)*, pages 768–771, Edinburgh, Scotland, September 13-16 1994. Elsevier.
- [10] A. Belouchrani and J.-F. Cardoso. Maximum likelihood source separation by the expectation-maximization technique: Deterministic and stochastic implementation. In *Proc. NOLTA*, pages 49–53, Las Vegas, Nevada, USA, December 10-14 1995.
- [11] M. S. Brandstein. On the use of explicit speech modeling in microphone array applications. In *Proc. ICASSP*, pages 3613–3616, Seattle, WA, May 12-15 1998.
- [12] H. Broman, U. Lindgren, H. Sahlin, and P. Stoica. Source separation: A TITO system identification approach. Technical Report CTH-TE-33, Chalmers University of Technology, September 27 1995.
- [13] V. Capdevielle, C. Serviere, and J. Lacoume. Blind separation of wide-band sources in the frequency domain. In *Proc. ICASSP*, pages 2080–2083, Detroit, MI, May 9-12 1995.
- [14] J.-F. Cardoso and S.-I. Amari. Maximum likelihood source separation: equivariance and adaptivity. In *Proc. of SYSID'97, 11th IFAC symposium on system identification*, pages 1063–1068, Fukuoka, Japan, 1997.
- [15] J.-F. Cardoso and B. Laheld. Equivariant adaptive source separation. *IEEE Transactions on Signal Processing*, 44(12):3017–3030, December 1996.
- [16] D. B. Chan, P. J. W. Rayner, and S. J. Godsill. Multi-channel signal separation. In *Proc. ICASSP*, Atlanta, GA, May 7-10 1996.
- [17] S. Choi and A. Cichocki. Adaptive blind separation of speech signals: Cocktail party problem. In *Proc. International Conference on Speech Processing (ICSP'97)*, pages 617–622, Seoul, Korea, August 26-28 1997.
- [18] S. Choi and A. Cichocki. Blind signal deconvolution by spatio-temporal decorrelation and demixing. In *IEEE Workshop on Neural Networks for Signal Processing*, Amelia Island, FL, USA, September 1997.
- [19] A. Cichocki, S. i Amari, and J. Cao. Blind separation of delayed and convolved signals with self-adaptive learning rate. In *Proceedings of the International Symposium on Nonlinear Theory and Applications (NOLTA)*, pages 229–232, Katsurahama-so, Kochi, Japan, October 7-9 1996.
- [20] P. Comon. Independent component analysis – a new concept? *Signal Processing*, 36(3):287–314, 1994.
- [21] P. Comon. Contrasts for multichannel blind deconvolution. *IEEE Signal Processing Letters*, 3(7):209–211, July 1996.
- [22] S. Cruces and L. Castedo. A Gauss-Newton method for blind source separation of convolutive mixtures. In *Proc. ICASSP*, pages 2093–2096, Seattle, WA, May 12-15 1998.
- [23] N. Delfosse and P. Loubaton. Adaptive blind separation of convolutive mixtures. In *Proc. ICASSP*, pages 2940–2943, Atlanta, GA, May 7-10 1996.
- [24] Y. Deville and N. Charkani. Analysis of the stability of time-domain source separation algorithms for convolatively mixed signals. In *Proc. ICASSP*, Munich, Germany, April 1997.
- [25] A. Dinc and Y. Bar-Ness. A forward/backward bootstrapped structure for blind separation of signals in a multi-channel dispersive environment. In *Proc. ICASSP*, pages 376–379, Minneapolis, MN, USA, April 27-30 1993.
- [26] S. C. Douglas and A. Cichocki. Neural networks for blind decorrelation of signals. *IEEE Trans. on Signal Processing*, 45(11):2829–2842, November 1997.
- [27] S. C. Douglas, A. Cichocki, and S. ichi Amari. Multichannel blind separation and deconvolution of sources with arbitrary distributions. In *IEEE Workshop on Neural Networks for Signal Processing*, Amelia Island, FL, USA, September 1997.
