Convolutive Blind Source Separation based on Multiple Decorrelation.

Lucas Parra, Clay Spence

Abstract— **Acoustic signals recorded simultaneously in a reverberant environment can be described as sums of differently convolved sources. The task of source separation is to identify the multiple channels and possibly to invert those in order to obtain estimates of the underling sources. We tackle the problem by explicitly exploiting the non-stationarity of the acoustic sources. Changing cross-correlations at multiple times give a sufficient set of constraints for the unknown channels. A least squares optimization allows us to estimate a forward model, identifying thus the multipath channel. In the same manner we can find an FIR backward model, which generates well separated model sources. Furthermore, for more than three channels we have sufficient conditions to estimate underlying additive sensor noise powers, which could be used for further signal enhancement.**

I. Introduction

A growing number of researchers have published in recent years on the problem of blind source separation. For one, the problem seems of relevance in various application areas such as speech enhancement with multiple microphones, crosstalk removal in multichannel communications, multipath channel identification and equalization, direction of arrival (DOA) estimation in sensor arrays, improvement over beam forming microphones for audio and passive sonar, and discovery of independent sources in various biological signals, such as EEG, MEG and others. Additional theoretical progress in our understanding of the importance of higher order statistics in signal modeling have generated new techniques to address the problem of identifying statistically independent signals - A problem which lays at the heard of source separation. This development has been driven not only by the signal processing community but also by machine learning research that has treated the issue mainly as a density estimation task.

The basic problem is simply described. Assume d_s statistically independent sources $\mathbf{s}(t) = [s_1(t), ..., s_{d_s}(t)]^T$. These sources are convolved and mixed in a linear medium leading to d_x sensor signals $\mathbf{x}(t) = [x_1(t), ..., x_{d_x}(t)]^T$ that may include additional sensor noise $\mathbf{n}(t)$,

$$\mathbf{x}(t) = \sum_{\tau=0}^{P} A(\tau)\mathbf{s}(t-\tau) + \mathbf{n}(t) \quad (1)$$

How can one identify the $d_x d_s P$ coefficients of the channels A and how can one find an estimate $\hat{\mathbf{s}}(t)$ for the unknown sources?

Sarnoff Corporation, CN-5300, Princeton, NJ 08543, lparra@sarnoff.com, cspence@sarnoff.com

Alternatively one may formulate an FIR inverse model W,

$$\mathbf{u}(t) = \sum_{\tau=0}^{Q} W(\tau)\mathbf{x}(t-\tau) \quad (2)$$

and try to estimate W such that the model sources $\mathbf{u}(t) = [u_1(t), ..., u_{d_u}(t)]^T$ are statistically independent.

Most of the previous work has concentrated on the property of statistical independence, and ignores the additive noise. We shall discuss this approaches in a separate section II.

An alternative approach to the statistical independence condition in the convolutive case has been touched on by Weinstein et al. in [1]. For non-stationary signals a set of second order conditions can be specified that uniquely determines the parameters A. No algorithm has been given in [1] nor has there been to our knowledge any results reported on this approach. A resent paper by Ehlers and Schuster [2] carries that spirit in attempting to solve for the frequency components of A by extending prior work of Molgedey and Schuster [3] on instantaneous mixtures into the frequency domain. They fall short however in carefully considering the issues at hand and mistakenly confuse this idea with simple decorrelation of multiple taps in the time domain, which is known to be insufficient [4], [5].

We take up this multiple decorrelations approach assuming quasi-stationary signals and use a least squares (LS) optimization to estimate A or W as well as signal and noise powers. In the following section we give a brief review of approaches taken in the literature for the instantaneous mixtures ($P = 1$, or $Q = 1$) and convolutive mixtures ($P > 1$). In section III we present our approach for the instantaneous case and point out the differences between estimating the forward model A and the backward model W. In addition to the source power one can estimate additive sensor noise powers. Computing estimates $\hat{\mathbf{s}}$ from a forward model A requires a further estimation step, in particular for the case of fewer sources than sensors, i.e. $d_s < d_x$. The least squares (LS), maximum likelihood (ML) or maximum a posteriori probability (MAP) estimates are given in section III-C. In a backward model W the LS optimization gives the inverse of the mixture and we obtain model sources $\hat{\mathbf{u}}$ directly. In section IV we carry over the concept of multiple decorrelation into the convolutive case by solving independent models for every frequency, thereby paying particular attention to the approximation of linear convolutions by circulant convolutions in section IV-A as well as the permutation issue in section IV-C. Since inverting a multichannel forward FIR model is in itself a challenging task we restrict ourself in the implementations to estimating the inverse model W. Finally we report some encouraging preliminary results in section V.

