Convolutive Source Separation and Signal Modeling with ML

Lucas C. Parra, Clay Spence, Bert de Vries
Sarnoff Corporation, CN-5300, Princeton, NJ 08543
lparra@sarnoff.com, devries@sarnoff.com, cspence@sarnoff.com

September 5, 1997

Abstract

In independent component analysis (ICA) an instantaneous mix of sources can be recovered using maximum Likelihood (ML). In Convolutive blind source separation (BSS) the mixture arises as a combination of differently convolved source signals due to time delays and a reverberating acoustic environment. Instead of modeling a particular time instant now a time window of the mixed signals has to be modeled. This allows to combine ICA with traditional ML signal modeling techniques. Here we use an auto-regressive (AR) model of the sources leading to a generalization of contextual ICA [18] to the convolutive case. This may improve source separation avoiding the typical whitening of the sources, and may allow us to incorporation simultaneous enhancing of the signal based on the AR models.

1 Introduction

Independent component analysis (ICA) aims to find statistical independent signals in an instantaneous linear mix. This concept was first introduced and formalized by Comon [7]. In recent years increasing interest in this concept arose in the neural network and signal processing community. There are wide range of possible applications. In the neural network community it has been suggested as a version of the well know concept of redundancy reduction that has been repeatedly considered as an underlying principle in sensory formation [5, 3, 8, 16]. For signal processing the interest arises in the context of sensor arrays, source localization, sonar applications, multidimensional blind channel equalization, spread-spectrum coding, QAM coding, and speech enhancement.

In the context of Neural Networks it can be formulated as finding a representation of the data with minimal mutual information in each output coordinate [8]. It can also be formulated as the representation that maximizes the information transmitted through a properly designed network [6, 11]. In the context of signal processing however its formulation as a Maximum Likelihood (ML) problem [19] allows the extension of ICA to Blind Source Separation (BSS) by properly adding the time dimension. In the instantaneous mixing case, modeling the time correlations of the signal improves the quality of the separation [18]. In contrast to the instantaneous mix, in convolutive BSS the mixture arises as a combination of differently convolved independent source signals due to time delays and a reverberating acoustic environment. ML allows us to combine ICA with traditional signal modeling techniques like the standard AR model, mixture models and others, which use specific models of the temporal properties of the sources. We will demonstrate here the combination of AR models and ICA which represents the extension of contextual ICA [18] to the convolutive case. This may improve source separation, avoiding the typical whitening of the sources. This may allow us also to combine source separation with ML signal enhancement based on AR models of the acoustic source signals, a problem on which we are currently working.

Earlier work on convolutive BSS has been presented in [23, 24]. It is based on a no-parametric approach whereby cumulants are used to capture the higher order statistic of the signal, leading to somewhat expensive and complicated objective functions. In contrast, in ML one tries to find a parametric model for the densities that captures higher statistic of the signals indirectly.

Higher order statistic can be also captured by using prior knowledge of the single source density functions and incorporating proper non-linearities into the objective functions. This is done in [12], where a multidimensional generalizations of well known blind equalization algorithms is presented. Very recently in [2] an alternative derivation for a similar algorithm has been given. There again, the underlying assumptions leads to signal equalization. A generalization of the instantaneous un-mixing based on maximum entropy has also been suggested [13], but an explicit derivation is so far lacking. The present work remedies that lack and represents an extension of previous work in the sense that is does not assume white sources (equalized model sources).

In the first section 2 we lay out the ML model for a time window of the detected convolutive mixtures in terms of statistical independent model sources. The following section 3 discussed some issues related to the modeling of temporal correlations of single sources and introduces the AR model into our framework. Section 4 derives a gradient expression of the likelihood function, leading to efficient update equations for the un-mixing filters. The expressions however can be only derived by making a circularity assumption on the signals, and filter, which cuts some of the generality of the first two sections.

