---

**Article:**
Peng, B., Liu, W. and Mandic, D.P. (2013) Design of oversampled generalised discrete Fourier transform filter banks for application to subband-based blind source separation. IET Signal Processing, 7 (9). pp. 843-853. ISSN 1751-9675

https://doi.org/10.1049/iet-spr.2012.0361

---

# Design of Oversampled GDFT Filter Banks for Application to Subband Based Blind Source Separation

Bo Peng and Wei Liu

Communications Research Group

Department of Electronic and Electrical Engineering

University of Sheffield, UK

Danilo P. Mandic

Communications and Signal Processing Research Group

Department of Electrical and Electronic Engineering

Imperial College London, UK

**Abstract.** A novel design of oversampled generalised discrete Fourier transform (GDFT) filter banks is proposed, with application to subband-based convolutive blind source separation (BSS), where either instantaneous BSS algorithms or joint BSS algorithms can be applied. Conventional filter banks design is usually focused on elimination of the overall aliasing error and the perfect reconstruction (PR) condition, which are required by traditional subband adaptive filtering applications. However, due to the unknown scaling factor, the traditional PR condition is not necessary in the context of subband BSS and can be relaxed in the design. Due to the increased degrees of design freedom, we can introduce an additional cost function to enhance the mutual information between adjacent subband signals. Together with a reduced subband aliasing level, it leads to an improved subband permutation alignment result for instantaneous BSS and an overall better performance for the joint BSS.

## 1 Introduction

Blind source separation (BSS) has been studied extensively in the past due to its wide range of applications, and various designs have been proposed to reconstruct a set of unknown signal sources from all kinds of their mixtures [1, 2, 3, 4, 5]. There are many effective algorithms available for

the instantaneous mixing problem. However, it is a simplified model that only considers magnitude attenuations at the transmitting channel and in practice, more complicated convolutive mixing models are often used, which also consider the effect of reverberations and delays during the transmission. However, the direct time-domain extension of BSS algorithms from instantaneous mixtures to the convolutive case is difficult and computationally very expensive. To tackle the convolutive mixing problem, transformation of the received signal from the time domain to the frequency domain is performed by discrete Fourier transform (DFT), and many separating algorithms for instantaneous mixtures can then be applied directly, since convolutive mixing in the time domain corresponds to an instantaneous one in the frequency domain [6, 7, 8].

The DFT and inverse DFT pair can be considered as a special class of filter banks and we can extend the frequency-domain approach to the more general subband-based one by employing a general filter banks system [9, 10]. The typical structure of a filter banks system with $M$ channels is shown in Fig. 1, where the fullband input signal $x[n]$ is split into $M$ subbands by the analysis filters $h_1[n], ..., h_M[n]$, and then decimated by a factor of $N$ due to reduced bandwidth. For $N = M$ it is a critically sampled system and it becomes an oversampled one if $N < M$. After the required processing, such as BSS, the subband signals are then upsampled by the same factor $N$ and combined together to form the fullband output $y[n]$ by a set of synthesis filters $f_1[n], ..., f_M[n]$. Not limited by the DFT operation, now we have the freedom of designing all kinds of filter banks to meet the specific requirements. To reduce the complexity in both design and implementation, we usually choose the modulated filter banks, where only one low-pass filter has to be designed as the prototype filter, which follows the criteria for eliminating the overall aliasing component and minimizing the reconstruction error [11].

After decomposition by the analysis filters, the original problem becomes $M$ sub-problems, which can be solved individually. However, due to the blind nature of the sub-problems and the limitation of the separation algorithms, the subband signals are separated up to unknown attenuation and permutation. Without synchronization between subbands, the synthesis process will remix the components, reverse the separation process and degrade the overall performance severely [12]. To overcome the permutation problem, many methods have been proposed in the past [12, 13]. For example, for sources like speech signals, there are strong dependencies between signals from different subbands [14]. So the alignment process based on the inter-subband correlation can be applied, which further assumes that different source signals are all uncorrelated and adjacent subband components of the same source signal are highly correlated with each other [7, 15].

The correlation based alignment is a synchronizing process applied immediately after the separation for each subband. Alternatively, we can avoid the permutation problem at the beginning of the separation process, by employing joint BSS algorithms [16]. A joint BSS algorithm exploits the mutual information between multiple data sets, and assumes that the data sets are correlated. Since this assumption is often valid for adjacent subband signals, we can apply joint BSS algorithms in subbands instead and their outputs will be separated and also aligned automatically.

Both methods need strong inter-subband correlation to meet their assumptions. However, when the number of subbands increases, the cross-correlation between adjacent subbands can be very small, and filter banks designed by conventional techniques are not optimum in terms of inter-subband correlation. To tackle this problem, a cost function to maximize the cross-correlation between adjacent subbands has been proposed for cosine modulated filter banks [17]. However, since the oversampled GDFT filter banks has a better performance in suppressing the in-band aliasing error, and therefore has a higher level of correlation between adjacent subbands, we will extend the design to the oversampled GDFT filter banks in this work.

As for the scaling effect, although it can be mitigated by normalization [7], it remains ambiguous because of the unknown mixing process, i.e., at each subband, each of the separated source signals will be subject to an arbitrary scaling factor, which can be different for different subbands. In the fullband domain, it is equivalent to passing the separated signal through a filter with an arbitrary frequency response, causing distortion to it, irrespective of whether a PR (perfect reconstruction) filter banks system is employed or not, as this distortion is unknown and can not be compensated by the design of the filter banks. So in view of the overall system response, the PR or near PR condition is not a definitive requirement in the context of BSS. Instead, a relaxed condition is proposed in this paper, which will provide extra degrees of design freedom for optimizing the stopband attenuation and the additional cost function to optimize the inter-subband correlation.

There are three major contributions in this work compared to [17]. Firstly, by realizing the in-band aliasing error in oversampled GDFT filter banks is much smaller than the cosine-modulated filter banks, and therefore a better candidate for inter-subband correlation, we extend the design in [17] to oversampled GDFT filter banks. Secondly, by a detailed analysis, we will show that the PR condition required in the traditional filter banks design is not necessary and by relaxing this condition, more degrees of freedom are available to meet other necessary criteria, and improved results can be obtained. Thirdly, we have extended the application to joint BSS by realizing the joint BSS algorithms will also benefit from the increased inter-subband correlation.

This paper is structured as follows: In Section 2, the subband-based BSS structure is introduced with two representative BSS algorithms: one for the traditional instantaneous BSS problem and one for the joint BSS problem. In Section 3, the different design criteria of the prototype filter for the proposed GDFT filter banks are discussed in detail, including a new analysis about the reconstruction condition. In Section 4, design examples and simulations results are provided and finally Section 5 concludes the paper.

# 2 Subband-Based BSS

## 2.1 Convolutive mixing model

In a real world scenario, the transmitting channels cause not only magnitude attenuation to the signals, but also all kinds of reverberations and delays. The effect is usually modelled by finite impulse response (FIR) filters, which leads to the following convolutive mixing model

$$x_j[n] = \sum_{i=1}^{N_s} h_{ji}[n] * s_i[n] \qquad j = 1, ..., N_s \ , \tag{1}$$

where $h_{ji}[n]$ denotes the channel impulse response from the $i$-th source $s_i[n]$ to the $j$-th sensor and $N_s$ is the source signal number. For simplicity, we also assume that the number of mixtures is the same. For convolutive mixtures, the direct time-domain extension of BSS algorithms from instantaneous mixtures to the convolutive case is difficult and computationally expensive. To circumvent this problem, it is convenient to transform the received sensor signals into the frequency domain or different subbands, where many separating algorithms for instantaneous mixtures can be applied.

Decomposing each of the mixed signals $x_j[n]$ into subbands, we then obtain the subband BSS structure shown in Fig. 2 for the case $N_s = 2$. In this structure, each of the convolutive mixtures is passed through $M$ analysis filters, followed by a decimation operation by a factor of $N$, after which the length of the mixing filter is reduced by approximately the same factor, and (1) is converted into $M$ shorter convolution problems at different subbands [18]. If the decimation factor is sufficiently large compared to the length of the channel impulse response $h_{ji}$, (1) can be simplified into [19]

$$\mathbf{x}^{(m)}[n] = \mathbf{H}^{(m)}\mathbf{s}^{(m)}[n] \ , \tag{2}$$

where $\mathbf{x}^{(m)}[n] = \left[x_1^{(m)}[n], ..., x_{N_s}^{(m)}[n]\right]^T$ is the $m-$th subband components of the fullband mixed signals, $\mathbf{s}^{(m)}[n] = \left[s_1^{(m)}[n], ..., s_{N_s}^{(m)}[n]\right]^T$ is the $m-$th subband components of the fullband source signals, and $\mathbf{H}^{(m)}$ is the corresponding $N_s \times N_s$ instantaneous mixing matrix.

