

Convolutional Blind Source Separation by Minimizing Mutual Information Between Segments of Signals

Kenneth E. Hild, II, *Member, IEEE*, David Pinto, *Member, IEEE*, Deniz Erdogmus, *Member, IEEE*, and Jose C. Principe, *Fellow, IEEE*

Abstract—A method to perform convolutional blind source separation of super-Gaussian sources by minimizing the mutual information between segments of output signals is presented. The proposed approach is essentially an implementation of an idea previously proposed by Pham. The formulation of mutual information in the proposed criterion makes use of a nonparametric estimator of Renyi's α -entropy, which becomes Shannon's entropy in the limit as α approaches 1. Since α can be any number greater than 0, this produces a family of criteria having an infinite number of members. Interestingly, it appears that Shannon's entropy cannot be used for convolutional source separation with this type of estimator. In fact, only one value of α appears to be appropriate, namely $\alpha = 2$, which corresponds to Renyi's quadratic entropy. Four experiments are included to show the efficacy of the proposed criterion.

Index Terms—Convolutional blind source separation (BSS), information theoretic learning, nonparametric entropy estimator, Renyi's entropy.

I. INTRODUCTION

CONVOLUTIONAL blind source separation (BSS) involves the transformation of a set of observations, where each is a different mixture of a common set of sources, in an attempt to recover the original (unmixed) sources. The observations in (linear) convolutional BSS are related to the sources in the following manner:

$$x_i(n) = \sum_{j=1}^{N_s} \sum_{k=0}^{L_h-1} h_{ij}(k) s_j(n-k) + e_i(n) \quad (1)$$

for $i = 1, 2, \dots, N_s$, where $x_i(n)$ denotes the i th observation at time n , $s_j(n)$ denotes the j th source at time n , $e_i(n)$ is the signal at time n that accounts for model error and additive noise, N_s is the number of sources and observations, and L_h is the length of the individual mixing filters $h_{ij}(n)$. It is also commonly assumed that at most 1 of the N_s sources is Gaussian-distributed

and all sources are mutually statistically independent, which makes independent component analysis (ICA) an appropriate method to find a solution for BSS. Equation (1) may be written compactly as $\mathbf{x}(n) = \mathbf{H}(z)\mathbf{s}(n) + \mathbf{e}(n)$, where $\mathbf{x}(n)$, $\mathbf{s}(n)$, and $\mathbf{e}(n)$ are $N_s \times 1$ vectors and $\mathbf{H}(z)$ is a $N_s \times N_s$ matrix of Z -transforms of the individual mixing filters. Applications for which the model of (1) are appropriate include speech enhancement for hearing aids, hands-free telephony, speech recognition [1], [2], sonar signal processing [3], [4], and communications systems [5], [6], [7].

In the BSS paradigm the receiver knows neither the mixing filters nor the sources. Consequently, BSS algorithms must separate the sources by training demixing filters using only the observations. One structure for time-domain demixing is the feed-forward (FF) architecture represented by

$$y_i(n) = \sum_{j=1}^{N_s} \sum_{k=0}^{L_w-1} w_{ij}(k) x_j(n-k) \quad (2)$$

where $y_i(n)$ is the i th output at time n , L_w is the length of the demixing filters $w_{ij}(n)$ and the individual demixing filters are described by an impulse response the implementation of which can have one of several different (local) structures. Note that the acronyms FF and FB (feedback) are used here in a global sense while the individual demixing filters are free to be locally feed-forward or feedback. A partial list of local filters includes the autoregressive (AR) [8], moving average (MA, also known as FIR) [8], autoregressive-moving average (ARMA) [8], Laguerre [8], [9], and Gamma filters [10]. Using the compact notation, (2) is given by $\mathbf{y}(n) = \mathbf{W}(z)\mathbf{x}(n)$, where $\mathbf{y}(n)$ is the $N_s \times 1$ vector of outputs at time n and $\mathbf{W}(z)$ is the $N_s \times N_s$ matrix of Z -transforms of the individual demixing filters.

II. PUBLISHED TECHNIQUES

The existing convolutional source separation algorithms could be classified in any number of different ways. Fig. 1 shows one possible way to categorize most of the published methods using characteristics of the demixing filters (i.e., FF time domain, FB time domain, or frequency domain) and the cost function (e.g., whether it uses second order statistics, higher order statistics, or is information theoretic). The structure determines what parameters need to be adapted, whereas the criterion determines how they are adapted. For example, the parameter updates for the Type II JBD algorithm [11] are found after applying an FFT to the observations; hence it is a frequency domain criterion. However, the demixing is performed by the time domain structure that results from applying an inverse FFT to the previously

Manuscript received June 12, 2004. This work was supported in part by the National Science Foundation under ECF 9900394. This paper was recommended by Associate Editor V. E. DeBrunner.

K. E. Hild, II is with the Department of Radiology, University of California, San Francisco, CA 94122 USA (e-mail: k.hild@ieee.org).

D. Pinto is with the Exploratory Real-Time Systems Department, Sandia National Labs, Albuquerque, NM 87123 USA (e-mail: dpinto@sandia.gov).

D. Erdogmus is with the Departments of Computer Science and Engineering and Biomedical Engineering, Oregon Health & Science University, Beaverton, OR 97006 USA (e-mail: derdogmus@ieee.org).

J. C. Principe is with the Department of Electrical Engineering and Computer Science, University of Florida, Gainesville, FL 32611 USA (e-mail: principe@cnel.ufl.edu).

Digital Object Identifier 10.1109/TCSI.2005.852915

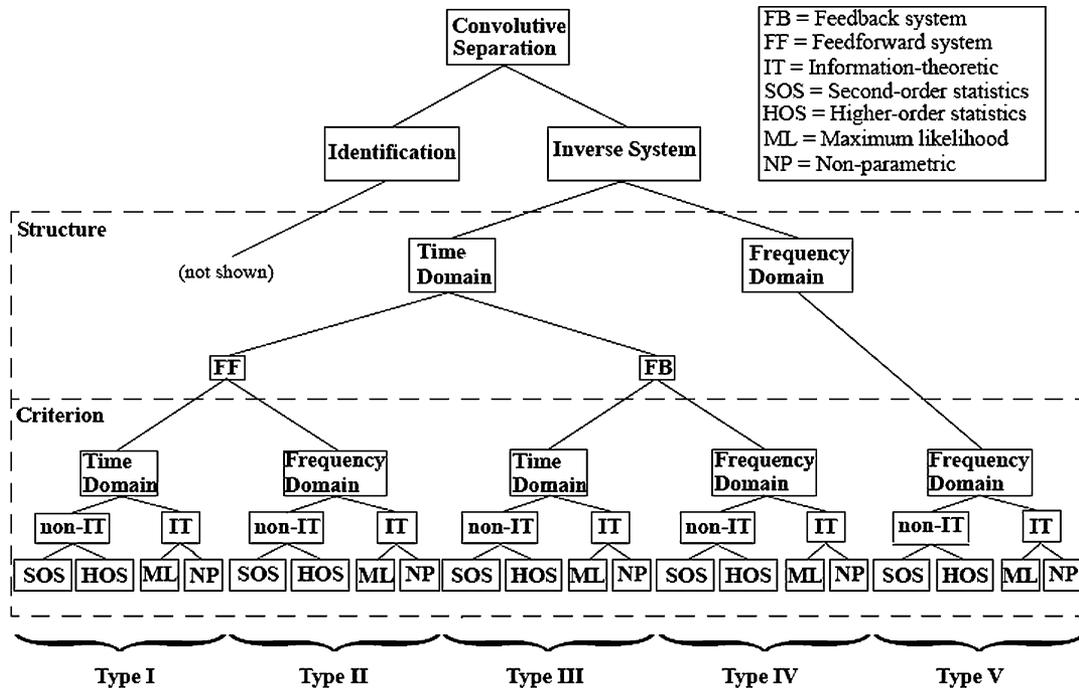


Fig. 1. Method for categorizing convolutive BSS systems.