2 Convolutive Mixture Model

Assume N independent sources $(s_1(t)...s_N(t))^\top = \mathbf{s}(t)$ mixed in a unknown linear medium,

$$x_i(t) = \sum_{j=1}^{N} \sum_{\tau=0}^{\infty} h_{ij}(\tau) s_j(t-\tau) \tag{1}$$

We observe the mixtures $(x_1(t)...x_N(t))^\top = \mathbf{x}(t)$. To undo the effect of this causal filtering and mixing we require (eventually infinite size) non-causal finite impulse response (FIR) filters. We wish to find N statistically independent signals $(y_1(t)...y_N(t))^\top = \mathbf{y}(t)$ with a multidimensional, non-causal FIR filter $w(-K)...w(K)$ from the convolutive mixtures, where we limit us to a finite filter size K. Every $w(\tau)$ here represents a $N \times N$ unximing matrix for the time lag τ,

$$y_i(t) = \sum_{j=1}^{N} \sum_{j=-K}^{K} w_{ij}(\tau) x_j(t-\tau) \tag{2}$$

Note that we are not explicitly aiming to recover the original signals $\mathbf{s}(t)$ that lead to the mixtures $\mathbf{x}(t)$. We will merely try to model the mixtures by independent model sources $\mathbf{y}(t)$. The true sources $\mathbf{s}(t)$ may differ from these recovered independent signals $\mathbf{y}(t)$ by an arbitrary convolution and permutation [2]. We will try however to match the statistics of the sources by using linear prediction or signal subspace modeling techniques.

For the ML approach we require a density function of the observed signals as a function of the model parameters, which we will for now generically denote Φ. We will formulate the density function for a time window of the mixture signals $X(t) = (\mathbf{x}(t), ..., \mathbf{x}(t+T))$. This stands in contrast to previous formulations of the problem that have considered the likelihood of a single time instance only [18, 15, 2].

In order to express the density function in the space of the model sources we will consider the conditional density of the signals within the window condition on the signals outside the window which we will denote by $X \setminus X(t) = \mathbf{x}(-\infty), ..., \mathbf{x}(t-1), \mathbf{x}(t+T+1), ..., \mathbf{x}(\infty)$.

$$p(X(t)|X \setminus X(t); \Phi) = \left| \frac{\partial Y(t)}{\partial X(t)} \right| p(Y(t)|X \setminus X(t); \Phi) \tag{3}$$

Here $Y(t) = (\mathbf{y}(t), ..., \mathbf{y}(t+T))$ is the corresponding window in the model source space[1]. The Jacobian $\frac{\partial Y(t)}{\partial X(t)}$ is a $NT \times NT$ matrix with coefficients,

$$\frac{\partial y_i(t)}{\partial x_l(r)} = \sum_{j=1}^{N} \sum_{\tau=-K}^{K} w_{ij}(\tau) \frac{\partial x_j(t-\tau)}{\partial x_l(r)} = w_{il}(t-r) \tag{4}$$

where $i, l = 0...T$ and $r = 0..K$, and $w(\tau)$ vanishes for values outside $-K \leq \tau \leq K$. It is instructive to arrange the coefficients of the Jacobian such that the matrix $w(0)$ lies on the diagonal blocks,

$$\frac{\partial Y(t)}{\partial X(t)} = \begin{pmatrix} w(0) & w(-1) & ... & w(-T) \\ w(1) & w(0) & ... & w(1-T) \\ ... & ... & ... & ... \\ w(T) & w(T-1) & ... & w(0) \end{pmatrix} \equiv W \tag{5}$$

For a causal FIR the upper block triangle vanishes and the determinant in (3) is given by the determinant of $w(0)$,

$$\left| \frac{\partial Y(t)}{\partial X(t)} \right| = |w(0)|^T \text{ if } w(\tau) = 0 \text{ for } \tau < 0 \tag{6}$$

Although some have made these simplifying assumptions [4, 22], we wish to keep a non-causal filter, and will instead restrict ourself in section 4 to a circulant W in order to arrive to an efficient algorithm that can be implemented using the fast Fourier transform (FFT).