By estimating a separating matrix $\mathbf{W}^{(m)}$ at the $m-$th subband, the corresponding separated signal vector $\mathbf{u}^{(m)}[n]$ is obtained by

$$\mathbf{u}^{(m)}[n] = \mathbf{W}^{(m)}\mathbf{x}^{(m)}[n] \tag{3}$$

with $\mathbf{u}^{(m)}[n] = \left[u_1^{(m)}[n], ..., u_{N_s}^{(m)}[n]\right]^T$.

## 2.2 Subband based natural gradient algorithm

Depending on the statistics of the signals, any instantaneous BSS algorithm can be employed in the subband. As an example in this paper we use an algorithm based on minimizing the mutual information and the learning equation for the separation matrix is obtained by the natural gradient [1]. At the $p-$th iteration, the resultant separation matrix for the $m-$th subband is given by

$$\mathbf{W}_{p+1}^{(m)} = \mathbf{W}_p^{(m)} + \mu \left[\mathbf{I} - \varphi\left(\mathbf{u}^{(m)}[p]\right)\left(\mathbf{u}^{(m)}[p]\right)^T\right]\mathbf{W}_p^{(m)} \tag{4}$$

$$\mathbf{u}^{(m)}[p] = \mathbf{W}_p^{(m)}\mathbf{x}^{(m)}[p]\ , \tag{5}$$

where $\varphi\left(\mathbf{u}^{(m)}[p]\right)$ is the nonlinear function chosen based on the source signal's statistical properties. Since speech signals can be approximated by a Laplacian distribution, the following nonlinear function can be used [20]

$$\varphi(\mathbf{u}^{(m)}[p]) = \left[\frac{u_1^{(m)}[p]}{\left|u_1^{(m)}[p]\right|}, \cdots, \frac{u_{N_s}^{(m)}[p]}{\left|u_{N_s}^{(m)}[p]\right|}\right]^T\ . \tag{6}$$

After estimating the subband separation matrix, we can express its transfer function as

$$\mathbf{W}^{(m)} \cdot \mathbf{A}^{(m)} = \mathbf{P}^{(m)} \cdot \mathbf{D}^{(m)}\ , \tag{7}$$

where $\mathbf{P}^{(m)}$ is the permutation matrix which reveals the correspondence between the source and the separated signals and $\mathbf{D}^{(m)}$ is the scaling matrix that only has non-zero elements at the diagonal.

For each subband, there are uncertainties in the values of $\mathbf{P}^{(m)}$ and $\mathbf{D}^{(m)}$. The indeterminacy of $\mathbf{D}^{(m)}$ causes the local scaling problem at the $k-$th subband, and the uncertainties of the permutation across all the subbands cause the global permutation problem. To mitigate the scaling ambiguity problem, we can multiply the separated components with the inverse of the separation matrix at each subband as follows [7],

$$v_i^{(m)}[n] = \mathbf{1}^T \cdot \left(\mathbf{W}^{(m)}\right)^{-1} [\underbrace{0, \cdots, 0}_{i-1}, u_i^{(m)}[n], \underbrace{0, \cdots, 0}_{N_s - i}]^T\ , \tag{8}$$

where **1** is an all-one column vector, $u_i^{(m)}[n]$ is the $i-$th output of an instantaneous BSS at the $m-$th subband.

After applying (8), we can now focus on the permutation ambiguity between subbands. For mixtures with $N_s$ sources, there are $N_p = (N_s!)$ possible combinations between any two subbands. The problem with correlation-based permutation alignment is that when the number of subbands/frequencies increases, the mutual information between two subbands could be very small, rendering this approach less effective. To improve the result, a novel design based on cosine-modulated filter banks for maximizing adjacent subband correlation can be employed [21], which will be extended to oversampled GDFT filter banks in this work.

### 2.3 Subband based M-CCA

When we have multiple data sets available and each data set is derived from a set of its own source signals, where different sets of the source signals are related in some way, we can recover all of the source signals jointly taking into consideration the multivariate nature of the multiple data sets.

In many applications, if the source signals are coloured like speech signals, their subband components are correlated especially for those neighbouring subbands. This feature can be exploited by the multiset canonical correlation analysis (M-CCA) [16, 22], which estimates the linear relationship of data sets by maximizing their correlation [23]. It only relies on the second-order statistics of the signals and has been proved to be an efficient algorithm for separation [24, 25, 26].

After passing through the analysis bank, the subband signals will be pre-processed by a whitening operation; then the M-CCA based on maximizing the sum of squared correlation (SSQCOR) is employed. At the $k-$th stage, the criterion to recover the $k-$th source is given by [16]

$$[\mathbf{w}_k^{(0)}, \cdots, \mathbf{w}_k^{(M-1)}] = \underset{\mathbf{w}_k}{\text{argmax}} \{ \sum_{m,n=1}^{M} |\hat{r}_k^{(m,n)}|^2 \} \ , \tag{9}$$

subject to

$$\mathbf{w}_k^{(m)} \perp \left\{ \mathbf{w}_1^{(m)}, \cdots, \mathbf{w}_{k-1}^{(m)} \right\} \ , \tag{10}$$

$$\left\| \mathbf{w}_k^{(m)} \right\| = 1 \ , \quad \text{for} \quad m = 0, \cdots, M-1 \ , \tag{11}$$

where

$$\hat{r}_k^{(m,n)} = \text{corr} \left( \mathbf{w}_k^{(m)} \mathbf{x}^{(m)}, \mathbf{w}_k^{(n)} \mathbf{x}^{(n)} \right) \ . \tag{12}$$

In the context of BSS, $\mathbf{w}_k^{(m)}$ denotes the $k-$th row vector of the separation matrix applied to the $m-$th subband. The above orthogonality condition for $\mathbf{w}_k^{(m)}$ is to make sure that at the $m-$th subband, the $k-$th separated signal is not correlated with any of the $k-1$ signals separated earlier. The objective

function (9) with two constraints (10) and (11) can be solved by forming a Lagrangian function with respect to the separation matrix for each of the subbands. The optimum values of $\mathbf{w}_k$ is then obtained by setting its partial derivative function to zero, which leads to the solution to a generalised eigenvalue problem that is updated for each stage [22]. The procedure is repeated until the last signal is recovered.

Equation (12) can be further derived as

$$\begin{aligned}\hat{r}_k^{(m,n)} &= \text{corr}\left(\mathbf{w}_k^{(m)}\mathbf{A}^{(m)}\mathbf{s}^{(m)}, \mathbf{w}_k^{(n)}\mathbf{A}^{(n)}\mathbf{s}^{(n)}\right) \\ &= \text{corr}\left(\mathbf{t}_k^{(m)}\mathbf{s}^{(m)}, \mathbf{t}_k^{(n)}\mathbf{s}^{(n)}\right) = \mathbf{t}_k^{(m)}\Lambda^{(m,n)}\mathbf{t}_k^{(n)} \ , \end{aligned} \tag{13}$$

where $\Lambda^{(m,n)}$ is the correlation matrix of the source signals $\mathbf{s}^{(m)}$ and $\mathbf{s}^{(n)}$, and $\mathbf{A}^{(m)}$ is the equivalent instantaneous mixing matrix of $\mathbf{A}$ for the $m-$th subband. We use $\mathbf{T}^{(m)}$ to denote the global mixing-demixing matrix at the $m-$th subband as

$$\mathbf{T}^{(m)} = \mathbf{W}^{(m)}\mathbf{A}^{(m)} \ . \tag{14}$$

For a satisfactory separation result, the M-CCA would require $\Lambda^{(m,n)}$ having a form close to a diagonal matrix, whose diagonal entries are the correlation values between the matched sources from $s_i^{(m)}$ and $s_i^{(n)}$, $i = 1, \cdots, N_s$. For speech signals decomposed by filter banks, this assumption can be enhanced by using the prototype filter optimised for the inter-subband correlation, which will be shown in the next section.

