IWSSIP 2010 - 17th International Conference on Systems, Signals and Image Processing

Subband Blind Source Separation with Critically Sampled Filter Banks

Paulo Bulkool Batalheiro
State University of Rio de Janeiro
FEN/DETEL/PEL/PROSAICO
20559-900, Rio de Janeiro, Brazil
Email: pbb@uerj.br

Mariane Rembold Petraglia
Federal University of Rio de Janeiro
PEE/COPPE - CP 68504
21945-970, Rio de Janeiro, Brazil
Email: mariane@pads.ufrj.br

Diego Barreto Haddad
Federal Center for Technological Education
CEFET-RJ - CP 26041-271
Nova Iguaçu, RJ, Brazil
Email: diego@pads.ufrj.br

Abstract— Subband blind source separation methods have been recently proposed with the objective of reducing the computational complexity and improving the convergence rate of online adaptive algorithms. Oversampled subband structures, employing uniform DFT filter banks, are usually employed in order to avoid aliasing effects and keep enough samples to estimate the statistics of the subband signals. In this paper we present a critically sampled subband structure, composed of real-coefficients uniform filter banks and reduced-order adaptive subfilters, for the blind separation of convolutive mixtures. Through experimental results, we evaluate the impact of the filter banks and lengths of the adaptive filters on the source separation for different reverberation characteristics.

Keywords - subband Blind Source Separation; convolutive mixtures; second-order statistics.

I. Introduction

Blind source separation (BSS) techniques have been extensively investigated in the last decade, for applications in several areas, such as audio systems, image enhancement and digital communications. These emerging techniques allow the extraction of a desired source signal from a mixture of several observed signals without knowledge of the sources (location, spectrum, etc.) nor of the mixture system. This paper considers determined convolutive linear mixtures of speech signals (where the number of sensors P is equal to the number of sources Q), which takes into account the reverberation in echoic ambient. In such cases, typically finite impulse response (FIR) separation filters of large orders are required, making the separation task very complex. In order to solve such problem, several methods based on independent component analysis (ICA), which assume independent non-gaussian sources, have been proposed in the literature.

Some of these solutions employ FIR separation filters, whose coefficients are estimated by an ICA algorithm directly in the time-domain. In real applications, the separation filters have thousands of coefficients and, therefore, such algorithms present large computational complexity and slow convergence in the source estimations [1]. In order to mitigate such difficulties, frequency-domain BSS methods were proposed, where the convolutions become products, and the convolutive mixtures can be treated as instantaneous mixtures in each frequency bin [2]. The disadvantages of such methods are the scaling and permutation problems among the bins, besides the need of using long windows of data for implementing high-order filters. Owing to the non-stationarity of the speech signals and mixing systems, the estimates of the needed statistics for each bin might not be correct for long window data. Such disadvantages can severely degrade the performance of frequency-domain algorithms. Some of these problems can be lessened in off-line implementation by minimal distortion principle, multivariate score function and direction of arrival (DOA) estimation for sources [1]. In this scenery, subband methods can be employed mainly due to their characteristics of breaking the high-order separation filters into independent smaller-order ones and allowing the reduction of the data sampling rate.

In this paper we present an online subband BSS method that employs critically sampled uniform filter banks and reduced-order separation FIR filters. The coefficients of the subband separation filters are adjusted independently by a time-domain adaptive algorithm using second-order statistics [3], making it possible to assume different reverberation conditions in several subbands. The proposed structure applies multirate processing to reduce the computational complexity and extra filters to cancel aliasing among adjacent bands [4]. Another advantage of the proposed algorithm is the use of real-coefficients filters, which is suitable for DSP implementations. An experimental evaluation of the performance of the proposed method for speech signal is conducted, considering the filter bank design and the reverberation characteristics.

