Microsoft Research Technical Report
MSR-TR2005-116
September 2005

The Design of Nonuniform Lapped Transforms

Byung-Jun Yoon and *Henrique S. Malvar*
Microsoft Research

Abstract

We propose a new method for designing nonuniform lapped transforms. The proposed method provides a simple yet effective way of designing lapped transforms with nonuniform frequency resolution and good time localization. This method is a generalization of an approach proposed by Princen, where the nonuniform filter bank is obtained by joining uniform cosine-modulated filter banks using transition filters. In the proposed approach, we use several transition filters to obtain a near perfect-reconstruction nonuniform lapped transform with significantly reduced overall distortion. This method has several advantages over existing ones, especially in producing nonuniform filter banks with good stopband attenuation and short filters; the resulting reduced delay makes the proposed method preferable for applications such as real-time audio coding.

Microsoft Research
One Microsoft Way
Redmond, WA 98052
http://research.microsoft.com

1 Introduction

Lapped transforms such as the LOT (lapped orthogonal transform) and the MLT (modulated lapped transform) have been widely used in various applications such as image processing and audio coding [2]. For example, the MLT is a special kind of cosine-modulated filter bank (CMFB) that has been very popular in audio processing due to its computational efficiency and good frequency discrimination. For this reason, most modern audio coders such as the MPEG-2 Layer III (MP3) [3, 4], Dolby AC-3, and MPEG-4 AAC [5] are based on MLT and its variants.

Although MLT has many advantages, there are cases where a different time-frequency resolution is more desirable. One such example is audio coding at low bit rates. When there are high-frequency transients in the input signal, the poor time resolution of the basis functions give rise to the so-called pre-echo. In order to alleviate this problem, many modern audio codecs adopt the window switching scheme which uses a shorter window when the high-frequency transient sound has significant amount of energy, hence better time resolution is desired. An alternative approach would be to use a nonuniform filter bank or a lapped transform with a nonuniform subband decomposition. The shorter impulse responses of the high-frequency subbands reduce the pre-echo effect [7], and there's no need to use look ahead to determine window switching points, thus reducing processing delay.

Until now, various methods have been proposed for designing nonuniform filter banks [1]–[17]. A number of schemes are based on the subband merging approach [7]-[15], while others are based on a tree-structured design [2], [6] or constructed from joining uniform filter banks by a transition filter [1], and so forth. In this report, we are especially interested in designing a nonuniform lapped transform that has a large number of subbands, low system delay and good time localization. The existing design methods have their own advantages, but none of them are suitable for designing such a transform. In the following sections, we review several existing design schemes and propose a new method that can be used to construct a lapped transform with the aforementioned properties.

The report is organized as follows. In Section 2, we briefly investigate various methods for designing nonuniform filter banks that have been proposed so far. In Section 3, we consider possible extensions of the existing methods. Section 4 proposes a new method for designing nonuniform lapped transforms and the report is concluded in Section 5.

2 Review of Existing Methods

In this section we present a brief overview of the existing techniques for the design of nonuniform filter banks based on CMFBs.

2.1 Malvar (NMLBT)

The nonuniform modulated lapped biorthogonal transform (NMLBT) combines the high-frequency basis functions of the MLT (or MLBT) using $+1/-1$ butterflies [7]. It provides a simple way of obtaining a lapped transform with nonuniform bands. The resulting basis functions are more localized in the time-domain and the PR property is preserved after the combination. However, the time-domain separation of the combined basis functions are less than ideal, and there still remains significant time-domain aliasing.

2.2 Niamut et al.

The sub-band merging approach proposed by Niamut et al. [10] can be viewed as a generalization of the idea behind the NMLBT [7]. The proposed method allows the combination of arbitrary number of subbbands using a principal submatrix of a Hadamard matrix. It is claimed that the proposed method is the optimal way for obtaining a nonuniform filter bank with a good frequency selectivity and a flat passband response. As the NMLBT, this method preserves the PR property of the original filter bank. However, it suffers from the same problem as the NMLBT as the resulting basis functions have significant amount of time-domain aliasing terms.

2.3 Li et al.

In [15] Li et al. proposed a method for designing a filter bank with nonuniform decimation ratios. Firstly, a uniform CMFB with a larger number of subbands is designed, and then a set of analysis filters (and the corresponding synthesis filters) are combined to obtain the nonuniform filter bank with the target decimation ratios. Li et al. provides a simple method for designing a nonuniform filter bank, since there are a number of efficient methods for designing good CMFBs [6], [18], [19]. However, this method has the disadvantage that the PR property is not retained after combining the filters. Therefore it is suitable for designing nonuniform near-PR filter banks only when the filters have high stopband attenuation such that the increase in the distortion level due to the combination of filters is negligible.

2.4 Argenti et al.

Unlike the previous approaches, Argenti et al. [16] designs the the NUFB using multiple prototype filters. Initially, a prototype filter with the narrowest passband is designed with the specified band-edge, where the filter is constrained to satisfy power complementarity in the transition band. This prototype filter is cosine-modulated to obtain a uniform section of subbands in the specified frequency region. The prototype filter with a broader passband (corresponding to a smaller decimation ratio)is derived from the previous prototype filter. This is repeated until the complete nonuniform filter bank is obtained. Although this approach provides an interesting way of designing multiple prototypes for a NUFB, they assume that the length of all the prototype filters are identical. Therefore, although we have nonuniform frequency resolution, the time resolution remains uniform, which is certainly undesirable. Another disadvantage is that the subsequent prototype filters are obtained by frequency sampling of the desired frequency response and taking the IDFT of it, and it is not obvious whether the resulting filters will always have good characteristics (unless a large number of samples are used.)

2.5 Purat et al.

In order to obtain a frequency-varying MLT, Purat and Noll combine subbands of a MLT using another MLT of a smaller size [9]. A number of bands of the MLT analysis bank is combined using an IMLT of a smaller size. Similarly, the transform coefficients obtained from the combined subbands is fed into the smaller MLT where the subbands are again connected to the IMLT of the larger MLT. As the other subband-merging schemes [7], [10], the basis functions resulting from this approach are also periodically time-varying and they are not well-localized in the time-domain.

2.6 Nayebi et al.

A general procedure for designing a nonuniform filter bank is elaborated in [17]. The reconstruction error is defined in terms of the time-domain expression of the filters and iterative optimization routines are used to minimize the reconstruction error. The main focus of the paper is on a general theory for designing NUFBs, and there is no consideration for a efficient modulated structure. Therefore, it cannot be directly used for designing NUFBs with large number of subbands, although the proposed theory can be useful in formulating objective functions and constraints for optimizing NUFBs based on various frameworks.

2.7 Chan et al.

The design method proposed by Chan et al. is also based on the subband merging approach [12], [13], [14]. Subbands in the larger CMFB are combined using a trans-multiplexer that has a smaller number of subbands than the CMFB. The stopband edges of the filters in the CMFB and the trans-multiplexer (T-MUX) are adjusted in order to reduce artifacts that can arise from the mismatch in the transition bands of those filters [12]. By careful designing the T-MUX, the filters in the merged bands can have high stopband attenuation. The disadvantage of this method is that the T-MUX introduces additional system delay unlike the simpler methods [7], [10], [15]. Moreover, the filters in the combined bands will have a longer response although the passband has become wider. As we expect the basis functions in the wider subbands to be more localized in the time-domain, this is certainly undesirable.

2.8 Princen

In [1], Princen proposed an interesting method for designing a nonuniform filter bank, which joins uniform CMFBs using a transition filter. The transition filter is designed to match the transition bands of the adjacent filters that belong to different CMFBs with unequal decimation ratios. The main aliasing components between adjacent bands are cancelled by the use of an optimized transition filter. In general, the transition filters are much longer than the filters in the adjacent uniform sections. For example, when joining two CMFBs with decimation ratios $M_1 = 8$ and $M_2 = 4$, the length of the prototype filters were 40 and 88, where the length of the transition filter was 96.