# 3 Design of GDFT Filter Banks

## 3.1 GDFT filter banks

The analysis filters and the synthesis filters of the GDFT filter banks are derived by modulating a prototype filter $p_0[n]$,

$$h_m[n] = p_0[n] \cdot e^{j\frac{2\pi}{M}(m+m_0)(n+n_0)} \ , \tag{15}$$

$$f_m[n] = h_m^*[L_p - n] \ , \tag{16}$$

$$\text{for} \quad n = 0, \cdots, L_p - 1 \ , m = 0, \cdots, M-1 \ ,$$

where $m_0$ and $n_0$ are offsets for the frequency and time indices, respectively. When $m_0 = 0.5$ and $M$ is even, we will have a special case where the first $M/2$ subbands are all located within the frequency

range $[0, \pi]$, as shown in Fig. 3. The centre of each analysis filter is located at $(\frac{2m\pi}{M} + \frac{\pi}{M})$ and filter banks with this arrangement is often referred to as odd-stacked filter banks [27].

Because of the symmetry of the frequency responses imposed by the odd-stacked arrangement, the first and the last $M/2$ analysis filters are conjugately related [28], i.e.

$$h_m[n] = (h_{M-m}[n])^* \ . \tag{17}$$

So in case of real-valued input, only the first $\frac{(2m+1)\pi}{M}$ subbands need to be calculated. And a good choice for the time offset is $n_0 = \frac{L_p-1}{2}$, where the linear phase property can be kept for all the analysis and synthesis filters if the prototype filter has a linear phase too.

Another class of modulated filter banks is the cosine-modulated filter banks, whose coefficients are real-valued, and the decimation rate is restricted by the theory of bandpass sampling [29]. In contrast, the GDFT filter banks can choose any decimation ratio that $N \leq M$, and suffer less from aliasing errors if each subband is oversampled with $N > M$.

### 3.2 Reducing subband aliasing errors

At the $m-$th subband, the signal after decimation can be formulated by the following equation

$$X^{(m)}(z) = \frac{1}{N} F_m(z^{1/N}) X(z^{1/N}) + \frac{1}{N} \sum_{n=1}^{N-1} F_m(z^{1/N} e^{-j2\pi n/N}) X(z^{1/N} e^{-j2\pi n/N}) \ , \tag{18}$$

where $N$ is the decimation factor, $X^{(m)}(z)$ is the z-transform of the frequency decomposed signal at the $m-$th subband, and $F_m(z)$ is the z- transform of the $m-$th analysis filter. The first term at the right hand of (18) denotes the desired subband signal, and the second term denotes the sum of $(N-1)$ aliasing components, which are the frequency-shifted versions of the original subband signal after decimation.

For the oversampled GDFT filter banks, the frequency response of the prototype filter is illustrated in Fig. 4. For an $M-$channel filter banks system, the cut-off frequency has to be at least $\omega_p = \pi/M$ to cover the fullband, and the transition band is between $\frac{\pi}{M}$ and $\omega_s = \frac{\pi}{N}$. In order to minimize the overlapping of the aliasing components with the baseband signal at each subband, the stopband energy of the prototype filter has to be minimized, which can be written as

$$\begin{aligned} E_s &= \int_{\omega_s}^{\pi} \left| P_0(e^{j\omega}) \right|^2 \, d\omega \\ &= \int_{\omega_s}^{\pi} \left| \sum_{n=0}^{L_p-1} p_0[n] e^{-j\omega n} \right|^2 \, d\omega \ . \end{aligned} \tag{19}$$

### 3.3 Reconstruction condition

Based on the expression of the decimated subband signals (18), we can further derive the response for the whole subband-based BSS system, given by

$$\begin{aligned} Y_i(z) =& \frac{1}{N}\sum_{m=0}^{M-1} G_m(z)\hat{\mathbf{w}}_i^{(m)} F_m(z)\mathbf{X}(z) \\ &+\frac{1}{N}\sum_{m=0}^{M-1} G_m(z)\hat{\mathbf{w}}_i^{(m)} \times \sum_{l=1}^{N-1} F_m(ze^{\frac{-j2\pi l}{N}})\mathbf{X}(ze^{\frac{-j2\pi l}{N}}) \ , \end{aligned} \tag{20}$$

for $i = 1, ..., N_s$, where $\mathbf{X} = [X_1(z), \cdots, X_{N_s}(z)]$ is the z-transform of received signals, $Y_i(z)$ denotes the $i-$th separated signal, and $G_m(z)$ is the z-transform of the $m-$th synthesis filter. The vector $\hat{\mathbf{w}}_i^{(m)}$ is the $i-$th row of the matrix $\hat{\mathbf{W}}^{(m)}$, which is the equivalent separation matrix after the scaling normalisation and permutation alignment at the $m-$th subband. The first part on the right hand side of (20) is the transfer function between the source and the output and the second part represents the aliasing components from all the frequencies.

When the stopband energy is minimised in (19) and we adopt the oversampling structure to reduce the aliasing component, the distortion of the subband-based BSS will be governed by the first part of (20), given by

$$E_d = \frac{1}{2\pi}\int_{-\pi}^{\pi} \left| \frac{1}{N}\sum_{m=0}^{M-1} G_m(e^{j\omega})\hat{\mathbf{w}}_i^{(m)} F_m(e^{j\omega})\mathbf{X}(e^{j\omega}) - S_i(e^{j\omega}) \right|^2 d\omega \ , \tag{21}$$

where $S_i(e^{j\omega})$ is the $i-$th source signal.

In BSS, the mixing filter $\mathbf{A}^{(m)}$ is unknown, and each of the separated signals is always subject to an arbitrary filtering effect. In addition, the separation vector $\mathbf{w}_i^{(m)}$ will not always converge to the ideal coefficients, and thus the separated subband signals will retain residues from other sources. Using $X_{int}(e^{j\omega})$ to denote the interference components and the scaler $\beta_i(e^{j\omega})$ for the attenuation caused by the overall filtering effect between the $i-$th source and the $i-$th receiver at frequency $\omega$, we have

$$\hat{\mathbf{w}}_i^{(m)}\mathbf{X}(e^{j\omega}) = \beta_i(e^{j\omega})S_i(e^{j\omega}) - X_{int}(e^{j\omega}) \ . \tag{22}$$

Since the analysis and the synthesis filters are derived by the same low-pass filter, we can substitute

$\left|P_0(e^{j(\omega-\omega_k)}\right|^2 = G_k(e^{j\omega})F_k(e^{j\omega})$ and (22) into (21). Therefore,

$$\begin{aligned}
E_d &= \frac{1}{2\pi}\int_{-\pi}^{\pi}\left|\frac{1}{N}\sum_{m=0}^{M-1}\left|P_0(e^{j(\omega-\omega_k)}\right|^2\hat{\mathbf{w}}_i^{(m)}\mathbf{X}(e^{j\omega}) - S_i(e^{j\omega})\right|^2 d\omega \\
&= \frac{1}{2\pi}\int_{-\pi}^{\pi}\left|\left(\frac{1}{N}\sum_{m=0}^{M-1}\left|P_0(e^{j(\omega-\omega_k)})\right|^2 - 1\right)\left(\beta_i(e^{j\omega})S_i(e^{j\omega}) - X_{int}(e^{j\omega})\right)\right. \\
&\quad \left. -X_{int}(e^{j\omega}) - (1-\beta_i(e^{j\omega}))S_i(e^{j\omega})\right|^2 d\omega \\
&\leq \frac{1}{2\pi}\int_{-\pi}^{\pi}\left|\left(\frac{1}{N}\sum_{m=0}^{M-1}\left|P_0(e^{j(\omega-\omega_k)})\right|^2 - 1\right)\left(\beta_i(e^{j\omega})S_i(e^{j\omega}) - X_{int}(e^{j\omega})\right)\right|^2 \\
&\quad + \left|X_{int}(e^{j\omega})\right|^2 + \left|(1-\beta_i(e^{j\omega}))S_i(e^{j\omega})\right|^2 d\omega\ , \qquad (23)
\end{aligned}$$

where $H(e^{j\omega})$ is the frequency response of the prototype filter, $\omega_m = \frac{2\pi(m+1/2)}{M}$, and $\beta_i$ is an unknown scaling coefficient, determined by the mixing filters. Thus, the value of $\left|(1-\beta_i(e^{j\omega}))S_i(e^{j\omega})\right|^2$ is also unknown.

