16th European Signal Processing Conference (EUSIPCO 2008), Lausanne, Switzerland, August 25-29, 2008, copyright by EURASIP

BLIND SOURCE SEPARATION FOR CONVOLUTIVE MIXTURES USING A NON-UNIFORM OVERSAMPLED FILTER BANK

Mariane R. Petraglia [1], *Paulo B. Batalheiro* [2], *Diego B. Haddad* [1]

[1]Federal University of Rio de Janeiro, PEE/COPPE
CP 68504, 21945-970, Rio de Janeiro, Brazil
Emails: mariane@pads.ufrj.br, diego@pads.ufrj.br
[2]State University of Rio de Janeiro, CTC/FEN/DETEL
20550-013, Rio de Janeiro, Brazil
Email: bulkool@pads.ufrj.br

ABSTRACT

Adaptive subband structures have been proposed with the objective of increasing the convergence speed and/or reducing the computational complexity of adaptation algorithms for applications which require a large number of adaptive coefficients. In this paper we propose a blind source separation method for convolutive mixtures which employs a real-coefficient non-uniform filter bank and a new normalization scheme for the adaptation algorithm. Since the separation filters in the subbands work at reduced sampling rates, the proposed method presents smaller computational complexity and faster convergence rate when compared to the corresponding fullband algorithm.

1. INTRODUCTION

Blind source separation (BSS) techniques have been extensively investigated in the last decade, allowing the extraction of the signal of a desired source $s_q(n)$ from mixed signals of more than one source $x_p(n)$ without any other knowledge of the original sources, such as their positions or spectral contents, nor of the mixing process. Examples of applications of BSS are speech enhancement/recognition (cocktail party problem) and digital communication, among others. The mixtures can be classified as linear or non-linear and instantaneous or convolutive. This paper considers convolutive mixtures of speech signals, which takes into account the reverberation in echoic ambients. 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 time-domain and frequency-domain methods based on independent component analysis (ICA) have been proposed in the literature.

Some of these solutions employ FIR separation filters and estimate their coefficients with 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, slow convergence and undesired whitening effect in the sources estimations [1, 2]. In order to ease 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 [3, 4]. 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. Due 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 degrade severely the performance of the frequency-domain algorithms. There are also techniques which combine the time and frequency domain solutions to improve the BSS performance and reduce its computational complexity [5]. In this scenery, subband methods have been proposed mainly due to their characteristics of breaking the high-order separation filters into independent smaller-order filters and of allowing the reduction of the sampling rate. Such methods usually employ complex-coefficients oversampled uniform filter banks [6].

In this paper we propose a subband BSS method which employs real-coefficients oversampled non-uniform filter banks and reduced order separation FIR filters. The coefficients of the subband separation filters are adjusted independently by a time-domain adaptation algorithm [1], which employs second-order statistics and a new gradient normalization scheme. The proposed normalization scheme results in faster convergence and reduced complexity when compared to the original normalization scheme. The proposed structure employs multirate processing, with smaller sampling rates at the lower frequency bands, where the speech signal energy is larger. Another advantage of the proposed algorithm is the use of real-coefficients filters, which is attractive for DSP implementations.

In Section 2 the BSS problem for linear convolutive mixtures is presented. In Section 3 the fullband time-domain method proposed in [1], as well as a new gradient normalization scheme, are presented. The proposed non-uniform subband structure is described in Section 4. Section 5 presents experimental results, comparing the performance of the fullband and subband BSS algorithms. In Section 6 the concluding remarks are presented.

2. BSS FOR CONVOLUTIVE MIXTURES

In teleconference systems, the original sources are speech signals and the convolutive mixtures of the sources are caused by the auditorium reverberation. Fig. 1 illustrates a blind source separation system, where the number of sources is equal to the number of microphones. Considering that the unknown mixture system can be modeled by a set of finite impulse response (FIR) filters of length U (convolutive linear mixtures), the signals captured by the microphones $x_p(n)$ can be written as