TABLE I
CATEGORIZATION OF CONVOLUTIVE BSS SYSTEMS

Type I Systems	
SOS	[26],[28]-[41]
HOS	[42]-[57]
ML	[13],[15],[24],[58]-[67]
NP	MRMI-SIG (described in next section)
Type II Systems	
SOS	[11],[38],[68]-[71], (see below)
HOS	(see below)
ML	[72], (see below)
Type III Systems	
SOS	[26]
HOS	[43],[73]-[77]
ML	[13],[78]-[81]
Type II, IV, V Systems	
SOS	[41],[82]-[91]
HOS	[25],[92]-[98]
ML	[58],[72],[87],[99]-[106]
Type VI Systems	
	[107]-[110]
Type VII Systems	
	[111]-[114]

found solution. Several methods that did not fit into this particular classification include subspace methods and linear prediction methods, which are referred to as Type VI and Type VII, respectively. Using this approach, a total of approximately 90 papers are categorized in Table I. On occasion the task of categorizing a particular method proved formidable. In fact, the impetus for creating the categories is precisely due to the difficulty encountered when attempting to decipher some of the algorithms. The principal complication stems from the subtle difference that exists between Types II, IV, and V. Consequently, many of them are grouped together near the bottom of Table I. The difficulty arises from the fact that each uses a frequency-domain criterion, and the conversion to time-domain filters, re-

quired by Types II and IV, is very straightforward. Even though the distinction is subtle, there are important differences in the performance of each. Because of the qualitative differences in performance, Types II, IV, and V are kept as three distinct types. The reader is referred to the first author’s dissertation [12] for more details on this particular approach to categorization, which is hoped to facilitate the succinct description of BSS algorithms by authors of future papers.

Generally speaking, Types II, IV, and V require less time/computation to produce a solution than the other types due to their use of frequency-domain criteria and/or structures. Types III and IV are unable to implement general acausal solutions, which will cause a large drop in performance for certain types of mixtures. However, some criteria need to use FB structures since they do not work well with FF structures, i.e., it has been noted that the Bell and Sejnowski InfoMax algorithm prefers to temporally whiten rather than separate the outputs when used with a FF structure [13]. Finally, Type V methods produce outputs that sound artificial due to the lack of a “time-domain constraint” [8]. Supposing that the computation time is not a limiting factor, Types I and II appear to be the most promising.

Pham discusses several variants of an information-theoretic (IT) criterion suitable for a Type I implementation, which makes use of spatial information across multiple lags [14], [15]. This criterion involves the minimization of Shannon’s mutual information between segments of processes and is, in fact, the canonical contrast for finding independent components. However, neither of these papers discusses the implementation of this criterion. More specifically, the criterion in both papers includes an entropy function but no reference is given to the means by which the entropy is estimated. Herein, the criterion discussed by Pham is implemented (approximately) by means of a nonparametric entropy estimator. The details of this criterion are given in Section III.

III. PROPOSED CRITERION

The (Shannon) mutual information [16] between segments of processes, which is a contrast for jointly stationary processes [15], may be approximated by a function of Renyi's entropies [17] as follows:

$$I(\vec{\mathbf{y}}_1(n)\vec{\mathbf{y}}_2(n)\dots\vec{\mathbf{y}}_{N_s}(n)) \cong \sum_{i=1}^{N_s} H_{R_\alpha}(\vec{\mathbf{y}}_i(n)) - H_{R_\alpha}(\vec{\mathbf{y}}(n)) \quad (3)$$

where each $\vec{\mathbf{y}}_i(n) = [y_i(n), y_i(n-1), y_i(n-2), \dots, y_i(n-L_t+1)]^T$ is an $L_t \times 1$ vector, L_t is the extent of the temporal information used in the criterion, $H_{R_\alpha}(\mathbf{v}(n))$ is Renyi's α -entropy of the random vector from which $\mathbf{v}(n)$ is drawn, and $\vec{\mathbf{y}}(n)$ is an $L_t \times 1$ vector given by, $\vec{\mathbf{y}}(n) = [\mathbf{y}(n)^T, \mathbf{y}(n-N_s)^T, \dots, \mathbf{y}(n-L_t+N_s)^T]^T$. To show that (3) is a valid approximation, it is sufficient to show that the following two approximations are valid

$$H_{R_\alpha}(\vec{\mathbf{y}}(n)) \cong H_{R_\alpha}(\mathbf{y}'(n)) \quad (4)$$

$$H_{R_\alpha}(\mathbf{v}(n), \vec{W}) \cong aH_{R_1}(\mathbf{v}(n), \vec{W}) + b \quad (5)$$

where $\mathbf{y}'(n) = [\mathbf{y}(n)^T, \mathbf{y}(n-1)^T, \dots, \mathbf{y}(n-L_t+1)^T]^T$, $H_{R_\alpha}(\mathbf{v}(n), \vec{W})$ is Renyi's entropy of the output (marginal or joint; i.e., $\mathbf{v}(n) = \vec{\mathbf{y}}_i(n)$ or $\mathbf{v}(n) = \vec{\mathbf{y}}(n)$) as a function of the set of demixing parameters, \vec{W} , a is a positive constant, b is any constant, and $H_{R_1}(\cdot)$ is used to denote Shannon's entropy. The use of $\alpha = 1$ to denote Shannon's entropy is the natural choice since Renyi's entropy becomes Shannon's entropy in the limit as α approaches 1 [17].

Concerning the approximation of (4), equality is obtained when the sampling frequency is large enough to guarantee that $\mathbf{y}(n)^T$ becomes indistinguishable from $\mathbf{y}(n-1)^T, \dots, \mathbf{y}(n-N_s+1)^T$. This is easily shown by substituting $\mathbf{y}(n-kN_s)^T$ for $\mathbf{y}(n-1-kN_s)^T, \dots, \mathbf{y}(n-N_s+1-kN_s)^T$ (for $k = 0, 1, \dots, L_t/N_s - 1$) into the expression for $H_{R_\alpha}(\mathbf{y}'(n))$. With this substitution the joint entropy becomes $H_{R_\alpha}(\mathbf{y}(n)^T, \mathbf{y}(n)^T, \dots, \mathbf{y}(n)^T, \mathbf{y}(n-N_s)^T, \mathbf{y}(n-N_s)^T, \dots, \mathbf{y}(n-N_s)^T, \dots, \mathbf{y}(n-L_t+N_s)^T, \mathbf{y}(n-L_t+N_s)^T, \dots, \mathbf{y}(n-L_t+N_s)^T)$, which is equal to $H_{R_\alpha}(\mathbf{y}(n)^T, \mathbf{y}(n-N_s)^T, \dots, \mathbf{y}(n-L_t+N_s)^T)$ as shown in Cover and Thomas for Shannon's entropy [16] and which is also easily shown to be true in the case of Renyi's entropy. Notice that this latter representation of the joint entropy is identically $H_{R_\alpha}(\vec{\mathbf{y}}(n))$. Needless to say, it is not possible to obtain the equality for any practical sampling frequency. Consequently, (4) remains an approximation the quality of which is assessed in this paper only in terms of how well the overall criterion separates sources. The need for this approximation will be explained later.

The approximation of (5) simply states that the α parameter should be chosen in a manner such that Renyi's entropy is approximately linearly related to Shannon's entropy in the space of the demixing parameters. If this is the case, the criteria based on these two definitions of entropy will produce the same solution since an additive constant and a positive multiplicative constant have no effect on the location of the minimum value. Since two entropy measures can have a nonlinear relationship yet still produce the same solution, (5) represents a sufficient but unnecessary requirement. No mathematical proof is known that demonstrates the ex-

istence of a value of α , other than 1, that has this linear relationship. However, extensive experience with this criterion indicates this condition is met for Renyi's quadratic entropy ($\alpha = 2$) when the signals are i.i.d. and have a super-Gaussian distribution. In particular, in Hild *et al.* [14] a plot of Renyi's quadratic entropy and Shannon's entropy as a function of the demixing parameter is included where the sources are super-Gaussian and the values of a and b have been properly selected. In this plot, the curves representing $\alpha = 1$ and $\alpha = 2$ are indistinguishable. A three-dimensional (3-D) plot showing similar results for a higher-dimensional case may be found in the first author's dissertation [12]. Equation (5) does not hold for sub-Gaussian sources, although the need for this approximation is easily circumvented for the case of instantaneous mixtures [14].