Now assume for a perfect separation, i.e., $\beta_i = 1$ and the interference component $\left|X_{int}(e^{j\omega})\right|^2$ is eliminated. Then only the first part of the final expression of (23) remains, which can be further transformed into

$$\begin{aligned}
E_{d_1} &= \frac{1}{2\pi}\int_{-\pi}^{\pi}\left|\left(\frac{1}{N}\sum_{m=0}^{M-1}\left|P_0(e^{j(\omega-\omega_m)})\right|^2 - 1\right)\left(S_i(e^{j\omega}) - X_{int}(e^{j\omega})\right)\right|^2 \\
&\leq \max_{\omega}\left|S_i(e^{j\omega}) - X_{int}(e^{j\omega})\right|^2\frac{1}{2\pi}\int_{-\pi}^{\pi}\left(\frac{1}{N}\sum_{k=0}^{M-1}\left|P_0(e^{j(\omega-\omega_m)})\right|^2 - 1\right)d\omega\ . \qquad (24)
\end{aligned}$$

It defines the upper bound of the reconstruction error, and forms the classic power complimentary condition for the prototype filter and $E_{d_1}$ can be minimised by adopting the PR condition

$$\frac{1}{N}\sum_{m=0}^{M-1}\left|P_0(e^{j(\omega-\omega_m)})\right|^2 = 1. \qquad (25)$$

However, as the separating matrix $W^{(m)}$ can only be approximated by the inverse of the mixing filter at each subband subject to an arbitrary scaling function by the BSS algorithm, the assumption of $\beta_i = 1$ and $\left|X_{int}(e^{j\omega})\right|^2 = 0$ is not practical and the PR condition is not really necessary in the context of subband-based BSS.

However, instead of removing the PR condition completely, we can adopt a relaxed condition on the passband energy of the prototype filter, given by

$$\begin{aligned}
E_p &= \frac{1}{N_p}\sum_{k=1}^{N_p}\left|\left|P_0(e^{jw_k})\right|^2 - 1\right|^2 \\
&= \frac{1}{N_p}\sum_{k=1}^{N_p}\left|\left|\sum_{n=0}^{n=L_p-1}h_0[n]e^{-j\omega_k n}\right|^2 - 1\right|^2\ , \qquad (26)
\end{aligned}$$

where $N_p$ is the number of frequency points selected, and the frequency points $[w_1, \cdots, w_{N_p}] \in (0, \frac{\pi}{M})$. During optimization, only a small value for $N_p$ is needed.

### 3.4 Inter-subband correlation

The relaxed PR condition requires much fewer number of constraints at the passband of the prototype filter. The additional design freedom provides chance to further reduce the energy at the stopband, and also more space to introduce a new optimization criterion specific for the BSS application.

As mentioned, the mutual information between subbands is important to permutation alignment and the joint BSS by M-CCA. In [21], the cost function of the inter-subband correlation is proposed, in which the correlation $\overline{r}$ over all $M$ channels is calculated by (27), (28) and (29).

$$r^{(m,m+1)} = \arg \max_{l \in [-p,\ldots,p]} \left\{ |\lambda^{(m,m+1)}(l)| \right\} , \tag{27}$$

$$\lambda^{(m,m+1)}(l) = \frac{\sum_{n=0}^{\infty} \left[ q^{(m)}[n+l] \right] \left[ q^{(m+1)}[n] \right]}{\sigma_q^{(m)} \cdot \sigma_q^{(m+1)}} , \tag{28}$$

$$\overline{r} = \frac{1}{M-1} \sum_{m=1}^{M-1} r^{(m,m+1)} . \tag{29}$$

where $p$ is a small positive integer defining the range of the time lag over which the correlation is considered, $\lambda^{(m,m+1)}(l)$ is the normalised correlation between the $m-$th and the $(m+1)-$th subbands with an offset $l$, $q^{(m)}[n+l]$ is the $m-$th channel decimated signal for a general input signal $q[n]$ at time index $n+l$, $q[n]$ is modelled as a zero-mean wide sense stationary white Gaussian signal, and $\sigma_q^{(m)}$ is the standard deviation of $q^{(m)}[n]$.

Because the magnitude of the normalised correlation is always smaller than 1, the objective function about the inter-subband correlation for minimization can be formulated as

$$\Phi_{\text{corr}} = 1 - \overline{r} . \tag{30}$$

For the proposed design of the GDFT prototype filter, the optimization of $p_0[n]$ is formulated in (31), which minimises both the stopband energy $E_s$ given in (19) and $\Phi_{\text{corr}}$, constrained by the frequency response at the passband defined in (26)

$$\min_{h[n], 0 \le n \le L_p} (1-\alpha) E_s + \alpha \cdot \Phi_{\text{corr}} \text{ subject to } \quad E_p < \epsilon_p , \tag{31}$$

where $\epsilon_p$ is a small value set to be the upper limit of the passband distortion error $E_p$ and $\alpha$ is the weighting factor between stopband attenuation $E_s$ and subband correlation $\Phi_{\text{corr}}$.

Equation (31) is similar to the design of cosine-modulated filter banks proposed in [21]. However, for the reason stated in the previous section, the original PR condition is replaced by a soft constraint on the passband response of the prototype filter. As a result, the aliasing error is expected to be reduced significantly by replacing the cosine modulation with GDFT modulation, which translates into further increased inter-subband correlation, so that an improved performance can be obtained.

Moreover, for the two components in the cost function of the proposed design, if we want to increase the level of cross-correlation, the stopband attenuation for the designed prototype filter has to be smaller, which may increase the aliasing level after downsampling and as a result reduce the cross-correlation between the adjacent subbands after downsampling. On the other hand, smaller attenuation at the stopband also undermines the assumption that after subband decomposition, the convolutive mixing problem has been transformed into an instantaneous one. One important note is that, even if we have the same PR condition, the same stopband attenuation and the same overlapped area between adjacent subbands as the existing designs, the proposed method will at least have an effect of re-distributing the correlation value among different time lags and focusing the overall correlation at a specific time lag, so that we can use the correlation at that time lag for more effective permutation alignment.

### 3.5 Discussions

One issue with the choice of the oversampled GDFT filter banks is the values of $M$ and $N$. In theory, there are mainly two factors to consider in determining the values of $M$ and $N$. First, they should be large enough to make sure that after subband decomposition, the convolutive mixing problem has been transformed into a series of instantaneous mixing problems. In this case, their values are actually determined by the complexity of the unknown fullband mixing filters in the original convolutive mixing problem. However, a large value for $M$ and $N$ increases the computational complexity of the system and reduces the data length of the decomposed subband signals, with the latter one leading to less accurate estimation of their statistics and cross-correlation, and as a result a degraded overall performance. It is extremely difficult, if not impossible, to determine their optimum values and for now they can only be chosen empirically. The same problem exists in the frequency-domain BSS method, i.e., how to choose the right length of the DFT operation.

For oversampled GDFT filter banks, another problem is the ratio between $M$ and $N$. A larger ratio $M/N$ gives more overlapped area between adjacent subbands, and leaves more degrees of freedom for cross-correlation maximization. However, this also results in higher computational complexity

| | |
|---|---|
| $M = 64$ | $N = 48$ |
| $w_s = 1.96\pi/N$ | $w_p = 1.9\pi/M$ |
| $l = 0$ or 2 | $L_p = 384$ |
| $\epsilon_p = 10^{-3}$ | $\alpha = 10^{-2}$ |
| $N_p = 4$ | |

Table 1: Parameters of the design example for the proposed filter banks.

for the same value of $M$.

# 4 Design Examples and Simulation Results

## 4.1 Design examples

Two example prototype filters are designed based on the proposed method with the design parameters listed in Table 1 and the resultant frequency response shown in Figs. 5(b) and 5(c). As for the M-CCA based joint BSS, the inter-subband correlation is calculated based on the zero lag, with $l = 0$, and for permutation alignment, lags around zeros are considered, with $l = 2$.