Now we introduce the independence assumption for the model sources by replacing in (3) the joint density of the model source by the product of the density of the individual sources,

$$p(X(t)|X \setminus X(t); \Phi) = |W(t)| \prod_{i=1}^{N} p(y_i(t)...y_i(t+T)|X \setminus X(t); \Phi) \tag{7}$$

3 Source Modeling

To our knowledge all current BSS algorithm make at this point for each ith model source a time independence assumption in (7) for the joint density of $y_i(t)...y_i(t+T)$ [12, 13, 15, 2, 4] This is for any reasonable acoustic signal not an appropriate model, and leads in their experiments to a whitened signal recovery.

[1]For the ML approach one requires the density of the observations $X(t)$ as a function of $X(t)$ itself and some model parameters Φ. Therefore, we have to replace $Y(t)$ in (3) by its definition (2). Note that is not possible to write $Y(t)$ as a function of $X(t)$ only. The model source values in the window at time t will depend by definition (2) on mixture values before and after the current frame. The conditioning of the probability on $X \setminus X(t)$ is therefore a crucial step in order to make that substitution. In section 4 however, we consider periodic signals and the conditioning becomes superfluous.

The field of signal modeling in particular for speech enhancement offers a variety of ML approaches to single channel modeling. At this point many of these approaches can be combined to source separation by inserting the corresponding model probability into (7). We note however that all efficient algorithms are based on a linear dependency of the the variables $(y(t)...y(t+T))^\top = \mathbf{y}(t)$ and mostly a Gaussian density[2].

For source separation however we require a non-Gaussian model since statistical independence is not uniquely defined for more than one Gaussian component in the mixture [25].

If one uses linear time correlation as described by a covariance matrix or linear prediction coefficients (LPC) the parameters introduced are equivalent to the parameters of the convolutions of the un-mixing FIR. The hope is however that the parameters describing the un-mixing and the parameters describing the source signal have different stationarity time scales. Speech for example will be stationary only within some 20ms - 40ms time frame, while the un-mixing coefficients should remain constant at least on a seconds scale, assuming that the location of the sources and the environment remains constant over that time. Single channel algorithms that adapt to varying statistics on a millisecond rate, as required by any single microphone speech enhancement algorithm, will extract to a certain extend the rapid varying portion of the linear correlations, while the slower converging source separation will pick up the slow varying time correlation due to the linear medium that mixes the source signals.

For short times on the order of 100-200 samples at 8kHz sampling rate the second order statistics of speech is well described with a multivariate Gaussian density. The covariance matrix however will change for larger time periods. The overall density will therefore be an accumulation or mixture of the instantaneous statistics. The net result of such a mixture is that the overall joint distribution will have high kurtosis, i.e. a strong mass at low amplitudes due to silence periods and long tales for high amplitude peeks. In the BSS literature the signal distribution has been therefore approximated by non-Gaussian distributions. The strongest approximation ignores time correlations, and assumes a high kurtosis one time step accumulated density $f(y)$,

$$p(y(t)...y(t+T)) \approx \prod_{\tau=0}^{T} f(y(t+\tau)) \tag{8}$$

A generalized Gaussian [13] has been used for $f(y)$ or a mixture of two zero mean Gaussian with variances describing the silence and signal amplitudes [15]. A better approximation might result if we avoid the time independence assumption, and capture linear time correlations with a matrix $\Sigma(t)$ estimated in some window around t,

$$p(y(t)...y(t+T)) \approx f(\mathbf{y}^\top(t)\Sigma(t)\mathbf{y}(t)) \tag{9}$$

One might use also a multiple Gaussian model that allows for different covariance $\Sigma(t)$ for the silence, voiced and unvoiced, states up to a full hidden Markov model that incorporates state transition probabilities of the different sounds. These are routinely used for speech recognition resulting however in rather expensive models that require prior training.