- [28] S. C. Douglas and S. Haykin. On the relationship between blind deconvolution and blind source separation. In *31st Annual Asilomar Conf. on Signals, Systems, and Computers*, Pacific Grove, CA, USA, November 1997.
- [29] F. Ehlers and H. Schuster. Blind separation of convolutive mixtures and an application in automatic speech recognition in noisy environment. *IEEE Trans. on Signal Processing*, 45(10):2608, 1997.
- [30] A. M. Engebretson. Acoustic signal separation of statistically independent sources using multiple microphones. In *Proc. ICASSP*, volume II, pages 343–346, Minneapolis, MN, USA, April 27-30 1993.
- [31] M. Girolami. Noise reduction and speech enhancement via temporal anti-Hebbian learning. In *Proc. ICASSP*, Seattle, WA, USA, May 12-15 1998.
- [32] M. Girolami and C. Fyfe. A temporal model of linear anti-Hebbian learning. *Neural Processing Letters*, 4(3):139–148, 1996.
- [33] M. Girolami and C. Fyfe. Generalised independent component analysis through unsupervised learning with emergent Bussgang properties. In *Proc. ICNN*, Houston, TX, USA, 1997.
- [34] A. Gorokhov, P. Loubaton, and E. Moulines. Second order blind equalization in multiple input multiple output FIR systems: A weighted least squares approach. In *Proc. ICASSP*, pages 2417–2420, Atlanta, GA, May 7-10 1996.
- [35] Y. Guo, F. Sattar, and C. Koh. Blind separation temporomandibular joint sound signals. In *Proc. ICASSP*, Phoenix, AZ, April 1999. (submitted).
- [36] S. Haykin. *Neural Networks, A Comprehensive Foundation*. Macmillan Publishing and IEEE press, New York, 1998.
- [37] J. Herault and C. Jutten. Space or time adaptive signal processing by neural network models. In *Neural Networks for Computing, AIP Conf. Proc.*, volume 151, pages 206–211, Snowbird, UT, USA, 1986.
- [38] A. Hyvärinen. New approximations of differential entropy for independent component analysis and projection pursuit. In *Advances in Neural Information Processing Systems 10 (Proc. of NIPS'97)*. MIT Press, 1998. Also published as Report A47, Helsinki University of Technology, Laboratory of Computer and Information Science, August 1997.
- [39] S. Icart and R. Gautier. Blind separation of convolutive mixtures using second and fourth order moments. In *Proc. ICASSP*, pages 3018–3021, Atlanta, GA, May 7-10 1996.
- [40] C. Jutten, L. Nguyen Thi, E. Dijkstra, E. Vittoz, and J. Caelen. Blind separation of sources: An algorithm for separation of convolutive mixtures. In J. L. Lacoume, editor, *Higher order statistics: Proc. of the Int. Signal Processing Workshop on Higher Order Statistics*, pages 275–278, Chamrousse, France, July 10-12 1992. Elsevier.
- [41] B.-U. Kohler, T.-W. Lee, and R. Orglmeister. Improving the performance of Infomax using statistical techniques. In *Proc. Int. Conf on Artificial Neural Networks (ICANN)*, Lausanne, Switzerland, 1997.
- [42] R. H. Lambert. A new method for source separation. In *Proc. ICASSP*, pages 2116–2119, Detroit, MI, May 9-12 1995.
- [43] R. H. Lambert. *Multichannel blind deconvolution: FIR matrix algebra and separation of multipath mixtures*. PhD dissertation, University of Southern California, Department of Electrical Engineering, May 1996.
- [44] R. H. Lambert. Difficulty measures and figures of merit for source separation. In *Proc. ICA and BSS*, Aussois, France, January 11-15 1999.