For comparison, the prototype filter for conventional GDFT filter banks of $M = 64$ and $N = 48$ is also designed, and the frequency response is shown in Fig. 5(a). The frequency response in Fig. 5(c) has a small ripple around the passband edge, as the PR condition is relaxed. In return, it has a wider bandwidth for signal to pass and a steeper transition band before reaching the aliasing margin at $\pi/N$. The improvement due to the new design can be evaluated by calculating the signal to aliasing ratio (SAR) [30], given by

$$\text{SAR} = \frac{\int_0^{\pi/N} \left|P_0(e^{j\omega})\right|^2 d\omega}{\int_{\pi/N}^{\pi} \left|P_0(e^{j\omega})\right|^2 d\omega} \,. \tag{32}$$

The proposed prototype filter achieves a ratio of $29.80$ dB while the conventional one is 26.99 dB.

## 4.2 Joint BSS using M-CCA

First, we consider a BSS problem with three speakers and three receivers. Nine randomly generated FIR filters are used for mixing, as shown in Fig. 6 and three recorded speech signals are used as source signals, sampled at $8$ KHz. We attempt to solve the problem by using the subband based M-CCA algorithm explained in Section 2.3.

During the experiments, the oversampled GDFT filter banks with 64 channels and a decimation ratio of 48 are employed with the optimised prototype filter. The iterative M-CCA algorithm is applied to the 64 subband signals, which uses the SSQCOR criterion to optimise the correlation between different data sets. For each subband, a $3 \times 3$ separation matrix and three separated subband signals are then obtained. Unlike the existing frequency-domain or subband-based BSS, these separation matrices are optimised jointly, and post-processing for permutation alignment is not required.

Since the mixing filters and source signals are available for evaluation, we can obtain the overall impulse response $\mathbf{T}[n]$ of the mixing-demixing system, which is shown in Fig. 7, and the outputs are given by

$$y_j[n] = \sum_{i=1}^{K=3} t_{ji}[n] \cdot s_i[n] \;\; i = 1, 2, 3 \; , \tag{33}$$

where $t_{ji}[n]$ is the $(j, i)$th entry of $\mathbf{T}[n]$ and $y_j[n]$ is the $j-$th output. As the magnitudes in $A_{11}$, $A_{22}$ and $A_{33}$ are much larger than the others, and have a similar shape as the impulse, the resultant outputs are well separated signals corresponding to the three source signals.

For comparison, we also used the filter banks whose prototype filter is designed by the conventional method [28]. We can calculate the output SIR for each output and a comparison of the results is shown in Fig. 8, where an improvement for the optimised system can be observed. Due to the increase in mutual information between subbands, the M-CCA algorithm has provided more accurate estimation of the linear relationship between subband data, and avoided the permutation misalignment problem, which occurred at the 3rd output using the conventional filter banks and adversely affected the overall separation result.

### 4.3 Natural gradient algorithm with permutation alignment

In this simulation, two sets of mixing filters are randomly generated, which are 10 taps long and 20 taps long, respectively, as shown in Figs. 9(a) and 9(b). The conventional GDFT filter banks and the proposed GDFT filter banks in Simulation I are applied for subband decomposition and same speech signals are used as the sources.

Figs. 10 and 11 show the subband SIR result when the mixing filters of length $= 10$ is used, and both the proposed and the conventional GDFT filter banks have produced good results at lower frequencies, where an SIR level around $5$ dB is achieved. However, as we can see at subband $m = 21$ of the conventional system, misalignment occurs and the subband SIRs for the three outputs are $(-20.1, 5.24, -4.72)$ dB. Comparing the separated signals with the source signals, we can obtain

the SIRs for correct permutation, which are $(18.05, 5.24, 2.07)$ dB. When the subband SIRs are at a low level, i.e., $2.07$ dB, a misalignment occurs because of the presence of interference signals. Similarly for $m = 22$, the subband SIRs are $(-23.09. - 7.07, -20.20)$ dB and the fixed SIRs are $(21.199.401.42)$ dB, which caused a second misalignment because of the same reason.

In contrast, when the proposed design is employed, the subband SIR has changed to $(12.75, 7.63, 4.2)$ dB and $(17.07, 6.69, 2.84)$ dB; however, correct alignment is obtained as the proposed design can enhance the inter-subband correlation between matched subband signals.

In Figs. 12 and 13, mixing filters of length $= 20$ are used for the convolutive mixtures. As the mixing filters become longer with a more complicated frequency response, the separation is expected to become more difficult as the number of estimated coefficients has been increased. This change can be observed from the two figures, as the subband separation performance becomes worse and more misalignments are presented for both GDFT filter banks. For the conventional design, the misalignment appeared at $(m = 5, 9, 14, 15, 23, 40, 48, 49, 54, 58)$, and permutation errors propagate between these subbands, which has severely distorted the separation results. However, as the proposed design is more robust to the reduction of subband SIR, misalignment has only occurred at subbands with lowest SIR (at $m = 15, 23, 40, 48$). Since the first misalignment occurred at a higher subband, and the energy of speech signals is normally focused on lower frequencies, the impact from the permutation errors is less significant.

The fullband overall SIR values can be obtained by passing the subband components $t_{ii}^{(m)} \cdot s_i^{(m)}$ and $t_{ij}^{(m)} \cdot s_j^{(m)}$ through the synthesis filters and calculating the ratio after the summation. Table 2 summarises the results of Simulation I and Table 2 summarises the results of Simulation II.

The number of misalignments and errors is obtained for each simulation. Cosine-modulation without correlation optimization produced the worst results, and by introducing correlation optimization, the number of permutation errors is largely reduced. The GDFT filter banks have much better separation results, which reveals that the subband aliasing components have a significant negative impact on the performance of the subband-based BSS.

The proposed and the conventional GDFT filter banks have similar results when the subband SIRs are high, where permutation alignment can be correctly achieved. However, when separation difficulty increases, which is usually because of the changes in the source signals or the mixing filters, permutation misalignment may occur. For the proposed design, it is quite robust and correct alignment can still be obtained even when the subband SIR is relatively low.

The permutation error is a good performance indicator of the subband-based BSS, and when the subband system has zero permutation error, the BSS algorithm can reach its full potential; while the proposed prototype filter is designed to improve the alignment result, misalignment may still occur at a few subbands. Although occurrence of the misalignment is hardly predicted and a few misalignment may propagate to a larger number of subbands, the proposed design can generally improve the overall separation result, and if the misalignment occurs at higher frequencies, the majority of the source speech signals can still be separated.

Table 2: Results of simulation I: averaged fullband SIR values for each output, the number of permutation errors and permutation misalignments.

| | SIR1 (dB) | SIR2 (dB) | permutation error | misalignment |
|---|---|---|---|---|
| conventional CMFB | 2.1 | -1.9 | 26/64 | 17/64 |
| correlation maximized CMFB | 12.19 | 6.1 | 0 | 0 |
| conventional GDFT | 14.26 | 10.5 | 20/64 | 2/64 |
| proposed GDFT | 20.18 | 10.8 | 0 | 0 |

Table 3: Results of simulation II: averaged fullband SIR values for each output, the number of permutation errors and permutation misalignments.

| | SIR1 (dB) | SIR2 (dB) | SIR3 (dB) | perm error | misalignment |
|---|---|---|---|---|---|
| mixing tap = 10 | | | | | |
| conventional GDFT | 12.55 | 15.1 | 17.9 | 24/64 | 4/64 |
| proposed GDFT | 20.4 | 16.8 | 17.24 | 0 | 0 |
| mixing tap = 20 | | | | | |
| conventional GDFT | 2.98 | 1.99 | 13.22 | 20/64 | 10/64 |
| proposed GDFT | 12.56 | 6.33 | 13.53 | 16/64 | 2/64 |

# 5 Conclusions

In this paper, an oversampled GDFT filter banks design with correlation optimization has been proposed. By relaxing the traditional PR condition, we can focus on stopband energy minimization and inter-subband correlation optimization. Meanwhile, subband based separation using the M-CCA algorithm and the natural gradient algorithm has been studied, which can be used to solve the convo-

lutive mixing BSS problem. As both methods rely on good inter-subband correlation, an improved performance has been achieved by the proposed design for both BSS algorithms, as demonstrated by our simulations.