II. Previous work on blind source separation

Early work in the signal processing community had suggested decorrelating the measured signals, i.e. diagonalizing measured correlations for multiple time delays [6], [4]. For an instantaneous mix, also referred to as the constant gain case, it has been shown that for non-white signal decorrelation of multiple taps are sufficient to recover the sources [7], [8]. Early on however it was clear that for convolutive mixtures of wide-band signals this solution is not unique [4], [5], and in fact may generated source estimates that are decorrelated but not statistically independent. As clearly pointed out by Weinstein et.al. in [1] additional conditions are required. In order to find separated sources it seems one would have to capture more than second order statistic, since indeed statistical independence requires that not only second but all higher cross moments vanish.

Comon formulated the problem of an instantaneous linear mix, clearly defining the term Independent Component Analysis, and presented an algorithm that measures independence by capturing higher statistics of the signals [9], [10]. Previous work on DOA estimation had already suggested higher order statistics [11], [12], [13]. Cardoso suggested to consider the eigenstructure of 4th order cumulants for blind separation [14]. Herault and Jutten [15] were the first to capture higher statistic by decorrelating nonlinear transformations of the signals. Pham [16] and later Bell and Sejnowski [17] presented a simple algorithmic architecture which in effect density estimation [18], [19] and is based on prior knowledge of the cumulative density function of the source signals. Amari make modifications to the update equations to dramatically improve convergence and computational costs [20].

In the convolutive case Yellin and Weinstein [21], [22] established conditions on higher order multi-tap cross moments that allow convolutive cross talk removal. Nguyen Thi and Jutten [23] gave simpler algorithms to estimate the forward model A based on multi-tap third and forth order cross moments. Although the optimization criteria extend naturally to higher dimensions, these researchers have concentrated on the two dimensional case since there a multichannel FIR forward model can be inverted with a properly chosen architecture using the estimated forward filters. For higher dimensions however finding a stable approximation of the forward model remains an open question.

In contrast the density estimation approaches mentioned before generalize to the convolutive case by estimating an FIR backward model W that directly tries to generate independent model sources [24], [25]. They resemble equations obtained from multidimensional extensions of the Busgang blind equalization method [24]. Maximum likelihood density estimation derivations of this type of algorithm are given by [26], [27].

All these techniques are shown to work satisfactorily in computer simulations but perform poorly for real recordings. One could speculate that the densities may not have the hypothesized structures, the higher order statistics may lead to estimation instabilities, or the violation of the stationarity condition cause problems.

Our present work makes no assumptions about the cumulative densities of the signals and limits itself to more robust second order statistics wile explicitly exploiting nonstationarity and a colored signal spectra.

III. Instantaneous Mixture

As laid out in the previous section the instantaneous case has been worked out for some time now and a multitude of approaches have been suggested. We present it here in order to lay out some basic ideas, which will be used again in the convolutive case. Part of our treatment of additive sensor noise estimates may go beyond previous work.

A. Forward model estimation

For an instantaneous mixture, i.e. $P = 1$, the forward model (1) simplifies to,

$$\mathbf{x}(t) = A\mathbf{s}(t) + \mathbf{n}(t) \tag{3}$$

We can formulate the covariance $R_x(t)$ of the measured signals at time t with the assumption of independent noise as

$$\begin{aligned} R_x(t) \equiv& \left\langle \mathbf{x}(t)\mathbf{x}^T(t) \right\rangle \\ =& A \left\langle \mathbf{s}(t)\mathbf{s}^T(t) \right\rangle A^T + \left\langle \mathbf{n}(t)\mathbf{n}^T(t) \right\rangle \\ \equiv& A\Lambda_s(t)A^T + \Lambda_n(t) \end{aligned} \tag{4}$$

Since we assume decorrelated sources at all times we postulate diagonal covariance matrixes $\Lambda_s(t)$. We also assume independent noise at each sensor, i.e. diagonal $\Lambda_n(t)$.

Note that any scaling and permutation of the coordinates of $\Lambda_s(t)$ can be absorbed by A. Therefore we see that the solution is only specified up to an inherently arbitrary permutation and scaling, as is well known. We are free to choose the scaling of the coordinates in $\mathbf{s}$. For now we choose $A_{ii} = 1 \, ; i = 1, ..., d_s$, which places d_s conditions on our solutions.

