International Journal of Information, Control and Computer Sciences
ISSN: 2517-9942
Vol:8, No:5, 2014

An Efficient Separation for Convolutive Mixtures

Salah Al-Din I. Badran, Samad Ahmadi, Dylan Menzies, Ismail Shahin

***Abstract*—**This paper describes a new efficient blind source separation method; in this method we uses a non-uniform filter bank and a new structure with different sub-bands. This method provides a reduced permutation and increased convergence speed comparing to the full-band algorithm. Recently, some structures have been suggested to deal with two problems: reducing permutation and increasing the speed of convergence of the adaptive algorithm for correlated input signals. The permutation problem is avoided with the use of adaptive filters of orders less than the full-band adaptive filter, which operate at a sampling rate lower than the sampling rate of the input signal. The decomposed signals by analysis bank filter are less correlated in each sub-band than the input signal at full-band, and can promote better rates of convergence.

***Keywords*—**Blind source separation (BSS), estimates, full-band, mixtures, Sub-band.

I. Introduction

THE blind separation separates different signal sources statistically. The work of this paper is based on the structure in [1]. In the real world, due to reverberant environment, the signals of the original sources are filtered by a linear Multiple Input Multiple Output (MIMO) system before being captured by the microphones. The second order statistics (SOS) is used where we use the same number of sources and microphones [2].

In BSS problem, we are interested in the system that dissolves the mixture, it is described by

$$y_{ns}(n) = \sum_{nm=1}^{Nm} \sum_{k=0}^{S-1} \omega_{nm,ns}(k) x_{nm}(n-k) \quad (1)$$

Extending the formulation of the output signals into a matrix form, we can describe the signal of the $(ns)^{th}$ output at time n as.

$$y_{ns}(n) = \sum_{nm=1}^{Nm} x_{nm}^{T}(n) \omega_{nm,ns} \quad (2)$$

where $x_{nm}(n)$ is with $2S$ updated samples captured by p^{th} microphones and $\omega_{nm.ns}(n) = [\omega_{nm.ns}(0), \omega_{nm.ns}(1), \dots, \omega_{nm.ns}(2S-1)]^T$ is the vector containing $2S$ coefficients of the FIR filter that models the route of the $(nm)^{th}$ sensor and the $(ns)^{th}$ output. Two new parameters are needed for generalization of the formulation, are the delays in time (lag) taken into consideration the calculation of the correlation ($1 \le \text{lag} \le 2S$) and 2N size block output signal. From (2) it can be described that the vector that contains a block of 2N samples of size (ns)th output at time k as

$$y_{ns}(k) = \sum_{mn}^{2Nm} \hat{X}_{nm}^{T}(k) \bar{\omega}_{nm,ns} \quad (3)$$

where

$$\hat{X}_{ns}(k) = [x_{nm}(2kS) \cdots x_{nm}(2kS + 2N - 1)]^T \quad (4)$$

is Toeplitz matrix of dimension $2S \times 2N$ containing the $2S$ blocks with delayed versions of the samples of the signal captured by the $(nm)^{th}$ sensor.

Then, (3) can be extended to include samples of (lag) blocks of times. Thus, the matrix with the data of $(ns)^{th}$ output of dimension $2N \times 2D$, is given by

$$Y_{ns}(k) = \sum_{nm=1}^{1Nm} X_{nm}^{T}(k) \widehat{W}_{nm,ns} \quad (5)$$

To ensure the linear convolution of $Y_{ns}(m)$ the delay time should be: $lag = S$ [3], it takes four input blocks of X_{nm}^{T}. Therefore, the dimensions of $X_{nm}(m)$ is $2N \times 4S$ and $\widehat{W}_{nm,ns}$ is $4S$ x lag. $X_{nm}(k)$ matrices are attained by doubling $\hat{X}_{nm}$,

$$X_p(k) = \left[\hat{X}_p^T(k), \hat{X}_p^T(k-1)\right] \quad (6)$$

where $\hat{X}_{nm}^{T}(k-1)$ represents also Toeplitz matrix, so that the first row of the matrix $X_{nm}(k)$ contains $4S$ samples of nm^{th} input signal and each subsequent row is obtained by shifting the previous row to the right a sample containing a new sample per row. The Sylvester matrix $\widehat{W}_{nm,ns}$ is of dimension $4S$ x lag, defined as