All that is needed to implement the criterion of (3) is an estimator for Renyi's α -entropy. A paper by Erdogmus *et al.* [18] introduces one such estimator, which is given by

$$\hat{H}_{R_\alpha}(v(n)) = \frac{1}{1-\alpha} \log \frac{1}{N} \times \sum_{n=1}^N \left(\frac{1}{N} \sum_{m=1}^N G(v(n) - v(m), \sigma^2) \right)^{\alpha-1} \quad (6)$$

where N is the block size, $G(v(n), \sigma^2)$ is the Gaussian function evaluated at $v(n)$ and having variance σ^2 . This is based on a previously published entropy estimator [19] for $\alpha = 2$, but was generalized by Erdogmus in order to allow for any positive value of α . The nonparametric estimator in (6) (as well as the estimator on which it is based) has $O(N^2)$ complexity, uses Parzen Window density estimation [20], and is computed directly from the data by considering all pair-wise interactions of the output samples. Using a change of variables (6) can also be represented as

$$\hat{H}_{R_\alpha}(v(n)) = \frac{1}{1-\alpha} \log \frac{1}{N} \sum_{p=1-N}^{N-1} \left(\frac{1}{N} \sum_{n \in I_p} G(v(n) - v(n-p), \sigma^2) \right)^{\alpha-1} \quad (7)$$

where I_p represents the valid set of values for n , which is necessarily a function of p for finite N . Using this representation the inner summation uses a fixed lag of p and the outer summation computes an average over all possible lags. This entropy estimator is asymptotically unbiased for i.i.d. data due to the consistency of the Parzen window density estimator on which it is based. A modification to this entropy estimator, referred to as the Stochastic Information Gradient (SIG) [21], was introduced 1 yr later. The modification amounts to the removal of the outer summation in (7) such that the new $O(N)$ estimator is a function of only a single lag. This estimator is given by

$$\hat{H}_{R_2}(v(n)) = -\log \frac{1}{N} \sum_{n=1}^N G(v(n) - v(n-p), \sigma^2) \quad (8)$$

where p is commonly chosen to be 1. This entropy estimator is also asymptotically unbiased for i.i.d. data; however, it only estimates Renyi's quadratic entropy ($\alpha = 2$) [12]. A simple argument for this claim is that, for i.i.d. data and as N goes to infinity, the argument of the log in (6) and (7) becomes an expectation

that equals the argument of the log in (8) only when $\alpha = 2$. Equations (6)–(8) are given for the univariate case. The extension to the multi-variate case only requires that the scalar σ^2 be replaced by the covariance matrix, $\sigma^2 I$, where I is the identity matrix whose size is commensurate with the random vector in question, i.e., $\vec{\mathbf{y}}_i(n)$ or $\vec{\mathbf{y}}(n)$. It should also be mentioned that these entropy estimators are a function of the dimensionality of the constituent random vector when the amount of data is finite [22]. As a result, if the dimensionalities of the random vectors used in the marginal and joint entropies are not identical then one or the other (i.e., the sum of marginal entropies or the joint entropy) will be more heavily weighted in the criterion of (3), which is suboptimal. The approximation represented by (4) was introduced in order to ameliorate this problem. Although the dimensionalities of the marginal entropies and the joint entropy are identical only if L_t is an integer multiple of N_s , it appears to be sufficient in practice if L_t is approximately an integer multiple of N_s .

Not only is the $O(N)$ entropy estimator of (8) much more practical for training long demixing filters due to the significantly reduced computational complexity, but it has also been experimentally determined to be much more robust to the i.i.d. assumption when estimating joint entropies. This is critical since $y_i(n)$ is almost never i.i.d. for convolutive mixtures due to the very nature of the problem (the problem of robustness is not encountered when mutual information can be expressed in terms of only marginal entropies, which is possible for instantaneous demixing [14]). On the other hand, if the desire is to use the $O(N^2)$ criterion the authors recommend using an architecture that includes a deconvolving filter at each demixer output. The lack of robustness to the i.i.d. assumption reduces the accuracy of the $O(N^2)$ entropy estimator for convolutive demixing. Consequently, for the minimization of the mutual information by means of the family of entropy estimators given in (6)–(8), $\alpha = 2$ is the preferred choice. Notice that this precludes the use of Shannon's entropy.

The proposed method for convolutive BSS involves the structure defined by (2) and the criterion formed by plugging (8) into (3). The resulting criterion to be minimized is given by

$$\begin{aligned} I(\vec{\mathbf{y}}_1(n); \dots; \vec{\mathbf{y}}_{N_s}(n)) \\ \cong - \sum_{i=1}^{N_s} \log \frac{1}{N} \sum_{n=1}^N G(\vec{\mathbf{y}}_i(n) - \vec{\mathbf{y}}_i(n-1), \sigma^2 I) \\ + \log \frac{1}{N} \sum_{n=1}^N G(\vec{\mathbf{y}}(n) - \vec{\mathbf{y}}(n-1), \sigma^2 I). \end{aligned} \quad (9)$$

Without loss of generality the $W_{ii}(z)$ filters are constrained to be z^{-L} . The $W_{ij}(z)$ filters, for i not equal to j , are found using gradient descent, which requires the derivative of (9) with respect to each $w_{ij}(n)$. This is given by (10) at the bottom of the page, where the square brackets are used to denote a row vector formed from the constituent signals. For convenience a scalar gain is also applied to each output prior to the computation of (10) such that each output has unit variance [this variable is not shown in (10)].

Due to the similarities of the underlying criterion, the proposed method retains the same designation, MRMI-SIG, used previously for feature extraction of static data [23] and instantaneous BSS [14]. There are four variables that must be chosen for this method. Two of these, the window size N and the length of the demixing filters L_w , are common for any time-domain structure. The remaining two are the temporal extent of the criterion L_t and the kernel size σ . If the variance of the output is fixed at 1, it has been experimentally determined that the kernel size should be approximately 0.25. Concerning L_t , Pham implicitly uses the value of $L_t = \infty$ [15]. However, since the sources are assumed to be independent, the mixing and demixing filters are the only cause of the dependencies between the outputs. As a result, L_t need not be any larger than $L_h + L_w - 1$, which equals the length of the impulse response of the combined mixing and demixing. Even if L_h and/or L_w equal infinity, a finite-valued L_t may produce good separation results as long as its value is greater than or equal to the effective length of the combined mixing and demixing. However, L_h is unknown (as is the effective L_h). Therefore, the suggestion is to use $L_t = 2L_w - 1$. Performance is expected to decrease if L_t is chosen too small.

IV. PERFORMANCE

Table II shows a representative set of convolutive source separation algorithms, which are used to demonstrate how the separation performance of the proposed method fares. Each time-domain structure is given in terms of {global structure}/ {local structure}. Notice that one or two methods from each of the five categories of Fig. 1 are included. Bingham and Hyvarinen describe the Fixed Point method [25], although they did not use the FB/FIR structure. Fixed point was chosen for the Type IV method based on its speed, which is imperative since as many as 8000 (complex) instantaneous BSS solutions are required for each separation task using a BSS algorithm of this category. For JADE, the minimum Frobenius norm between the separating and the estimated solution was used to correct for (local) permutations at each frequency [12].

$$\begin{aligned} \frac{\partial I(\vec{\mathbf{y}}_1(n); \dots; \vec{\mathbf{y}}_{N_s}(n))}{\partial w_{jk}(m)} \cong \frac{1}{\sigma^2} \frac{\sum_{n=1}^N e^{\frac{-1}{2\sigma^2}} (\vec{\mathbf{y}}_j(n) - \vec{\mathbf{y}}_j(n-1))^T (\vec{\mathbf{y}}_j(n) - \vec{\mathbf{y}}_j(n-1)) (\vec{\mathbf{y}}_j(n) - \vec{\mathbf{y}}_j(n-1))^T (\vec{\mathbf{x}}_k(n-m) - \vec{\mathbf{x}}_k(n-m-1))}{\sum_{n=1}^N e^{\frac{-1}{2\sigma^2}} (\vec{\mathbf{y}}_j(n) - \vec{\mathbf{y}}_j(n-1))^T (\vec{\mathbf{y}}_j(n) - \vec{\mathbf{y}}_j(n-1))} \\ - \frac{1}{\sigma^2} \left[\sum_{n=1}^N e^{\frac{-1}{2\sigma^2}} (\vec{\mathbf{y}}(n) - \vec{\mathbf{y}}(n-1))^T (\vec{\mathbf{y}}(n) - \vec{\mathbf{y}}(n-1))^T [y_j(n) y_j(n - N_s) \dots y_j(n - L_t + N_s)] [x_k(n-m) x_k(n-m - N_s) \dots x_k(n-m - L_t + N_s)]^T \right] / \left[\sum_{n=1}^N e^{\frac{-1}{2\sigma^2}} (\vec{\mathbf{y}}(n) - \vec{\mathbf{y}}(n-1))^T (\vec{\mathbf{y}}(n) - \vec{\mathbf{y}}(n-1)) \right] \end{aligned} \quad (10)$$