II. Online Time-domain Subband BSS

Fig. 1 shows the kth subband of the linear TITO (two inputs and two outputs) configuration for the M-band BSS. In this structure the qth observed signal ($x_q(n)$) is decomposed by the direct-path filters ($h_{k,k}(n)$) and the extra filters ($h_{k,k-1}(n)$ and $h_{k,k+1}(n)$). Assuming that the filter prototype used in the implementation of the analysis filters are sufficiently selective, we can consider that aliasing effect only exists

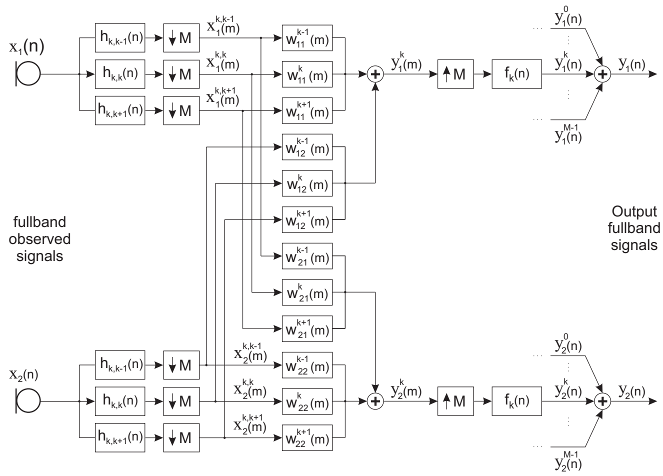

Figure 1. k-th subband of the linear TITO configuration for BSS in M-band.

among adjacent bands being extra filters responsible for its cancellation [4]. The resulting signals ($x_q^{k,k}(n)$, $x_q^{k,k-1}(n)$ and $x_q^{k,k+1}(n)$) are downsampled by the critical decimation factor (M) and applied to the separation filters ($w_{qp}^{k}(n)$, $w_{qp}^{k-1}(n)$ and $w_{qp}^{k+1}(n)$, respectively). The corresponding output signals are up-sampled and recombined by the synthesis filters ($f_k(n)$) to restore the fullband output signals (estimated sources). With perfect reconstruction (PR) filter banks, including extra filters for cancelling aliasing among adjacent subbands, this structure is able of exactly modeling any FIR unmixing system [4].

Assuming that $h_p(n)$ is the impulse response of a prototype filter of length N_P of a cosine modulated multirate system having M bands [5], we conclude that the number of coefficients of each separation subfilter at the kth subband should be at least [4]

$$K = \left\lceil \frac{S + N_P}{M} \right\rceil, \tag{1}$$

where S is length of the fullband separation filter. The filters $h_{k,i}(n)$ of Fig. 1, which decompose the observed signals $x_q(n)$, have impulse responses $h_{k,i}(n) = h_k(n) * h_i(n)$.

The observed subband signals $x_q^{k,i}(m)$ of Fig. 1 can be expressed as

$$x_q^{k,i}(m) = \mathbf{x}_p^T(m)\mathbf{h}_{k,i}, \tag{2}$$

where $\mathbf{x}_p(m) = [x_p(mM), x_p(mM-1), \ldots, x_p(mM - N_H + 1)]^T$ is the vector that contains the latest $N_H = 2N_p - 1$ samples of the pth sensor signal and $\mathbf{h}_{k,i} = [h_{k,i}(0), h_{k,i}(1), \ldots, h_{k,i}(N_H{-}1)]^T$ is the vector that contains the N_H coefficients of the analysis filter $h_{k,i}(n)$.

For colored and non-stationary signals, such as speech signals, the BSS problem can be solved by diagonalizing the output correlation matrix considering multiple blocks in different time instants. The method derived in [3] is based on second order statistics and explores two characteristics of the source signals simultaneously: nonwhiteness, and nonstationarity. This approach, differently of the ICA, is robust to the permutation problem among sources [3], being interesting for an implementation in subbands. In this context, we have extended such algorithm to the subband time-domain, with multirate processing.