$$\widehat{W}_{nm,ns} = \begin{bmatrix} \bar{\omega}_{nm,ns}(0) & 0 & \cdots & 0 \\ \bar{\omega}_{nm,ns}(1) & \bar{\omega}_{nm,ns}(0) & \ddots & \vdots \\ \vdots & \bar{\omega}_{nm,ns}(1) & \ddots & 0 \\ \bar{\omega}_{nm,ns}(2S-1) & \vdots & \ddots & \bar{\omega}_{nm,ns}(0) \\ 0 & \bar{\omega}_{nm,ns}(2S-1) & \ddots & \bar{\omega}_{nm,ns}(1) \\ \vdots & 0 & \ddots & \vdots \\ 0 & \cdots & 0 & \bar{\omega}_{nm,ns}(2S-1) \\ 0 & \cdots & 0 & 0 \\ \vdots & \cdots & \ddots & \vdots \\ 0 & \cdots & 0 & 0 \end{bmatrix} \quad (7)$$

which has the latest $2S$ - lag + 1 rows padded by zeroes in order to be deal mathematically with the $X_{nm}(k)$. The general case is $1 \le lag \le 2S$. We can rewrite (5) for a more compact form, i.e.

$$Y(k) = X(k)\widehat{W} \quad (8)$$

where

S. B. is affiliated with Sohar University, P.O. Box: 44, P. Code 311, Sohar Sohar, Oman (e-mail: sbadran@soharuni.edu.om).

S. A. is with the School of Computer Science and Informatics, De Montfort University, The Gateway, Leicester, LE1 9BH, , UK (e-mail: sahmadi@dmu.ac.uk).

D. M. is with the School of Engineering, De Montfort University, The Gateway, Leicester, LE1 9BH, UK (e-mail: rdmg1@dmu.ac.uk).

I. S. is with the Electrical and Computer Engineering Department, University of Sharjah, Sharjah, UAE.

$$Y(k) = \left[Y_1(k) \cdots Y_p(k)\right] \quad (9)$$

is a matrix of $2N \times (P)(lag)$ dimension containing the building blocks of the output signals of all channels,

$$X(k) = \left[X_1(k) \cdots X_{Nm}(k)\right] \quad (10)$$

is a matrix of order $2N \times 4SP$ containing all the blocks behind the times of all sensors, and

$$\widehat{W} = \begin{bmatrix} \widehat{W}_{11} & \cdots & \widehat{W}_{1,Nm} \\ \vdots & \ddots & \vdots \\ \widehat{W}_{Nm,1} & \cdots & \widehat{W}_{Nm,Nm} \end{bmatrix} \quad (11)$$

is a matrix of dimension $4SP$ x $lagP$ containing all coefficients of all filters of separation.

II. Cost Function and Updating

Similar to the separation system described by (8), the mixing system can be modeled as $X(m) = S(m)\tilde{G}$, where $S(k)$ is a matrix of $2N \times (Nm)$ $(U + 2S - 1)$ dimension containing the backward versions of the sources signals and $\tilde{G}$ is the mixing matrix of Sylvester type of order $(Nm)(U + 2S - 1) \times 4(Nm)S$ containing the coefficients of the impulse response of all filters. These dimensions result, again, the condition of linearity of convolutions performed. It is therefore possible to obtain a block diagonal matrix $B = \tilde{G}\widehat{W}$, such that $B - bdiag\ B = 0$.

The *Bdiag* operator operates on a matrix formed by sub-matrices, zeroing all sub-matrices that do not belong to the main diagonal.

We can define

$$R_{xx}(k) = X^H(k)X(k) \quad (12)$$

and

$$R_{yy}(k) = Y^H(k)Y(k) \quad (13)$$

having dimensions $2(Nm)S$ x $2(Nm)S$ and $(Nm)(lag)$ x $(Nm)(lag)$, respectively. For (13) has full rank, it is necessary the size of the output block to be $N \geq (lag)$.