TABLE II
REPRESENTATIVE SET OF CONVOLUTIVE SOURCE SEPARATION METHODS

Method	Type	Structure	Criterion
Douglas' Nonholonomic [24]	I-ML	FF/FIR	Min. mutual information
Hild's MRMI-SIG (proposed)	I-NP	FF/FIR	Min. mutual information
Parra's JBD [11]	II-SOS	FF/FIR	Min. cross-power spectra
Van Gerven's SAD [26]	III-SOS	FB/FIR	Min. cross-correlation
Bell's InfoMax [13]	III-ML	FB/FIR	Max. entropy
Bingham's Fixed Point [25]	IV-HOS	FB/FIR	Max. sum of HOS
Cardoso's JADE [27]	V-HOS	Freq. Domain	Jointly diagonalize cumulant matrices

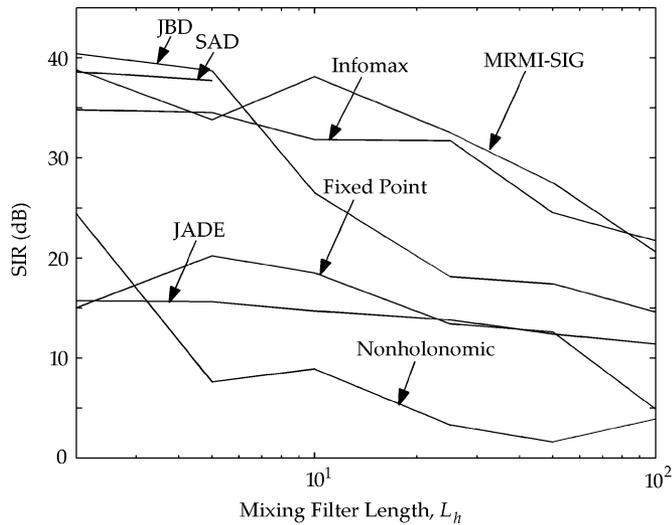


Fig. 2. SIR as a function of L_h where $L_w = L_h$.

Therefore, the results for JADE represent a best-case scenario for Type V methods. For the Nonholonomic, MRMI-SIG, SAD, and InfoMax algorithms, the stepsize was adjusted in order to optimize separation performance. Likewise, optimization of the block size (which determines the number of frequency bins) was performed for the Fixed Point and JADE algorithms and optimization of both the stepsize and the block size was done for JBD.

The figure of merit that is used is the signal-to-interference ratio (SIR), which for a given permutation is given by

$$\text{SIR} = \frac{1}{N_S} \sum_{i=1}^{N_s} 10 \log_{10} \left(\frac{P_{n_i}}{P_i - P_{n_i}} \right) \quad (11)$$

where n_i , for $i = 1, 2, \dots, N_s$, is an element of $\{1, 2, \dots, N_s\}$ which is used to define the particular permutation, n_i not equal to n_j for i not equal to j , P_{n_i} is the power of source n_i in output i , and P_i is the total power in output i . In words, the SIR for the i th output, SIR_i , is defined as 10 times the log of the ratio of the signal power to the interference power found in the specified output and the overall SIR of (11) is the mean of the individual SIR values.

Figs. 2–5 show the SIR as a single parameter of a synthetic mixture is varied while all other parameters assume the following default values, except as noted. The default number of sources/sensors is $N_s = 2$, the diagonal elements of the Z-transform of the mixing matrix are z^{-L} , the acausal factor for the mixing is $L = 0$, the length of the data is $N = 140 \cdot 10^3$, the length of the mixing filters is $L_h = 25$, the length of the demixing filters is $L_w = 25$, and the diagonal elements of

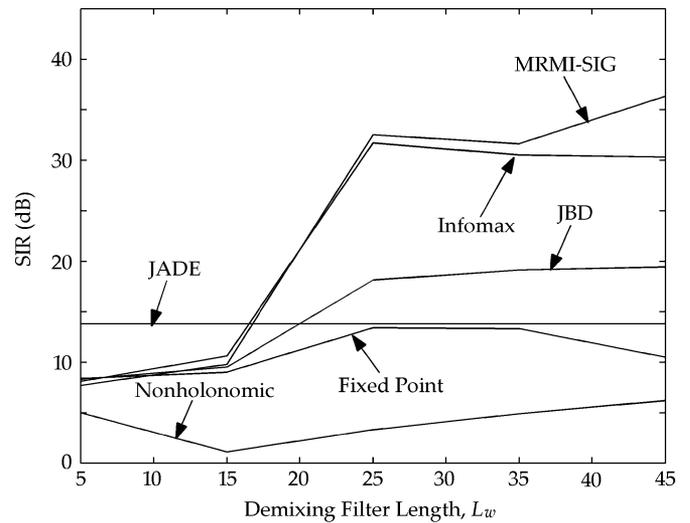


Fig. 3. SIR as a function of L_w for $L_h = 25$.

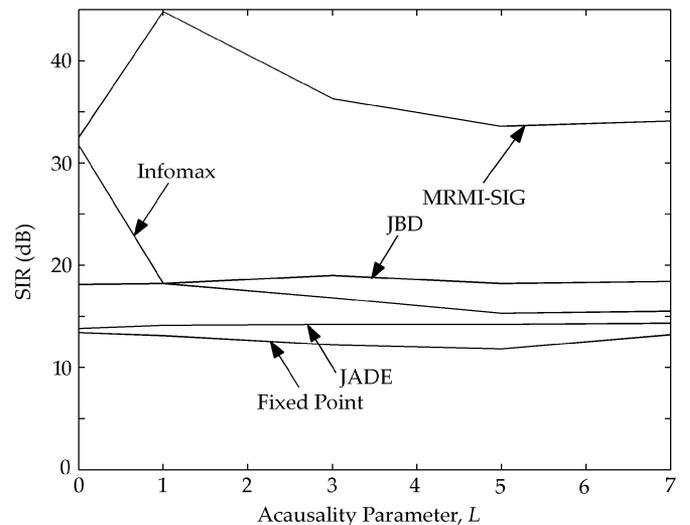


Fig. 4. SIR as a function of L .

the Z-transform of the demixing matrix are constrained to be $z^{-0} = 1$ for all methods except for the Nonholonomic method. Furthermore, the sources are speech signals and the off-diagonal components of the Z-transform of the mixing matrix are

$$\begin{aligned} H_{12}(z) &= 0.2z^{-1} - 0.2z^{-4} + 0.2z^{-9} + 0.4z^{-24} \\ &\quad + 0.1z^{-30} - 0.2z^{-49} + 0.1z^{-70} + 0.1z^{-99} \\ H_{21}(z) &= 0.3z^{-1} - 0.3z^{-4} + 0.4z^{-9} + 0.3z^{-24} \\ &\quad + 0.4z^{-30} - 0.1z^{-49} + 0.3z^{-70} - 0.2z^{-99} \end{aligned}$$

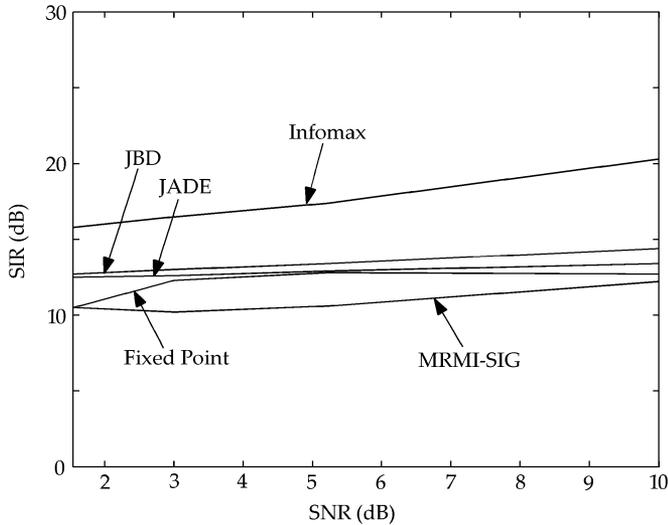


Fig. 5. SIR as a function of SNR.

where $H_{ij}(z)$ is the element of the $H(z)$ matrix located at the i th row and the j th column and where both transforms given above are modified so that the corresponding impulse response is truncated to the desired length L_h . For $N_s = 2$ and the mixing matrix defined above, a perfect separating solution is possible for: (1) all the time-domain algorithms in Table I whenever $L_w \geq L_h$ and $L = 0$ and (2) all the FF time-domain algorithms whenever $L_w \geq L_h + M$, $L > 0$, and the acausality parameter of the demixing is M larger than L , the acausality parameter of the mixing. These conditions are met for the default set of parameters used in the simulation. The second condition above is valid only for FF systems since FB systems are very restricted in their ability to implement acausal solutions.