For non-stationary signals a set of K equations (4) for different times $t_1, ...t_K$ gives then a total of $Kd_x(d_x + 1)/2 + d_s$ constraint on $d_s d_x + d_s K + d_x K$ unknown parameters $A, \Lambda_s(t_1), ..., \Lambda_s(t_K), \Lambda_n(t_1), ..., \Lambda_n(t_K)$.[1]. Assuming all conditions are linearly independent we will have sufficient conditions if,

$$K d_x(d_x + 1)/2 + d_s \geq d_s d_x + d_s K + d_x K \tag{5}$$

[1] We will write in the remainder in brief $\Lambda_s(k) = \Lambda_s(t_k)$ and $\Lambda_s = \Lambda_s(t_1), ..., \Lambda_s(t_K)$ whenever possible. The same applies to $\Lambda_n(t)$ and $R_x(t)$

It is interesting to note that in the square case, $d_s = d_x$, there are not sufficient constraints to determine the additional noise parameters unless $d_x \geq 4$. If we assume zero additive noise in principle $K = 2$ is sufficient to specify the solution up to arbitrary permutations.

In the square case the solutions can be found as a non-symmetric eigenvalue problem as outlined in [3]. The difficulty with such algebraic solutions however is that one does not have perfect estimates of $R_x(t)$. At best one can assume quasi-stationary signals and measure the sample estimates $\hat{R}_s(t)$ within the stationarity time. If we interpret the inaccuracy of that estimation as measurement error

$$E(k) = \hat{R}_x(k) - \Lambda_n(t) - A\Lambda_s(k)A^T \tag{6}$$

it is reasonable to estimate the unknown parameters by minimizing the total measurement error for a sufficiently large K,

$$\hat{A}, \hat{\Lambda}_s, \hat{\Lambda}_n = \arg\min_{A,\Lambda_s,\Lambda_n,A_{ii}=1} \sum_{k=1}^{K} \|E(k)\|^2 \tag{7}$$

The matrix norm here is the sum of the absolute values. Note that $\|E\|^2 = \mathrm{Tr}(EE^H)$. This represents a least squares (LS) estimation. To find the extrema of the LS cost $E = \sum_{k=1} \left|^K E(k)\right|$ in (7) let us compute the gradients with respects to its parameters[2]

$$\frac{\partial E}{\partial A} = -4 \sum_{k=1}^{K} E(k)A\hat{\Lambda}_s(k) \tag{8}$$

$$\frac{\partial E}{\partial \hat{\Lambda}_s(k)} = -2\,\mathrm{diag}\left(AE(k)A^T\right) \tag{9}$$

$$\frac{\partial E}{\partial \hat{\Lambda}_n(k)} = -2\,\mathrm{diag}\left(E(k)\right) \tag{10}$$

While we can explicitly solve (10)=0 and for square and invertible A also (9)=0 we will have to use an iterative algorithm to find the extrema with respect to A using the gradients in (8).

B. Normalization conditions

In the previous section we proposed to fix the arbitrary scaling by fixing the diagonal parameters $A_{ii} = 1$. For the non-square case however this normalization may seem somewhat arbitrary. One could in such a case demand instead that $\|\mathbf{a}_j\| = 1, j = 1, ..., d_s$ with $\mathbf{a}_j = [A_{1j}, ..., A_{d_x j}]^T$. Instead of the gradients given in (8) one has to then consider their projections onto the hyper-planes defined by $\|\mathbf{a}_j\| = 1$. The projections operator for the jth column of $\partial E/\partial A$ is $P_j^{(1)} = I - \mathbf{a}_j\mathbf{a}_j^T$. Or we can write a constraint gradient

$$\left.\frac{\partial E}{\partial A}\right|_{\|\mathbf{a}_j\|=1} = \frac{\partial E}{\partial A} - A\,\mathrm{diag}\left(A^T \frac{\partial E}{\partial A}\right) \tag{11}$$

[2]The diagonalization operator here zeros the off-diagonal elements, i.e. $\mathrm{diag}(M)_{ij} = \begin{cases} M_{ij}, & i = j \\ 0, & i \neq j \end{cases}$

C. Estimation of source signals

In the case of a square and invertible mixing $\hat{A}$ the signal estimates are trivially computed to be $\hat{\mathbf{s}} = \hat{A}^{-1}\mathbf{x}$. In the non-square case for $d_s < d_x$ we can compute the LS estimate,

$$\hat{\mathbf{s}}_{LS}(t) = \arg\min_{\mathbf{s}(t)} \|\hat{A}\mathbf{x}(t) - \mathbf{s}(t)\| = (\hat{A}^T\hat{A})^{-1}\hat{A}^T\mathbf{x}(t) \tag{12}$$