## References


[1] A. Cichocki and S. Amari, *Adaptive Blind Signal and Image Processing*. New York: John Wiley & Sons, Inc., 2003.

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

[3] B. Gao, W. L. Woo, and S. S. Dlay, "Single channel blind source separation using emd-subband variable regularized sparse features," *IEEE Transactions on Acoustics, Speech, and Language Processing*, vol. 19, no. 4, pp. 961–976, May 2011.

[4] Y. N. Guo, S. H. Huang, and Y. T. Li, "Single-mixture source separation using dimensionality reduction of ensemble empirical mode decomposition and independent component analysis,," *Circuits, Systems, and Signal Processing*, vol. 31, pp. 2047–2060, December 2012.

[5] B. Gao, W. L. Woo, and S. S. Dlay, "Unsupervised single channel separation of non-stationary signals using gammatone filterbank and Itakura-Saito nonnegative matrix two-dimensional factorizations," *IEEE Transactions on Circuits & Systems I: Regular Papers*, vol. 60, no. 3, pp. 662–675, 2013.

[6] P. Smaragdis, "Blind separation of convolved mixtures in the frequency domain," *Neurocomputing*, vol. 22, no. 1-3, pp. 21–34, 1998.

[7] N. Murata, S. Ikeda, and A. Ziehe, "An approach to blind source separation based on temporal structure of speech signals," *Neurocomputing*, vol. 41, no. 1–4, pp. 1–24, Oct. 2001.

[8] S. Araki, R. Mukai, S. Makino, T. Nishikawa, and H. Saruwatari, "The fundamental limitation of frequency domain blind source separation for convolutive mixtures of speech," *IEEE Transactions on Speech and Audio Processing*, vol. 11, no. 2, pp. 109 – 116, Mar. 2003.

[9] S. Y. Low, S. Nordholm, and R. Togneri, "Convolutive blind signal separation with post-processing," *IEEE Transactions on Speech and Audio Processing*, vol. 12, no. 5, pp. 539–548, 2004.

[10] N. Grbic, X.-J. Tao, S. Nordholm, and I. Claesson, “Blind signal separation using overcomplete subband representation,” *IEEE Transactions on Speech and Audio Processing*, vol. 9, no. 5, pp. 524 –533, July 2001.

[11] S. Weiss and R. W. Stewart, *On Adaptive Filtering in Oversampled Subbands*. Aachen, Germany: Shaker Verlag, 1998.

[12] M. Ikram and D. Morgan, “Permutation inconsistency in blind speech separation: investigation and solutions,” *IEEE Transactions on Speech and Audio Processing*, vol. 13, no. 1, pp. 1 – 13, Jan. 2005.

[13] H. Sawada, R. Mukai, S. Araki, and S. Makino, “A robust and precise method for solving the permutation problem of frequency-domain blind source separation,” *IEEE Transactions on Speech and Audio Processing*, vol. 12, no. 5, pp. 530 – 538, Sept. 2004.

[14] T. Kim, H. T. Attias, S. Y. Lee, and T. W. Lee, “Blind source separation exploiting higher-order frequency dependencies,” *IEEE Transactions on Acoustics, Speech, and Language Processing*, vol. 15, no. 1, pp. 70 –79, Jan. 2007.

[15] H. M. Park, C. S. Dhir, S. H. Oh, and S. Y. Lee, “A filter bank approach to independent component analysis for convolved mixtures,” *Neurocomputing*, vol. 69, no. 16-18, pp. 2065–2077, Oct. 2006.

[16] Y. O. Li, T. Adalı, W. Wang, and V. D. Calhoun, “Joint blind source separation by multi-set canonical correlation analysis,” *IEEE Transactions on Signal Processing*, vol. 57, no. 10, pp. 3918–3929, October 2009.

[17] B. Peng, W. Liu, and D. Mandic, “Reducing permutation error in subband-based convolutive blind separation,” *IET Signal Processing*, vol. 6, no. 1, pp. 34 –44, Feb. 2012.

[18] P. P. Vaidyanathan, *Multirate Systems and Filter Banks*. Englewood Cliffs: Prentice Hall, 1993.

[19] S. Y. Low, S. Nordholm, and R. Togneri, “Convolutive blind signal separation with post-processing,” *IEEE Transactions on Speech and Audio Processing*, vol. 12, no. 5, pp. 539–548, Sept. 2004.

[20] H. Sawada, R. Mukai, S. Araki, and S. Makino, "Polar coordinate based nonlinear function for frequency-domain blind source separation," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 1, May 2002, pp. 1001–1004.

[21] B. Peng, W. Liu, and D. P. Mandic, "An improved solution to the subband blind source separation permutation problem based on optimized filter banks," in *Proc. International Symposium onCommunications, Control and Signal Processing*, March 2010, pp. 1–4.

[22] J. R. Kettenring, "Canonical analysis of several sets of variables," *Biometrika*, vol. 58, no. 3, pp. 433–451, 1971.

[23] H. Hotelling, "Relations between two sets of variates," *Biometrika*, vol. 28, no. 3-4, pp. 321–377, 1936.

[24] W. Liu, D. P. Mandic, and A. Cichocki, "Analysis and online realization of the CCA approach for blind source separation," *IEEE Transactions on Neural Networks*, vol. 18, no. 5, pp. 1505 –1510, Sept. 2007.

[25] ——, "Blind source separation based on generalised canonical correlation analysis and its adaptive realization," in *Proc. International Congress on Image and Signal Processing*, vol. 5, Hainan, China, May 2008, pp. 417–421.

[26] ——, "A dual-linear predictor approach to blind source extraction for noisy mixtures," in *Proc. IEEE Workshop on Sensor Array and Multichannel Signal Processing*, Darmstadt, Germany, July 2008, pp. 515–519.

[27] R. E. Crochiere and L. R. Rabiner, *Multirate Digital Signal Processing*. Englewood Cliffs, NJ: Prentice Hall, 1983.

[28] M. Harteneck, S. Weiss, and R. W. Stewart, "Design of near perfect reconstruction oversampled filter banks for subband adaptive filters," *IEEE Transactions on Circuits and Systems — II: Analog and Digital Signal Processing*, vol. 46, pp. 1081–1085, Aug. 1999.

[29] R. G. Vaughan, N. L. Scott, and D. R. White, "The Theory of Bandpass Sampling," *IEEE Transactions on Signal Processing*, vol. 39, no. 9, pp. 1973–1984, Sept. 1991.

[30] S. Weiss, R. W. Stewart, A. Stenger, and R. Rabenstein, "Performance limitations of subband adaptive filters," in *Proc. European Signal Processing Conference*, vol. III, 1998, pp. 1245–1248.

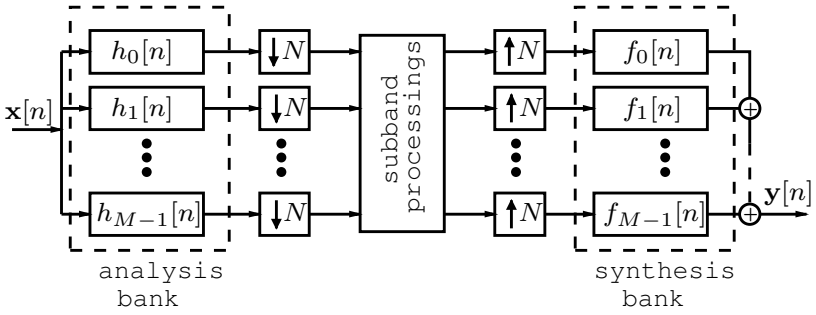



Figure 1: A general M-channel filter banks structure with a decimation factor N.