Fig. 2 shows the results as L_h is varied from 2 to 100. In this experiment $L_w = L_h$ so that a perfect separating solution is always possible for all the time-domain methods. Values of SIR above 15 dB represent fairly good separation. The performance for SAD is only shown for small L_h due to stability problems encountered whenever $L_h \geq 10$, which also explains the exclusion of SAD from the subsequent figures. Fig. 3 shows how the SIR varies as a function of L_w for $L_h = 25$. For this experiment a perfect separating solution for the time-domain methods is possible only when $L_w \geq 25$. Notice that JBD, InfoMax, and MRMI-SIG are robust to a factor of two overestimation of L_w . The performance for JADE in this plot is shown as a straight line since it is not a function of L_w .

Fig. 4 shows how the convolutive source separation methods perform as L is increased, which causes the separating solution to become increasingly acausal. No additional results are shown for the Nonholonomic method due to poor separation performance for these specific datasets. In this experiment the diagonal elements of the demixing matrix are fixed at z^{-L} for the FF methods and 1 for the FB methods, which are unable to implement the required acausal solution for $L > 0$. Due to the choice of the mixing filters it is expected that the performance of the FB methods will fall as L transitions from 0 to 1, and as it transitions from 3 to 4. Fig. 4 indicates that the performance of InfoMax does indeed fall for the transition between 0 and 1, but the second transition is not very noticeable.

Fig. 5 shows how the separation performance is affected by the power of the spatially and temporally uncorrelated Gaussian noise that is added to the observations. The power of the noise is not removed in the calculation of the denominator of the SIR. Consequently, the SIR values are reduced from those shown in the previous plots, although this is inconsequential for relative performance comparisons. Notice that the proposed method, MRMI-SIG, is the least robust to additive noise while InfoMax performs quite well.

V. CONCLUSION

The proposed method MRMI-SIG implements Pham's suggested approach to convolutive BSS. The stochastic gradient coupled with a nonparametric estimation of Renyi's quadratic entropy of segments of signals is able to exploit both spatial and temporal information about the mixing process, thus allowing its use in convolutive source separation. Simulations of the method (not shown) demonstrated a very noticeable increase in performance by applying the approximation of (4), the idea of which came from insights provided by Morejon [22]. One item that is particularly interesting is that Shannon's entropy does not appear to be robust for convolutive BSS when used in a nonparametric entropy estimator. Moreover, only the choice of $\alpha = 2$ produces an efficient $O(N)$ criterion. Much work is still required to determine if there are additional conditions for which the proposed method might fail. The possible advantages of the proposed criterion are that it is IT, and it allows the use of a FF structure. This last item is important since there is some evidence that the (global) FF structure provides an improved separation performance than the alternatives [12], it is inherently stable unlike the FB structure, and it can easily implement arbitrary acausal solutions unlike the FB structure. The weaknesses of this criterion, which are left as open problems, are that it only works for super-Gaussian sources and it does not work well when the observations are noisy.

REFERENCES

- [1] S. Araki, S. Makino, T. Nishikawa, and H. Saruwatari, "Fundamental limitation of frequency domain blind source separation for convolutive mixture of speech," in *Proc. Int. Conf. Acoustics, Speech, Signal Processing*, vol. 5, Salt Lake City, UT, May 2001, pp. 2737–2740.
- [2] M. Knaak, S. Araki, and S. Makino, "Geometrically constraint ICA for convolutive mixtures of sound," in *Proc. Int. Conf. Acoustics, Speech, Signal Processing*, vol. 2, Hong Kong, Apr. 2003, pp. 725–728.
- [3] M. Gaeta, F. Briolle, and P. Esparcieux, "Blind separation of sources applied to convolutive mixtures in shallow water," in *IEEE Signal Process. Workshop Higher Order Statistics*, Banff, Canada, Jul. 1997, pp. 340–343.
- [4] Z. Xinhua, Z. Anqing, F. Jianping, and Y. Shaoqing, "Study on blind separation of underwater acoustic signals," in *Proc. Int. Conf. Signal Processing*, vol. 3, Beijing, China, Aug. 2000, pp. 1802–1805.
- [5] E. S. Warner and I. K. Proudler, "Single-channel blind signal separation of filtered MPSK signals," in *Proc. IEE-Radar, Sonar, Navig.*, vol. 150, Dec. 2003, pp. 396–402.
- [6] C. Andrieu, A. Doucet, and S. Godsill, "Bayesian blind marginal separation of convolutively mixed discrete sources," in *Proc. Neural Networks for Signal Processing*, Cambridge, MA, Aug. 1998, pp. 43–52.
- [7] G. Leus, P. Vandaele, and M. Moonen, "Deterministic blind modulation-induced source separation for digital wireless communications," *IEEE Trans. Signal Process.*, vol. 49, no. 1, pp. 219–227, Jan. 2001.
- [8] S. Haykin, *Adaptive Filter Theory*, 4th ed. Englewood Cliffs, NJ: Prentice-Hall, 2001.

