University of Wollongong
Research Online

Faculty of Informatics - Papers (Archive)

Faculty of Engineering and Information Sciences

2004

Integration of DFT and cosine-modulated filter banks with blind separation of convolutively mixed non-stationary sources

I. Russell
University of Wollongong

J. Xi
University of Wollongong, jiangtao@uow.edu.au

A. Mertins
University of Oldenburg, Germany

J. F. Chicharo
University of Wollongong, chicharo@uow.edu.au

Publication Details

This paper originally appeared as : Russell, I, Xi, J, Mertins, A and Chicharo, JF, Integration of DFT and cosine-modulated filter banks with blind separation of convolutively mixed non-stationary sources, Sensor Array and Multichannel Signal Processing Workshop Proceedings, 18-21 July 2004, 441-445. Copyright IEEE 2004.

Research Online is the open access institutional repository for the University of Wollongong. For further information contact the UOW Library: research-pubs@uow.edu.au

Integration of DFT and cosine-modulated filter banks with blind separation of convolutively mixed non-stationary sources

Abstract

In this paper, oversampled M channel FIR filter banks using both DFT modulation and cosine modulation designs are used in conjunction with a time domain blind source separation (BSS) algorithm I. Russell et al., (2003). This BSS algorithm has been shown to blindly separate the fullband versions of non-stationary convolutively mixed sources in the time domain. However further savings on convergence and computational complexity can be made by using subband decomposition on the mixed signals before implementation of the time domain BSS algorithm in each subband. An extended lapped transform (ELT) prototype is modulated using a cosine-modulated (CM) FIR filter bank and then with a DFT modulated FIR filter bank. Both of these designs are compared to the typical frequency domain BSS approach to solving these convolutive non-stationary BSS problems such as in L. Parra et al., (2000). The signal to interference ratio (SIR) is used as the performance metric to evaluate and analyse the comparison of the three separation methods.

Keywords

FIR filters, blind source separation, channel bank filters, computational complexity, convolution, discrete Fourier transforms, integration, modulation, signal sampling, time-domain analysis

Disciplines

Physical Sciences and Mathematics

Publication Details

This paper originally appeared as : Russell, I, Xi, J, Mertins, A and Chicharo, JF, Integration of DFT and cosine-modulated filter banks with blind separation of convolutively mixed non-stationary sources, Sensor Array and Multichannel Signal Processing Workshop Proceedings, 18-21 July 2004, 441-445. Copyright IEEE 2004.

This conference paper is available at Research Online: http://ro.uow.edu.au/infopapers/150

INTEGRATION OF DFT AND COSINE-MODULATED FILTER BANKS WITH BLIND SEPARATION OF CONVOLUTIVELY MIXED NON-STATIONARY SOURCES

Iain Russell, Jiangtao Xi*, Alfred Mertins**, and Joe Chicharo**

* School of Elec., Comp., and Tele. Eng., University of Wollongong,
Wollongong, N.S.W. 2522, Australia, Email: {iainr,jiangtao,chicharo}@uow.edu.au
** Signal Processing Group, Institute of Physics, University of Oldenburg,
26111 Oldenburg, Germany, Email: alfred.mertins@uni-oldenburg.de

ABSTRACT

In this paper, oversampled $\tilde{M}$ channel FIR filter banks using both DFT modulation and cosine modulation designs are used in conjunction with a time domain blind source separation (BSS) algorithm [1]. This BSS algorithm has been shown to blindly separate the fullband versions of non-stationary convolutively mixed sources in the time domain. However further savings on convergence and computational complexity can be made by using subband decomposition on the mixed signals before implementation of the time domain BSS algorithm in each subband. An extended lapped transform (ELT) prototype is modulated using a cosine-modulated (CM) FIR filter bank and then with a DFT modulated FIR filter bank. Both of these designs are compared to the typical frequency domain BSS approach to solving these convolutive non-stationary BSS problems such as in [2]. The signal to interference ratio (SIR) is used as the performance metric to evaluate and analyse the comparison of the three separation methods.

1. INTRODUCTION

Multiple acoustic signals including speech, recorded simultaneously in a reverberant environment by multiple microphones, have a mixing system that can be modelled as a convolutive multiple-input-multiple-output (MIMO) system of FIR filters. The problem of blind source separation (BSS) is to identify the respective multiple convolutive unmixing channels using a FIR backward model, which generates the separated non-stationary acoustic sources. There are numerous techniques to achieve this including the use of higher order statistics (HOS), maximum likelihood and mutual information and an extensive review can be found on these methods in [3]. The time domain convolutive BSS algorithm in [1, 4] uses the non-stationarity property of the input sources. The process involves joint diagonalization of output correlation matrices with time varying second order statistics (SOS).