- [45] R. H. Lambert and A. J. Bell. Blind separation of multiple speakers in a multipath environment. In *Proc. ICASSP*, pages 423–426, Munich, Germany, April 21-24 1997.
- [46] R. H. Lambert and C. Nikias. Polynomial matrix whitening and application to the multichannel blind deconvolution problem. In *Proc. IEEE MILCOM*, San Diego, CA, Nov. 5-8 1995.
- [47] T.-W. Lee, A. Bell, and R. H. Lambert. Blind separation of delayed and convolved sources. In *Advances in Neural Information Processing Systems 9*, pages 758–764. MIT Press, 1997.
- [48] T.-W. Lee, A. Bell, and R. Orglmeister. Blind source separation of real world signals. In *Proc. ICNN*, Houston, TX, June 9-12 1997.
- [49] T.-W. Lee, A. Bell, and R. Orglmeister. A contextual blind separation of delayed and convolved sources. In *Proc. ICASSP*, pages 1199–1202, Munich, Germany, April 21-24 1997.
- [50] T.-W. Lee, M. Girolami, A. Bell, and T. Sejnowski. A unifying information-theoretic framework for independent component analysis. *International Journal on Mathematical and Computer Modeling*, 1999. (in press).
- [51] T.-W. Lee, A. Ziehe, R. Orglmeister, and T. Sejnowski. Combining time-delayed decorrelation and ica: towards solving the cocktail party problem. In *Proc. ICASSP*, pages 1249–1252, Seattle, WA, May 1998.
- [52] S. Li and T. J. Sejnowski. Adaptive separation of mixed broad-band sound sources with delays by a beamforming Héroult-Jutten network. *IEEE Journal of Oceanic Engineering*, 20(1):73–79, January 1995.
- [53] U. Lindgren and H. Broman. Monitoring the mutual independence of the output of source separation algorithms. In *Proc. ISITA'96*, Victoria, B.C., Canada, September 1996.
- [54] U. Lindgren and H. Broman. On the identifiability of a mixing channel based on second order statistics. In *Proc. Radiovetenskap och Kommunikation*, pages 420–424, Lulea, Sweden, May 1996.
- [55] U. Lindgren, H. Sahlin, and H. Broman. Source separation using second order statistics. In *Signal Processing IX: Theories and Applications (Proc. of the EUSIPCO-96)*, Trieste, Italy, September 1996. Elsevier.
- [56] U. Lindgren and A.-J. van der Veen. Source separation based on second order statistics – An algebraic approach. In *Proc. of the IEEE 8th Workshop on Statistical Signal and Array processing*, pages 324–327, Corfu, Greece, June 24-26 1996. IEEE.
- [57] U. Lindgren, T. Wigren, and H. Broman. On local convergence of a class of blind separation algorithms. *IEEE Transactions on Signal Processing*, 43(12):3054–3058, December 1995.
- [58] K. Matsuoka and M. Kawamoto. Blind signal separation based on mutual information criterion. In *Proc. NOLTA*, pages 85–90, Las Vegas, Nevada, USA, December 10-14 1995.
- [59] K. Matsuoka, M. Ohya, and M. Kawamoto. A neural net for blind separation of nonstationary signals. *Neural Networks*, 8(3):411–419, 1995.
- [60] C. Mejuto and J. C. Principe. A second-order method for blind separation of convolutive mixtures. In *Proc. ICA and BSS*, Aussois, France, January 11-15 1999.
- [61] L. Molgedy and H. Schuster. Separation of independent signals using time-delayed correlations. *Physical Review Letters*, 72(23):3634–3637, June 6 1994.

- [62] E. Moreau and J.-C. Pesquet. Generalized contrasts for multichannel blind deconvolution of linear systems. *IEEE Signal Processing Letters*, 4(6):182–183, June 1997.
- [63] E. Moreau and N. Thirion. Multichannel blind signal deconvolution using high order statistics. In *Proc. of the IEEE 8th Workshop on Statistical Signal and Array processing*, pages 336–339, Corfu, Greece, June 24–26 1996. IEEE.