- [9] T. Oliveira e Silva, "On the equivalence between gamma and laguerre filters," in *Proc. Int. Conf. Acoustics, Speech, Signal Processing*, vol. 4, Adelaide, Australia, Apr. 1994, pp. 385–388.
- [10] J. C. Principe, N. R. Euliano, and W. C. Lefabvre, *Neural and Adaptive Systems*. New York: Wiley, 1999.
- [11] L. Parra and C. Spence, "Convolutional blind separation of nonstationary sources," *IEEE Trans. Speech Audio Process.*, vol. 8, no. 3, pp. 320–327, May 2000.
- [12] K. E. Hild II, "Blind separation of convolutional mixtures using Renyi's divergence," Ph.D. dissertation, Dept. Elect. Comp. Eng., Univ. Florida, Gainesville, Nov. 2003.
- [13] K. Torkkola, "Blind separation of convolved sources based on information maximization," in *Proc. Neural Networks for Signal Processing*, Kyoto, Japan, Sep. 1996, pp. 1–10.
- [14] K. E. Hild II, D. Erdogmus, and J. C. Principe, "Blind source separation using information theoretic learning," *Signal Process.*, to be published.
- [15] D. Pham, "Mutual information approach to blind separation of stationary sources," *IEEE Trans. Inform. Theory*, vol. 48, no. 7, pp. 1935–1946, Jul. 2002.
- [16] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.
- [17] A. Rényi, *Probability Theory*. Amsterdam, The Netherlands: North-Holland, 1970.
- [18] D. Erdogmus, K. Hild II, and J. C. Principe, "Blind source separation using Renyi's alpha-marginal entropies," *Neurocomp.*, vol. 49, no. 1, pp. 25–38, Dec. 2002.
- [19] J. Principe, D. Xu, and J. Fisher, "Information theoretic learning," in *Unsupervised Adaptive Filtering*, S. Haykin, Ed. New York: Wiley, 2000, pp. 265–319.
- [20] E. Parzen, "On estimation of a probability density function and mode," *Ann. Math. Stat.*, vol. 33, no. 3, pp. 1065–1076, Sep. 1962.
- [21] D. Erdogmus, K. E. Hild II, and J. C. Principe, "On-line entropy manipulation: Stochastic information gradient," *IEEE Signal Process. Lett.*, vol. 10, no. 8, pp. 242–245, Aug. 2003.
- [22] R. A. Morejon, "An information-theoretic approach to sonar automatic target recognition," Ph.D. dissertation, Dept. Elect. Comp. Eng., Univ. Florida, Gainesville, 2003.
- [23] K. E. Hild II, D. Erdogmus, K. Torkkola, and J. C. Principe, "Sequential feature extraction using information-theoretic learning," *IEEE Trans. Pattern Anal. Mach. Intell.*, submitted for publication.
- [24] S. Douglas and X. Sun, "Blind separation of acoustical mixtures without time-domain deconvolution or decorrelation," *Proc. Neural Networks for Signal Processing*, pp. 323–332, Sep. 2001.
- [25] E. Bingham and A. Hyvarinen, "A fast fixed-point algorithm for independent component analysis of complex valued signals," *Int. J. Neural Syst.*, vol. 10, no. 1, pp. 1–8, Feb. 2000.
- [26] S. Van Gerven and D. Van Compernelle, "Signal separation by symmetric adaptive decorrelation: Stability, convergence, and uniqueness," *IEEE Trans. Signal Process.*, vol. 43, no. 7, pp. 1602–1612, Jul. 1995.
- [27] J. F. Cardoso and A. Souloumiac, "Blind beamforming for non-Gaussian signals," in *Proc. IEE-Radar, Signal Process.*, vol. 140, Dec. 1993, pp. 362–370.
- [28] F. Ehlers and H. Schuster, "Blind separation of convolutional mixtures and an application in automatic speech recognition in a noisy environment," *IEEE Trans. Signal Process.*, vol. 45, no. 10, pp. 2608–2612, Oct. 1997.
- [29] U. Lindgren and H. Broman, "Source separation using a criterion based on second-order statistics," *IEEE Trans. Signal Process.*, vol. 46, no. 7, pp. 1837–1850, Jul. 1998.
- [30] H. Sahlin and H. Broman, "Separation of real-world signals," *Signal Process.*, vol. 64, no. 1, pp. 103–113, Jan. 1998.
- [31] C. Fancourt and L. Parra, "The coherence function in blind source separation of convolutional mixtures of nonstationary signals," *Proc. Neural Networks for Signal Processing*, pp. 303–312, Sep. 2001.
- [32] E. Weinstein, M. Feder, and A. Oppenheim, "Multi-channel signal separation by decorrelation," *IEEE Trans. Speech Audio Process.*, vol. 1, no. 4, pp. 405–413, Oct. 1993.
- [33] R. Aichner, S. Araki, S. Makino, T. Nishikawa, and H. Saruwatari, "Time domain blind source separation of nonstationary convolved signals by utilizing geometric beamforming," in *Proc. Neural Networks for Signal Processing*, Martigny, Switzerland, Sep. 2002, pp. 445–454.
- [34] B. Krongold and D. Jones, "Blind source separation of nonstationary convolutedly mixed signals," in *Proc. IEEE Workshop Statistical Signal Array Processing*, Pocono Manor, PA, Aug. 2000, pp. 53–57.
- [35] M. Kawamoto, A. Barros, A. Mansour, K. Matsuoka, and N. Ohnishi, "Real world blind separation of convolved nonstationary signals," in *Proc. Int. Workshop ICA Signal Separation*, Jan. 1999, pp. 347–352.
- [36] M. Kawamoto, K. Matsuoka, and N. Ohnishi, "Real world blind separation of convolved speech signals," in *Proc. Int. Joint Conf. Neural Networks*, vol. 2, Washington, DC, Jul. 1999, pp. 993–997.
- [37] H. Wu and J. Principe, "A unifying criterion for blind source separation and decorrelation: Simultaneous diagonalization of correlation matrices," in *Proc. Neural Networks for Signal Processing*, Sep. 1997, pp. 496–505.
- [38] L. Parra and C. Spence, "On-line convolutional blind source separation of nonstationary signals," *J. VLSI Signal Process. Syst. Signal Image Video Technol.*, vol. 26, no. 1/2, pp. 39–46, Aug. 2000.
- [39] H. Bousbia-Salah, A. Belouchrani, and K. Abed-Meriam, "Jacobi-like algorithm for blind signal separation of convolutional mixtures," *Electron. Lett.*, vol. 37, no. 16, pp. 1049–1050, Aug. 2001.
- [40] —, "Blind separation of convolutional mixtures using joint block diagonalization," in *Proc. Int. Symp. Signal Process. Applications*, Aug. 2001, pp. 13–16.
- [41] H. Buchner, R. Aichner, and W. Kellermann, "A generalization of blind source separation algorithms for convolutional mixtures based on second-order statistics," *IEEE Trans. Speech Audio Process.*, vol. 13, no. 1, pp. 120–134, Jan. 2005.
- [42] X. Sun and S. Douglas, "Adaptive paraunitary filter banks for contrast-based multichannel blind deconvolution," in *Proc. Int. Conf. Acoustics, Speech, Signal Processing*, vol. 5, May 2001, pp. 2753–2756.
- [43] N. Charkani and Y. Deville, "Self-adaptive separation of convolutional mixed signals with a recursive structure. Part I: Stability analysis and optimization of asymptotic behavior," *Signal Process.*, vol. 73, no. 3, pp. 225–254, Jan. 1999.
- [44] M. Stanacevic, M. Cohen, and G. Cauwenberghs, "Blind separation of linear convolutional mixtures using orthogonal filter banks," in *Proc. Int. Workshop ICA Signal Separation*, Dec. 2001, pp. 260–265.
- [45] L. De Lathauwer, B. De Moor, and J. Vandewalle, "An algebraic approach to blind MIMO identification," in *Proc. Int. Workshop ICA Signal Separation*, Jun. 2000, pp. 211–214.
- [46] P. Comon, E. Moreau, and L. Rota, "Blind separation of convolutional mixtures, a contrast-based joint diagonalization approach," in *Proc. Int. Workshop ICA Signal Separation*, Dec. 2001, pp. 686–691.
- [47] M. Kawamoto, Y. Inouye, A. Mansour, and R.-W. Liu, "Blind deconvolution algorithms for MIMO FIR systems driven by fourth-order colored signals," in *Proc. Int. Workshop ICA Signal Separation*, Dec. 2001, pp. 692–697.
- [48] D. Yellin and E. Weinstein, "Criteria for multichannel signal separation," *IEEE Trans. Signal Process.*, vol. 42, no. 8, pp. 2158–2167, Aug. 1994.
- [49] R. Liu and Y. Inouye, "Blind equalization of MIMO-FIR channels driven by white but higher order colored source signals," *IEEE Trans. Inform. Theory*, vol. 48, no. 5, pp. 1206–1214, May 2002.
- [50] I. Fijalkow and P. Gaussier, "Self-organizing blind MIMO deconvolution using lateral-inhibition," in *Proc. Int. Workshop ICA Signal Separation*, Aussois, Jan. 1999, pp. 221–226.
- [51] C. Simon, P. Loubaton, C. Vignat, C. Jutten, and G. d'Urso, "Blind source separation of convolutional mixtures by maximization of fourth-order cumulants: The non-IID case," in *Proc. Asilomar Conf. Signals Systems Computers*, vol. 2, Pacific Grove, CA, Nov. 1998, pp. 1584–1588.
- [52] Y. Inouye and K. Tanebe, "Super-exponential algorithms for multichannel blind deconvolution," *IEEE Trans. Signal Process.*, vol. 48, no. 3, pp. 881–888, Mar. 2000.
- [53] J. Tugnait, "On blind separation of convolutional mixtures of independent linear signals in unknown additive noise," *IEEE Trans. Signal Process.*, vol. 46, no. 11, pp. 3117–3123, Nov. 1998.
- [54] E. Moreau and J. Pesquet, "Generalized contrasts for multichannel blind deconvolution of linear systems," *IEEE Signal Process. Lett.*, vol. 4, no. 6, Jun. 1997.
- [55] A. Touzni, I. Fijalkow, M. Larimore, and J. Treichler, "A globally convergent approach for blind MIMO adaptive deconvolution," *IEEE Trans. Signal Process.*, vol. 49, no. 6, pp. 1166–1178, Jun. 2001.
- [56] J. Tugnait, "Adaptive blind separation of convolutional mixtures of independent linear signals," in *Int. Conf. Acoust., Speech, Signal Process.*, vol. 4, Seattle, WA, May 1998, pp. 2097–2100.
- [57] P. Comon, "Contrasts for multichannel blind deconvolution," *IEEE Signal Process. Lett.*, vol. 3, no. 7, pp. 209–211, Jul. 1996.
- [58] H. Attias and C. E. Schreiner, "Blind source separation and deconvolution: The dynamic component analysis algorithm," *Neur. Comput.*, vol. 10, no. 6, pp. 1373–1424, Aug. 1998.
- [59] M. Ohata and K. Matsuoka, "Stability analyzes of information-theoretic blind separation algorithms in the case where the sources are nonlinear processes," *IEEE Trans. Signal Process.*, vol. 50, no. 1, pp. 69–77, Jan. 2002.