Applications in which this problem is prevalently applied include speech enhancement with multiple microphones for improved speech recognition, high-quality hearing aids, hands free telephony, EEG, MRI and other biomedical and neurological signals, cross-media retrieval in multimedia modelling [5] and multipath channel identification and equalization in wireless-communications [6].

For BSS problems that have convolutive mixing systems that model real environments using MIMO FIR filters, the number of unknown variables that must be estimated is in the order of several thousand. Traditionally these convolutive BSS models are solved by transforming to the frequency domain such as in [2]. As an alternative, we are motivated to investigate different methods of separation by including a subband preprocessor before implementing the time domain BSS algorithm given in [1, 4]. To reduce the convergence time for solving the total number of unknown parameters in the fullband model, subband decomposition is performed as a preprocessor to the time domain BSS algorithm thus solving in the subband domain as opposed to the fullband. This is implemented using oversampled uniform filter bank models satisfying perfect reconstruction (PR), including DFT and CM FIR filter banks such as in [7, 8]. These two models are then compared to the traditional frequency domain BSS method given in [2] and the separation performance for each model is measured.

In Section 2 we briefly describe the convolutive BSS model and summarise the fullband version of the time domain BSS algorithm used in [1, 4]. In Section 3 the oversampled uniform filter bank models including the CM and DFT modulated filter banks based on an ELT prototype are defined. Section 4 integrates the filter banks with the fullband time domain convolutive BSS algorithm to allow subband BSS. Additionally, a brief review of a traditional frequency domain BSS approach used in [2] is given. Section 5 gives a comparative analysis of the subband based BSS models with the traditional method with focus on the separation performance using the SIR BSS metric. The real mixing response of a typical office room is measured and identified. This identified system is mixed synthetically with segments of real speech signals taken from the TIMIT corpus of speech to produce some mixed signals. These mixed signals are used as input to each of the three convolutive BSS models and initialization of the unknown unmixing system to be identified is a perturbed version of the known unmixing system. Finally, a conclusion is provided in Section 6.

The following notations are used in this paper. We use bold upper and lowercase letters to show matrices and vectors, respectively in the time and frequency domains, e.g., $\mathbf{A}(t), \mathbf{A}(\omega)$ for matrices and $\mathbf{a}(t)$ for vectors. Matrix and vector transpose, complex conjugation, and Hermitian transpose are denoted by $(\cdot)^T, (\cdot)^*$, and $(\cdot)^H \triangleq ((\cdot)^*)^T$, respectively. $E\{\cdot\}$ means the expectation operation. $\|\cdot\|_F$ is the Frobenius norm of a matrix. $\otimes$ is the kronecker product and $Trace(\mathbf{A})$ is the Trace of matrix $\mathbf{A}$. With $\mathbf{a} = diag(\mathbf{A})$ we obtain a vector whose elements are the diagonal elements of $\mathbf{A}$ and $diag(\mathbf{a})$ is a square diagonal matrix which contains the elements of $\mathbf{a}$. $ddiag(\mathbf{A})$ is a diagonal matrix where its diagonal elements are the same as the diagonal elements of $\mathbf{A}$ and

$$off(\mathbf{A}) \triangleq \mathbf{A} - ddiag(\mathbf{A}). \quad (1)$$

$1_{N\times N}$ is an $N \times N$ matrix of ones, $0_{N\times N}$ is an $N \times N$ matrix of zeros, and $\mathbf{I}_N$ is the $N \times N$ identity matrix. $vec(\mathbf{A})$ forms a column vector by stacking the columns of $\mathbf{A}$. The operator $mat_{N,MQ}(\mathbf{a})$ reshapes a vector $\mathbf{a}$ of length NMQ to an $N \times MQ$ matrix. The matrices $\mathbf{P}_{off}$, $\mathbf{P}_{diag}$, and $\mathbf{P}_{vec}^{(N,L)}$ in Table 1 are defined as follows. $\mathbf{P}_{off}$ and $\mathbf{P}_{diag}$ are given by

$$\mathbf{P}_{off} = diag(vec(off(1_{N\times N}))), \tag{2}$$

$$\mathbf{P}_{diag} = diag(vec(\mathbf{I}_N)). \tag{3}$$

The matrix $\mathbf{P}_{vec}^{(N,L)}$ is the permutation matrix defined by

$$\mathbf{P}_{vec}^{(N,L)} vec(\mathbf{A}^T) = vec(\mathbf{A}), \tag{4}$$

for $N \times L$ matrices $\mathbf{A}$. Note that for $N \neq L$ the matrix $\mathbf{P}_{vec}^{(N,L)}$ is, in general, not self-inverse.