If we assume the additive noise to be Gaussian, not necessarily white, nor stationary we can compute the maximum likelihood (ML) estimate

$$\begin{aligned} \hat{\mathbf{s}}_{ML}(t) &= \arg\max_{\mathbf{s}(t)} p(\mathbf{x}(t)|\mathbf{s}(t); \hat{A}, \hat{\Lambda}_n(t)) \\ &= (\hat{A}^T\hat{\Lambda}_n(t)^{-1}\hat{A})^{-1}\hat{A}^T\hat{\Lambda}_n(t)^{-1}\mathbf{x}(t) \end{aligned} \tag{13}$$

If we further assume the signal to be Gaussian, again not necessarily white, nor stationary we can compute the maximum a posteriori probability (MAP) estimate

$$\begin{aligned} \hat{\mathbf{s}}_{MAP}(t) &= \arg\max_{\mathbf{s}(t)} p(\mathbf{x}(t)|\mathbf{s}(t); \hat{A}, \hat{\Lambda}_n(t), \hat{\Lambda}_s(t)) \\ &= (\hat{A}^T\hat{\Lambda}_n(t)^{-1}\hat{A} + \hat{\Lambda}_s(t)^{-1})^{-1}\hat{A}^T\hat{\Lambda}_n(t)^{-1}\mathbf{x}(t) \end{aligned} \tag{14}$$

Note however that the resulting estimates may not be decorrelated. Assuming the model is correct and we found the correct estimate $\hat{A} \approx A$,

$$\left\langle \hat{\mathbf{s}}_{LS}\hat{\mathbf{s}}_{LS}^T \right\rangle \approx \left\langle \mathbf{s}\mathbf{s}^T \right\rangle + (\hat{A}^T\hat{A})^{-1}\hat{A}^T\Lambda_n\hat{A}(\hat{A}^T\hat{A})^{-1} \tag{15}$$

Since the second term may not be diagonal the resulting estimates can be correlated. However this should be no major concern since the correlation is entirely due to correlated noise and the signal portion of the estimates remains decorrelated.

D. Backward model

Instead of estimating a forward model and then from that further estimating the source signal one may directly try to estimate a backward model in the form of (2) with the objective of generating separated model sources $\hat{\mathbf{s}}(t)$, which we define as

$$\hat{\mathbf{s}}(t) \equiv WA\mathbf{s}(t) \tag{16}$$

We are looking therefore for a W that inverts A. This will be in particular relevant for the convolutive case which we will discuss in the following section. In analogy with the previous discussions and using definition (16) and assuming (3) we have,

$$\left\langle \hat{\mathbf{s}}(t)\hat{\mathbf{s}}(t)^T \right\rangle = W\left(R_x(t) - \Lambda_n(t)\right)W^T \tag{17}$$

We will search for W such that $\langle \hat{\mathbf{s}}(t)\hat{\mathbf{s}}(t)^T \rangle$ diagonalizes simultaneously for K different times[3]. The LS estimate is then,

$$E(k) = W(\hat{R}_x(k) - \Lambda_n(k))W^T - \Lambda_s(k)$$
$$\hat{W}, \hat{\Lambda}_s, \hat{\Lambda}_n = \underset{W,\Lambda_s,\Lambda_n,W_{ii}=1}{\arg\min} \sum_{k=1}^{K} \|E(k)\|^2 \tag{18}$$

In analogy to the discussion in section III-A we can find the solutions with an iterative gradient algorithm.

IV. Convolutive Mixture

In the previous section we described how one can treat the case of instantaneous mixtures by decorrelating the covariance matrices at several times. This approach requires non-stationary sources. The problem can also be treated by decorrelating the cross-correlation at different taps. This requires that the signals be non-white rather than non-stationary. This is the approach traditionally take in the literature [4], [3], [8], [1]. For the convolutive case neither one by itself is sufficient due to the larger number of parameters one wishes to estimate. The crucial insight of this work is that in the convolutive case one can consider multiple taps *and* multiple times and assume non-white *and* non-stationary signals. Then again we obtain sufficient conditions to estimate all parameters.

As suggested for other source separation algorithms our approach to the convolutive case is to transform the problem into the frequency domain and to solve simultaneously a separation problems for every frequency [28], [29], [25], [2]. The solution for each frequency would seem to have an arbitrary permutation. The main issues to be addressed here are how to obtain equations equivalent to (4) or (17) in the frequency domain, and how to choose the arbitrary permutations for all individual problems consistently. We will take up these issues in the following sections.