- [60] B. Pearlmutter and L. Parra, "Maximum likelihood blind source separation: A context-sensitive generalization of ICA," *Adv. Neur. Inform. Process. Syst.*, pp. 613–619, Dec. 1996.
- [61] X. Sun and S. Douglas, "A natural gradient convolutive blind source separation algorithm for speech mixtures," in *Proc. Int. Workshop ICA Signal Separation*, Dec. 2001, pp. 59–64.
- [62] D. Pham, "Mutual information approach to blind separation of stationary sources," in *Proc. Int. Workshop ICA Signal Separation*, Aussois, France, Jan. 1999, pp. 215–220.
- [63] H. Attias and C. Schreiner, "Blind source separation and deconvolution by dynamic component analysis," in *Proc. Neural Networks for Signal Processing*, Amelia Island, FL, Sep. 1997, pp. 456–465.
- [64] J. Reilly and L. Mendoza, "Blind signal separation for convolutive mixing environments using spatial-temporal processing," in *Proc. Int. Conf. Acoustics, Speech, Signal Processing*, vol. 3, Mar. 1999, pp. 1437–1440.
- [65] I. Sabala, A. Cichoki, and S. Amari, "Relationships between instantaneous blind source separation and multichannel blind deconvolution," in *Proc. Int. Joint Conf. Neural Networks*, vol. 1, Anchorage, AK, May 1998, pp. 39–44.
- [66] S. Amari, S. Douglas, A. Cichocki, and H. Yang, "Multichannel blind deconvolution and equalization using the natural gradient," in *Proc. Workshop Signal Processing Advances in Wireless Communication*, Apr. 1997, pp. 101–104.
- [67] L. Zhang, A. Cichocki, and S. Amari, "Multichannel blind deconvolution of nonminimum phase systems using information backpropagation," in *Proc. Int. Conf. Neural Information Processing*, vol. 1, Nov. 1999, pp. 210–216.
- [68] L. Parra, C. Spence, and B. De Vries, "Convolutive blind source separation based on multiple decorrelation," in *Proc. Neural Networks for Signal Processing*, Cambridge, MA, Aug. 1998, pp. 23–32.
- [69] J. Principe and H. Wu, "Blind separation of convolutive mixtures," in *Proc. Int. Joint Conf. Neural Networks*, vol. 2, Washington, DC, Jul. 1999, pp. 1054–1058.
- [70] H. Wu and J. Principe, "Simultaneous diagonalization in the frequency domain (SDIF) for source separation," in *Proc. Int. Work. ICA Signal Separation*, Aussois, France, Jan. 1999, pp. 245–250.
- [71] M. Ikram and D. Morgan, "A multiresolution approach to blind separation of speech signals in a reverberant environment," in *Proc. Int. Conf. Acoustics, Speech, Signal Processing*, vol. 5, May 2001, pp. 2757–2760.
- [72] T. Lee and R. Orglmeister, "A contextual blind separation of delayed and convolved sources," in *Proc. Int. Conf. Acoustics, Speech, Signal Processing*, vol. 2, Munich, Germany, Apr. 1997, pp. 1199–1202.
- [73] H. N. Thi and C. Jutten, "Blind separation for convolutive mixtures," *Signal Process.*, vol. 45, no. 2, pp. 209–229, Aug. 1995.
- [74] J. Platt and F. Faggin, "Networks for the separation of sources that are superimposed and delayed," *Adv. Neur. Inform. Process. Syst.*, pp. 730–737, Dec. 1991.
- [75] N. Charkani, Y. Deville, and J. Herault, "Stability analysis and optimization of time-domain convolutive source separation algorithms," in *Proc. Workshop Signal Processing Advanced Wireless Communication*, Apr. 1997, pp. 73–76.
- [76] N. Charkani and Y. Deville, "A convolutive source separation method with self-optimizing nonlinearities," in *Proc. Int. Conf. Acoustics, Speech, Signal Processing*, vol. 5, Mar. 1999, pp. 2909–2912.
- [77] —, "Self-adaptive separation of convolutively mixed signals with a recursive structure. Part II: Theoretical extensions and application to synthetic and real signals," *Signal Process.*, vol. 75, no. 2, pp. 117–140, Jun. 1999.
- [78] M. Girolami, "Symmetric adaptive maximum likelihood estimation for noise cancellation and signal separation," *Electron. Lett.*, vol. 33, no. 17, pp. 1437–1438, Aug. 14, 1997.
- [79] N. Kanlis, J. Simon, and S. Shamma, "Complete training analysis of feedback architecture networks that perform blind source separation and deconvolution," in *Proc. Int. Workshop ICA Signal Separation*, Jun. 2000, pp. 139–144.
- [80] K. Torkkola, "Blind separation of delayed sources based on information maximization," in *Proc. Int. Conf. Acoustics, Speech, Signal Processing*, vol. 6, Atlanta, GA, May 1996, pp. 3509–3512.
- [81] T. Lee, A. Bell, and R. Lambert, "Blind separation of delayed and convolved sources," *Adv. Neur. Inform. Process. Syst.*, pp. 758–764, Dec. 1996.
- [82] K. Rahbar and J. Reilly, "Blind source separation algorithm for MIMO convolutive mixtures," in *Proc. Int. Workshop ICA Signal Separation*, Dec. 2001, pp. 242–247.
- [83] M. Handa, T. Nagaim, and A. Kurematsu, "Frequency domain multi-channel speech separation and its applications," in *Proc. Int. Conf. Acoustics, Speech, Signal Processing*, vol. 5, May 2001, pp. 2761–2764.
- [84] L. Parra and C. Alvino, "Geometric source separation: Merging convolutive source separation with geometric beamforming," *IEEE Trans. Speech Audio Process.*, vol. 10, no. 6, pp. 352–362, Sep. 2002.
- [85] K. Rahbar and J. Reilly, "Blind source separation of convolved sources by joint approximate diagonalization of cross-spectral density matrices," in *Proc. Int. Conf. Acoustics, Speech, Signal Processing*, vol. 5, Salt Lake City, UT, May 2001, pp. 2745–2748.
- [86] A. Koutras, E. Dermatas, and G. Kokkinakis, "Continuous speech recognition in a multi-simultaneous-speaker environment using decorrelation filtering in the frequency domain," in *Proc. Int. Workshop Speech Computers*, St. Petersburg, Russia, Oct. 1998, pp. 253–256.
- [87] —, "Blind signal separation and speech recognition in the frequency domain," in *Proc. IEEE Int. Conf. Electronics Circuits Systems*, vol. 1, Sep. 1999, pp. 427–430.
- [88] J. Anamuler and B. Kollmeier, "Amplitude modulation decorrelation for convolutive blind source separation," in *Proc. Int. Workshop ICA Signal Separation*, Jun. 2000, pp. 215–220.
- [89] M. Kawamoto, A. Barros, K. Matsuoka, and N. Ohnishi, "A method of real-world separation implemented in frequency domain," *Proc. Int. Workshop ICA Signal Separation*, pp. 267–272, Jun. 2000.
- [90] C. Mejuto, "A second-order method for blind source separation of convolutive mixtures," in *Proc. Int. Workshop ICA Signal Separation*, Aussois, France, Jan. 1999, pp. 395–400.
- [91] W. Wang, S. Sanei, and J. A. Chambers, "Penalty function-based joint diagonalization approach for convolutive blind separation of nonstationary sources," *IEEE Trans. Signal Process.*, vol. 53, no. 5, pp. 1654–1669, May 2005.
- [92] W. Baumann, B. Kohler, D. Kolossa, and R. Orglmeister, "Real time separation of convolutive mixtures," in *Proc. Int. Workshop ICA Signal Separation*, San Diego, CA, Dec. 2001, pp. 65–69.
- [93] A. Dapena and C. Serviere, "A simplified frequency-domain approach for blind separation of convolutive mixtures," in *Proc. Int. Workshop ICA Signal Separation*, San Diego, CA, Dec. 2001, pp. 569–574.
- [94] D. Yellin and E. Weinstein, "Multichannel signal separation: Methods and analysis," *IEEE Trans. Signal Process.*, vol. 44, no. 1, pp. 106–118, Jan. 1996.
- [95] J. Pesquet, B. Chen, and A. Petropulu, "Frequency-domain contrast functions for separation of convolutive mixtures," in *Proc. Int. Conf. Acoustics, Speech, Signal Processing*, vol. 5, May 2001, pp. 2765–2768.
- [96] A. Dapena, M. Bugallo, and L. Castedo, "Separation of convolutive mixtures of temporally-white signals: A novel frequency-domain approach," in *Proc. Int. Workshop ICA Signal Separation*, Dec. 2001, pp. 179–184.
- [97] N. Mitianoudis and M. Davies, "New fixed-point algorithms for convolved mixtures," *Proc. Int. Workshop ICA Signal Separation*, pp. 633–638, Dec. 2001.
- [98] C. Serviere, "Blind source separation of convolutive mixtures," in *Proc. IEEE Workshop Statistical Signal and Array Processing*, Corfu, Greece, Jun. 1996, pp. 316–319.
- [99] R. Lambert and A. Bell, "Blind separation of multiple speakers in a multipath environment," in *Proc. Int. Conf. Acoustics, Speech, Signal Processing*, vol. 1, Munich, Germany, Apr. 1997, pp. 423–426.
- [100] F. Asano, S. Ikeda, M. Ogawa, H. Asoh, and N. Kitawaki, "A combined approach of array processing and independent component analysis for blind separation of acoustic signals," in *Proc. Int. Conf. Acoustics, Speech, Signal Processing*, vol. 5, Salt Lake City, UT, May 2001, pp. 2729–2732.
- [101] T. Lee, A. Bell, and R. Orglmeister, "Blind source separation of real world signals," in *Proc. Int. Conf. Neural Networks*, vol. 4, Houston, TX, Jun. 1997, pp. 2129–2134.
- [102] P. Smaragdis, "Efficient blind separation of convolved sound mixtures," in *Proc. IEEE ASSP Workshop Applications of Signal Processing Audio and Acoustics*, New Paltz, NY, Oct. 1997, pp. 19–22.
- [103] M. Joho, H. Mathis, and G. Moschytz, "An FFT-based algorithm for multichannel blind deconvolution," in *Proc. IEEE Int. Symp. Circuits Systems*, vol. 3, Orlando, FL, Jul. 1999, pp. 203–206.
- [104] C. Mejuto, A. Dapena, and L. Castedo, "Frequency-domain infomax for blind separation of convolutive mixtures," in *Proc. Int. Workshop ICA Signal Separation*, Jun. 2000, pp. 315–320.
- [105] P. Smaragdis, "Blind separation of convolved mixtures in the frequency domain," *Neurocomp.*, vol. 22, no. 1–3, pp. 21–34, Nov. 1998.
- [106] N. Mitianoudis and M. E. Davies, "Audio source separation of convolutive mixtures," *IEEE Trans. Speech Audio Process.*, vol. 11, no. 5, pp. 489–497, Sep. 2003.