2. BSS CONVOLUTIVE MODEL

The convolutive BSS model assumes N statistically independent sources, $\mathbf{s}(t) = [s_1(t), ..., s_N(t)]^T$. Due to multipath propagation in a reverberant environment, these signals are convolutively mixed to provide M observed signals, $\mathbf{x}(t) = [x_1(t), ..., x_M(t)]^T$. The relationship between the source and observed signals can be written as:

$$\mathbf{x}(t) = \sum_{\tau=0}^{P-1} \mathbf{H}(\tau)\mathbf{s}(t-\tau) \tag{5}$$

where $\mathbf{H}(\tau)$ is a $M \times N$ matrix of FIR filters of length P. The BSS algorithm in [1, 4] uses a backward model for separation and so we are more interested in the demixing system which couples the M observed signals to N reconstructed signals $\hat{\mathbf{s}}(t) = [\hat{s}_1(t), ..., \hat{s}_N(t)]^T$. This relationship can be written as:

$$\hat{\mathbf{s}}(t) = \sum_{\tau=0}^{Q-1} \mathbf{W}(\tau)\mathbf{x}(t-\tau). \tag{6}$$

The demixing system is an $N \times M$ matrix $\mathbf{W}(\tau)$, with each matrix element being a FIR filter of length Q.

The fullband convolutively mixed BSS time domain algorithm based on nonstationary sources is summarized as follows. For a more detailed description of the algorithm and notation refer to [1, 4] and references therein. The objective function of the algorithm is written as:

$$\mathcal{J}_1 \triangleq \sum_{\tau=-\tau_{min}}^{\tau_{max}} \sum_{k=1}^{K} \beta_{k,\tau} \| off(\mathcal{W}\mathbf{R}_{\mathcal{X}\mathcal{X},k}(\tau)\mathcal{W}^H)\|_F^2. \tag{7}$$

Separation of the N reconstructed nonstationary signals $\hat{\mathbf{s}}(t)$ from the M convolutively mixed signals $\mathbf{x}(t)$, up to an arbitrary global permutation and scaling factor, is obtained when the scalar objective value $\mathcal{J}_1$ from Equation (7) is minimised. In Equation (7), $\beta_{k,\tau}$ is a normilization factor defined in [1, 4]. Ideally $\mathcal{J}_1$ should have a value of 0. Basically we perform joint diagonalization to minimise or effectively zero the off-diagonal elements of the correlation matrices for the recovered sources, at time frame k, over all necessary time lags τ. Each value of k represents a different time window frame where the SOS are considered stationary over that particular time frame. In adjacent non-overlapping time frames

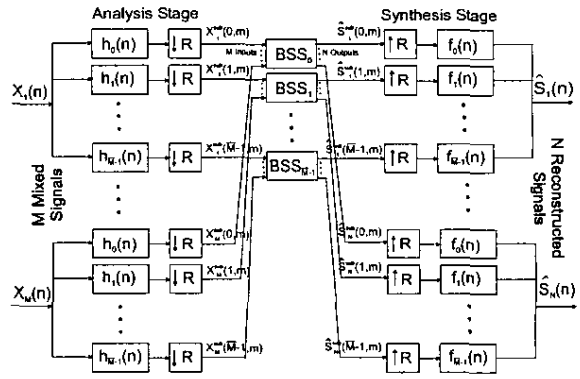

Figure 1: General subband MIMO BSS model with oversampling factor $\frac{\tilde{M}}{R}$.

k and $k+1$, the SOS are changing due to the non-stationarity assumption [4]. The correlation matrices of the recovered sources are given as:

$$\begin{aligned}\mathbf{R}_{\hat{s}\hat{s},k}(\tau) &= \mathcal{W}E\{\mathcal{X}(k)\mathcal{X}^H(k+\tau)\}\mathcal{W}^H \\ &= \mathcal{W}\mathbf{R}_{\mathcal{X}\mathcal{X},k}(\tau)\mathcal{W}^H\end{aligned}, \tag{8}$$

where $\mathcal{W}$ is a $(N \times QM)$ matrix given by $\mathcal{W} = [\mathbf{W}(0), \mathbf{W}(1), ..., \mathbf{W}(Q-1)]$, and $\mathcal{X}$ is a $(QM \times 1)$ vector used to perform convolution in the time domain using matrix multiplication and is defined in [1, 4]. $\mathcal{W}$ represents the unknown demixing system which must be solved. To avoid a trivial solution we must solve the nonlinearly constrained optimization problem given by Equation (9).

$$\begin{aligned}\mathcal{W}_{opt} &= \arg\min_{\mathcal{W}} \mathcal{J}_1(\mathcal{W}) \\ & s/t \ \ \|ddiag(\mathcal{W}\mathcal{W}^H - \mathbf{I})\|_F^2 = 0.\end{aligned} \tag{9}$$

In Table 1, closed form analytical expressions of the gradient and Hessian matrices of the objective function given in Equation (7) and the nonlinear constraint specified in Equation (9) are provided. Note the $\mathcal{J}_2$ defines the constraint given in Equation (9) and expresses the unit energy of the rows of $\mathcal{W}$. Table 2 describes the Newton method of optimization used for the BSS time domain algorithm. The steps in this algorithm will be described when integrating the subband domain in Section 4.