The objective function is given by [2]

$$\varsigma(k) = \sum_{i=0}^{\infty} \beta(i,k)\{log[bdiag(Y^H(i)Y(i))] - log[\det(Y^H(i)Y(i))]\} \quad (14)$$

where β represents a normalized constant according to $\sum_{i=0}^{\infty} \beta(i,k) = 1$. Using the matrix formulation of (8) to calculate the reduced temporal correlation matrices of (13), the objective function contains *lag* time delays of autocorrelations and cross-correlations of output signals.

Considering an algorithm based on gradient method, the recursive equation for updating the coefficients of the filters that extract the mixture is written as

$$\widehat{W}(k+1) = \widehat{W}(k) - \mu\nabla_{\widehat{W}}\varsigma(l) \quad (15)$$

Using the formulation of the natural gradient [4] which is more robust and less computationally complexity, we obtain the following recursive equation for updating the coefficients:

$$\widehat{W}(k+1) = \widehat{W}(k) - \mu\nabla_{\widehat{W}}^{NG}\varsigma(k) \quad (16)$$

where the natural gradient of the objective function (14):

$$\nabla_{\widehat{W}}^{NG}\varsigma(k) = \widehat{W}\widehat{W}^H\nabla_{\widehat{W}}\varsigma(k) \quad (17)$$

$$= 2\sum_{i=0}^{\infty} \beta(i,k)\widehat{W}\left\{R_{yy}(i) - bdiagR_{yy}(i)\right\}bdiag^{-1}R_{yy}(i) \quad (18)$$

and μ is the step of adapting the algorithm.

The operator *bdiag*(•) interprets the matrix to which is applied as a composition of sub-matrices, zeroing all sub-matrices that do not belong to its main diagonal. To illustrate this operator, assume a system with 3 sources.

The array $R_{yy}(k)$ is written as:

$$R_{yy}(k) = \begin{bmatrix} Y_1^H(k)Y_1(k) & Y_1^H(k)Y_2(k) & Y_1^H(k)Y_3(k) \\ Y_2^H(k)Y_1(k) & Y_2^H(k)Y_2(k) & Y_2^H(k)Y_3(k) \\ Y_3^H(k)Y_1(k) & Y_3^H(k)Y_2(k) & Y_3^H(k)Y_3(k) \end{bmatrix} \quad (19)$$

where $Y_i^H(k)Y_i(k)$are matrices (with i = 1, 2 and 3) are the autocorrelation matrices of the i^{th} output, while the matrices $Y_i^H(k)Y_j(k)$, with $i \neq j$, are the matrices of cross-correlation between the i^{th} and j^{th} output. It is natural to subdivide $R_{yy}(k)$ matrix into sub-matrices, and the autocorrelation sub-matrices belonging to the main diagonal of the matrix of sub-matrices. So *bdiag* $R_{yy}(k)$ yields the following result [5]:

$$bdiagR_{yy}(k) = \begin{bmatrix} Y_1^H(k)Y_1(k) & 0 & 0 \\ 0 & Y_2^H(k)Y_2(k) & 0 \\ 0 & 0 & Y_3^H(k)Y_3(k) \end{bmatrix} \quad (20)$$

where zero is a matrix that has dimension as $Y_i^H(k)Y_j(k) = R_{y_iy_j}(k)$.

During updating the coefficients it is necessary to ensure the structure of Sylvester matrix $\widehat{W}(k+1)$. The indiscriminate use of a gradient that acts on the entire array can destroy this characteristic by removing the redundancy that allows a two-way relationship between the matrices $\widehat{W}_{pq}$ and the corresponding filters $(\widehat{W}_{nm,ns})$ [6]. This is easily imposed by selecting one of the columns of matrices $\widehat{W}_{nm,ns}$ which contains all coefficients of the filters $\widehat{\omega}_{nm,ns}(kk)$ (for kk = 0, ..., 2S - 1) and generate $\nabla_{\widehat{W}}^{NG}\varsigma(k)$ according to (7). In [2] it is shown that the choice of the first S elements of column $\widehat{W}_{nm,ns}$ is the best choice for optimization purposes. Consider a system with two sources and two sensors (TITO, Two Input Two Output):