From Fig. 1, assuming that there is no overlap between the frequency responses of non-adjacent filters $h_k(n)$ [4], we observe that the qth output signals at the kth subband are given by

$$y_q^k(m) = \sum_{p=1}^{P} \sum_{i=k-1}^{k+1} [\mathbf{w}_{qp}^i]^T \mathbf{x}_p^{k,i}(m), \tag{3}$$

where $\mathbf{x}_p^{k,i}(m)$ is the vector that contains the latest K samples of the pth sensor subband signal $x_p^{k,i}(m)$, and $\mathbf{w}_{qp}^i = [w_{qp}^i(0), w_{qp}^i(1), \ldots, w_{qp}^i(K{-}1)]^T$ is the vector containing the K coefficients of the subband separation subfilters $w_{qp}^i(m)$. The vector $\mathbf{x}_p^{k,i}(m)$ can be written as

$$\mathbf{x}_p^{k,i}(m) = \mathbf{X}_p(m)\mathbf{h}_{k,i}, \tag{4}$$

where the $K \times N_H$ matrix $\mathbf{X}_p(m)$ is given by

$$\mathbf{X}_p(m) = \begin{bmatrix} \mathbf{x}_p(mM)^T \\ \mathbf{x}_p((m-1)M)^T \\ \vdots \\ \mathbf{x}_p((m-K+1)M)^T \end{bmatrix}. \tag{5}$$

In the generic block time-domain subband BSS algorithm, defining N as the block size and D as the number of blocks which are used in the correlation estimates ($1 \leq D \leq K$), we can write the kth output vectors of block index ℓ as

$$\mathbf{y}_q^k(\ell) = \sum_{p=1}^{P} \sum_{i=k-1}^{k+1} [\mathbf{w}_{qp}^i]^T \; \hat{\mathbf{X}}_p^{k,i}(\ell), \tag{6}$$

with the $K \times N$ matrix $\hat{\mathbf{X}}_p^{k,i}(\ell)$ expressed as

$$\hat{\mathbf{X}}_p^{k,i}(\ell) = \left[\mathbf{X}_p(\ell K), \mathbf{X}_p(\ell K+1), \cdots, \mathbf{X}_p(\ell K+N-1)\right] \mathbf{H}_{k,i}\ , \tag{7}$$

where the $N_H N \times N$ matrix $\mathbf{H}_{k,i}$ has the first column formed by the coefficients of $h_{k,i}(n)$ followed by $(N-1)N_H$ zeros, and the remaining columns are circularly shifted by N_H positions from one column to the next. The $D \times N$ matrices $\mathbf{Y}_q^k(m)$ contain D subsequent output vectors and can be written as

$$\mathbf{Y}_q^k(\ell) = \sum_{p=1}^{P} \sum_{i=k-1}^{k+1} \mathbf{W}_{qp}^i(\ell) \; \mathbf{X}_p^{k,i}(\ell)\ , \tag{8}$$

with

$$\mathbf{X}_p^{k,i}(\ell) = \left[\hat{\mathbf{X}}_p^{k,i}(\ell),\ \hat{\mathbf{X}}_p^{k,i}(\ell-1)\right]^T \tag{9}$$

and $\mathbf{W}_{pq}^i(\ell)$ is a $D \times 2K$ Sylvester-type matrix given by

$$\mathbf{W}_{qp}^i(\ell) = \begin{bmatrix} w_{qp,0}^i & w_{qp,1}^i & \cdots & w_{qp,k-1}^i & 0 & \cdots & 0 & 0 \\ 0 & w_{qp,0}^i & w_{qp,1}^i & \cdots & w_{qp,k-1}^i & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & 0 & 0 \\ 0 & \cdots & 0 & w_{qp,0}^i & w_{qp,1}^i & \cdots & w_{qp,k-1}^i & 0 \end{bmatrix}. \tag{10}$$