3. MODULATED FIR FILTERBANKS WITH PERFECT RECONSTRUCTION

To utilize BSS in the subband domain we must perform subband decomposition using some type of uniform or non-uniform FIR filter bank. As Figure 1 shows, we have chosen an oversampled uniform $\tilde{M}$ channel modulated FIR filter bank in direct form based on the modulation of an ELT prototype function $h(n)$ [9] given by:

$$h(n) = -\frac{1}{2\sqrt{2}} + \frac{1}{2}\cos[(n+\frac{1}{2})\frac{\pi}{2\tilde{M}}]. \tag{10}$$

Two designs of modulated FIR filter banks that exhibit perfect reconstruction (PR) are investigated for subband BSS and these are cosine modulated (CM) filter banks and discrete Fourier transform (DFT) modulated filter banks. Without performing any subband processing on the decomposed mixed signals $\mathbf{x}^{sub}_{\{1,\ldots,M\}}(p,m)$, where m is the time index and $p = 0,1,\ldots,\tilde{M}-1$ is the subband index, we obtain PR. Performing subband BSS using the algorithm referred to in Section 2, aliasing is introduced and so we must oversample by the factor $\frac{\tilde{M}}{R}$ to minimise this. Direct form versions of the filter banks are used here for simplicity however equivalent polyphase structures for both types filter banks can improve efficiency [7, 8].

3.1 Cosine Modulated FB

A filter bank is said to be cosine modulated if all analysis and synthesis filters are generated by cosine modulation of one or two prototype filters. The prototype lowpass filter has a cutoff of $\pm\pi/2\tilde{M}$ for $\tilde{M}$ filters. Individual analysis and synthesis filters have real coefficients and are of equal length. The impulse response of the synthesis FIR filter is defined as:

$$f_p(n) = h(n)\sqrt{\frac{2}{\tilde{M}}}\cos[(n+\frac{\tilde{M}+1}{2})(p+\frac{1}{2})\frac{\pi}{\tilde{M}}] \qquad (11)$$

and the analysis filters are related as:

$$f_p(n) = h_p(L-1-n), \qquad (12)$$

where $p = 0,1,\ldots,\tilde{M}-1$, and $n = 0,1,\ldots,L-1$. For the ELT prototype defined in Equation (10), $L = 4\tilde{M}$. Due to the oversampling factor $\frac{\tilde{M}}{R}$ to obtain PR of the filter bank a scalar of $\sqrt{\frac{R}{\tilde{M}}}$ must be multiplied with each $f_p(n)$.

3.2 DFT Modulated FB

This filter bank uses exponential modulation. The individual analysis and synthesis filters have complex coefficients in the DFT filter design. The prototype lowpass filter has a cutoff of $\pm\pi/\tilde{M}$ for $\tilde{M}$ filters. Note that we consider a $2\tilde{M}$-band DFT and $\tilde{M}$ band cosine-modulated filter bank so that the subbands are of equal spectral width in both filter bank designs. The impulse response of the synthesis FIR filter is defined as:

$$f_p(n) = h(n)e^{j(\frac{2\pi}{\tilde{M}})p(n-\frac{L-1}{2})} \qquad (13)$$

and the analysis filters are related as:

$$f_p(n) = h_p(n), \qquad (14)$$

where $p = 0,1,\ldots,\tilde{M}-1$, and $n = 0,1,\ldots,L-1$ with $L = 2\tilde{M}$. The scalar factor $\sqrt{\frac{2R}{\tilde{M}^2}}$ again must be added due to the oversampling factor.