$$x_p(n) = \sum_{q=1}^{P} \sum_{k=0}^{U-1} g_{qp}(k) s_q(n-k) \qquad (1)$$

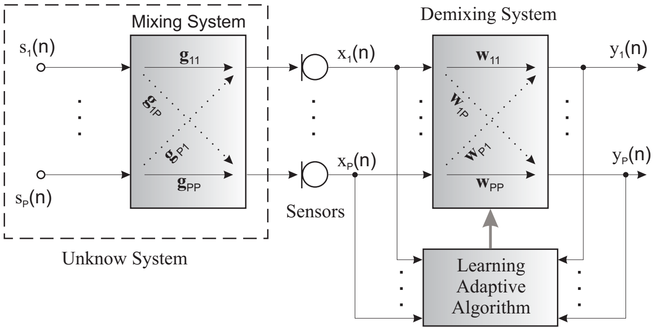

Figure 1: Linear MIMO configuration for fullband BSS.

where g_{qp} is the filter that models the echo path from the qth source to the pth sensor, and P is the number of sources and sensors (determined BSS).

In the BSS problem, the coefficients of the separation filters w_{pq} (of length S) are estimated through an adaptive algorithm, based on the independent component analysis (ICA) technique, so that their output signals $y_q(n)$, given by

$$y_q(n) = \sum_{p=1}^{P}\sum_{k=0}^{S-1} w_{pq,k}x_p(n-k) \quad \text{for } q = 1,\dots,P \tag{2}$$

become mutually independent.

3. BLOCK TIME-DOMAIN BSS ALGORITHM

For colored and non-stationary signals, such as speech signals, the BSS problem can be solved diagonalizing the output correlation matrix considering multiple blocks in different time instants (TDD - *Time-Delayed Decorrelation*). In this section we review the wideband solution based on second order statistics proposed in [1], that explores three caracteristics of the source signals simultaneously: nongaussianity, nonwhiteness, and nonstationarity.

In the generic block time-domain BSS algorithm, defining N as the block size and D as the number of blocks which are used in the correlation estimates ($1 \le D \le S$), the output vectors of block index m are given by

$$\mathbf{y}_q(m) = [y_q(mS), y_q(mS+1), \dots, y_q(mS+N-1)]^T \tag{3}$$

and the $N \times D$ matrices $\mathbf{Y}_q(m)$ containing D subsequent output vectors can be expressed as [1]

$$\mathbf{Y}_q(m) = \sum_{p=1}^{P} \mathbf{X}_p(m)\mathbf{W}_{pq}, \tag{4}$$

with

$$\mathbf{X}_p(m) = \left[\hat{\mathbf{X}}_p^T(m), \hat{\mathbf{X}}_p^T(m-1)\right], \tag{5}$$

$$\hat{\mathbf{X}}_p^T(m) = \begin{bmatrix} x_p(mS) & \cdots & x_p(mS-S+1) \\ x_p(mS+1) & \cdots & x_p(mS-S+2) \\ \vdots & \ddots & \vdots \\ x_p(mS+N-1) & \cdots & x_p((m-1)S+N) \end{bmatrix}. \tag{6}$$

The matrix $\mathbf{W}_{pq}$ is a $2S \times D$ Sylvester-type matrix defined as

$$\mathbf{W}_{pq} = \begin{bmatrix} w_{pq,0} & 0 & \cdots & 0 \\ w_{pq,1} & w_{pq,0} & \ddots & \vdots \\ \vdots & w_{pq,1} & \ddots & 0 \\ w_{pq,S-1} & \vdots & \ddots & w_{pq,0} \\ 0 & w_{pq,S-1} & \ddots & w_{pq,1} \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & w_{pq,S-1} \\ 0 & \cdots & 0 & 0 \\ \vdots & \cdots & \vdots & \vdots \\ 0 & \cdots & 0 & 0 \end{bmatrix}. \tag{7}$$

Combining all channels, Eq. (4) can be expressed concisely as

$$\mathbf{Y}(m) = \mathbf{X}(m)\mathbf{W}, \tag{8}$$

where

$$\mathbf{Y}(m) = [\mathbf{Y}_1(m), \dots, \mathbf{Y}_P(m)], \tag{9}$$

$$\mathbf{X}(m) = [\mathbf{X}_1(m), \dots, \mathbf{X}_P(m)], \tag{10}$$

$$\mathbf{W} = \begin{bmatrix} \mathbf{W}_{11} & \cdots & \mathbf{W}_{1P} \\ \vdots & \ddots & \vdots \\ \mathbf{W}_{P1} & \cdots & \mathbf{W}_{PP} \end{bmatrix}. \tag{11}$$

The above matrices have dimensions $N \times PD$, $N \times 2SP$ and $2SP \times PD$, respectively.