3 Several Design Approaches

In this section, we consider extensions and generalizations of the previous methods in a MLT-based framework.

3.1 MLBT

The prototype filter in the MLT is defined as

$$h(n) = -\sin\left[\left(n + \frac{1}{2}\right)\frac{\pi}{2M}\right] \qquad (n = 0, 1, \ldots, 2M-1), \tag{1}$$

where M is the number of subbands. By using a different prototype filter, we can reduce the time-domain aliasing of the combined filters that are obtained by applying the subband merging

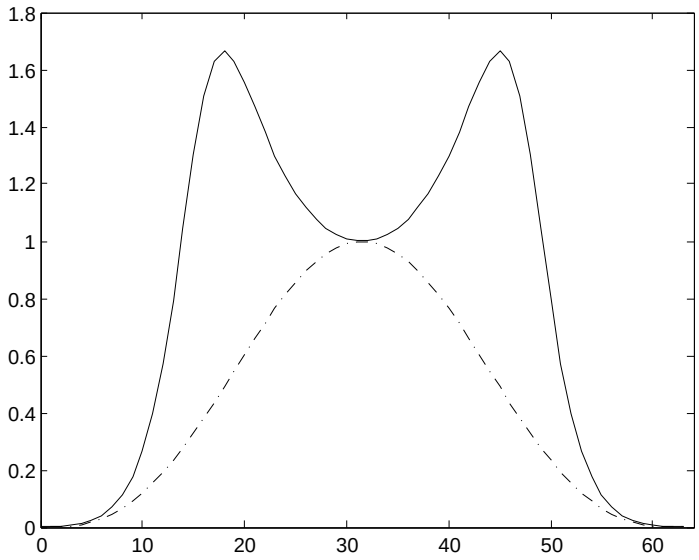

Figure 1: The synthesis window $h_s(n) = \sin^3\left[(n+1/2)\,\pi/2M\right]$ (dashed line) and the corresponding analysis window $h_a(n)$ (solid line).

scheme in [7] and [10]. For example, we may use the following windows.

$$h_s(n) = \sin^2\left[\left(n+\frac{1}{2}\right)\frac{\pi}{2M}\right] \tag{2}$$

$$h_s(n) = \sin^3\left[\left(n+\frac{1}{2}\right)\frac{\pi}{2M}\right] \tag{3}$$

Using these windows in the synthesis bank reduces the time-domain aliasing significantly. However, the corresponding analysis window, which is defined as

$$h_a(n) = \frac{h_s(n)}{h_s^2(n) + h_s^2(M-1-n)} \tag{4}$$

to ensure PR [2], will have undesirably large ripples in the passband as shown in Fig. 1.

3.2 Optimizing the Window

Instead of using biorthogonal windows, we may consider optimizing the orthogonal window function $h(n)$. The goal is to find the optimal window that results in better time-domain behavior, when the subband merging technique is applied [7], [10]. To be more precise, we want to make the following functions

$$h_1(n) = h_k(n) + h_{k+1}(n) \tag{5}$$

$$h_2(n) = h_k(n) - h_{k+1}(n) \tag{6}$$

more localized in the time-domain, where $h_k(n)$ is defined as

$$h_k(n) = h(n)\sqrt{\frac{2}{M}}\cos\left[\left(n+\frac{M+1}{2}\right)\left(k+\frac{1}{2}\right)\frac{\pi}{M}\right]. \tag{7}$$

However, it turns out that optimizing a single window function $h(n)$ is not very helpful in reducing the time-domain aliasing of $h_1(n)$ and $h_2(n)$ at the same time.

As using a single optimized window function is not very helpful in improving the time localization of the combined basis functions, one natural extension would be to use different optimized windows in adjacent bands. Let us consider using the cosine-modulated filters of $h_o(n)$ in the odd-numbered subbands and the cosine-modulated filters of $h_e(n)$ in the even-numbered subbands. So, we have

$$h_k(n) = h_o(n)\cos\left[\left(n+\frac{M+1}{2}\right)\left(k+\frac{1}{2}\right)\frac{\pi}{M}\right] \qquad \text{(odd } k\text{)} \tag{8}$$

$$h_k(n) = h_e(n)\cos\left[\left(n+\frac{M+1}{2}\right)\left(k+\frac{1}{2}\right)\frac{\pi}{M}\right] \qquad \text{(even } k\text{)}. \tag{9}$$

Now, the question is whether it is possible to jointly optimize $h_o(n)$ and $h_e(n)$ such that $h_1(n) = h_k(n) + h_{k+1}(n)$ and $h_2(n) = h_k(n) - h_{k+1}(n)$ have better time localization. Without loss of generality, let us assume that k is odd. Then we have

$$h_1(n) = (h_o(n)+h_e(n))\cos\beta\cos\alpha + (h_o(n)-h_e(n))\sin\beta\sin\alpha \tag{10}$$

$$h_2(n) = (h_o(n)-h_e(n))\cos\beta\cos\alpha + (h_o(n)+h_e(n))\sin\beta\sin\alpha, \tag{11}$$

where

$$\alpha = \left(n+\frac{M+1}{2}\right)\left(k+\frac{1}{2}\right)\frac{\pi}{M} \tag{12}$$

$$\beta = \left(n+\frac{M+1}{2}\right)\frac{\pi}{2M} \tag{13}$$

If $h_o(n) = h_e(n) = h(n)$ as in the original MLT, (10) and (11) are either cosine-modulated or sine-modulated. However, if we let $h_o(n)$ to be different from $h_e(n)$, then $h_1(n)$ and $h_2(n)$ have both cosine-modulated and sine-modulated terms, which makes it nearly impossible to jointly optimize $h_o(n)$ and $h_e(n)$ so that the time-domain aliasing is cancelled.

3.3 Applying Li et al. [15] to MLT

The idea in Li et al. [15] can be easily incorporated into a MLT-based framework. We have to note that this method is different from other subband merging techniques proposed in [7], [8], [9] and

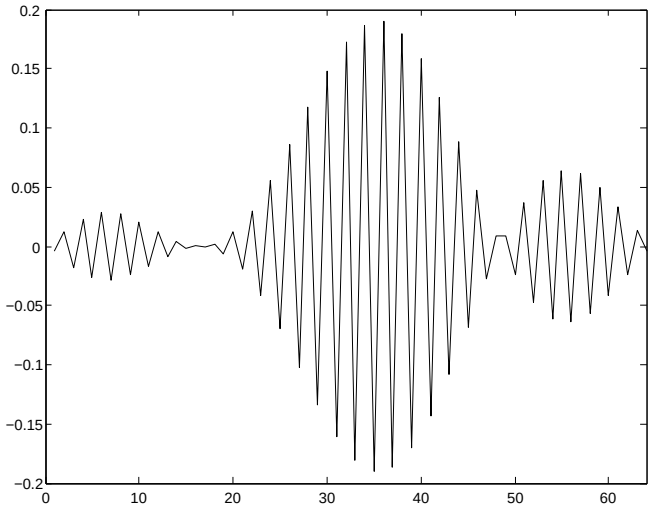

Figure 2: The basis function obtained by merging the last two subbands of MLT ($M = 32$).

[12]. Li et al. combine the filter outputs before decimation, while others combine the transform-domain coefficients after decimation. Therefore, if we want to apply Li et al.'s idea to MLT, we have to do the following. Let us assume that the number of subbands of the larger MLT is M, and we want to combine the last two high-frequency subbands. In order to do this, we have to compute the MLT coefficients of the last two subbands for every $M/2$ block-shift, while other coefficients are computed for every M block-shift as usual. This certainly increases the computational cost of the overall transform, but it has the advantage of using time-invariant basis function in the combined subbands. Note that other subband-merging approaches basically use periodically time-varying basis functions in the combined bands.