4. SUBBAND BSS ALGORITHM

There are three stages to the model as shown in Figure 1. Firstly we decompose the M fullband mixed signals $\mathbf{x}(t)$ into $\tilde{M}$ subbands via the analysis stage of the filter bank to obtain the subband signals $\mathbf{x}^{sub}_{\{0,1,\ldots,M\}}(p,m)$ where m is the time index and $p = 0,1,\ldots,\tilde{M}-1$ is the subband index. With the cosine modulated design, the fullband mixed signals are convolved with the respective impulse responses of the analysis filters defined in Equation (12) and then oversampled by the factor $\frac{\tilde{M}}{R}$ while convolution with the impulse responses defined in Equation (13) provides the DFT modulated result for each subband also after subsampling by R. In [1] we solve a problem in the fullband domain where there exist 'NMQ' free parameters. Shorter FIR filters of length Q_p can be solved for each subband which effectively reduces the overall convergence time of the algorithm to find the unknown demixing system. Note that each subband BSS problem is a MIMO problem where there are M input signals from each respective subband of the mixed signals and N separated output signals for each respective subband. In the second stage, integration of the fullband time domain BSS algorithm given in [1, 4] is simply made by substituting the subband versions of the mixed signals $\mathbf{x}^{sub}_{\{0,1,\ldots,M\}}(p,m)$ and the unknown demixing system $\mathscr{W}^{sub}(p,m)$, for the fullband versions of the mixed signals $\mathbf{x}(t)$ and the unknown demixing system $\mathscr{W}$, and solve p separation problems where $p = 0,1,\ldots,\tilde{M}-1$. For simplicity, $\mathbf{x}^{sub}_{0,1,\ldots,M}(p,m)$ is denoted as $\mathbf{x}^p$, $\mathscr{W}^{sub}(p,m)$ is denoted as $\mathscr{W}^p$, and $\mathbf{R}^p_{\mathscr{X}\mathscr{X},k}(\tau)$ is denoted as $\mathbf{R}^{\tau,p}_{\mathscr{X}\mathscr{X},k}$. Substituting $\mathscr{W}^p, Q_p$ and $\mathbf{R}^{\tau,p}_{\mathscr{X}\mathscr{X},k}$ for $\mathscr{W}, Q$ and $\mathbf{R}^{\tau}_{\mathscr{X}\mathscr{X},k}$ in all expressions in Table 1 respectively, will provide correct subband expressions for $\mathscr{J}_1{}^p, \mathbf{G}_1{}^p, \mathbf{H}_1{}^p, \mathscr{J}_2{}^p, \mathbf{G}_2{}^p$ and $\mathbf{H}_2{}^p$ in Table 2. It should be noted that the value of Q_p will be determined by the decided value of Q in the fullband domain, the number of subbands $\tilde{M}$, the length of the analysis FIR filters L and the oversampling ratio $\frac{\tilde{M}}{R}$. The final stage of the model is the synthesis stage and involves upsampling the separated subband signals $\hat{\mathbf{s}}^{sub}_{\{0,1,\ldots,N\}}(p,m)$ by R and convolving this result with the respective impulse responses of the synthesis filters defined in Equation (11,14) for each design. This will provide the N fullband separated signals $\hat{\mathbf{s}}(t)$. To avoid the local permutation problem within the adjacent subbands, initialization using a perturbed version of the true unmixing system after subband decomposition will be used to mitigate this. Where this is unknown in a practical environment, geometric beamforming and/or other *priori* knowledge of either the mixing system or the input signals is made. Alternatively global optimization techniques can be used to solve each subband without initialization and a dyadic sorting routine used to align all subbands to the same permutation could be used to avoid this inherent problem although the use of global optimization remains an open problem. Obviously when dealing with blind separation in any transform domain we will always need to address the permutation issue. Separation algorithms which incorporate sorting routines that utilize correlation between adjacent frequency bins are available with one example given in [10]. However the focus of this paper is not on the permutation solution but rather the quality of separation between the proposed separation algorithm and a typically used frequency domain separation algorithm. For our comparison we use a perturbed system which is close to the ideal system and use the same initialization for each algorithm. Admittedly the frequency domain algorithm described below has to work harder as it solves the permutation problem but the performance of separation in Section 5 indicates the proposed algorithm in this instance obtains better separation as measured with the performance metric defined also in Section 5.

Table 1: Closed form analytical expressions for the gradient and Hessian of the cost function and constraints.