In the matrix formulation, the BSS cost function is given by:

$$\begin{aligned} \Im &= \sum_{i=1}^{b} \frac{1}{b}\{\log(\det(\operatorname{bdiag}(\mathbf{Y}^T(i)\mathbf{Y}(i)))) \\ &\quad -\log(\det(\mathbf{Y}^T(i)\mathbf{Y}(i)))\}, \end{aligned} \tag{12}$$

where b is the number of blocks considered in the optimization and $\operatorname{bdiag}(\mathbf{A})$ is the operator which zeroes all the submatrices which are not located in the main diagonal of matrix $\mathbf{A}$.

Applying the natural gradient method to the cost function of Eq. (12), we get

$$\begin{aligned} \nabla_{\mathbf{W}}^{GN}\Im(m) &= \frac{2}{b}\sum_{i=1}^{b} \mathbf{W}\{\mathbf{R}_{yy}(m) - \operatorname{bdiag}(\mathbf{R}_{yy}(m))\} \\ &\quad \times \left\{\operatorname{bdiag}\left(\mathbf{R}_{yy}(m)\right)\right\}^{-1}, \end{aligned} \tag{13}$$

where

$$\mathbf{R}_{yy}(m) = \mathbf{Y}^H(m)\mathbf{Y}(m) \tag{14}$$

is a matrix of dimension $PD \times PD$.

The batch-type off-line algorithm for adjusting the coefficients of the separation filters, considering a TITO (two sources and two sensors) system, is given by

$$\mathbf{W}(i) = \mathbf{W}(i-1) - \frac{2\mu}{b}\sum_{m=1}^{b}\begin{bmatrix} \mathbf{W}_{12}\mathbf{R}_{21}\mathbf{R}_{11}^{-1} & \mathbf{W}_{11}\mathbf{R}_{12}\mathbf{R}_{22}^{-1} \\ \mathbf{W}_{22}\mathbf{R}_{21}\mathbf{R}_{11}^{-1} & \mathbf{W}_{21}\mathbf{R}_{12}\mathbf{R}_{22}^{-1} \end{bmatrix} \tag{15}$$

where $\mathbf{R}_{pq}$, of dimension $D \times D$, is a sub matrix of $\mathbf{R}_{yy}$ (Eq. (14)), i is the number of the iteration (off-line), and μ is the step-size of the adaptation algorithm.

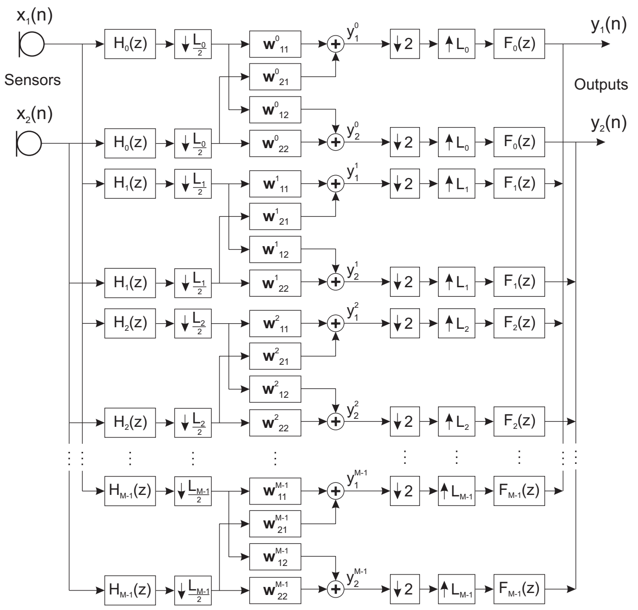

Figure 2: Linear TITO configuration for subband BSS.