- [64] E. Moulines, J.-F. Cardoso, and E. Gassiat. Maximum likelihood for blind separation and deconvolution of noisy signals using mixture models. In *Proc. ICASSP*, pages 3617–3620, Munich, Germany, April 21–24 1997.
- [65] N. Murata, S. Ikeda, and A. Ziehe. An approach to blind source separation based on temporal structure of speech signals. Technical Report BSIS Technical Reports No.98-2, RIKEN Brain Science Institute, Japan, 1998.
- [66] M. Nájjar, M. A. Lagunas, and I. Bonet. Blind wideband source separation. In *Proc. ICASSP*, volume IV, pages 65–68, Adelaide, Australia, April 19–22 1994.
- [67] J. T. Ngo and N. A. Bhadkamkar. Adaptive blind separation of audio sources by a physically compact device using second-order statistics. In *Proc. ICA and BSS*, Aussois, France, January 11–15 1999.
- [68] S. Ohno and Y. Inouye. A least-squares interpretation of the single-stage maximization criterion for multichannel blind deconvolution. In *Proc. ICASSP*, pages 2101–2104, Seattle, WA, May 12–15 1998.
- [69] H. Pan, D. Xia, S. Douglas, and K. Smith. A scalable vlsi architecture for multichannel blind deconvolution and source separation. In *Proc. IEEE Workshop on Signal Processing Systems*, Boston, MA, October 1998.
- [70] L. Parra and C. Spence. Convolutional blind source separation based on multiple decorrelation. In *Proc. of NNISP98*, Cambridge, UK, September 1998.
- [71] L. Parra and C. Spence. Temporal models in blind source separation. In *Adaptive Processing of Sequences and Data Structures - International Summer School on Neural Networks E.R. Caianiello, Vietri sul Mare, Salerno, Italy, September 6-13, 1997, Tutorial Lectures*. Springer, 1998.
- [72] L. Parra and C. Spence. Convolutional blind source separation based on multiple decorrelation. *IEEE Transactions on Speech and Audio Processing*, ?-?-?, 1999. (submitted).
- [73] L. Parra, C. Spence, and B. D. Vries. Convolutional source separation and signal modeling with ML. In *Proc. International Symposium on Intelligent Systems (ISIS'97)*, Regio Calabria, Italy, 1997.
- [74] B. A. Pearlmutter and L. C. Parra. A context-sensitive generalization of ICA. In *International Conference on Neural Information Processing*, Hong Kong, Sept. 24–27 1996. Springer.
- [75] D. Pham, P. Garat, and C. Jutten. Separation of a mixture of independent sources through a maximum likelihood approach. In J. Vandevallé, R. Boite, M. Moonen, and A. Oosterlink, editors, *Signal Processing VI: Theories and Applications*, pages 771–774. Elsevier, 1992.
- [76] J. C. Platt and F. Faggin. Networks for the separation of sources that are superimposed and delayed. In J. Moody, S. Hanson, and R. Lippmann, editors, *Advances in Neural Information Processing Systems 4*. Morgan-Kaufmann, 1992.
- [77] J. C. Principe, C. Wang, and H.-C. Wu. Temporal decorrelation using teacher forcing anti-Hebbian learning and its application in adaptive blind source separation. In *Neural Networks for Signal Processing VI (Proc. IEEE Workshop on Neural Networks for Signal Processing)*, pages 413–422, Kyoto, Japan, September 4–6 1996.
- [78] H. Sahlin and H. Broman. Signal separation applied to real world signals. In *Proceedings of 1997 Int. Workshop on Acoustic Echo and Noise Control (IWAENC97)*, London UK, September 11–12 1997.
- [79] H. Sahlin and H. Broman. A decorrelation approach to blind mimo signal separation. In *Proc. ICA and BSS*, Aussois, France, January 11–15 1999.