We suggest a short term estimation of the linear correlations according the subspace or linear prediction methods used in speech enhancement [10, 14, 20, 9]. Assuming a correlation time P for which $\Sigma_{ij} = 0$ if $i - j > P$ we can expand the density of a frame of size T as,

[2]We have ignored here and in the reminder of this section the model source index i since we are dealing with a individual channel. Bold notation now refers to the vector of signal values in the time window of size T for a single channel.

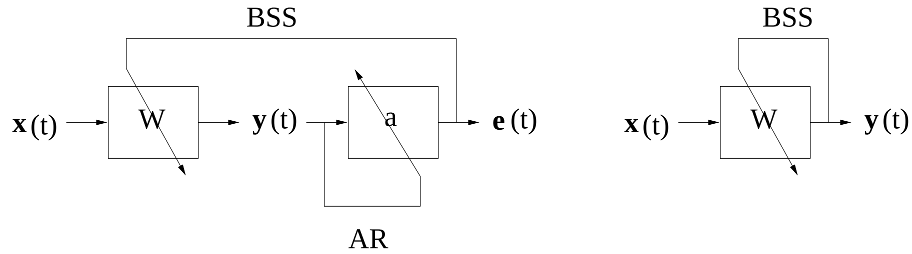

Figure 1: Schematic representation of the suggested signal modeling and its relation to previous convolutive BSS algorithms. Left: The AR model whitens the signal producing the error signal $\mathbf{e}(t)$ which is used in turn for the BSS update. Details can be seen in the final update equation (26). Right: In previous work no modeling of temporal correlations of the model signal leads to separation and equalization.

$$p(y(t)...y(t+T)) \quad = \quad p(y(t)...y(t+P-1)) \prod_{\tau=t+P}^{t+T} p(y(\tau)|y(\tau-1)...y(\tau-P)) \tag{10}$$

The auto-regressive (AR) model makes a linear prediction $\bar{y}(t)$ of $y(t)$ from the past P samples, $e(t) = y(t) - \bar{y}(t) = y(t) - \sum_{\tau=1}^{P} a(\tau)y(t-\tau)$, where $e(t)$ is considered to be the error of the prediction, and $a(\tau)$ the linear prediction coefficients (LPC). The corresponding density function is then,

$$p(y(t)|y(t-1)...y(t-P);\mathbf{a}) = p(a^\top y(t)), \text{ with } \mathbf{a} = (1, -a(1), .., -a(P))^\top \tag{11}$$

Combining this source signal model for each of the N sources i with the source independence model we obtain the overall logarithmic likelihood[3],

$$L(W, \mathbf{a}_i...\mathbf{a}_N) = \ln|W| + \sum_{i=1}^{N} \sum_{\tau=t+P}^{t+T} \ln p(\mathbf{a}_i^\top \mathbf{y}_i(\tau)|X \setminus X(\tau); W) + \sum_{i=1}^{N} \ln p(y_i(t)...y_i(t+P-1)|...) \tag{12}$$

We will assume the LPC parameter to be constant within a time frame but change from frame to frame[4]. The un-mixing filters we assume constant throughout time. For $P << T$ we can neglect the last term here, and initialize the sum in the second term at $\tau = t$.

The extension done in this section compared to previous work is schematically depicted in figure 1. The present formulation should avoid the whitening of the model sources if the order P of the AR model is sufficiently large and the window size T in which to compute the AR parameters is sufficiently small to capture the fast variation of speech.

If the densities can in fact be described in a short time frame of size P by a (zero mean) Gaussian we have,

[3]The conditionalization on $X \setminus X(\tau)$ means nothing else that we can now substitute all $y()$ using (2). In the next section however it will be necessary to to assume periodic signals in order to obtain an efficient algorithm for updating W. The conditioning on the previous frame becomes then superfluous.