- [107] A. Mansour, C. Jutten, and P. Loubaton, "Adaptive subspace algorithm for blind separation of independent sources in convolutive mixture," *IEEE Trans. Signal Process.*, vol. 48, no. 2, pp. 583–586, Feb. 2000.
- [108] W. Hachem, F. Desbouvries, and P. Loubaton, "On the identification of certain noisy FIR convolutive mixtures," in *Proc. Int. Workshop ICA Signal Separation*, Jan. 1999, pp. 401–405.
- [109] E. Moulines, P. Duhamel, J. F. Cardoso, and S. Mayrargue, "Subspace methods for the blind identification for multichannel FIR filters," *IEEE Trans. Signal Process.*, vol. 43, no. 2, pp. 516–525, Feb. 1995.
- [110] K. Abed-Meriam, Y. Hua, Ph. Loubaton, and E. Moulines, "Subspace method for blind identification of multichannel FIR systems in noise field with unknown spatial covariance," *IEEE Signal Process. Lett.*, vol. 4, no. 5, pp. 135–137, May 1997.
- [111] C. Simon, G. d'Urso, C. Vignat, P. Loubaton, and C. Jutten, "On the convolutive mixture source separation by the decorrelation approach," in *Proc. Int. Conf. Acoustics, Speech, Signal Processing*, vol. 4, Seattle, WA, May 1998, pp. 2109–2112.
- [112] N. Delfosse and P. Loubaton, "Adaptive blind separation of convolutive mixtures," in *Proc. Int. Conf. Acoustics, Speech, Signal Processing*, vol. 5, Atlanta, GA, May 1996, pp. 2940–2943.
- [113] ———, "Adaptive blind separation of convolutive mixtures," in *Proc. Asilomar Conf. Signals, Systems, Computers*, vol. 1, Pacific Grove, CA, Oct. 1995, pp. 341–345.
- [114] S. Icart and R. Gautier, "Blind separation of convolutive mixtures using 2nd and 4th order moments," in *Proc. Int. Conf. Acoustics, Speech, Signal Processing*, vol. 5, Atlanta, GA, May 1996, pp. 3018–3021.



Kenneth E. Hild, II (M'90) received the B.S.E.E. and M.S.E.E. degrees from The University of Oklahoma, Norman, with emphasis on signal processing and communications, and the Ph.D. degree in electrical engineering from The University of Florida, Gainesville, in 1992, 1996, and 2003, respectively.

He then worked at Seagate Technologies, where he served as an Advisory Development Engineer in the Advanced Concepts Group. He taught several classes on adaptive filter theory and stochastic processes and studied information theoretic learning and blind source separation in the Computational NeuroEngineering Laboratory, The University of Florida. He is currently with the Biomagnetic Imaging Laboratory, Department of Radiology, The University of California at San Francisco, where he is applying Variational Bayesian techniques for biomedical signal processing of electroencephalographic (EEG), magnetoencephalographic (MEG), and magnetocardiographic (MKG) data. He has also studied Biomedical Informatics at Stanford University, Palo Alto.

Dr. Hild is a member of Tau Beta Pi and Eta Kappa Nu, and the International Society for Brain Electromagnetic Topography.



David Pinto (M'02) received the M.S.E.E., B.S.E.E., and C.E.E.E. degrees from the University of Florida, Gainesville, all in 2003.

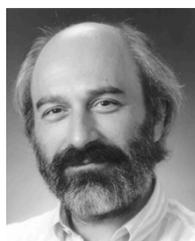
His scholastic emphasis was placed on digital signal processing, controls, adaptive and neural systems, pattern recognition, information theoretic algorithms, and computer information sciences. He has been with Sandia National Labs, Albuquerque, NM, since 2004 where he focuses on sensor networks and software architecture.



Deniz Erdogmus (M'02) received the B.S. degree in electrical and electronics engineering and mathematics in 1997 and the M.S. degree in electrical and electronics engineering, with emphasis on systems and control, in 1999, both from the Middle East Technical University, Ankara, Turkey. He received the Ph.D. degree in electrical and computer engineering from the University of Florida, Gainesville, in 2002.

He was a Research Engineer with the Defense Industries Research and Development Institute (SAGE), Ankara, from 1997 to 1999. From 1999 until 2004, he was with the Computational NeuroEngineering Laboratory, University of Florida, under the supervision of Dr. J. C. Principe, the last two years of which he was a Postdoctoral Fellow. He is currently an Assistant Professor with a joint appointment in the Department of Computer Science and Engineering and the Department of Biomedical Engineering, Oregon Health & Science University, Beaverton. His current research interests include information theory and its applications to adaptive systems and adaptive systems for signal processing, communications, and control.

Dr. Erdogmus is a member of Tau Beta Pi and Eta Kappa Nu.



Jose C. Principe (M'83–SM'90–F'00) received the Licenciatura in electrical engineering from the University of Porto, Portugal, in 1972, the M.Sc. and Ph.D. degrees in electrical engineering from the University of Florida, Gainesville, in 1974, and 1979, respectively, and the Agregado degree from the University of Aveiro, Portugal in 1985.

He is Distinguished Professor of Electrical and Biomedical Engineering with the University of Florida, Gainesville, where he teaches advanced signal processing and artificial neural networks (ANNs) modeling. He is BellSouth Professor and Founder and Director of the University of Florida Computational NeuroEngineering Laboratory (CNEL). He has been involved in biomedical signal processing, brain machine interfaces, nonlinear dynamics, and adaptive systems theory (information theoretic learning). He has more than 100 publications in refereed journals, 10 book chapters, and over 200 conference papers. He has directed 42 Ph.D. degree dissertations and 57 Master's degree theses.

Dr. Principe is Editor-in-Chief of the IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING, President of the International Neural Network Society, and formal Secretary of the Technical Committee on Neural Networks of the IEEE Signal Processing Society. He is also a member of the Scientific Board of the Food and Drug Administration, and a member of the Advisory Board of the University of Florida Brain Institute.