Combining all P outputs of each subband, we can rewrite Eq. (3) concisely as

$$\mathbf{Y}^k(\ell) = \left[\mathbf{Y}_1^k(\ell), \; \cdots, \; \mathbf{Y}_P^k(\ell)\right]^T = \sum_{i=k-1}^{k+1} \mathbf{W}^i(\ell)\mathbf{X}^{k,i}(\ell) \; , \quad (11)$$

where

$$\mathbf{X}^{k,i}(\ell) \;\; = \;\; \left[\mathbf{X}_1^{k,i}(\ell), \quad \ldots , \quad \mathbf{X}_P^{k,i}(\ell)\right]^T , \quad (12)$$

$$\mathbf{W}^i(\ell) \;\; = \;\; \begin{bmatrix} \mathbf{W}_{11}^i(\ell) & \ldots & \mathbf{W}_{1P}^i(\ell) \\ \vdots & \ddots & \vdots \\ \mathbf{W}_{P1}^i(\ell) & \ldots & \mathbf{W}_{PP}^i(\ell) \end{bmatrix} . \quad (13)$$

In matrix formulation, the online subband BSS cost function is given by

$$\Im^k(\ell) = \log(\det(\text{bdiag}(\mathbf{Y}^k(\ell)[\mathbf{Y}^k(\ell)]^T)) - \log(\det(\mathbf{Y}^k(\ell)[\mathbf{Y}^k(\ell)]^T)). \quad (14)$$

where bdiag$(\mathbf{A})$ is the operator that zeroes all the submatrices which are not located in the main diagonal of $\mathbf{A}$.

Applying the natural gradient method to the above cost function, considering a TITO system and omitting index ℓ of the submatrix correlation to simplify the notation, we obtain

$$\nabla^{GN}_{\mathbf{W}_k}\Im(\ell) = 2\begin{bmatrix} [\mathbf{R}^k_{y_1y_1}]^{-1}[\mathbf{R}^k_{y_2y_1}]^T\mathbf{W}^k_{21} & [\mathbf{R}^k_{y_1y_1}]^{-1}[\mathbf{R}^k_{y_2y_1}]^T\mathbf{W}^k_{22} \\ [\mathbf{R}^k_{y_2y_2}]^{-1}[\mathbf{R}^k_{y_1y_2}]^T\mathbf{W}^k_{11} & [\mathbf{R}^k_{y_2y_2}]^{-1}[\mathbf{R}^k_{y_1y_2}]^T\mathbf{W}^k_{12} \end{bmatrix}, \quad (15)$$

where $\mathbf{R}^k_{y_qy_p}(\ell)$ is a submatrix of dimension $D \times D$.

The online algorithm for adjusting the coefficients of the separation subfilters of each subband, considering a TITO system, is given by

$$\mathbf{W}^k(\ell) = \mathbf{W}^k(\ell-1) - \mu\left[\lambda\nabla^{GN}_{\mathbf{W}_k}\Im(\ell-1) + (1-\lambda)\nabla^{GN}_{\mathbf{W}_k}\Im(\ell)\right] \quad (16)$$

where μ is the step-size of the adaptation algorithm and λ is a forgetting factor [6].

III. Experimental Evaluation

In all experiments, the signals were sampled at F_s =16 kHz and the subband coefficients $w^i_{qp}(n)$ were initialized with zeros, except for $p = q$, in which the coefficients were initialized with 1. The number of data blocks used to estimate the correlations (Eq. (15)) was $D = K$, the length of the output blocks was $N = 2K$, the lengths of the separation filters were equal to the lengths of the mixture filters ($S = U$) and the forgetting factor was $\lambda = 0.2$. During our experiments we monitored the correlations among the estimates of the sources in the several subbands in order to avoid permutation problems.

A. Experiment 1

In this experiment we evaluate the impact of the filter bank (through the use of different prototype filters) in the following performance measures: SIR^G (Global Signal to Interference Ratio), SAR^G (Global Signal-to-Artifact Ratio) and SDR^G (Global Signal-to-Distortion Ratio). We will establish the better relation among the number of subbands (M) and the prototype length (N_p) for the BSS application.