One problem with this approach is that this works well only when the prototype filters have very good stopband attenuation. Combining subbands in this way does not preserve the PR property, and the original CMFB has to use a good prototype filter with a long length to minimize the increase in the distortion level that arise from combining filters. However, the MLT uses relatively short filters, with not too large stopband attenuation (around 23 dB), and if the PR property is lost due to the combination of subbands using this method, the distortion level can be quite large, because of the uncanceled aliasing components. Another problem is that although this method uses time-invariant filters in the combined bands, the basis function that comes from filter combination has a significant amount of time-domain aliasing, as with other subband merging methods. This is shown in Fig. 2, where the last two subbands of an MLT for $M = 32$ have been combined. These problems show that Li et al. [15] is still not the right approach to achieve our goal.

3.4 Various Optimization Approaches

We considered the design of a $2M \times M$ transform matrix via optimization techniques, based on several kinds of constructions and objective functions. These are described in the following.

3.4.1 Approach I

We first constructed the matrix

$$\mathbf{P} = \begin{bmatrix} p_{00} & p_{01} & r_{00} & r_{01} \\ p_{10} & p_{11} & r_{10} & r_{11} \\ p_{20} & p_{21} & r_{20} & r_{21} \\ p_{30} & p_{31} & r_{30} & r_{31} \\ p_{40} & p_{41} & r_{40} & r_{41} \\ p_{50} & p_{51} & r_{50} & r_{51} \\ p_{60} & p_{61} & r_{60} & r_{61} \\ p_{70} & p_{71} & r_{70} & r_{71} \end{bmatrix}, \tag{14}$$

where p_{nk} comes from the first two columns of MLT for M=4. We performend nonlinear constrained optimization with respect to $\mathbf{r}_0 = [r_{00} \cdots r_{70}]^T$ and $\mathbf{r}_1 = [r_{01} \cdots r_{71}]^T$, to minimize the stopband energy of $\mathbf{r}_0$ and $\mathbf{r}_1$ in $0 \leq \omega \leq \pi/2$ and the distance $\|r_0(n - M/2) - r_1(n)\|^2$. The constraints were

$$\begin{cases} \mathbf{P}_0\mathbf{P}_0^T + \mathbf{P}_1\mathbf{P}_1^T &= \mathbf{I} \\ \mathbf{P}_0\mathbf{P}_1^T &= \mathbf{0} \end{cases}, \tag{15}$$

where $\mathbf{P} = [\mathbf{P}_0^T \ \mathbf{P}_1^T]^T$, which is just the perfect reconstruction condition for MLTs.

Interestingly enough, the optimized filters r_0 and r_1 always converged to those that result from combining the MLT filters using $+1/-1$ butterflies, as in the NMLBT [7]. These results indicate that forcing the first two filters for $M = 4$ (or the first six filters for $M = 8$) to be the same as those in the MLT, imposes a stringent condition on the optimization process, where the only feasible solution that satisfies the PR property and makes the optimized filters highpass filters that look similar to the shifted version of each other is that of the subband merging approach.

3.4.2 Approach II

We started with the following matrix, where all r_{ij}'s are free to be optimized.

$$P = \begin{bmatrix} r_{00} & r_{01} & r_{03} & 0 \\ r_{10} & r_{11} & r_{13} & 0 \\ r_{20} & r_{21} & r_{23} & r_{03} \\ r_{30} & r_{31} & r_{33} & r_{13} \\ r_{40} & r_{41} & 0 & r_{23} \\ r_{50} & r_{51} & 0 & r_{33} \\ r_{60} & r_{61} & 0 & 0 \\ r_{70} & r_{71} & 0 & 0 \end{bmatrix}. \tag{16}$$

We performed nonlinear constrained optimization with the PR condition in (15), and tried to make the filter in the first column a LPF with a passband in $[0, \pi/4]$, the filter in the second column a BPF with a passband in $[\pi/4, \pi/2]$, and the filters in the third column (and also the fourth) a HPF with a passband in $[\pi/2, \pi]$. Perfect time-domain localization and time-invariance of the highpass filters are guaranteed by construction.

The optimization of the matrix $\mathbf{P}$ led to an interesting result, where $\mathbf{P}$ converged to a PR lapped transform with the desired frequency resolution. However, the price to pay was a significant decrease in the stopband attenuation. When compared to MLT ($M = 4$), the stopband attenuation was reduced from around 25 dB to 20 dB. This shows that there is a trade-off between time-domain aliasing and frequency separation.

The major problem with this approach is that it does not easily scale to the case when M is large. The number of coefficients to be optimized increases as $O(M^2)$, and brute-force optimization of the filter coefficients may not converge to the global optimum (or even a good local one) once M gets larger. As we are interested in a design method that easily scales to the case of very large M, this approach is not so practical.

3.4.3 Approach III

Instead of using a constrained optimization with a PR condition, we may guarantee the PR condition by construction and use an unconstrained optimization algorithm. Let us consider two $M \times M$ orthogonal matrices $\mathbf{Q} = [\mathbf{Q}_0\ \mathbf{Q}_1]$ and $\mathbf{R} = [\mathbf{R}_0^T\ \mathbf{R}_1^T]^T$, where $\mathbf{Q}_i$ is a $M \times M/2$ matrix and $\mathbf{R}_j$ is a $M/2 \times M$ matrix. Now, let

$$\mathbf{P}_0 = \mathbf{Q}_0\mathbf{R}_0, \qquad \mathbf{P}_1 = \mathbf{Q}_1\mathbf{R}_1 \tag{17}$$

and define $\mathbf{P} = [\mathbf{P}_0^T\ \mathbf{P}_1^T]^T$. This matrix P is guaranteed to satisfy the PR condition in (15).

Now, assume that we want to choose the last $M/2$ column vectors using the last $M/4$ columns of a smaller MLT ($M/2$ subbands) and the same column vectors shifted by $M/2$. In this way, the dimensionality of the optimization space can be reduced, since many components in $\mathbf{Q}$ and $\mathbf{R}$ are immediately decided based on this setting.

However, simulation results show that by imposing both the PR condition and perfect time-localization of the highpass filters, the stopband attenuations of the optimized filters are decreased significantly. This can be clearly seen Fig. 3. Moreover, although the optimization space has become much smaller than the previous approach, the number of coefficients to be optimized still

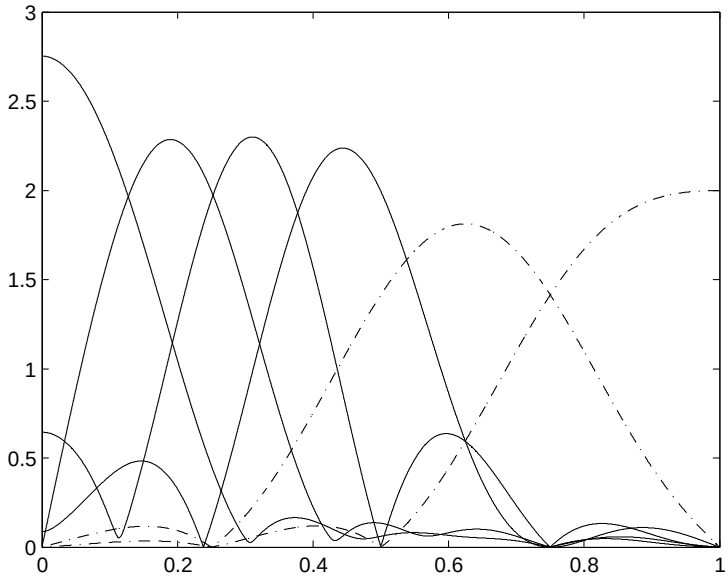

Figure 3: Magnitude frequency response (amplitude vs. normalized frequency) of a nonuniform transform designed using Approach III. The optimized filters are in solid lines and the filters obtained from MLT are shown in dashed lines.

increases as $O(M^2)$ which is undesirable. This implies that this method may not be applicable when we need a large number of subbands.

3.4.4 Approach IV

Based on extensive simulation results, we tried to come up with a formula for constructing a $2M \times M$ transform matrix that guarantees perfect time-localization of the high-frequency basis functions and near PR property.