Cost function - $\mathscr{J}_1$
$\mathscr{J}_1 \triangleq \sum_{\tau=-\tau_{min}}^{\tau_{max}} \sum_{k=1}^{K} \beta_{k,\tau} \lVert off(\mathscr{W}\mathbf{R}^{\tau}_{\mathscr{X}\mathscr{X},k}\mathscr{W}^H)\rVert_F^2$
Gradient - $\mathbf{G}_1$
$\mathbf{G}_1 = 2\sum_{\tau=-\tau_{min}}^{\tau_{max}} \sum_{k=1}^{K} \beta_{k,\tau}\{off(\mathscr{W}\mathbf{R}^{\tau}_{\mathscr{X}\mathscr{X},k}\mathscr{W}^H)\mathscr{W}\mathbf{R}^{\tau}_{\mathscr{X}\mathscr{X},k}{}^H$ $+off(\mathscr{W}\mathbf{R}^{\tau}_{\mathscr{X}\mathscr{X},k}{}^H\mathscr{W}^H)\mathscr{W}\mathbf{R}^{\tau}_{\mathscr{X}\mathscr{X},k}\}$
Hessian - $\mathbf{H}_1$
$\mathbf{H}_1 = 2\sum_{\tau=-\tau_{min}}^{\tau_{max}} \sum_{k=1}^{K} \beta_{k,\tau}\{(\mathbf{R}^{\tau}_{\mathscr{X}\mathscr{X},k}{}^* \otimes off(\mathscr{W}\mathbf{R}^{\tau}_{\mathscr{X}\mathscr{X},k}\mathscr{W}^H))$ $+(\mathbf{R}^{\tau}_{\mathscr{X}\mathscr{X},k}{}^T \otimes off(\mathscr{W}\mathbf{R}^{\tau}_{\mathscr{X}\mathscr{X},k}{}^H\mathscr{W}^H))$ $+(\mathbf{R}^{\tau}_{\mathscr{X}\mathscr{X},k}{}^T\mathscr{W}^T \otimes \mathbf{I}_N)\mathbf{P}_{off}(\mathscr{W}^*\mathbf{R}^{\tau}_{\mathscr{X}\mathscr{X},k}{}^* \otimes \mathbf{I}_N)$ $+(\mathbf{R}^{\tau}_{\mathscr{X}\mathscr{X},k}{}^*\mathscr{W}^T \otimes \mathbf{I}_N)\mathbf{P}_{off}(\mathscr{W}^*\mathbf{R}^{\tau}_{\mathscr{X}\mathscr{X},k}{}^T \otimes \mathbf{I}_N)$ $+(\mathbf{R}^{\tau}_{\mathscr{X}\mathscr{X},k}\mathscr{W}^H \otimes \mathbf{I}_N)\mathbf{P}^{(N,N)}_{vec}\mathbf{P}_{off}(\mathscr{W}^*\mathbf{R}^{\tau}_{\mathscr{X}\mathscr{X},k}{}^* \otimes \mathbf{I}_N)$ $+(\mathbf{R}^{\tau}_{\mathscr{X}\mathscr{X},k}{}^H\mathscr{W}^H \otimes \mathbf{I}_N)\mathbf{P}_{off}\mathbf{P}^{(N,N)}_{vec}(\mathscr{W}^*\mathbf{R}^{\tau}_{\mathscr{X}\mathscr{X},k}{}^T \otimes \mathbf{I}_N)\}$
Row-normalized Constraint $\mathscr{J}_2$
$\mathscr{J}_2 = \lVert ddiag(\mathscr{W}\mathscr{W}^H - \mathbf{I}_N)\rVert_F^2$
Constraint Gradient $\mathbf{G}_2$
$\mathbf{G}_2 = 4ddiag(\mathscr{W}\mathscr{W}^H - \mathbf{I}_N)\mathscr{W}$
Constraint Hessian $\mathbf{H}_2$
$\mathbf{H}_2 = 4(\mathbf{I}_{MQ} \otimes ddiag(\mathscr{W}\mathscr{W}^H - \mathbf{I}_N))$ $+4(\mathscr{W}^T \otimes \mathbf{I}_N)\mathbf{P}_{diag}(\mathscr{W}^* \otimes \mathbf{I}_N)$ $+2\mathbf{P}^{(N,MQ)}_{vec}(\mathbf{I}_N \otimes \mathscr{W}^H)\mathbf{P}_{diag}(\mathscr{W}^* \otimes \mathbf{I}_N)$ $+2(\mathscr{W}^H \otimes \mathbf{I}_N)\mathbf{P}_{diag}(\mathbf{I}_N \otimes \mathscr{W}^*)(\mathbf{P}^{(N,MQ)}_{vec})^T$

To obtain the performance of the subband based BSS algorithm using the two designs described above, we compare them with a typical BSS frequency domain algorithm used in [2]. This algorithm directly estimates a stable multipath backward FIR model for a MIMO system where there is assumed to be at least as many sensors as sources present. The algorithm uses gradient descent to find the optimal value which minimises the weighted constrained Least Squares (LS) cost function given below:

$$E(\omega,k) = \mathbf{W}(\omega)\bar{R}_x(\omega,k)\mathbf{W}^H(\omega) - \Lambda_s(\omega,k) \qquad (15)$$

$$\hat{\mathbf{W}},\hat{\Lambda}_s = \underset{\substack{\mathbf{W},\Lambda_s,\\ \mathbf{W}(\tau)=0,\tau>Q\ll T,\\ W_{ii}(\omega)=1}}{\arg\min} \sum_{\omega=1}^{T}\sum_{k=1}^{K}\lVert E(\omega,k)\rVert^2 \qquad (16)$$

This algorithm essentially attempts to minimise the difference between the cross-power-spectra of the estimated recovered sources and the actual diagonalised recovered sources. It is essentially performing the same task in the frequency domain as the joint diagonalization used in our time domain BSS algorithm [1, 4]. For this algorithm only consistent permutations for all frequencies will correctly reconstruct the sources and the first constraint on filter size Q versus the frequency resolution $1/T$ given in Equation (16) links the otherwise independent frequencies and solves this problem.