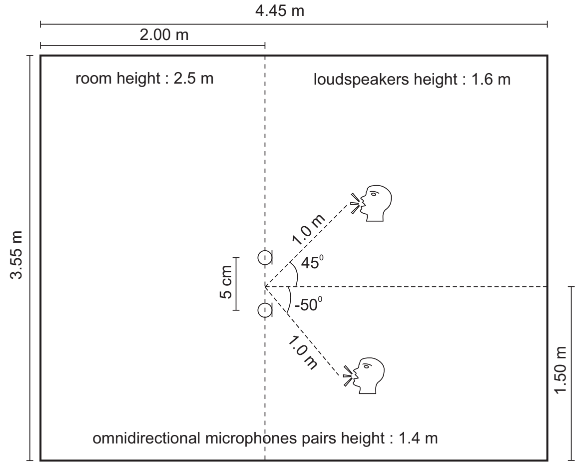

Figure 2. Virtual scenario used in the experiments.

TABLE I. Global performance measures for Experiment 1

	$U = 128$			$U = 256$			$U = 512$		
N_p	SIR^G	SAR^G	SDR^G	SIR^G	SAR^G	SDR^G	SIR^G	SAR^G	SDR^G
$4M$	15.5	20.4	14.0	12.3	19.7	11.6	9.9	9.9	6.6
$8M$	15.3	21.2	14.3	12.8	20.2	12.0	10.5	10.2	7.0
$16M$	12.3	0.5	-1.2	12.4	0.3	-0.2	9.9	0.6	-0.6

A uniform subband structure was implemented using critically-sampled cosine modulated filter banks with $M = 8$ and prototype filters of lengths $N_p = 4M$, $8M$ and $16M$, which yield perfect reconstruction. The adaptation step-size was set equal to $\mu = 2 \times 10^{-3}$. Four speech signals were used: two female and two male, being two of them spoken in English and the other two spoken in Portuguese. Theses signals were convolved with synthetic impulse responses obtained by simulation of the virtual scenario of Fig. 2, which was proposed for the 2006 Signal Separation Evaluation Campaign (SiSEC 2006) [7], presenting reverberation time of 250 ms. Even though the original scenario had four distinct sources in different directions, our simulations employed only two sources, as illustrated in Fig. 2, in order to reduce the computational complexity of the BSS methods. Each pair of speech signals were mixed, resulting in 6 different combinations. Table I presents the average values of the SIR^G, SAR^G and SDR^G, considering different reverberation conditions: $U = 128$, 256 and 512. These measures were obtained using the toolbox for the calculation of the performance measures of BSS available in [8].

From Table I we observe that the use of very selective prototype filters ($N_p = 16M$) sometimes does not result in the expected performance in terms of SIR^G, SAR^G and SDR^G, with observed distortions in the source estimates. On the other hand, the use of less selective filter banks ($N_p = 4M$) might degrade the separation, since the mean-square error of

TABLE II. COMPLEXITY COMPUTATIONAL FOR EXPERIMENT 1

$M = 8$			
N_p	$U = 128$	$U = 256$	$U = 512$
$4M$	3.52×10^5	2.05×10^6	13.83×10^6
$8M$	6.08×10^5	2.82×10^6	16.42×10^6
$16M$	1.44×10^6	4.86×10^6	22.53×10^6
$M = 1$ (full band)			
	$U = 128$	$U = 256$	$U = 512$
	33.55×10^6	2.68×10^8	2.15×10^9

the structure is increased [4]. Table II presents complexity computational from Table I simulations and compares with fullband algorithm ($M = 1$) presented in [3].

Therefore, according to Tables I and II, the best global measures were obtained with the prototypes of lengths $N_p = 8M$, which also reduces the complexity computational when compared to fullband implementation. Such relation between N_p and M will be adopted in the next experiment.

B. Experiment 2

The reverberation time is an important factor for the sound quality of the signals captured by the microphones. In [9], several experiments in rooms of different dimensions were conducted, demonstrating that the average reverberation time decreases as the frequencies of the signals increases, mainly due to propagation losses caused by the air viscosity for frequencies above 1 kHz [10]. Based on these results, we explore the flexibility of the subband BSS structure, which allows the independent adaptation of the separation subfilters and the use of different lengths for such filters. The lengths of the high-frequency subfilters were reduced and the results of the separation procedure were compared to those of the fullband algorithm presented in [3], for mixture filters of lengths $U = 256$, 512 and 1024. The subband structure was first simulated with all subfilters of same length, given in Eq. (1). Subsequently, the length of the highest-band subfilter was reduced to half of its initial value. A last experiment was conducted reducing the lengths of the one fourth of bands of highest frequencies to half of their initial values. The results of these three experiments are presented in Table III. From such results, we conclude that the reduction in the number of coefficients in the high-frequency bands does not affect the performance of the separation algorithm, and still contributes to reduce the complexity computational.