The suggested formula is as follows. For $0 \leq k < M/2$

$$p_{nk} = \begin{cases} 0 & 0 \leq n < M/2 \\ \sqrt{\frac{2}{M}} h(n - M/2) \cos\left[\left(n + \frac{M+2}{4}\right)\left(k + \frac{1}{2}\right)\frac{\pi}{M}\right] & M/2 \leq n < 2M \end{cases} \tag{18}$$

and

$$h(n) = \sqrt{\sin\left[\left(n + \frac{1}{2}\right)\frac{2\pi}{3M}\right]}. \tag{19}$$

For $M/2 \leq k < 2M$, we use the last $M/4$ basis functions of a smaller MLT (order $M/2$) and their shifted versions by $M/2$. Fig. 4 shows the magnitude response for $M = 16$.

For small M, this designs is closer to PR than the simple combination of the first $M/2$ low-frequency basis functions of MLT (order M) and the last $M/4$ high-frequency basis functions and their shifted version by $M/2$ of a smaller MLT ($M/2$ subbands). However, as shown in Fig. 5 there

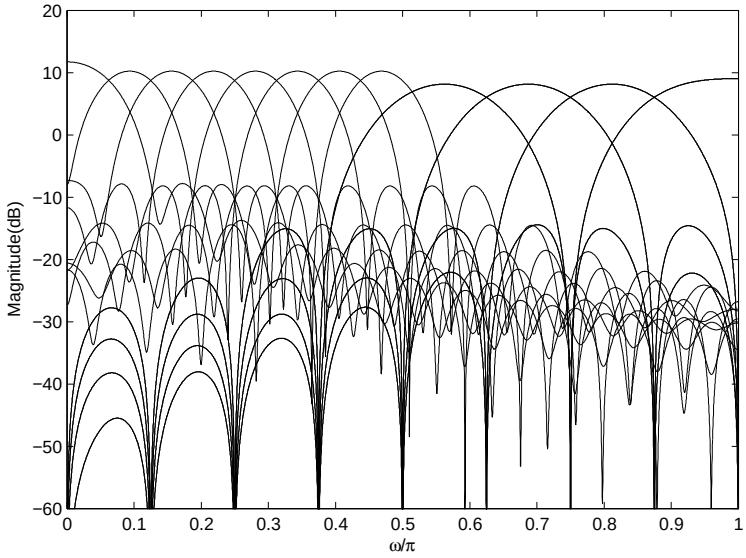

Figure 4: Magnitude frequency response of a nonuniform transform designed using Approach IV.

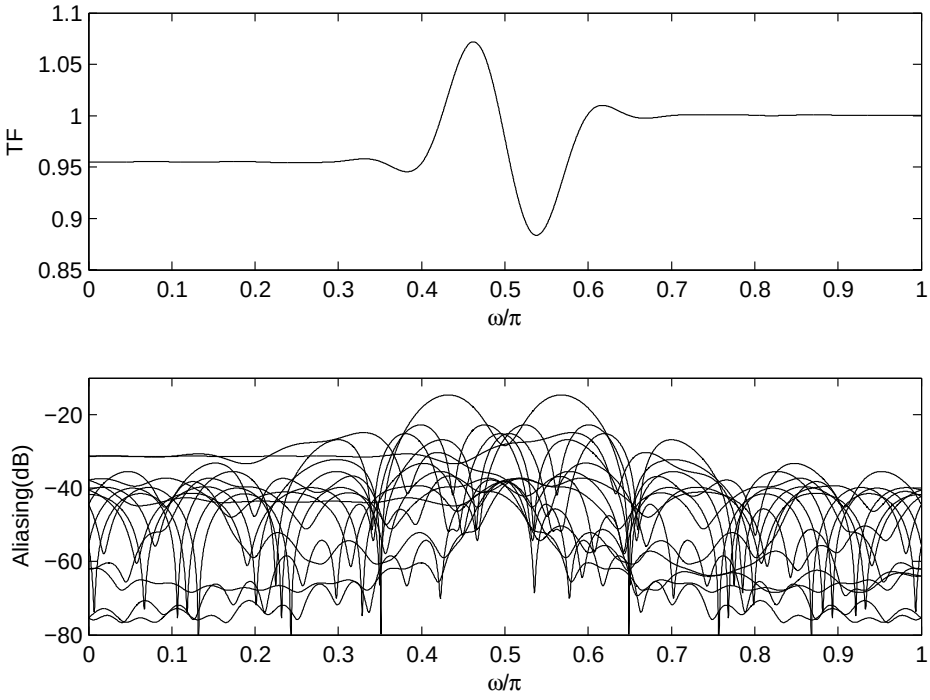

Figure 5: Nonuniform lapped transform designed using Approach IV. (Top) Transfer function. (Bottom) Aliasing components.

are large ripples in the overall transfer function and the aliasing terms are also significant. As we can see in Fig. 5, this transform is not close enough to PR to be of practical use.

4 Proposed Design

Until now, we have reviewed several nonuniform filter bank design methods that have been proposed so far, and also considered several extensions of these ideas based on a MLT framework. In this section we are going to propose a method for designing nonuniform lapped transform that can be used for constructing transforms with a large number of subbands, and good time-localization of the high-frequency basis functions. The proposed method is an extension of Princen [1].

4.1 Shortcomings of Princen's Approach

In his work in [1], Princen proposed an interesting idea for designing nonuniform filter banks. The NUFB is designed by joining two uniform sections that come CMFBs of different decimation ratios using a transition filter. The transition filter is derived from a complex (hence asymmetric) prototype filter, which is cosine-modulated such that its passband is located between the two uniform sections. The prototype filter is optimized such that it minimizes the aliasing between itself and the adjacent filters. Aliasing between other filters are assumed to be negligible during this optimization procedure. As the transition filter is designed such that the aliasing in both sides are cancelled at the same time, it is considerably longer than the adjacent filters, in general. This is obviously undesirable, because it increases the overall system delay.

Another problem is the following. The method starts from the assumption that the original CMFBs have very good stopband attenuation, hence there exists aliasing only between adjacent filters Thus, the method does not give satisfactory results when this assumption is not met. For example, let us consider constructing a nonuniform lapped transform by joining two MLTs of order M and $M/2$. These MLTs use relatively short filters whose lengths are $2M$ and M. Unlike the basic assumption in [1], there exists significant aliasing between non-adjacent filters in this case. Therefore designing a single transition filter to join MLTs do not result in a good lapped transform that is close to PR. In the following subsection we propose a modified design approach that achieves superior results to that of Princen's approach. As will be shown later, the proposed method can be also applied to the case when the original CMFB does not use filters with a very high stopband attenuation.

4.2 Design Procedure

We can avoid the problems in the original method [1], by designing three transition filters $g_0(n)$, $g_1(n)$ and $g_2(n)$ instead of using only $g_1(n)$. This is illustrated in Fig. 6 along with the original design approach. Without loss of generality, we assume that the CMFB in the low frequency region has a larger decimation ratio than the CMFB in the high frequency region, so that we have better frequency resolution for low frequency signals and better time localization for high frequency signals. As shown in Fig. 6, the filters in the original CMFBs do not have sharp cut-offs, which give rise to aliasing between non-adjacent bands. This is the case when we use a CMFB with relatively short filters such as the MLT. In such a situation, using a single optimized filter $g_1(n)$ in the transition region is not enough, and using additional filters $g_0(n)$ and $g_2(n)$ on both sides is helpful in reducing the overall distortion of the filter bank considerably.

Another advantage of using multiple transition filters is that the two additional filters $g_0(n)$ and $g_2(n)$ relieve the burden on the center filter $g_1(n)$, which makes it possible to use shorter transition filters compared to the original design [1]. This is a considerable advantage from a practical point of view, because it is possible to keep the overall delay of the nonuniform lapped transform to be the same as that of the uniform CMFB with the longest response.