- [80] H. Sahlin and U. Lindgren. The asymptotic cramer-rao lower bound for blind signal separation. In *Proc. of the IEEE 8th Workshop on Statistical Signal and Array processing*, pages 328–331, Corfu, Greece, June 24–26 1996. IEEE.
- [81] D. Schobben and P. Sommen. A new blind signal separation algorithm based on second order statistics. In *Proc. IASTED International Conference on Signal and Image Processing*, Las Vegas, USA, October 27–31 1998.
- [82] D. Schobben and P. Sommen. Transparent communication. In *Proc. IEEE Benelux Signal Processing Chapter Symposium*, pages 171–174, Leuven, Belgium, March 26–27 1998.
- [83] D. Schobben, K. Torkkola, and P. Smaragdis. Evaluation of blind signal separation methods. In *Proc. ICA and BSS*, Aussois, France, January 11–15 1999.
- [84] C. Serviere. Blind source separation of convolutive mixtures. In *Proc. of the IEEE 8th Workshop on Statistical Signal and Array processing*, pages 316–319, Corfu, Greece, June 24–26 1996. IEEE.
- [85] C. Serviere. Feasibility of source separation in frequency domain. In *Proc. ICASSP*, pages 2085–2088, Seattle, WA, May 12–15 1998.
- [86] C. Serviere and V. Capdevielle. Blind adaptive separation of wide-band sources. In *Proc. ICASSP*, Atlanta, GA, May 7–10 1996.
- [87] S. Shamsunder and G. B. Giannakis. Multichannel blind signal separation and reconstruction. *IEEE Transactions on Speech and Audio Processing*, 5(6):515–528, 1997.
- [88] C. Simon, G. d'Urso, C. Vignat, and P. Loubaton. On the convolutive mixture source separation by the decorrelation. In *Proc. ICASSP*, pages 2109–2112, Seattle, WA, May 12–15 1998.
- [89] P. Smaragdis. Efficient blind separation of convoluted sound mixtures. In *Proceedings of IEEE 1997 Workshop on Applications of Signal Processing to Audio and Acoustics*, Mohonk Mountain House, New Paltz, NY, Oct 19–22 1997.
- [90] P. Smaragdis. Information theoretic approaches to source separation. Master's thesis, Massachusetts Institute of Technology, 1997.
- [91] P. Smaragdis. Blind separation of convoluted sound mixtures in the frequency domain. In *Proc. International Workshop on Independence & Artificial Neural Networks*, Tenerife, Spain, February 9–10 1998.
- [92] V. Soon, L. Tong, Y. Huang, and R. Liu. A wideband blind identification approach to speech acquisition using a microphone array. In *Proc. ICASSP*, volume 1, pages 293–296, San Francisco, California, USA, March 23–26 1992.
- [93] V. Soon, L. Tong, Y. Huang, and R. Liu. A robust method for wideband signal separation. In *Proc. ISCAS*, pages 703–706, 1993.
- [94] H.-L. N. Thi and C. Jutten. Blind source separation for convolutive mixtures. *Signal Processing*, 45(2), 1995.
- [95] H.-L. N. Thi, C. Jutten, and J. Caelen. Speech enhancement: Analysis and comparison of methods on various real situations. In *Signal Processing VI: Theories and Applications (Proc. of the EUSIPCO-92)*, pages 303–306, Bruxelles, Belgium, 1992. Elsevier.
- [96] L. Tong. Identification of multivariate FIR systems using higher-order statistics. In *Proc. ICASSP*, Atlanta, GA, May 7–10 1996.
- [97] L. Tong, R. Liu, V. Soon, and Y. Huang. Indeterminacy and identifiability of blind identification. *IEEE Transactions on Circuits and Systems*, 38(5):499–509, May 1992.