Due to redundancies in $\mathbf{W}_{pq}$ (Eq. (7)) and for convergence reasons [1], we update at each iteration only the first S elements of the first-column of this matrix (which are sufficient to form a Sylvester-type matrix). In order to reduce the computational complexity of the algorithm, the normalization factor $\mathbf{R}_{qq}^{-1}(m)$ can be simplified considering $\mathbf{R}_{qq}(m)$ a diagonal matrix [7], that is,

$$\mathbf{R}_{qq}(m) \approx \mathrm{diag}\{\mathbf{R}_{qq}(m)\} = \sigma^2_{\mathbf{Y}_q}(m). \tag{16}$$

In this way, its inverse can be obtained by inverting each element of its diagonal.

With the objective of further reducing the computational cost, we propose to simplify the normalization factor considering it a scalar. In such case,

$$\mathbf{R}_{qq}(m) \approx \mathbf{y}_q^T(m)\mathbf{y}_q(m)\mathbf{I} \tag{17}$$

with $\mathbf{y}_q(m)$ given in Eq. (3), where its inverse is obtained inverting the power of a single block of the output signal $y_q(n)$.

4. SUBBAND BSS METHOD

In this section we investigate the use of a subband structure in conjunction with the block time-domain BSS algorithm presented in the last section. The idea is to exploit the characteristics of better convergence rate and reduced computational complexity of such structures. In [6], a uniform subband structure was used in the BSS algorithm. In this paper, we propose to use a non-uniform subband structure, based on [8], which employs an octave-band frequency decomposition which results in narrower bands for low frequencies, where the speech signals have most of their energy.

Figure 2 shows a TITO system with subband BSS considering an M-channel non-uniform filter banks. This structure is a modified version of that of [8], where the input signals of the separation filters of each subband $\mathbf{w}_{pq}^k$ are down-sampled by half of the critical decimation factor in order to reduce the aliasing effects during the adaptation process. The decimation of the signals at the outputs of the separation filters by 2 restores the critical downsampling rate, before the output signal reconstruction. For M-channel octave-band filter banks, the decimation factors are $L_0 = 2^{M-1}$ and $L_k = 2^{M-k}$ for $k = 1, \cdots, M-1$, and the equivalent analysis filters are

$$\begin{aligned} H_0(z) &= \prod_{j=0}^{M-2} H^{0,j}(z^{2^j}), \\ H_k(z) &= H^{1,M-1-k}(z^{2^{M-1-k}}) \prod_{j=0}^{M-k-2} H^{0,j}(z^{2^j}), \end{aligned} \tag{18}$$

where $H^{0,j}(z)$ and $H^{1,j}(z)$ are the lowpass and high-pass filters, respectively, of the jth stage of the tree structure, and are designed to produce perfect reconstruction (PR) [9]. The number of coefficients of each separation filter at the kth subband should be at least [8]

$$S_k = 2\left\lfloor \frac{S-1+N_{F_k}}{L_k} \right\rfloor + 1, \tag{19}$$

where N_{F_k} is the order of the kth synthesis filter.

To adjust the coefficients of each separation subfilter, we employ the algorithm of Eq. (15). The update equation for the coefficients of the kth band is given by

$$\mathbf{W}^k(i) = \mathbf{W}^k(i-1) - \frac{2}{b_k}\sum_{m=1}^{b_k}\begin{bmatrix} \mathbf{W}_{12}^k\mathbf{R}_{21}^k\mathbf{R}_{11}^{k^{-1}} & \mathbf{W}_{11}^k\mathbf{R}_{12}^k\mathbf{R}_{22}^{k^{-1}} \\ \mathbf{W}_{22}^k\mathbf{R}_{21}^k\mathbf{R}_{11}^{k^{-1}} & \mathbf{W}_{21}^k\mathbf{R}_{12}^k\mathbf{R}_{22}^{k^{-1}} \end{bmatrix} \times \begin{bmatrix} \mu_1^k\mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mu_2^k\mathbf{I} \end{bmatrix} \tag{20}$$

where

$$\mathbf{R}_{pq}^k(m) = \mathbf{Y}_p^{k^H}(m)\mathbf{Y}_q^k(m) \tag{21}$$

and

$$\mathbf{Y}_q^k(m) = \begin{bmatrix} y_q^k(mS_k) & \dots & y_q^k(mS_k - D_k + 1) \\ y_q^k(mS_k + 1) & \ddots & y_q^k(mS_k - D_k + 2) \\ \vdots & \ddots & \vdots \\ y_q^k(mS_k + N_k - 1) & \dots & y_q^k(mS_k - D_k + N_k) \end{bmatrix}. \tag{22}$$