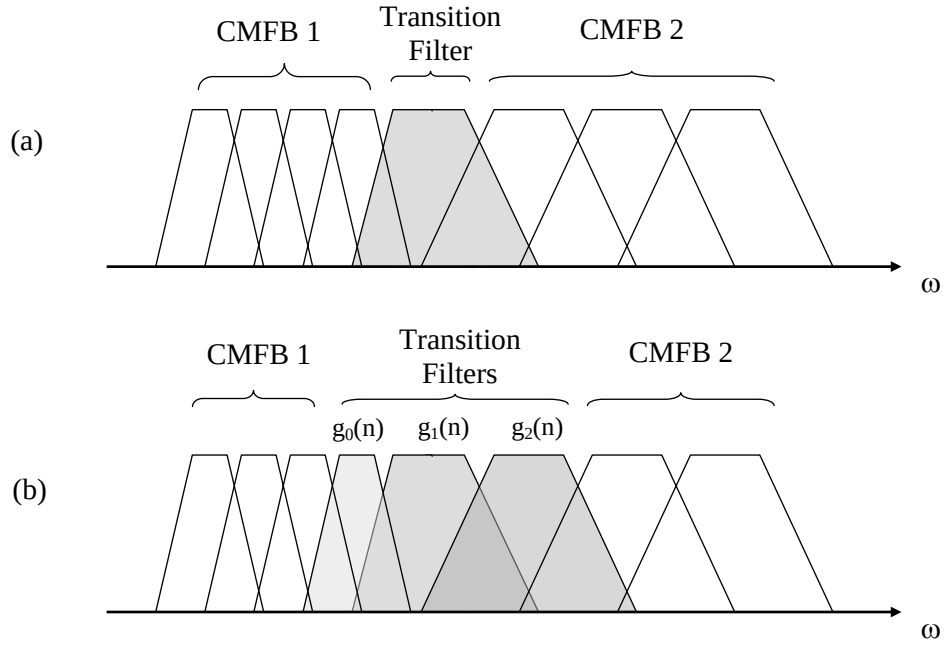

Figure 6: Joining two uniform CMFBs using transition filters. (a) Princen's approach. (b) Proposed approach.

In optimizing the transition filters, we try to make the system as close as possible to being PR. For this purpose, we define the following error metric, which can serve as an indicator of how close the given lapped transform is to a PR system. Given a $2MK \times M$ transform matrix $\mathbf{P}$ with M subbands,

$$\mathbf{P} = \left[\begin{array}{cccc} \mathbf{P}_0^T & \mathbf{P}_0^1 & \cdots & \mathbf{P}_{2K-1}^T \end{array} \right]^T, \tag{20}$$

let us define

$$\mathbf{E}_k = \sum_{i=0}^{2K-1-k} \mathbf{P}_i \mathbf{P}_{i+k}^T - \delta(k)\mathbf{I}, \tag{21}$$

where $\mathbf{I}$ is a $M \times M$ identify matrix. Now, the error metric is defined as the sum of the Frobenius norm[1] of $\mathbf{E_k}$

$$\eta = \sum_{k=0}^{2K-1} \|\mathbf{E}_k\|^2. \tag{22}$$

Note that if $\eta = 0$, the lapped transform $\mathbf{P}$ is PR [2], [17]. Based on this definition, we use unconstrained minimization techniques to minimize η in order to obtain a good near-PR nonuniform lapped transform.

In the proposed design, we choose the decimation ratio of subband that uses $g_0(n)$ to be the same as that of the CMFB in the left-hand side. For $g_1(n)$ and $g_2(n)$, the decimation ratios are chosen to be identical to that of the CMFB in the right-hand side. Now, the three transition filters are optimized one by one in a recursive manner. The optimization procedure can be summarized as follows.

Step 1 Combine the two uniform sections without using any transition filter.

Step 2 Optimize transition filter $g_0(n)$.

Step 3 Optimize transition filter $g_1(n)$.

Step 4 Optimize transition filter $g_2(n)$.

Step 5 If the $\eta < \eta_0$ for a specified η_0, terminate optimization. Otherwise, proceed to **Step 6**.

Step 6 Optimize transition filter $g_1(n)$.

Step 7 Optimize transition filter $g_0(n)$.

Step 9 If the $\eta < \eta_0$, terminate optimization. Otherwise, go to **Step 3**.

[1]The Frobenius norm of a matrix $\mathbf{A}$ is defined as $\|\mathbf{A}\|^2 = \sum_{i,j} |a_{ij}|^2 = \text{trace}(\mathbf{A}\mathbf{A}^H)$.

	$\bar{\eta}$	δ_T	$\mathcal{E}_{\text{alias}}$	γ (dB)
MLT (Princen)	0.000028415303	0.05602620	0.00130647	28.51
MLT (Proposed)	0.000014034253	0.02599753	0.00071550	31.38
ELT (Princen)	0.000000349104	0.00047347	0.00002225	46.50
ELT (Proposed)	0.000000223282	0.00030012	0.00001423	48.53

Table 1: Simulation results.

The optimization routine terminates once the transform $\mathbf{P}$ is close to PR. At the end of the routine, we can obtain a good near-PR lapped transform with approximately flat overall transfer function and negligible aliasing component, whose magnitude is comparable to the stopband attenuation of the filters in the original CMFBs.

4.3 Examples

In this section, we present design examples that clearly show the advantages of the proposed approach.

4.3.1 Example 1: Joining MLTs

Let us first consider joining two MLTs. We join an MLT $\mathbf{R}_1$ of order $M_1 = 32$ and a smaller MLT $\mathbf{R}_2$ with $M_2 = M_1/2 = 16$, according to the design procedure described in Sec. 4.2. Using these transforms, we construct a $2M_1 \times M_1$ transform matrix $\mathbf{P}$ as follows. The first $M_1/2 - 1$ column vectors of $\mathbf{P}$ are obtained from the first $M_1/2 - 1$ column vectors of $\mathbf{R}_1$. The last $M_1/2 - 4$ column vectors are obtained from the last $M_2/2 - 2$ column vectors of R_2 and the same vectors shifted by M_2. The transition filters are located in $M_1/2 - 1 \leq k \leq M_1/2 + 3$. The transition filter $g_0(n)$ is located at $k = M_1/2 - 1$ and has length $2M_1$. The transition filter $g_1(n)$ is located at $k = M_1/2, M_1/2 + 1$, where the vector at $k = M_1/2 + 1$ is a shifted version of $g_1(n)$ by M_2. The length of $g_1(n)$ is set to $3M_2 = 48$. Similarly, the filter $g_2(n)$ is located at $k = M_1/2 + 2, M_1/2 + 3$. The structure of the transform matrix is shown in Fig. 7

Fig. 8 shows the magnitude response, and Fig. 9 shows the overall transfer function and the aliasing components of the nonuniform lapped transform that was obtained using the proposed method. For comparison, the transform designed using Princen's approach is shown in Fig. 10 and Fig. 11. By comparing Fig. 9 and Fig. 11, we can easily see that the peak distortion of the overall transfer function and the magnitude of the aliasing components have been significantly reduced.