Table 2: Newton-type subband BSS algorithm for the joint-diagonalization task with a weighted constraint.

For $p = 0,1,...,\tilde{M}-1$ subbands

Initialization $(r=0)$: $\mathscr{W}_0^p$

For $r = 1,2,...$

$\mathbf{w}_r^p = \mu(\mathbf{H}_1{}^p + \alpha\mathbf{H}_2^p)^{-1}\mathrm{vec}(\mathbf{G}_1{}^p + \alpha\mathbf{G}_2^p)$
$\triangle\mathscr{W}_r^p = \mathrm{mat}_{N,MQ_p}(\mathbf{w}_r^p)$
$\mathscr{W}_{r+1}^p = \mathscr{W}_r^p - \triangle\mathscr{W}_r^p$

5. SIMULATION RESULTS

In this section we report the results of separation of two mixed signals in a realistic environment such as an office room, with dimensions $2.28m \times 5.21m \times 3.45m$, using the three different models described in Sections 3 and 4. As in [4] we identify the MIMO convolutive mixing impulse responses $\mathbf{H}_{known}(\tau)$ coupling two loudspeakers and two microphones in a reverberant environment. The technique used to obtain the corresponding known demixing impulse responses for separation $\mathbf{W}_{known}(\tau)$, or equivalently $\mathscr{W}_{known}$, is described in [4]. Using $8kHz$ as the sampling rate, $\mathscr{W}_{known}$ has a FIR filter length of $Q = 2048$ for a response time of $T_R = 250ms$. The two input signals $\mathbf{s}(t)$ are speech segments taken from the TIMIT corpus of speech. These signals are convolutively mixed with $\mathbf{H}_{known}(\tau)$ and provide the mixed signals $\mathbf{x}(t)$ which are observed by the two cardioid microphones that have an inter-element spacing of $38cm$. These mixed signals are used for each particular algorithm.

For the cosine modulated FIR filter bank model we decompose the unknown fullband demixing system $\mathscr{W}$ into $\tilde{M} = 256$ subbands with a subsampling factor of $R = 64$. This will mean that instead of trying to solve the nonlinearly constrained optimization problem given in Equation (5) for $'NMQ' = 8196$ unknown variables we have in each subband only 192 variables to solve for. Similarly for the DFT modulated model we decompose the unknown fullband demixing system into $\tilde{M} = 512$ subbands with a subsampling factor of $R = 128$. This ensures that the spectral width of the analysis and synthesis filters in the DFT modulated case is the same as that for the cosine modulated case. The Newton method time domain BSS algorithm for convolutive mixtures [1, 4] was used to solve for each subband unknown demixing system. To achieve this the fullband mixed signals $\mathbf{x}(t)$ were passed through the analysis stage shown in Figure 1 for each subband model to obtain the $\mathbf{x}^{sub}_{\{0,1,...,M\}}(p,m)$ subband signals respectively. Initial values of each subband unknown demixing system were set to a perturbed version of the known demixing subband system $\mathscr{W}^{sub}_{known}(p,m)$. This is derived by adding Gaussian random variables with standard deviation $\sigma = 0.1$ to the coefficients of the known fullband demixing system $\mathscr{W}_{known}$ and passing this through the analysis stage of the filter bank to obtain the perturbed subband demixing systems. In most cases information on the demixing system is unknown and geometric beamforming [11] is used to provide initialization information for the optimization process

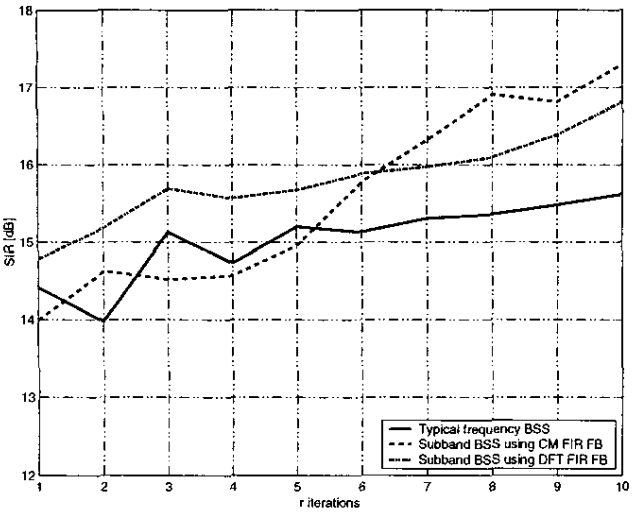

Figure 2: Separation performance using three different BSS techniques for two TIMIT speech segments recorded with two cardioid microphones in a reverberant office environment.