The above matrices have dimensions $D_k \times D_k$ (with $1 \leq D_k \leq S_k$) and $N_k \times D_k$ (with $N_k \geq D_k$) , respectively, b_k is the number of blocks, N_k is the block size, μ_k is the kth band adaptation step-size, i is the (off-line) iteration number, and y_q^k is the qth output in subband k.

5. EXPERIMENTAL RESULTS

In all experiments, two speech signals (of 10s length) sampled at $F_s = 8kHz$ were used: a female English voice and a male Portuguese voice. Such signals were convolved with artificial impulse responses [10], obtained considering the room of Fig. 3 of dimensions 3.55 m × 4.55 m × 2.5 m (with reverberation time around 250 ms). Such impulse responses were truncated, considering only their first S samples. The distance between the two microphones was 5 cm

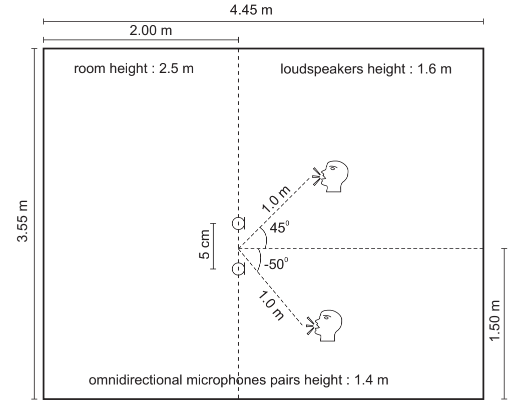
4.45 m
2.00 m
room height : 2.5 m
loudspeakers height : 1.6 m
3.55 m
1.0 m
45°
5 cm
-50°
1.0 m
1.50 m
omnidirectional microphones pairs height : 1.4 m

Figure 3: Virtual scenery used in the experiments.

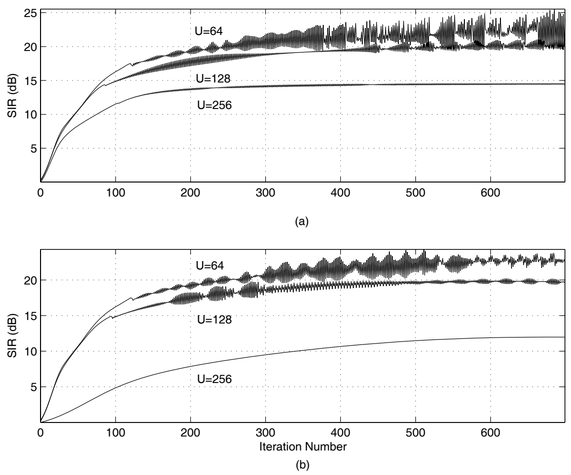

Figure 4: SIR evolution for the two normalization schemes: (a) NN and (b) ON.

and the sources were positioned at 1 m of distance from the center point of the microphones, at directions of -50^0 and 45^0. The length of the separation filters was fixed at the same value of the mixing filters ($U = S$). Both algorithms were implemented in Matlab and ran on an Intel Core 2 Duo 2GHz PC. We adopted, for performance evaluation, the signal to interference ratio (SIR), defined as

$$\text{SIR} = 10\log_{10}\left(\frac{\text{SIR}_1 + \text{SIR}_2}{2}\right) \tag{23}$$

where SIR_i is the ratio of the power of the output signal y_i when only the source s_i is active and the power of the output signal y_i when only the source s_i is inactive.

5.1 Experiment 1

In this experiment we compare the performance of the fullband algorithm with the two different normalization schemes: old normalization (ON - Eq. (16)) and new normalization (NN - Eq. (17)). Figure 4 shows the SIR evolution considering mixture filters of different lengths: $U =$ 64, 128 and 256. The adaptation step-size in all cases was $\mu = 5\times10^{-3}$, except for the old normalization with $U = 256$, where we used $\mu = 10^{-3}$ for convergence reasons. Table 1 shows the processing time for the simulations of Fig. 4.