The characteristics of the two transforms are summarized in Table 1. The value $\bar{\eta} = \eta/M^2$ is a

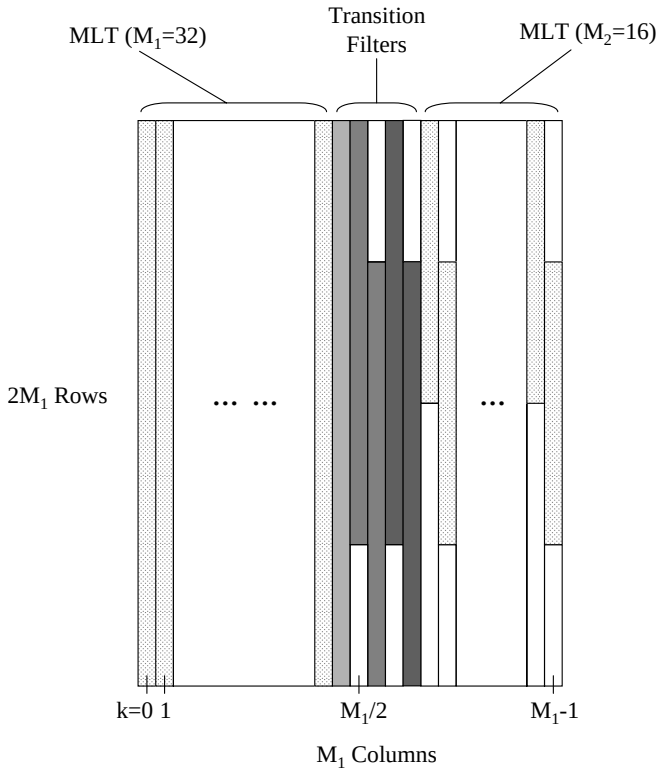

Figure 7: The structure of the transform matrix.

measure that shows how close the transform $\mathbf{P}$ is to a PR system. As mentioned earlier, $\bar{\eta} = \eta = 0$ implies that $\mathbf{P}$ is a PR transform. The overall transfer function and the aliasing terms can be defined as

$$T_\ell(z) = \frac{1}{M} \sum_{k=0}^{M-1} H_k(zW_M^\ell) F_k(z), \tag{23}$$

where $H_k(z)$ and $F_k(z)$ are respectively the analysis filter and the synthesis filter in the k-th band [6]. We define the maximum ripple size of the overall transfer function $T_0(z)$ as

$$\delta_T = \max_\omega \left| T_0(e^{j\omega}) - 1 \right|. \tag{24}$$

The energy of the aliasing terms is defined as

$$\mathcal{E}_{\text{alias}} = \sum_{\ell=1}^{M-1} \left(\frac{1}{\pi} \int_0^\pi |T_\ell(z)|^2 d\omega \right). \tag{25}$$

Finally, the SNR γ is the ratio between the input signal and the reconstruction error, which is defined as

$$\gamma = 10 \log_{10} \frac{\sigma^2}{\sigma_\epsilon^2}, \tag{26}$$

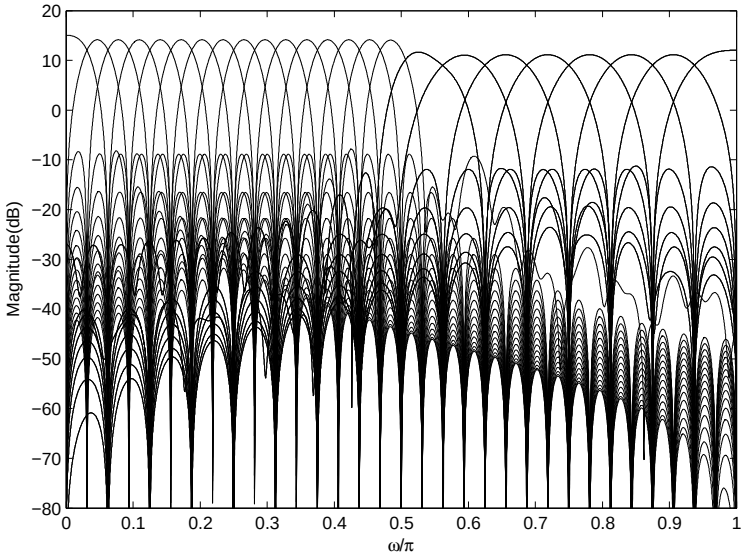

Figure 8: The magnitude response of the nonuniform lapped transform based on joining MLTs using the proposed approach.

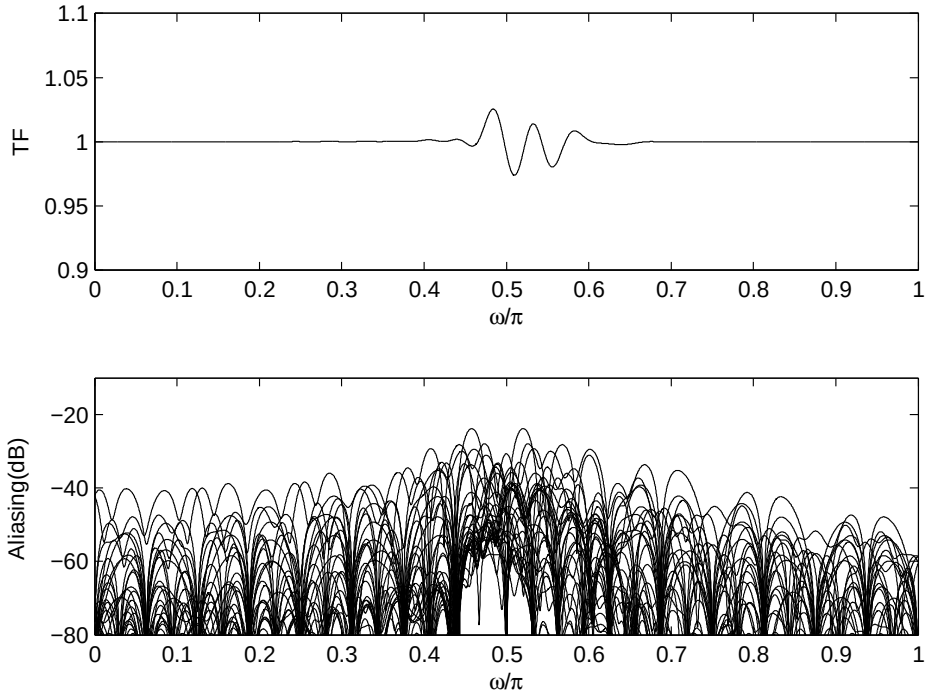

Figure 9: Nonuniform lapped transform based on joining MLTs using the proposed approach. (Top) Transfer function. (Bottom) Aliasing components.

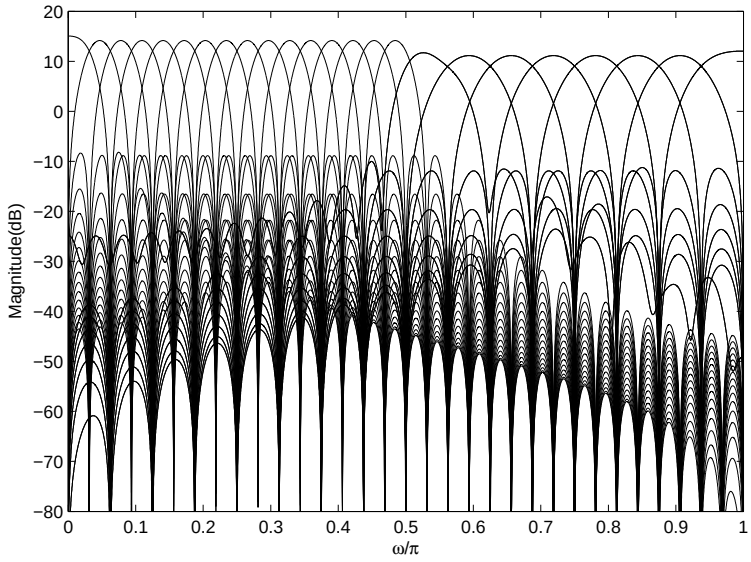

Figure 10: The magnitude response of the nonuniform lapped transform based on joining MLTs using Princen's approach.