A. Cross-correlations, circular and linear convolution

First consider the cross-correlations $R_x(t, t+\tau) = \langle \mathbf{x}(t)\mathbf{x}(t+\tau)^T \rangle$. For stationary signals the absolute time does not matter and the correlations depend on the relative time, i.e. $R_x(t, t+\tau) = R_x(\tau)$. Denote with $R_x(z)$ the z-transform of $R_x(\tau)$. We can then write

$$R_x(z) = A(z)\Lambda_s(z)A(z)^H + \Lambda_s(z) \tag{19}$$

where $A(z)$ represents the matrix of z-transforms of the FIR filters $A(\tau)$, and $\Lambda_s(z)$, and $\Lambda_n(z)$ are the z-transform of the auto-correlation of the sources and noise. Again they are diagonal due to the independence assumptions.

For practical purposes we have to restrict ourself to a limited number of sampling points of z. Naturally we will take T equidistant samples on the unit circle such that we can use the discrete Fourier transform (DFT). For periodic signals the DFT allows us to express circular convolutions as products such as in (19). However, in (1) and (2) we assumed linear convolutions. A linear convolution can be approximated by a circular convolution if $P \ll T$ and we can write approximately

$$\mathbf{x}(t,\nu) \approx A(\nu)\mathbf{s}(t,\nu) + \mathbf{n}(t,\nu), \text{ for } P \ll T \tag{20}$$

where $\mathbf{x}(t,\nu)$ represents the DFT of the frame of size T starting at t, $[\mathbf{x}(t), ..., \mathbf{x}(t+T)]$, and is given by $\mathbf{x}(t,\nu) = \sum_{\tau=0}^{T-1} e^{-\mathbf{i}2\pi\nu\tau}\mathbf{x}(t+\tau)$ and corresponding expressions for $\mathbf{s}(t,\nu)$ and $A(\nu)$.

We limit the estimates of the cross-correlations to a given estimation time. For non-stationary signals those estimates will be dependent on the absolute time and will indeed vary from one estimation segment to the next.

$$\hat{R}_x(t,\nu) = \frac{1}{N}\sum_{n=0}^{N-1} \mathbf{x}(t+nT,\nu)\mathbf{x}^T(t+nT,\nu) \tag{21}$$

We can then write for such estimates

$$\hat{R}_x(t,\nu) = A(\nu)\Lambda_s(t,\nu)A^H(\nu) + \Lambda_n(t,\nu) \tag{22}$$

If N is sufficiently large we can assume that $\Lambda_s(t,\nu)$ and $\Lambda_n(t,\nu)$ can be modeled as diagonal again due to the independence assumption. For equations (22) to be linearly independent for different times t and different ν it will be necessary that $\Lambda_s(t,\nu)$ changes over time and frequency, i.e. the signal are non-stationary and non-white.

B. Backward model

Given a forward model A it is not guaranteed that we can find a stable inverse. In the two dimensional square case the inverse channel is easily determined from the forward model [1], [23]. It is however not apparent how to compute a stable inversion for arbitrary dimensions. In this present work we prefer to estimate directly a stable multipath backward FIR model such as (2). In analogy to the discussion above and to section III-D we wish to find model sources with cross-power-spectra satisfying[4],

$$\Lambda_s(t,\nu) = W(\nu)\left(\hat{R}_x(t,\nu) - \Lambda_n(t,\nu)\right)W^H(\nu) \tag{23}$$

In order to obtain independent conditions for every time we choose the times such that we have non-overlapping estimation times for $\hat{R}_x(t_k,\nu)$, i.e. $t_k = kTN$. But if the signals vary sufficiently fast overlapping estimation times

[3] Similar considerations to those given for (15) show that decorrelating $\mathbf{u}(t)$ rather than $\hat{\mathbf{s}}(t)$ may not lead to the correct solution in the presence of sensor noise

[4] $W(\nu)$ represents the DFT with frame size T of the time domain $W(\tau)$. In what follows time and frequency domain are identified by their argument τ or ν.

could have been chosen. A multipath channel W that satisfies these equations for K times simultaneously can be found, again with an LS estimation[5]

$$E(k,\nu) = W(\nu)(\hat{R}_x(k,\nu) - \Lambda_n(k,\nu))W^H(\nu) - \Lambda_s(k,\nu)$$

$$\hat{W}, \hat{\Lambda}_s, \hat{\Lambda}_n = \underset{\substack{W, \Lambda_s, \Lambda_n, \\ W(\tau)=0, \tau > Q, \\ W_{ii}(\nu)=1}}{\arg\min} \sum_{\nu=1}^{T} \sum_{k=1}^{K} \|E(k,\nu)\|^2 \tag{24}$$

Note the additional constraint on the filter size in the time domain. Up to that constraint it would seem the various frequencies $\nu = 1, ..., T$ represent independent problems. The solutions $W(\nu)$ however are restricted to those filters that have no time response beyond $\tau > Q \ll T$. Effectively we are parameterizing $T d_s d_x$ filter coefficients in $W(\nu)$ with $Q d_s d_x$ parameters $W(\tau)$. The LS solutions can again be found with a gradient descent algorithm. We will first compute the gradients with respect to the complex valued filter coefficients $W(\nu)$ and discuss their projections into the subspace of permissible solutions in the following section.