Table 1: Processing time in minutes.

$S = U$	ON	NN
64	31	18
128	77	32
256	253	80

Table 2: Non-uniform structure parameters for $M = 4$ and $U = S = 1024$.

k	L_k	N_{H_k}	S_k	μ_1^k	μ_2^k
0	8	441	366	8.8×10^{-3}	8×10^{-3}
1	8	441	366	17.6×10^{-3}	16×10^{-3}
2	4	189	606	35.2×10^{-3}	32×10^{-3}
3	2	63	575	70.4×10^{-3}	64×10^{-3}

From Table 1 and Fig. 4, it can be observed that for mixing filters of large lengths (corresponding to highly reverberant ambients) the new normalization scheme reduces significantly the processing time and improves the convergence speed of the fullband algorithm.

5.2 Experiment 2

In this experiment we compare the performances of the fullband and subband algorithms with the normalization scheme of Eq. (17). The non-uniform subband structure was implemented using octave-band filter banks with M=4 subbands and yielding perfect reconstruction. Table 2 presents the decimation factors L_k, the orders of the analysis filters $H_k(z)$ (which are equal to the orders of the synthesis filters $F_k(z)$), the orders of the separation filters $W_{pq}^k(z)$, and the adaptation step-sizes used in the subband simulations. For the fullband algorithm, the adaptation step-size was the same as in Experiment 1, except for $U = 1024$ where $\mu = 3\times10^{-3}$ was used. These step-size values resulted in the fastest adaptation convergence and were obtained experimentally. In order to reduce the computational complexity without significant degradation on the separation process, the orders of the separation filters of the high-frequency band ($k = 3$) were reduced, with respect to Eq. (19), to $S_3 = 2\left\lfloor\frac{S/2-1+N_{F_k}}{L_k}\right\rfloor + 1$. Such reduction is possible due to the reverberation characteristics at high frequencies.

Figure 5 shows the SIR evolution for the fullband algorithm (Eq. (15)) and for the subband algorithm (Eq. (20)), with the lengths of the mixture filters equal to $U = 256$, 512 and 1024. Table 3 contains the maximum SIR in fullband and in each of the 4 bands of the non-uniform structure. Table 4 shows the processing time for the simulations of Fig. 5. From these tables and Fig. 5, it can be observed that as the order of the mixture system increases (more reverberant rooms), the advantages of the subband structure over the fullband structure become more evident, resulting in a significantly faster convergence rate and smaller processing time, which can be further reduced by using parallel processing for the subband implementation.

The behavior of the subband BSS algorithm in the highest frequency band ($k = 3$) was worse than in the other subbands (due to the use of smaller length separation filters in this band), causing a reduction in the final SIR. However, in such frequency band, the power of the speech signals is small and the audible results are considerably better for the

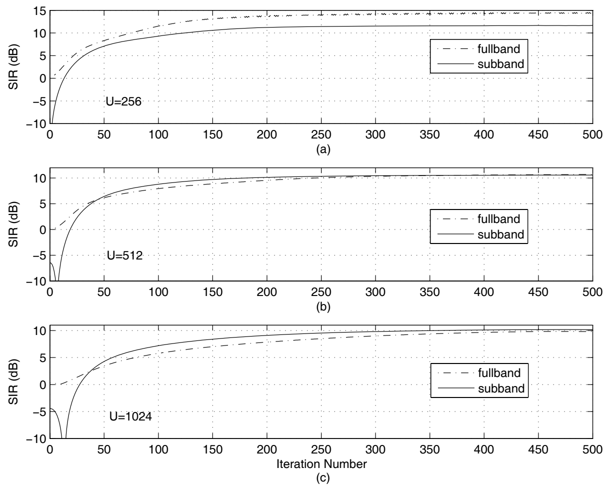

Figure 5: SIR evolution for fullband and subband algorithms with different mixing filters lengths: (a) U=256, (b) U=512, and (c) U=1024.

Table 3: Maximum SIR (in dB).