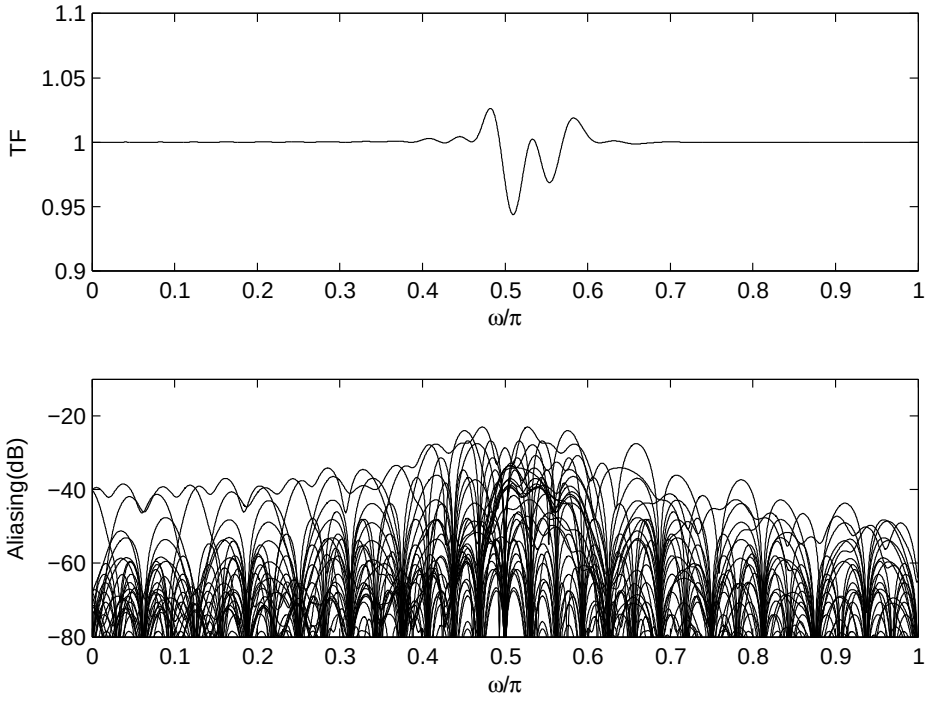

Figure 11: Nonuniform lapped transform based on joining MLTs using Princen's approach. (Top) Transfer function. (Bottom) Aliasing components.

where σ^2 is the variance of the input signal, and σ_ϵ^2 is the variance of the reconstruction error. In this example, we used a white gaussian noise as the input signal to evaluate the overall performance of the transform. As can be seen in Table 1, the proposed approach reduced all three metrics $\bar{\eta}$, δ_T and $\mathcal{E}_{\text{alias}}$ - which are indicators of the distortion level - by about 50%.

4.3.2 Example 2: Joining ELTs

In the next example, we consider joining two ELTs (extended lapped transforms) with $M_1 = 32$, $M_2 = M_1/2 = 16$ and $K = 4$ to construct a $2M_1K \times M_1$ transform matrix $\mathbf{P}$. The length of the filters in the CMFB in the low-frequency region is $2M_1K = 256$, and the length of the filters in the CMFB in the high-frequency region is $2M_2K = 128$. As in the first example, the transition filters $g_0(n)$, $g_1(n)$ and $g_2(n)$ are located in $M/2 - 1 \leq k \leq M/2 + 3$. The length of the transition filter $g_0(n)$ is $2M_1K = 256$, and the length of $g_1(n)$ and $g_2(n)$ is $2M_1K - M_2 = 240$.

The characteristics of the resulting NUFB are shown in Fig. 12 and Fig. 13, where those of the NUFB obtained from Princen's approach can be found in Fig. 14 and Fig. 15 for comparison. The properties of the two nonuniform lapped transforms are summarized in Table 1. As we can see from these results, using higher order CMFBs and longer transition filters improve the overall performance of the transform significantly. Now that the filters in the original CMFBs have better stopband characteristics, Princen's approach performs better than compared to the previous example. As Princen's approach now works better, the improvements obtained by the proposed design scheme is not as large as in the previous case. However, all three distortion metrics are still reduced by more than 35%.

4.4 Computational Cost

As we saw in the previous section, the use of three transition filters leads to improved performance when compared to the design proposed by Princen [1]. That was naturally to be expected, since there are more degree of freedom in optimizing three filters. That improvement comes at a small cost, though: increased computational complexity. The transition filters must be applied to the input signal frame in direct form, that is, they cannot be obtained from the fast transforms that are used to compute the MLT subbands [2]. The larger the number of subbands M_1 and M_2, though, the less significant the computational overhead of computing the transition filters. Consider the parameters in Example 1: $M_1 = 32$, $M_2 = 16$, and $K = 4$. For each input frame, the longer MLT can be computed via the FFT with 160 multiplications and 288 additions [2], and the shorter MLT with 72 multiplications and 120 additions, for a total of 232 multiplications and 408 additions. The

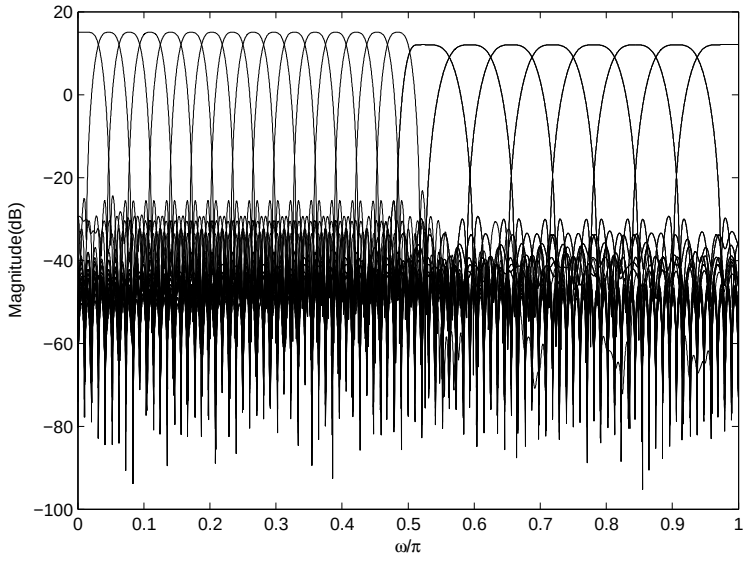

Figure 12: The magnitude response of the nonuniform lapped transform based on joining ELTs (K=4) using the proposed approach.

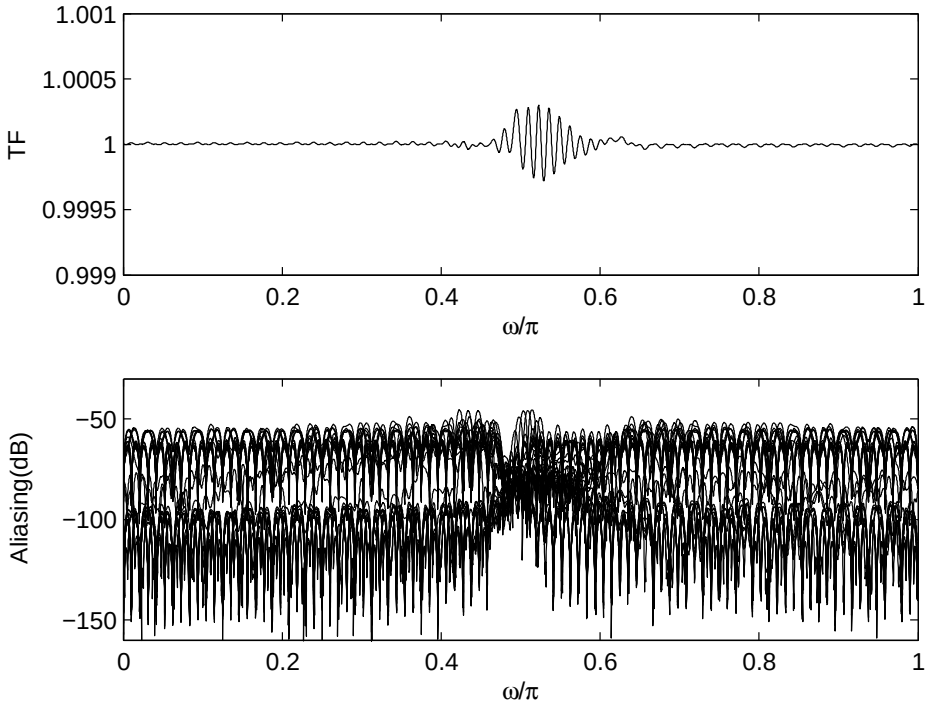

Figure 13: Nonuniform lapped transform based on joining ELTs (K=4) using the proposed approach. (Top) Transfer function. (Bottom) Aliasing components.