For any real valued function $f(\mathbf{z})$ of a complex valued variable $\mathbf{z}$ the gradients with respect to the real and imaginary part are obtained by taking derivatives formally with respect to the conjugate quantities $\mathbf{z}^*$ ignoring the nonconjugate occurrences of $\mathbf{z}^*$ [30].

$$\frac{\partial f(z)}{\partial \Re(z)} + i\frac{\partial f(z)}{\partial \Im(z)} = 2\frac{\partial f(z)}{\partial z^*} \tag{25}$$

Therefore the gradients of the LS cost in (24) are,

$$\frac{\partial E}{\partial W^*(\nu)} = \sum_{k=1}^{K} E(k,\nu)W(\nu)B^H(k,\nu) + E^H(k,\nu)W(\nu)B(k,\nu) \tag{26}$$

$$\frac{\partial E}{\partial \hat{\Lambda}_s^*(k,\nu)} = -\operatorname{diag}\left(E(k,\nu)\right) \tag{27}$$

$$\frac{\partial E}{\partial \hat{\Lambda}_n^*(k)} = -\operatorname{diag}\left(W^H(\nu)E(k,\nu)W(\nu)\right) \tag{28}$$

$$B(k,\nu) \equiv \hat{R}_x(k,\nu) - \Lambda_n(k,\nu) \tag{29}$$

With (27)=0 one can solve explicitly for parameters $\Lambda_s(k,\nu)$, while parameters $\Lambda_s(k,\nu), W(\nu)$ may be computed with a gradient descent rule.

C. Permutations and constraints

The above unconstrained gradients can not be used as such but have to be constrained to remain in the subspace of permissible solutions with $W(\tau) = 0$ for $\tau > Q \ll T$. This is important since it is a necessary condition for equations (23) to hold to a good approximation.

Additionally, and this is a crucial point that may have not been realized in previous literature, not all possible permutations of frequencies will lead to FIR filters which satisfy that constrain. Note that any permutation of the coordinates for every frequency will lead to exactly the same error $E(k,\nu)$. The total cost will therefore not change if we choose a different permutation of the solutions for every frequency ν. Obviously those solutions will not all satisfy the condition on the length of the filter. Effectively, requiring zero coefficients for elements with $\tau > Q$ will restrict the solutions to be smooth in the frequency domain, e.g., if $Q/T = 8$ the resulting DFT corresponds to a convolved version of the coefficients with a *sinc* function 8 times wider than the sampling rate.

It is therefore crucial to enforce that constraint by starting the gradient algorithm with an initial point that satisfies the constraints, and then following the constrained gradient. The normalization condition that avoid trivial solutions of the LS optimization have to be enforced simultaneously. The constrained gradients are obtained by applying the corresponding projection operators. The projection operator that zeros the appropriate delays for every channel $W_{ij} = [W_{ij}(0), ..., W_{ij}(\nu), ..., W_{ij}(T)]^T$ is

$$P^{(2)} = FZF^{-1} \tag{30}$$

where the DFT is given by $F_{ij} = 1/\sqrt{T}e^{-\mathrm{i}2\pi ij}$, and Z is diagonal with $Z_{ii} = 1$ for $i < Q$ and $Z_{ii} = 0$ for $i \geq Q$. The projection operator that enforces unit gains on diagonal filters $W_{ii}(\nu) = 1$ is applied simply by setting the diagonal terms of the gradients to zero. These projections are orthogonal and can be applied independently of each other. This stands in contrast to the normalization constraint outlined in section III-B. That projection operator is not orthogonal to $P^{(2)}$ and care has to be taken to apply a proper projection that maps the gradient to the joint subspace of $P^{(1)}$ and $P^{(2)}$. A simple, though admittedly inefficient, solution is to apply $P^{(1)}$ and $P^{(2)}$ successively and repeatedly to the gradients until convergence. In our simulations 3-5 iterations where sufficient. The so obtained constrained gradient can be used in a gradient update of the filter parameters.

V. Experimental results

The main difficulty in assessing the quality of a separation from real recordings is that the true sources are not available in general. And even if they were it would be difficult to relate the power of crosstalk given the scaling ambiguity of the solutions. One is therefore limited to listening experiments or to a performance measure of some subsequent processing such as automatic speech recognition.