Mixing Filters	SIR in each subband				SIR
$S = U$	$k = 0$	$k = 1$	$k = 2$	$k = 3$	$--$
256	21.72	11.06	12.57	6.63	14.56
512	13.81	9.41	10.37	6.31	10.70
1024	12.40	8.73	9.39	6.44	9.92

subband structure.

Figure 6 shows the power spectra of the original sources and of theirs estimates obtained with the fullband and subband algorithms, for different lengths of the mixture filters. These results show the robustness of the algorithms for the source whitening problem and for the scaling of the output signals. The permutation problem between the subband outputs did not appear in our experiments.

6. CONCLUSION

In this paper we propose a new subband blind source separation algorithm which employs non-uniform octave-band filter banks to decompose the signals from the sensors. Separating filters, of different lengths and working at reduced sampling rates, are applied to the subband signals. The adaptation is performed by a natural-gradient type algorithm, with a new normalization scheme which results in reduced computational cost. Computer simulations with speech signals were presented, showing the advantages of the new normalization scheme and of the subband structure with respect to processing time, adaptation convergence rate and final signal to noise interference ratio.

REFERENCES

[1] H. Buchner and 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, pp. 120–134, Jan. 2005.

[2] A. Hyvärinen and J. Karhunen and E. Oja, *Independent Component Analysis*. John Wiley & Sons, 2001.

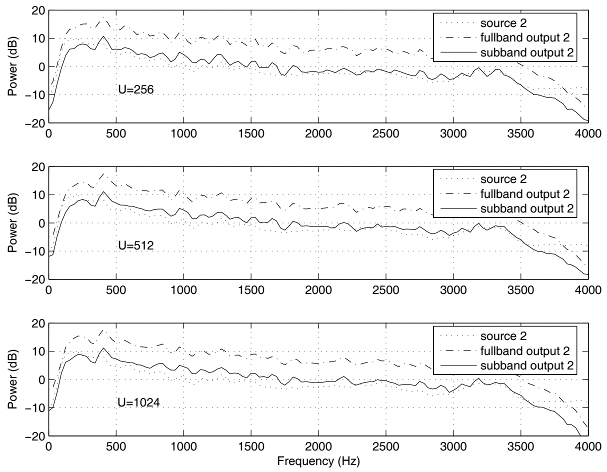

Figure 6: Spectrum of original male source (dotted lines), separated signal in fullband (dashdot lines), and subband (solid lines), with different mixing filters lengths: U=256, 512, and 1024.

Table 4: Processing time in minutes.

$S = U$	Fullband	Subband
256	57	45
512	127	83
1024	377	167

[3] H. Saruwatari and T. Kawamura and T. Nishikawa and A. Lee and K. Shikano, “Blind Source separation Based on a Fast-Convergence Algorithm Combining ICA and Beamforming,” *IEEE Trans. Audio, Speech, and Language Process.*, vol. 14, no. 2, pp. 666–678, Mar. 2006.

[4] P. Smaragdis, “Blind Separation of convolved mixtures in the Frequency Domain,” *Neurocomputing*, vol. 22, pp. 21–34, 1998.

[5] T. Nishikawa and H. Saruwatari and K. Shikano, “Comparison of Time-Domain ICA, Frequency-Domain ICA and Multistage ICA for Blind Source Separation,” in *Proc. European Signal Processing Conf.*, 2006, pp. 15–18.

[6] S. Araki and S. Makino and R.Aichner and T. Nishikawa and H. Saruwatari, “Subband-Based Blind Separation for Convolutive Mixtures of speech,” *IEICE Trans. Fundamentals*, vol. E88-A, pp. 3593–3603, Dec. 2005.

[7] R. Aichner and H. Buchner and W. Kellermann, “Exploiting Narrowband Efficiency for Broadband Convolutive Blind Source Separation,” *EURASIP Journal on Applied Signal*, vol. 2007, pp. 1–9, Sep. 2006.

[8] M. R. Petraglia and P. B. Batalheiro, “Nonuniform Subband Adaptive Filtering With Critical Sampling,” *IEEE Trans. on Signal Process.*, vol.56, no. 2, pp. 565-575, Feb. 2008.

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

[10] *http://sassec.gforge.inria.fr/*