$$\widehat{W}(i) = \widehat{W}(i-1) - \frac{2\mu}{b}\sum_{m=1}^{b} \begin{bmatrix} \widehat{W}_{12}R_{y2y1}R_{y1y1}^{-1} & \widehat{W}_{11}R_{y1y2}R_{y2y2}^{-1} \\ \widehat{W}_{22}R_{y2y1}R_{y1y1}^{-1} & \widehat{W}_{21}R_{y1y2}R_{y2y2}^{-1} \end{bmatrix} \quad (21)$$

where $R_{nm,ns}$, of dimension $lag \times lag$, a sub-matrix of R_{yy} (13), i is the number of iterations and μ is the step-size.

III. Convolutive Mixtures Problem

The algorithms described in the previous section do not perform well due to the reverberant environment. A simple way to improve the source separation to diagonalize the correlation matrix of outputs R_{YY} [7], which for a linear MIMO system $(Nm) \times Q$ with $(Nm) = Q$ is given by

$$R_{YY} = \begin{bmatrix} \langle \Phi(Y_1)Y_1^H \rangle & \langle \Phi(Y_1)Y_2^H \rangle & \cdots & \langle \Phi(Y_1)Y_p^H \rangle \\ \langle \Phi(Y_2)Y_1^H \rangle & \langle \Phi(Y_2)Y_2^H \rangle & \cdots & \langle \Phi(Y_2)Y_p^H \rangle \\ \vdots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ \langle \Phi(Y_p)Y_1^H \rangle & \langle \Phi(Y_p)Y_2^H \rangle & \cdots & \langle \Phi(Y_p)Y_p^H \rangle \end{bmatrix} \quad (22)$$

where $\langle . \rangle$ is the statistical average operator. The coefficients of the filters, $\omega_{ns,nm}(n)$, that extract the mixture should converge to values that minimize the mutual information between outputs, which correspond to elements that are outside the main diagonal of the correlation matrix, i.e.

$$\langle \Phi(Y_i)Y_j^H \rangle = 0 \quad for \ i \neq j \quad (23)$$

Already the main diagonal elements, which control the scaling of the outputs, must be restricted to appropriate constants b_i, i.e.:

$$\langle \Phi(Y_i)Y_j^H \rangle = b_i \quad (24)$$

The iterative equation for updating the filter coefficients based on the method of separation of the gradient is given by

$$W_{i+1} = W_i + \mu \Delta W_i \quad (25)$$

where

$$\Delta W_i = \begin{bmatrix} b_1 - \langle \Phi(Y_1)Y_1^H \rangle & \langle \Phi(Y_1)Y_2^H \rangle & \cdots & \langle \Phi(Y_1)Y_{nm}^H \rangle \\ \langle \Phi(Y_2)Y_1^H \rangle & b_2 - \langle \Phi(Y_2)Y_2^H \rangle & \cdots & \langle \Phi(Y_2)Y_{nm}^H \rangle \\ \vdots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ \langle \Phi(Y_{nm})Y_1^H \rangle & \langle \Phi(Y_{nm})Y_2^H \rangle & \cdots & b_p - \langle \Phi(Y_{nm})Y_{nm}^H \rangle \end{bmatrix} \quad (26)$$

We can use second order statistics (SOS) considering several blocks of samples of the output signals. This method is known as non-stationary decorrelation [8]. There is another method for colored sources, which also considers using SOS and Time-Delayed Decorrelation, i.e.

$$\langle \Phi(Y_i)Y_j^H \rangle = \langle Y_i(k)Y_j(k + \tau_i)^H \rangle = 0 \quad (27)$$

It can also solve the problem of BSS. Using these types of decorrelation has fair enough information for estimating the separating filter; there is no need for higher orders statistical information to ensure the independence between the sample estimates of the sources [9].

On the other hand, when we consider $\Phi(Y_i) = \tanh(Y_i)$ we have

$$\langle \Phi(Y_i)Y_j^H \rangle = \langle \tanh(Y_i)Y_j^H \rangle = 0 \quad (28)$$

which can be seen as a case of non-linear decorrelation.