[4]To be precise we should consider the Likelihood of multiple frames by adding frame index k to $\mathbf{a}$, and setting t=k*T, while adding over all frames

$$p(y(t)...y(t+P)) = \mathcal{N}(\mathbf{0}, \Sigma) \tag{13}$$

The conditional density of a time sample $y(t)$ given its past P samples is then described by a one dimensional normal distribution with mean $\bar{y}(t)$ and variance σ depending on the correlation matrix Σ,

$$p(y(t)|y(t-1)...y(t-P)) = \mathcal{N}(\bar{y}(t), \sigma) \tag{14}$$

The instantaneous prediction coefficients $\mathbf{a}$, which define the mean $\bar{y}(t)$, and the prediction accuracy, i.e. variance σ, are in this case explicitly determined by the covariance matrix Σ,

$$\mathbf{a} = \Sigma_{+-}\Sigma_{-}^{-1} \text{ , and } \sigma = \Sigma/\Sigma_{-} = \Sigma_{+} - \Sigma_{+-}\Sigma_{-}^{-1}\Sigma_{-+} \tag{15}$$

$$\Sigma = \left(\begin{array}{c|c} \Sigma_{+} & \Sigma_{+-} \\ \hline \Sigma_{-+} & \Sigma_{-} \end{array}\right) = \left(\begin{array}{c|ccc} \sigma_{11} & \sigma_{12} & ... & \sigma_{1P} \\ \hline \sigma_{21} & \sigma_{22} & ... & \sigma_{2P} \\ ... & ... & ... & ... \\ \sigma_{P1} & \sigma_{P2} & ... & \sigma_{PP} \end{array}\right) \tag{16}$$

We suggest to compute these coefficients from the sample autocorrelation matrix $\hat{\Sigma}$ estimated with the samples in the current window of since T. We may use the expressions above or more efficiently use Levinson-Durvin recursion on the signals $y(t)$ to compute the LPC from $\hat{\Sigma}$ [17]. This recursion gives an analytic solution for the LPC coefficients with a minimum least squares criterion, which is equivalent to the ML using the suggested signal model for a given frame [17].[5]

4 Stochastic ML gradient

In other to optimize the logarithmic likelihood (12) also for the un-mixing filters W we use gradient descent. The main difficulty in deriving a gradient expression of (12) is to find a feasible expression for the derivative of the Jacobian (5). In fact, it will be necessary to assume that W is a circulant matrix,

$$W = (w_{ij}(n,m)) \equiv \begin{pmatrix} w(0) & w(T) & ... & w(1) \\ w(1) & w(0) & ... & w(2) \\ ... & ... & ... & ... \\ w(T) & w(T-1) & ... & w(0) \end{pmatrix} \tag{17}$$

The coefficients n, m denote the block column and row index, while the subscript index i, j refer to the index within each block. Columns are indexed therefore by i, n and rows by j, m. With this notation we can write now the gradient as,

$$\frac{\partial \ln |W|}{\partial w_{ij}(n,m)} = \sum_{n'm'}^{TT} \sum_{i'j'}^{NN} \frac{\partial \ln |W|}{\partial w_{i'j'}(n',m')} \frac{\partial w_{i'j'}(n',m')}{\partial w_{ij}(n,m)} = \frac{1}{|W|} \sum_{n'm'}^{TT} |W_{ij}(n',m')| \delta_{n'-m'}^{n-m} \tag{18}$$

where $\delta_{z'}^{z} = 1$, if $modulo(z, (T+1)) = modulo(z', (T+1))$, and 0 otherwise. We have used

[5] To be precise the ML estimate of the LPC coefficients is computed with the so called covariance method which uses a somewhat different iteration [21].