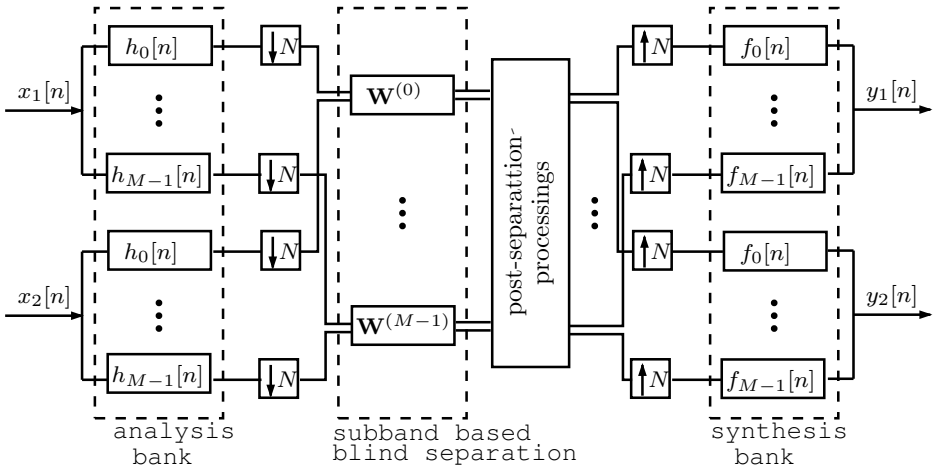



Figure 2: Structure of the subband-based BSS with two mixtures $(N_s = 2)$.

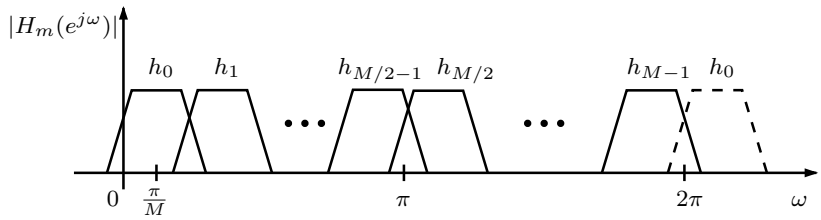



Figure 3: The arrangement of the analysis filters for $M-$channel generalised DFT filter banks.

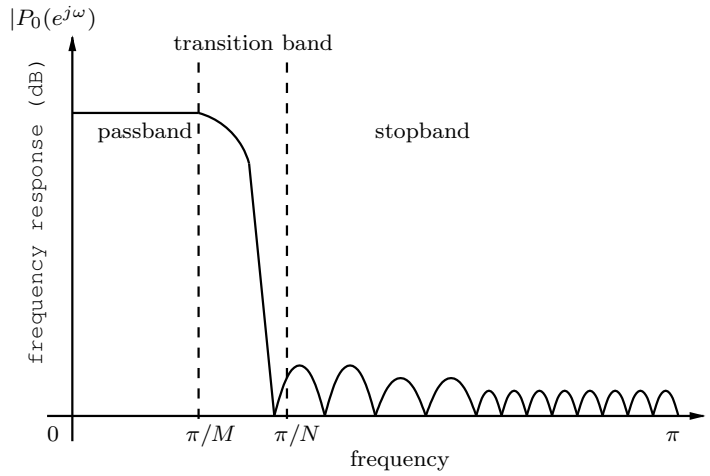
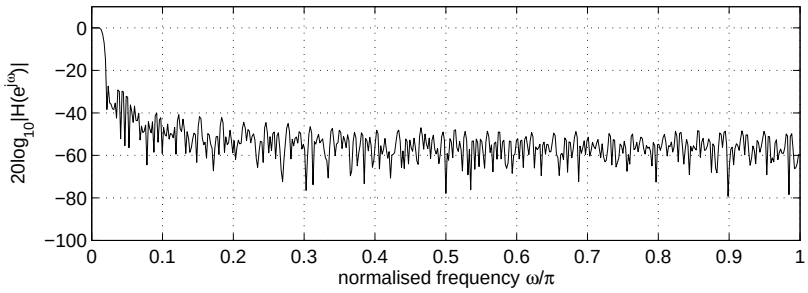
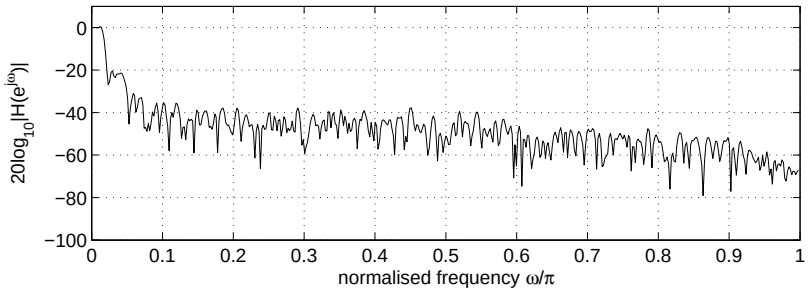


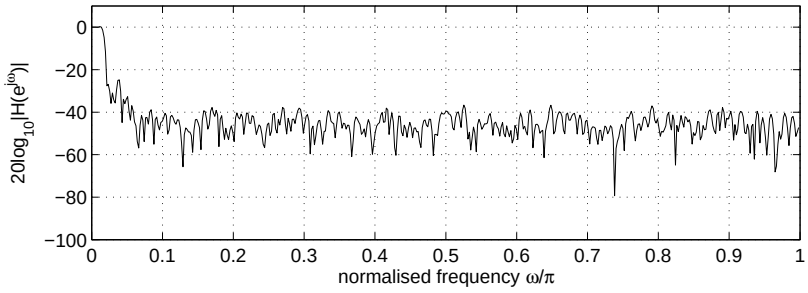

Figure 4: Required frequency response for a real-valued prototype filter $p_0[n]$ for $M-$channel oversampled GDFT filter banks with a decimation ratio $N$.




(a) The prototype filter designed by the conventional method.




(b) The prototype filter designed by the proposed method for the M-CCA algorithm with $l = 0$.




(c) The prototype filter designed by the proposed method for permutation alignment with $l = 2$.

Figure 5: Frequency response of three prototype filters of a 64-channel GDFT filter banks system, designed by the conventional method and the proposed method, respectively.

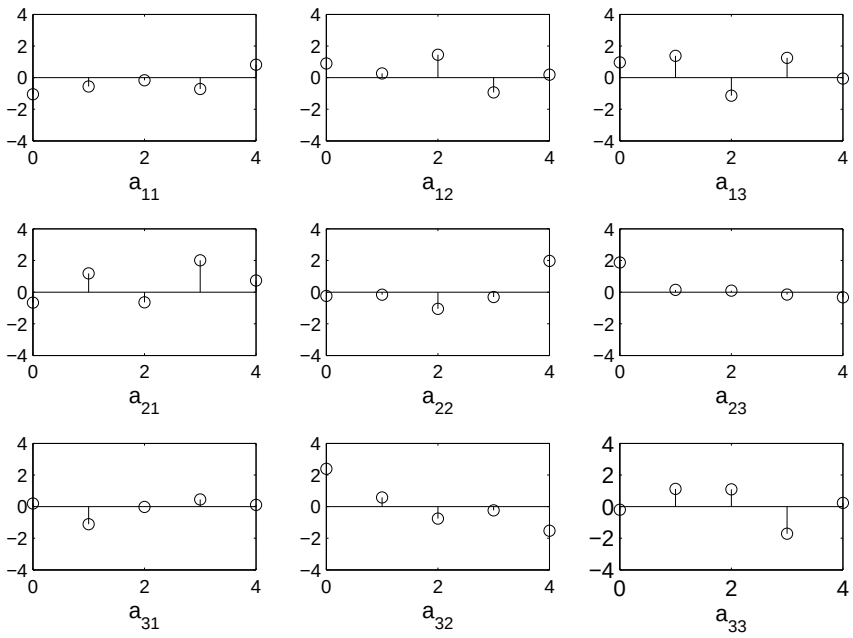



Figure 6: The 5-tap randomly generated mixing filters.