- [98] K. Torkkola. Blind separation of convoluted sources based on information maximization. In *IEEE Workshop on Neural Networks for Signal Processing*, pages 423–432, Kyoto, Japan, September 4–6 1996.
- [99] K. Torkkola. Blind separation of delayed sources based on information maximization. In *Proc. ICASSP*, pages 3510–3513, Atlanta, GA, May 7–10 1996.
- [100] J. K. Tugnait. Adaptive blind separation of convolutive mixtures of independent linear signals. In *Proc. ICASSP*, pages 2097–2100, Seattle, WA, May 12–15 1998.
- [101] D. Van Compernelle and S. Van Gerven. Signal separation in a symmetric adaptive noise canceler by output decorrelation. In *Proc. ICASSP*, volume 4, pages 221–224, San Francisco, California, USA, March 23–26 1992. IEEE.
- [102] A.-J. van der Veen. Algebraic methods for deterministic blind beamforming. *Proceedings of IEEE*, 86(10):1987–2008, October 1998.
- [103] S. Van Gerven and D. Van Compernelle. Feedforward and feedback in a symmetric adaptive noise canceler: Stability analysis in a simplified case. In J. Vandevallé, R. Boite, M. Moonen, and A. Oosterlink, editors, *Signal Processing VI: Theories and Applications*, pages 1081–1084. Elsevier, 1992.
- [104] S. Van Gerven and D. Van Compernelle. Signal separation by symmetric adaptive decorrelation: Stability, convergence, and uniqueness. *IEEE Transactions on Signal Processing*, 43(7):1602–1612, July 1995.
- [105] E. Weinstein, M. Feder, and A. Oppenheim. Multi-channel signal separation by decorrelation. *IEEE Transactions on Speech and Audio Processing*, 1(4):405–413, 1993.
- [106] A. Westner. Object-based audio capture: Separating acoustically-mixed sounds. Master's thesis, Massachusetts Institute of Technology, 1999.
- [107] A. Westner and J. V. Michael Bove. Blind separation of real world audio signals using overdetermined mixtures. In *Proc. ICA and BSS*, Aussois, France, January 11–15 1999.
- [108] H.-C. Wu and J. C. Principe. A unifying criterion for blind source separation and decorrelation: Simultaneous diagonalization of correlation matrices. In *Proc. of NNISP97*, pages 496–505, Amelia Island, FL, 1997.
- [109] H.-C. Wu and J. C. Principe. Simultaneous diagonalization algorithm for blind source separation based on subband filtered features. In *Proceedings of SPIE - Conference of The International Society for Optical Engineering*, pages 466–474, Orlando, Florida, 1998.
- [110] H.-C. Wu and J. C. Principe. Simultaneous diagonalization in the frequency domain (SDIF) for source separation. In *Proc. ICA and BSS*, Aussois, France, January 11–15 1999.
- [111] H.-C. Wu, J. C. Principe, and D. Xu. Exploring the time-frequency microstructure of speech for blind source separation. In *Proc. ICASSP*, pages 1145–1148, Seattle, WA, USA, 1998.
- [112] J. Xi and J. P. Reilly. Blind separation and restoration of signals mixed in convolutive environment. In *Proc. ICASSP*, pages 1327–1330, Munich, Germany, April 21–24 1997.
- [113] D. Yellin and B. Friedlander. Blind multi-channel system identification and deconvolution: Performance bounds. In *Proc. of the IEEE 8th Workshop on Statistical Signal and Array processing*, pages 582–585, Corfu, Greece, June 24–26 1996. IEEE.
- [114] D. Yellin and E. Weinstein. Multichannel signal separation: Methods and analysis. *IEEE Transactions on Signal Processing*, 44(1):106–118, January 1996.
- [115] K.-C. Yen and Y. Zhao. Robust automatic speech recognition using a multi-channel signal separation front end. In *Proc. 4th Int. Conf. on Spoken Language Processing (ICSLP'96)*, Philadelphia, PA, October 1996.