he fact that for any invertible, square matrix A, $\partial \ln |A| / \partial a_{ij} = |A_{ij}|/|A|$, where $|A_{ij}|$ is the the determinant of the matrix obtained after removing the ith row and jth column in A. Computing these determinants is an expensive operation of order $O(N^3T^3)$. To avoid this we will use the argument commonly used in ICA algorithms, which was first introduced by Amari [1]. We multiply the gradients with a positive definite matrix $W^\top W$, to obtain the so called natural gradient. First consider

$$
\begin{aligned}
\left(\frac{\partial \ln |W|}{\partial W} W^\top\right)_{iu}(n,l) &= \sum_{mj}^{TN} \frac{\partial \ln |W|}{\partial w_{ij}(n,m)} w_{uj}(l,m) \\
&= \sum_{mj}^{TN} \frac{1}{|W|} \sum_{n'm'}^{TT} |W_{ij}(n',m')| \delta_{n'-m'}^{n-m} w_{uj}(l,m) \\
&= \frac{1}{|W|} \sum_{j}^{N} \sum_{n'm'}^{TT} |W_{ij}(n',m')| w_{uj}(l-n+n',m') \\
&= \frac{1}{|W|} \sum_{n'}^{T} \begin{cases} |W| & \text{if } n' = l-n+n' \text{ and } i=u \\ 0 & \text{otherwise} \end{cases} = \frac{T}{|W|} I
\end{aligned} \tag{19}
$$

Read above $w_{ij}(n,m) = w_{ij}(modulo(n,(T+1)), modulo(m,(T+1)))$ if indexes n,m exceed their range $0..T$. Multiplying this identity matrix I with W finally leads to,

$$
\frac{\partial \ln |W|}{\partial W} W^\top W = TW \tag{20}
$$

Now we need to compute the gradient of the second term in (12).

$$
\begin{aligned}
\frac{\partial \sum_{k\tau} \ln p(\mathbf{a}_k^\top \mathbf{y}_k(\tau))}{\partial w_{ij}(z)} &= \frac{\partial}{\partial w_{ij}(z)} \sum_{k=1}^{N} \sum_{\tau=t}^{t+T} \ln p\Big(\sum_{\tau'=0}^{P} a_k(\tau') y_k(\tau-\tau')\Big) && (21) \\
&= \sum_{\tau=t}^{t+T} g(\mathbf{a}_i^\top \mathbf{y}_i(\tau)) \sum_{\tau'=0}^{P} a_i(\tau') \frac{\partial}{\partial w_{ij}(z)} y_i(\tau-\tau') && (22) \\
&= \sum_{\tau=t}^{t+T} g(\mathbf{a}_i^\top \mathbf{y}_i^\top(\tau)) \sum_{\tau'=0}^{P} a_i(\tau') x_j(\tau-\tau'-z) && (23)
\end{aligned}
$$

where $g(e) = \partial \ln p(e)/\partial e$. These coefficients for $z = 0..T$ represent the first column of the corresponding circulant matrix $\partial/\partial W$ arranged analogous to (17). In order to simplify the following multiplication with $W^\top W$ one has to assume periodic signals $\mathbf{x}(t)$, i.e. $\mathbf{x}(t) = \mathbf{x}(t+T+1)$. The model signals $\mathbf{y}(t)$ will then be periodic with period $T+1$ as well. This assumption not only simplifies the expressions but allows us to implement the convolutions with a discrete Fourier transform using a FFT. After some manipulations we obtain, again for the elements of the first column of a circulant matrix,

$$
\left(\frac{\partial \sum_{k\tau} \ln p(\mathbf{a}_k^\top \mathbf{y}_k(\tau))}{\partial W} W^\top W\right)_{ij}(z) = \sum_{\tau=t}^{t+T} \sum_{u=1}^{N} \sum_{\tau'=t}^{t+T} g(\mathbf{a}_i^\top \mathbf{y}_i(\tau)) \mathbf{a}_i^\top \mathbf{y}_u(\tau+\tau'-z) w_{uj}(\tau') \tag{24}
$$