The experiments performed this far are limited to hearing test. We used two channels recordings in a typi-

[5] In short we write again $\Lambda_s(k,\nu) = \Lambda_s(t_k,\nu)$ and $\Lambda_s = \Lambda_s(t_1,\nu), ..., \Lambda_s(t_K,\nu)$ whenever possible. The same applies to $\Lambda_n(t,\nu)$ and $R_x(t,\nu)$

cal office environment with ventilator noise in the background. We use approximately 10-15 sec of 16 bit recordings sampled at 8 KHz or 16 KHz. We applied parameters $T = 1024, Q = 128, N = 20$ or $T = 2048, Q = 256, N = 10$ which leads to about $K = 3...6$ windows depending on the amount of data. We assumed constant noise powers, i.e. $\Lambda_n(t,\nu) = \Lambda_n(t',\nu)$. In general we find the separation to work well at high signal to noise rations. In particular for the recording provided by Lee [25] we get a significantly improved separation of the two speakers as compared to the algorithms given in [25], [26] (see figure 1). Both these algorithms represent a multi-path equalization with some implementation differences. As opposed to that the present algorithm does not lead to whitened model sources, and no post-processing is required.

When the number of sources is larger than the number of microphones the algorithm fails as expected. The confounding additional sources can also just be background noise. Therefore we observe that in noisy environments the algorithms performs less favorably since the noise sources do not represent independent sensor noise.

Currently we are working on evaluating the result on a speech recognition task under the presence of a interfering source.

The restriction to stereo recordings is purely technical and not related to the algorithm and will be addressed next.

VI. Conclusion

A large body of work has been done in the last two decades on the problem of blind source separation. We have concentrated on the rather general case of recovering convolutive mixtures of wideband signals for less or equal number of sources than sensors. Most of the concepts in this work where borrowed from previous work. The main contributions are: We explicitly use the property of non-stationary and non-white signals. Careful considerations of how to measure second order statistics in the frequency domain allows us to obtain a constraint LS cost that is optimal at the desired solutions. The constraint on the filter size solves the permutation problem of wideband signals.

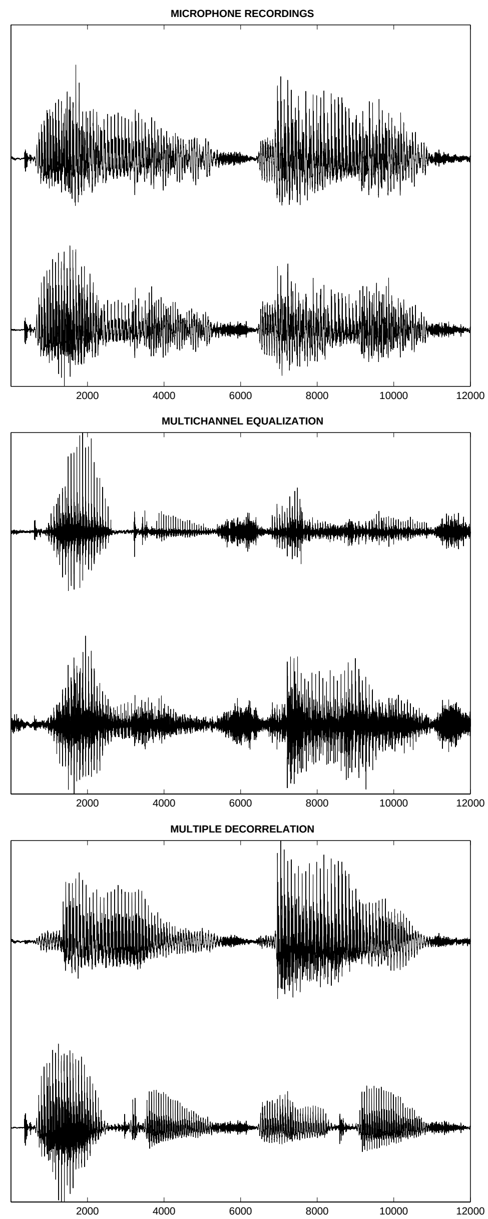

Fig. 1. Top panel: signals recorded in real room envoronment. Middle panel: results obtained in [25]. Results obtained with present algorithm. For comparison we use signals provided by Lee [25]

Acknowledgments

The authors would like to thank foremost Bert De Vries for educating the authors on may signal processing issues.

References

[1] E. Weinstein, M. Feder, and A.V. Oppenheim, "Multi-Channel Signal Separation by Decorrelation", *IEEE Transaction on Spreech and Audio Processing*, vol. 1, no. 4, pp. 405–413, 1993.