TABLE III. SIR WITH DIFFERENT LENGTH SUBFILTERS IN SOME BANDS

SIR without reduced length subfilters			
M	$U = 256$	$U = 512$	$U = 1024$
1	12.2	10.1	7.4
4	13.1	9.9	7.2
8	12.5	10.7	7.2
16	14.2	12.4	8.2
SIR with reduced length filter in the highest frequency band			
M	$U = 256$	$U = 512$	$U = 1024$
1	12.2	10.1	7.4
4	13.5	9.9	7.1
8	13.1	10.7	7.2
16	14.2	12.4	8.2
SIR with reduced length filters in the one fourth of bands of highest frequencies			
M	$U = 256$	$U = 512$	$U = 1024$
1	12.2	10.1	7.4
4	13.0	9.9	7.1
8	12.4	10.6	7.2
16	14.2	12.4	8.2

IV. CONCLUSIONS

In this paper we evaluated experimentally the performance of a subband structure applied to the BSS problem. The proposed subband structure, which cancels aliasing due to the spectral overlap of adjacent analysis filters, was first derived for the system identification application, where the mean-square value of the error, obtained comparing a reference signal and the filter ouptput, was used as cost function. In BSS applications, where a reference signal is not available, the cost function is composed by the correlations among the outputs, considering blocks of signals in different time instants. Computer simulations with speech signals were presented illustrating the influence of the acoustic reverberation and filter bank characteristics on the performance of the proposed BSS method. It was verified that a reduction in the number of coefficients of the subfilters which operate at the high-frequency components does not degrade the performance of the subband BSS algorithm.

ACKNOWLEDGMENT

The authors would like to thank CNPq, Brazil, for providing partial support.

REFERENCES

[1] A. Hyvarinen, H. Karhunen, and E. Oja, *Independent Component Analysis*. John Wiley & Sons, 2001.

[2] H. saruwatari, T. Kawamura, T. Nishikawa, A. Lee, and K. Shikano, "Blind source separation based on a fast-convergence algorithm combining ICA and beamforming," *IEEE Trans. Audio, Sppech, and Language Process.*, vol. 14, no. 2, pp. 666–678, Mar. 2006.

[3] H. Buchner, R. Aichner, and W. Kellermann, "A generalization of blind source separation algorithms for convolutive mixtures based on second-order statistics," *IEEE Trans. on Speech and Audio Process.*, vol. 13, no. 1, pp. 120–134, Jan. 2005.

[4] M. R. Petraglia and P. B. Batalheiro, "Filter bank design for a subband adaptive filtering structure with critical sampling," *IEEE Trans. on Circuits And Systems I - Fundamental Theory And Applications.*, vol. 51, pp. 1194–1202, Jun. 2004.

[5] T. Q. Nguyen, "Digital filter banks design - quadratic constrained formulation," *IEEE Trans. on Signal Processing*, vol. 43, pp. 2103–2108, Sep. 1995.

[6] S. Haykin, *Adaptive Filter Theory - 4th ed.* Englewood Cliffs, NJ: Prentice-Hall, 2001.

[7] http://www.irisa.fr/metiss/SASSEC07/?show=dev.

[8] C. Fevotte, R. Gribonval, and E. Vincent, "Bss_eval toolbox user guide," *IRISA Technical Report 1706*, Apr. 2005. http://www.irisa.fr/metiss/bss_eval/.

[9] F. A. Everest and K. Pohlmann, *Master HandBook of Acoustics*. McGraw-Hill, 2001.

[10] L. Beranek, *Music, Acoustics and Architecture*. John Wiley & Sons, 1962.