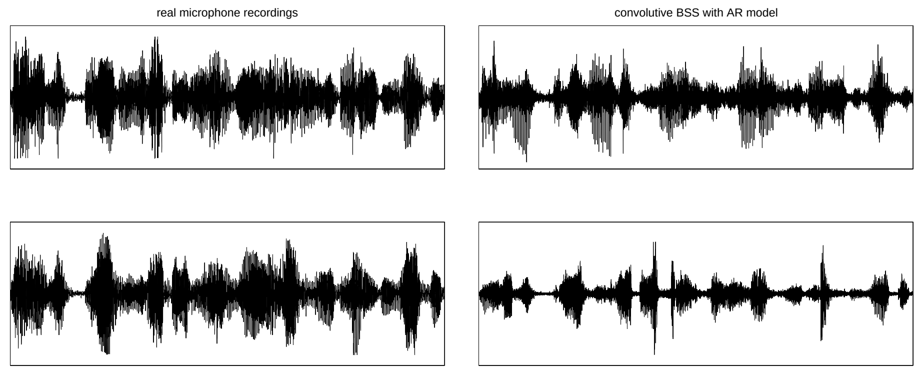

Figure 2: Separation of real recordings with two microphones in a reverberant environment (office room) with algorithm (26) with $T = 512$, which corresponds at 8kHz to 64ms. The AR parameter were computed in each frame with the Levinson-Durvin recursion with $P = 20$.

These convolutions now can best be performed in the frequency domain. Transforming the natural gradient of (12), as expressed by the sum of (20) and (24), into the frequency domain involves applying the orthonormal coordinate transformation expressed by a matrix F with elements $F_{\nu\tau} = \frac{1}{\sqrt{T+1}} e^{-\mathbf{i}\tau\nu 2\pi/(T+1)}$, $\mathbf{i} = \sqrt{-1}$, which results for a circulant matrix like W into[6],

$$FWF^{-1} = diag(\mathcal{W}(0), ..., \mathcal{W}(T)) \tag{25}$$

That is, the Fourier coefficients $\mathcal{W} = F\mathbf{w}$ of the circulant filter $\mathbf{w} = (w(0), ..., w(T))^\top$ represent the diagonal elements of a diagonal matrix. According to the convolution theorem the convolutions in (24) are performed by multiplying the Furrier coefficients independently. The overall gradients separate therefore in the frequency domain. Combining (20) and (24) and transforming the result into the frequency domain we obtain the total natural gradient $\partial\mathcal{W}(\nu)$ in each frequency ν.

$$\partial\mathcal{W}_{ij}(\nu) = \mathcal{W}_{ij}(\nu) + G\left[\mathcal{A}_i(\nu)\mathcal{Y}_i(\nu)\right] \sum_{u=1}^{N} \mathcal{A}_i^*(\nu)\mathcal{Y}_u^*(\nu)\mathcal{W}_{uj}(\nu) \tag{26}$$

where $G[]$ is an operator that applies the function $g()$ in the temporal domain, $G[\mathcal{Y}] = Fg\left(F^{-1}Y\right)$, and the Fourier coefficients are given by, $\mathcal{A}_i = F(a_i(0), a_i(1), ..., a_i(P), 0, 0, ..., 0)^\top$ and $\mathcal{Y}_i = F\mathbf{y}_i$.

This result represents the extension of contextual ICA [18] to the convolutive case. It also represents a generalization of the equations suggested in [12, 13], which one obtains for $P = 1$, i.e. a time independence assumption. Note also that very recently [2] gave a explicit derivation of a natural gradient algorithm with infinite size FIR for the un-mixing leading to equations similar again to the ones proposed in [12, 13]. They do however not report results on real room recordings. In figure 2 we see the results obtain for two speakers in a noisy office environment. The separation improves the signal to background ration. It does however not separate the signal completely. The results depend on the type of signal and the choice of the density $p()$. This is a current subject of experimentation and study in the context of higher order statistics.

[6] Here we are writing for simplicity only the one-dimensional case. The multi-dimensional case is a trivial extension.

References

[1] S. Amari, A. Cichocki, and Yang A.A. A new learning algorithm for blind signal separation. In *Advances in Neural Information Processing Systems 1995*, pages 752–763, Boston, MA, 1996. MIT Press.

[2] S. Amari, S.C. Douglas, A. Cichocki, and A.A. Yang. Multichannel blind deconvolution using the natural gradient. In *Proc. 1st IEEE Workshop on Signal Processing App. Wireless Comm.*, Paris, France, 1997. IEEE.