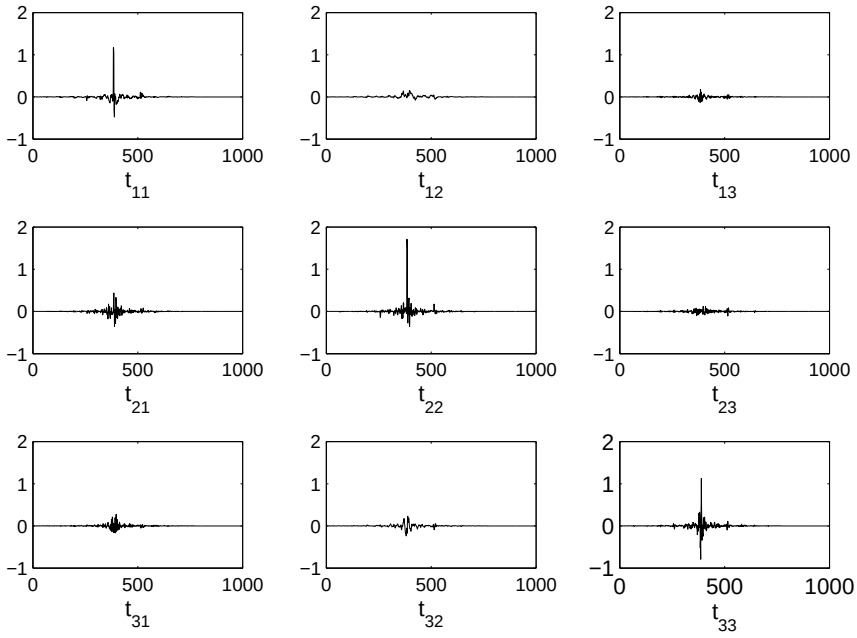



Figure 7: The overall impulse responses of the subband M-CCA system using the correlation optimised filter banks.

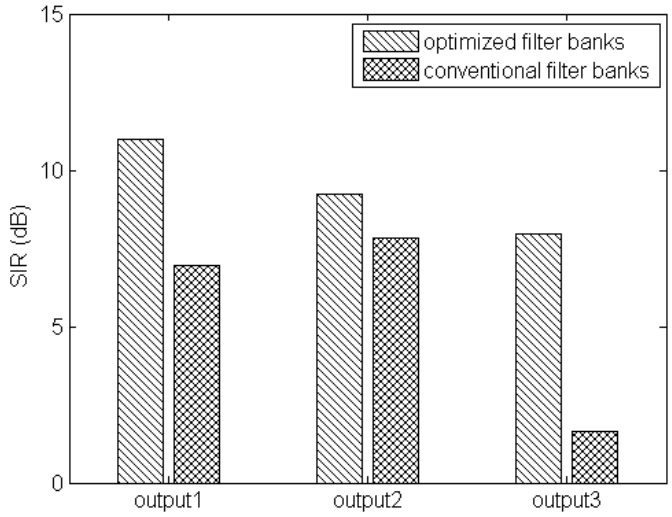
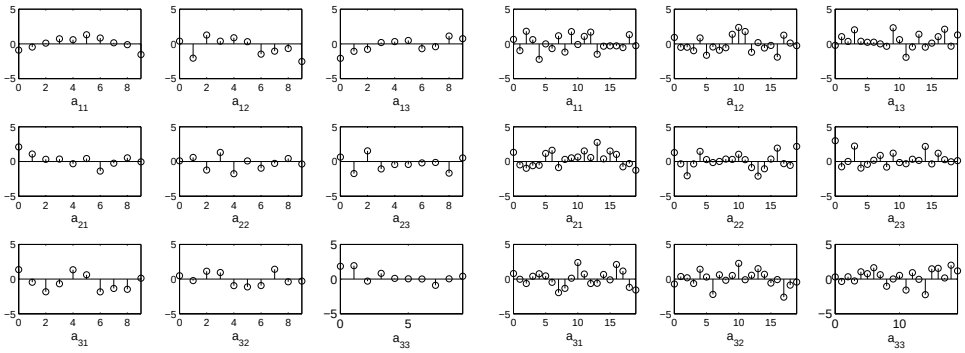



Figure 8: The output SIR for the 64-channel subband M-CCA, with comparison between correlation-optimised filter banks and the conventional ones.




(a) (b)

Figure 9: Impulse responses of the mixing filters for a three-speaker-three-receiver system. (a) mixing filter is 10 tap long (b)mixing filter is 20 tap long

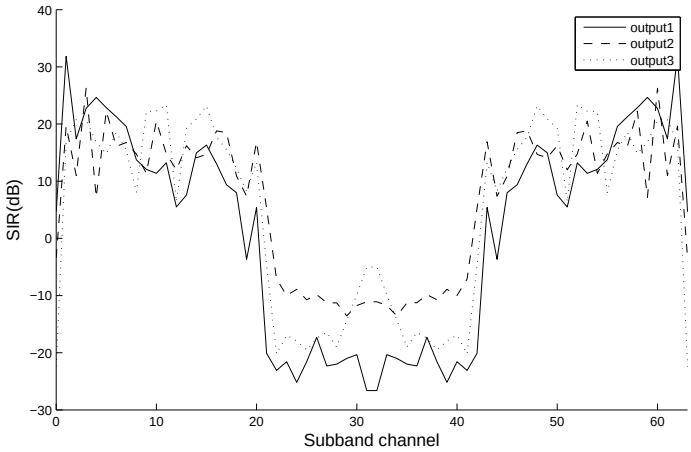



Figure 10: The subband SIR for the three outputs using the conventional oversampled GDFT modulated filter banks for mixing filters of length $= 10$.

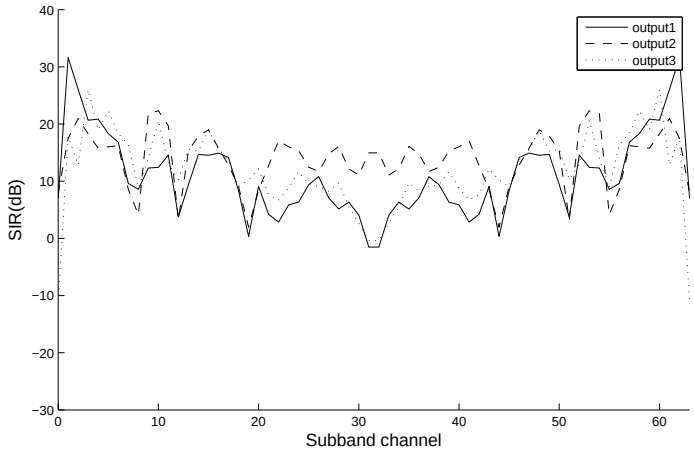



Figure 11: The subband SIR for the three outputs using the proposed oversampled GDFT modulated filter banks for mixing filters of length $= 10$.

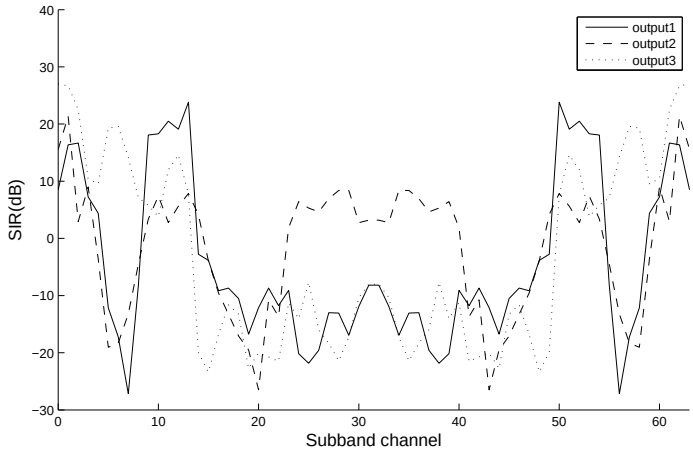



Figure 12: The subband SIR for the three outputs using the conventional oversampled GDFT filter banks for mixing filters of length $= 20$.

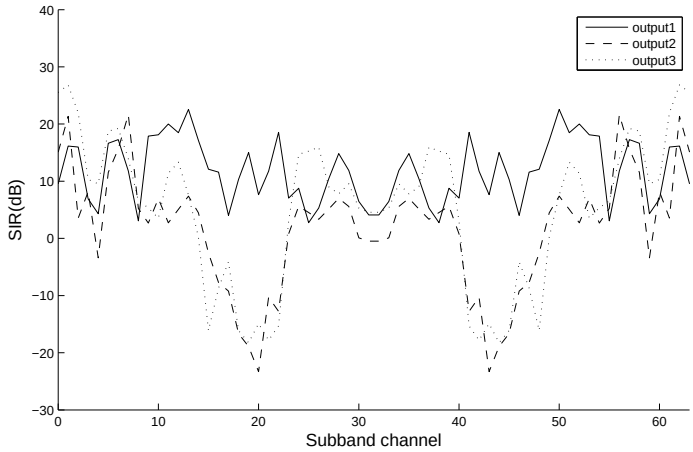



Figure 13: The subband SIR for the three outputs using the proposed oversampled GDFT filter banks for mixing filters of length $= 20$.