however in this case we are comparing the separation performance of the algorithms and initialization is the same for all three algorithms. The weighting factor for the penalty term for the constraint in the Newton update from Table 2 is set to $\alpha = 0.2$ and the learning coefficient is $\mu = 0.8$. The number of time frames over which joint diagonalization is performed is $K = 128$ which corresponds to a non-stationary time period for speech of $20-30$ ms. The typical BSS frequency domain approach also sets the required variables of the algorithm to be $Q = 2048, T = 4096, K = 128$ and an initial value for the unknown system in each frequency bin T that is derived by simply taking a T-point Fourier transform of the perturbed known fullband demixing system $\mathscr{W}_{known}$. In order to evaluate the performance of the proposed BSS methods we used the signal to interference ratio $SIR_i = SIR_{Oi} - SIR_{Ii}$, defined below as:

$$SIR_{Oi} = 10\log\frac{\sum_\omega |A_{ii}(\omega)S_i(\omega)|^2}{\sum_\omega |A_{ij}(\omega)S_j(\omega)|^2}, \qquad (17)$$

$$SIR_{Ii} = 10\log\frac{\sum_\omega |H_{ii}(\omega)S_i(\omega)|^2}{\sum_\omega |H_{ij}(\omega)S_j(\omega)|^2}, \qquad (18)$$

where $\mathbf{A}(\omega) = \mathbf{W}(\omega)\mathbf{H}(\omega)$ and $i \neq j$. SIR means the ratio of a target-originated signal to a jammer-originated signal [2]. For the subband BSS models the fullband converged solutions for $\mathscr{W}$ after the synthesis stage are then converted to the frequency domain via a T-point DFT to allow comparison with the BSS frequency domain approach using the SIR metric. Figure 2 shows the performance comparison of the two proposed methods with the typical frequency domain method. After each iteration through the algorithms we measure the SIR in decibels for each method. We only look at the first 10 iterations for the three methods. Initially we see that the subband based BSS that uses the DFT FIR filterbank has the highest SIR at 14.85 dB. After the 6th iteration the subband based BSS algorithm using the ELT prototype with CM is better with a higher SIR than the other two methods.

6. CONCLUSION

The main contributions of this paper demonstrate a general framework to approaching BSS utilizing the subband domain. Two oversampled uniform FIR filter bank designs using modulation with an ELT prototype have been used to decompose convolutively mixed non-stationary sources and perform BSS using a time domain algorithm exploiting changing SOS. In the fullband domain with a realistic mixing environment this algorithm would pose convergence problems due to the high order of variables, however using subband decomposition this problem is mitigated. The separation performance of subband BSS using the FIR filter bank designs defined in Section 3 shows better separation than the typical frequency domain method described in [2] as seen by the results in Figure 2.

REFERENCES

[1] I. Russell, J. Xi, A. Mertins, and J. Chicharo, "Blind seperation of non-stationary convolutively mixed signals in the time domain," *DSPCS03/WITSP03, Coolangatta, Gold Coast*, pp. 92–97, 8-11 December 2003.

[2] L. Parra and C. Spence, "Convolutive blind separation of non-stationary sources," *IEEE Trans., Speech and Audio Proc.*, vol. 8, pp. 320–327, May 2000.

[3] A. Cichocki and S.-I. Amari, *Adaptive Blind Signal and Image Processing*. Chichester: Wiley, 2002.

[4] I. Russell, J. Xi, A. Mertins and J.Chicharo, "Blind source separation of nonstationary convolutively mixed signals in the subband domain," *Accepted by ICASSP*, 17th-22nd May 2004.

[5] J. Larsen, L. Hansen, T. Kolenda, and F. Nielsen, "Independent Component Analysis in Multimedia Modelling," (Nara, Japan), pp. 687–696, ICA, April 1-4 2003.

[6] M. Feng and K. Kammeyer, "Blind Source separation for Communication Signals Using Antenna Arrays," (Florence, Italy), accepted by ICUPC, Oct. 1998.

[7] J. Kliewer, and A. Mertins, "Oversampled cosine-modulated filter banks with arbitrary system delay," *IEEE Trans., on Signal Processing*, vol. 46, pp. 941–955, April 1998.

[8] R. Koilpillai and P. Vaidyanathan, "Cosine-Modulated FIR Filter Banks Satisfying Perfect Reconstruction," *IEEE Trans. on Signal Processing*, vol. 40, pp. 770–783, April 1992.

[9] H. S. Malvar, *Signal Processing with Lapped Tranforms*. Norwood, MA: Artech House, 1992.

[10] K. Rahbar and J. Reilly, "A New Frequency Domain Method for Blind Source Separation of Convolutive Audio Mixtures," *Submitted to IEEE Trans. on Speech and Audio Processing*, January 2003.

[11] S. Araki, S. Makino, R. Aichner, T.Nishikawa and H. Sarawatari, "Subband based blind source separation with appropriate processing for each frequency band," *4th Int. Sym. on ICA and BSS*, pp. 499–504, April 2003.