[3] J. Atick and A. Redlich. Towards a theory of early visual processing. *Neural Computation*, 2:308–320, 1990.

[4] H. Attias and C.E. Schreiner. Blind source separation and deconvolution: Dynamic component analysis. *Neural Computations*, submitted, 1997.

[5] H. Barlow. Sensory mechanism, the reduction of redundancy, and intelligence. In *National Physical Laboratory Symposium*, volume 10. Her Majesty's Stationery Office, London, 1959. The Mechanization of Thought Processes.

[6] A. Bell and T. Sejnowski. An information maximization approach to blind separation and blind deconvolution. *Neural Computation*, 7:1129–1159, 1995.

[7] P. Comon. Independent comonent analysis, a new concept? *Signal Processing*, 36(3):287–314, 1994.

[8] G. Deco and Dragan Obradovic. *An Information Theoretic Approach to Neural Computing*. Perspective in Neural Computing. Springer, 1996.

[9] Yariv Ephraim and Harry Van Trees. A signal subspace approach fot speech enhancement. *IEEE Transactions on Speech and Audio Processing*, 3(4):251–266, 1995.

[10] H. Hayes, Monson. *Statistical Digital Signal Processing and Modeling*. Wiley, 1996.

[11] C. Jutten and J. Herault. Blind separation of sources part i: An adaptive algorithm based on neuromimetic architecture. *Signal Processing*, 24(1):1–10, 1991.

[12] R. Lambert and A. Bell. Blind separation of multiple speakers in a multipath environment. In *ICASP 97*, pages 423–426. IEEE, 1997.

[13] T. Lee, A. Bell, and R. Lambert. Blind separation of delayed and convolved sources. In *Advances in Neural Information Processing Systems 1996*, 1997.

[14] J.S. Lim and A.V. Oppenheim. All-pole modelling of degraded speech. *IEEE Transaction ofn Acoustics, Speech, and Signal Processing*, 26(3):197–210, 1979.

[15] E. Moulines, J.F. Cardoso, and E. Gassiat. Maximum likelihood for blind separation and deconvolution of noisy signals using mixture models. In *ICASP 97*, volume ?, pages 3617–3620. IEEE, 1997.

[16] B.A. Olhausen and D.J. Field. Emergence of simple-sell receptive field properties by learning sparce code for natural images. *Nature*, 381:607–609, 1996.

[17] A.V. Oppenheim and R.W. Schafer. *Discret-Time Signal Processing.* Prentice Hall, 1989.

[18] B. Pearlmutter and L. Parra. A context-sencitive generalization of independent component analysis. In *International Conf. on Neural Information Processing*, Hong Kong, 1996.

[19] D.T. Pham, Garrat P., and Jutten C. Separation of a mixture of independent sources through a maximum likelihood approach. In *Proc. EUSIPCO*, pages 771–774, 1992.

[20] ML subspace signal enhancement. Fast adaptive eigenvalue decomposition: A maxmum likelihood approach. In *ICASP 97*, volume ? IEEE, 1997.

[21] Charles W. Therrien. *Discrete Random Signals and Statistical Signal Processing.* Prentice Hall, 1992.

[22] Torkkola. Convolution using causal fir ?????????????? In *in some procedings?*, 1997.

[23] E. Weinstein, M. Feder, and A.V. Oppenheim. Multi-channel signal separation by decorrelation. *IEEE Transaction on Spreech and Audio Processing*, 1(4):405–413, 1993.

[24] D. Yellin and E. Weinstein. Multichannel signal separation: Methods and analysis. *IEEE Transaction on Signal Processing*, 44(1):106–118, 1996.

[25] J. Zhu, X.R. Cao, and R.W. Liu. Blind source separation based on output independence - theory and implementation. In *NOLTA 95*, volume 1, pages 97–102, Tokyo, Japan, 1995. NTA Research Society of IEICE.