[2] F. Ehlers and H.G. Schuster, "Blind Separation fo Convolutive Mixtures and an Application in Automatic Speech Recognition in a Noisy Environment", *IEEE Transaction on Signal Processing*, vol. 45, no. 10, pp. 2608–2612, 1997.

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

[4] Y. Bar-Ness, J. Carlin, and M. Steinberger, "Bootstrapping adaptive cross-pol canceller for satellite communications", in *IEEE Int. Conf. Comunications*, Philadelphia, PA, June 1982, pp. 4F.5.1–4F.5.5.

[5] R.L.L Tong and Y.H.V.C. Soon, "Interdeterminacy and Identifiability of blind identification", *IEEE Transaction on Circuits Systems*, vol. 38, pp. 499–509, 1991.

[6] Bradwood D., "Cross-coupled cancellation systems for improving cross-polarisation discrimination", in *IEEE Int. Conf. Antennas and Propagation*, London, England, November 1978, vol. Part I, pp. 41–45.

[7] Stefan Van Gerven and Van Compernolle Dirk, "Signal separation in a symmetric adaptive noise canceler by output decorrelation", in *ICASSP 92*, March 1992, vol. IV, pp. 221–224.

[8] Stefan Van Gerven and Van Compernolle Dirk, "Signal Separation by Symetric Adaptive Decorrelation: Stability, Convergence, and Uniqueness", *IEEE Transaction on Signal Processing*, vol. 43, no. 7, pp. 1602–1612, July 1995.

[9] P. Comon, "Blind separation of sources: Problem statement", *Signal Processing*, vol. 24, pp. 11–20, 1991.

[10] P. Comon, "Independent Comonent Analysis, a new concept?",

Signal Processing, vol. 36, no. 3, pp. 287–314, 1994.

[11] G. Giannakis, Y. Inouye, and J.M. Mendel, "Cumulant based identification of multichannel MA models", *IEEE Transaction on Automation and Control*, vol. 34, pp. 783–787, 1989.

[12] R. Pan and C.L. Nikias, "Harmonic decomposition in cumulant domains", in *ICASSP 88*, 1988, pp. 2356–2359.

[13] Sanyogita Shamsunder and B. Giannakis, Georgios, "Modeling of non-Gaussian array data using cumulants: DOA estimation of more sources with less sensors", *Signal Processing*, vol. 30, pp. 279–297, 1993.

[14] J.-F. Cardoso, "Eigenstructure of the 4th-order cumulant tensor with application to the blind source separation problem", in *ICASSP 89*, 1989, pp. 2109–2112.

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

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

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

[18] B. Pearlmutter and L. Parra, "Maximum Likelihood Source Separation: a context-sensitive generalization of ICA", in *NIPS 96*, Hong Kong, 1997.

[19] J.F. Cardoso, "Infomax and maximum likelihood for blind source separation", *Signal Processing Letters*, vol. to appear, 1997.

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

[21] Daniel Yellin and Ehud Weinstein, "Multichannel Signal Separation: Methods and Analysis", *IEEE Transaction on Signal Processing*, vol. 44, no. 1, pp. 106–118, January 1996.

[22] D. Yellin and E. Weinstein, "Multichannel Signal Separation: Methods and Analysis", *IEEE Transaction on Signal Processing*, vol. 44, no. 1, pp. 106–118, 1996.

[23] Hoang-Lan Nguyen Thi and Christian Jutten, "Blind source separation for convolutive mixtures", *Signal Processing*, vol. 45, pp. 209–229, 1995.

[24] R. Lambert and A. Bell, "Blind Separation of Multiple Speakers in a Multipath Environment", in *ICASSP 97*. IEEE, 1997, pp. 423–426.

[25] T. Lee, A. Bell, and R. Lambert, "Blind separation of delayed and convolved sources", in *Advances in Neural Information Processing Systems 1996*, 1997, Authors provide signals and results at http://www.cnl.salk.edu/∼tewon/.

[26] 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.

[27] Lucas Parra, *1997 International Summer School on Adaptive Processing of Sequences*, chapter Temporal Models in Blind Source Separation, Springer, 1998.

[28] C. Jutten, *Calcul neuromimetique et traitment du signal: Analyse en composantes independantes*, PhD thesis, UJF-INP Grenoble, 1987.

[29] V. Capdevielle, Ch. Serviere, and J.L. Lacoume, "Blind separation of wide-band sources in the frequency domain", in *ICASSP 1995*. IEEE, 1995, pp. 2080–2083.

[30] Klaus Jänich, *Einführung in die Funktionentheorie*, Springer-Verlag, 1977, see chaper 2 on Wirtinger-Kalkül.