IV. Results

In these experiments we used two speech signals with duration ranging from 15 to 30 seconds. The mixtures were carried out considering different reverberation conditions. The separating filter length is equal to the mixing filters, $Ls = S$. We have used the signal-to-interference ratio (SIR) in [2]

$$SIR_{ij} = \frac{\sum_n |b_{ji} * s_i(n)|^2}{\sum_{r=1, r \neq i}^{N} \sum_n |b_{jr} * s_r(n)|^2} \quad (29)$$

where $b_{jr}(n)$ is the sum of the convolutions of the filters of the j^{th} row of the matrix W with the filters of r^{th} column of the matrix H, i.e., $B = W * H$.

In these experiments we compare the performance of the algorithm presented in full-band [2] with the subbands algorithm proposed in this paper: Ls = 256, 512 and 1024, see Fig. 1.

Cosine modulated maximally decimated is used with M from one to sixteen. Table I shows the size of K separation sub-filters $w_{nm,ns}^{i}(k)$ and the steps in full band (M = 1) for different mixtures.

Table II contains the final SIR algorithms for the full-band and sub-band, and Table III shows the amount of multiplications each block according to (30), (31) [1], [5].

$$N_M(Subband) = \frac{P^2(12MK^3 - 8K^3)}{M} \quad (30)$$

and

$$N_M(Fullband) = 8P^2S^3 \quad (31)$$

Looking at Tables II and III we can see that with increasing order of the mixing scheme corresponds to longer echo. The benefits of sub-band configuration on full band come to be more apparent, causing considerably higher signal-to-interface ratio.

TABLE I
Different MS with Different LS for the Full-Band and Sub-Bands Algorithms

	K subfilters length of			Step size
M	*Ls*=256	*Ls* =512	*Ls*=1024	μ (*10^{-4})
2	132	260	516	5
4	68	132	260	10
8	36	68	132	20
16	20	36	68	30
32	12	20	36	40

TABLE II
Final SIR (dB)

	SIR Final				
S = *Ls*	*M* = 1	*M* = 2	*M* = 4	*M* = 8	*M* = 16
256	13.03	13.71	13.92	13.27	15.99
512	9.74	9.03	9.97	10.00	11.94
1024	7.25	6.59	7.45	7.51	8.84

TABLE III
NUMBER OF MULTIPLICATIONS PER BLOCK

$S = U$	$M = 1$	$M = 2$	$M = 4$	$M = 8$	$M = 16$
256	2.68×10^8	9.56×10^7	2.05×10^7	4.87×10^6	1.51×10^6
512	2.15×10^9	6.44×10^8	1.19×10^8	2.25×10^7	5.09×10^6
1024	1.72×10^{10}	4.71×10^9	8.05×10^8	1.31×10^8	2.35×10^7

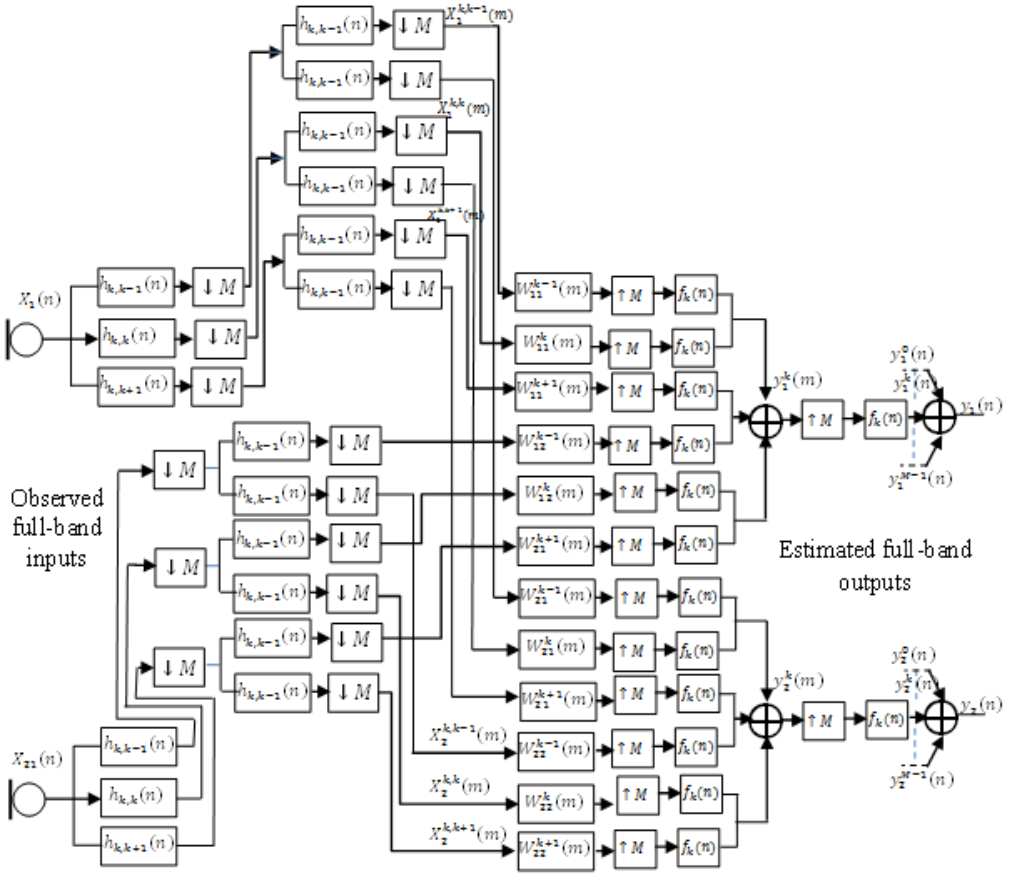

Fig. 1 Proposed configuration for two input-two output sub-band BSS

V. CONCLUSION

A new sub-band configuration is proposed for separation of sources, non-uniform structure that employs non-uniform decomposition of the signals observed by the sensors. We first tested K filter separators with different lengths. Second, the adaptation is performed by algorithms based on natural gradient with measures of the signal-to-interface ratios for different values of *S*. Finally we tested different number of blocks.

We observed the correlation between the outputs of different sub-bands to avoid problems of permutation, but the algorithms were very robust, not requiring any correction.

REFERENCES

[1] Batalheiro, P., Mariane R., Diego B., "Subband Blind Source Separation with Critically Sampled Filter Banks", IWSSIP 2010 - 17th International Conference on Systems, Signals and Image Processing.

[2] Buchner, H., Aichner, R., Kellermann, W., "A Generalization of Blind Source separation Algorithms for Convolutive Mixtures Based on Second-Order Statistics", IEEE Transaction on Speech and Audio Processing, Jan. 2005.

[3] Buchner, H., Aichner, R., Kellermann, W., "A Generalization of a Class of Blind Source Separation Algorithms for convolutive mixtures". In: Proc. Int. Symposium Independent Component Analysis Blind Signal Separation, April 2003.

[4] Nesta, F., Omologo, M., Svaizer, P., "Multiple TDOA estimation by using a state coherence transform for solving the permutation problem in frequency-domain BSS". In: Proc. Machine Learning for Signal Processing, October 2008.

[5] Aichner, R., Buchner, H., Kellermann,W., "Exploiting Narrowband Efficiency for Broadband Convolutive Blind Source Separation", EURASIP Journal on Applied Signal Processing , pp. 1–9, September 2006.

[6] Douglas, S. C., Malay Gupta, "Scaled Natural Gradient Algorithms for Instantaneous and Convolutive Blind Source Separation". In: Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing, v. 2, pp.II-637 – II-640, April 2007.

[7] Araki, S., Makino, S., R. Aichner, et al., "Subband-Based Blind Separation for Convolutive Mixtures of speech", IEICE Transaction Fundamentals, ver. E88-A, pp. 3593–3603, 2005

[8] Lee, I., Kim, T., Lee, T.W., "Independent vector analysis for convolutive blind speech separation". Signals and Communication Technology, pp. 169-192. Springer Netherlands, 2007.
[9] Aichner, R., Buchner, H., Araki, S., et al., "On-Line Time-Domain Blind Source Separation of Nonstationary Convolved Signals". In: Proc. Eur. Signal Processing Conf., pp. 987–992, April 2003.