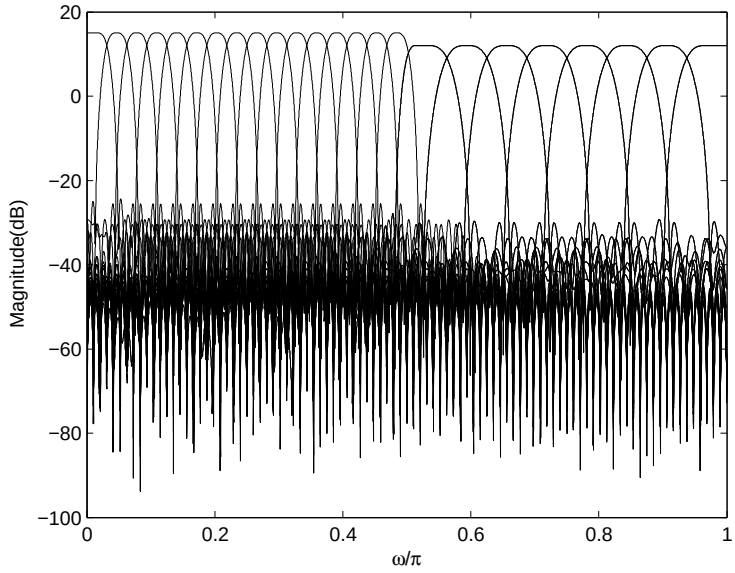

Figure 14: The magnitude response of the nonuniform lapped transform based on joining ELTs (K=4) using Princen's approach.

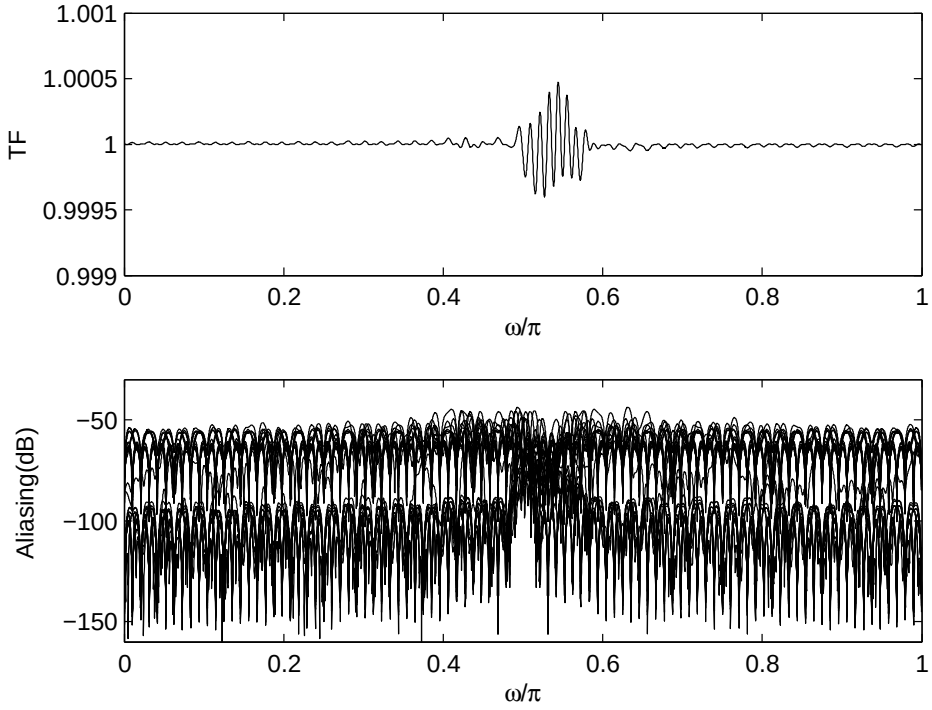

Figure 15: Nonuniform lapped transform based on joining ELTs (K=4) using Princen's approach. (Top) Transfer function. (Bottom) Aliasing components.

transition filters take a total of 112 multiplications and 112 additions. Thus, the transition filters lead to a 35% increase in computational cost. For example 2, the overhead increases because of the longer filteres required by an ELT-based design.

For applications such as audio coding, where $M_1 = 2,048$ is typical, the computational overhead of the MLT-based design is reduced a bit, to about 25%. Considering that computing the filter bank takes usually a small fraction of the total computational load, the 25% computational cost increase on the filter bank would typically be acceptable.

5 Conclusion

In this technical report, we have proposed an effective scheme for designing nonuniform lapped transforms. The proposed approach can be used to construct lapped transforms with large number of bands, nonuniform frequency resolution and good time localization of the high-frequency basis functions. The proposed approach is an extension of Princen [1], which uses multiple transition filters to join uniform CMFBs with different decimation ratios. As demonstrated in the included examples, this method reduces the distortion and the aliasing components in the original design significantly. It has also the advantages that it can be effectively used with CMFBs with relatively short filters (as the MLT), and that the overall system delay is identical to that of the longest delay of the original CMFBs that are joined together. These advantages make the proposed design approach especially useful in practical applications such as audio coding, as long as the increased computational complexity can be tolerated – which is usually the case.

Although we have used PR filter-banks in our design examples, it is also possible to use near-PR CMFBs instead [18], [19]. Such a design may have advantages in terms of an increased stop-band attenuation (hence, better frequency separation), at the expense of a small increase in the energy of the overall aliasing terms.

References

[1] J. Princen, "The design of nonuniform modulated filterbanks", *IEEE Trans. Signal Processing*, vol. 43, no. 11, pp. 2550-2560, Nov. 1995.

[2] H. S. Malvar, *Signal processing with lapped transforms*. Boston, MA: Artech House, 1992.

[3] MPEG, "Coding of moving pictures and associated audio for digital storage media at up to 1.5 Mbit/s, part 3: Audio", International Standard IS 11172-3, ISO/IEC JTC1/SC29 WG11, 1992.

[4] MPEG, "Information technology - generic coding of moving pictures and associated audio, part3: Audio", International Standard IS 13818-3, ISO/IEC JTC1/SC29 WG11, 1994.

[5] MPEG, "MPEG-2 advanced audio coding, AAC", International Standard IS 13818-7, ISO/IEC JTC1/SC29 WG11, 1997.

[6] P. P. Vaidyanathan, *Multirate systems and filter banks*. Englewood Cliffs, NJ: Prentice Hall, 1993.

[7] H. S. Malvar, "Biorthogonal and nonuniform lapped transforms for transform coding with reduced blocking and ringing artifacts", *IEEE Trans. Signal Processing*, vol. 46, no. 4, April 1998.

[8] Z. Xiong and H. S. Malvar, "A nonuniform modulated complex lapped transform", *IEEE Signal Processing Letters*, vol. 8, no. 9, Sep. 2001.

[9] M. Purat and P. Noll, "Audio coding with a dynamic wavelet packet decomposition based on frequency-varying modulated lapped transforms", *Proc. ICASSP*, vol. 2, pp. 1021-1024, May 1996.

[10] O. A. Niamut and R. Heusdens, "Subband merging in cosine-modulated filter banks", *IEEE Signal Processing Letters*, vol. 10, no. 4, pp. 111-114, April 2003.

[11] R. L. de Queiroz, "Uniform filter banks with nonuniform bands: post-processing design", *Proc. ICASSP*, vol. 3, pp. 1341-1344, 1998.

[12] S. C. Chan, X. M. Xie and T. I. Yuk, "Theory and design of a class of cosine-modulated nonuniforma filter banks", *Proc. ICASSP*, vol. 1, pp. 504-507, 2000.

[13] X. M. Xie, S. C. Chan and T. I. Yuk, "A class of biorthogonal nonuniform cosine-modulated filter banks with lower system delay", *Proc. ISCAS*, vol. 2, pp. 25-28, May 2001.

[14] X. M. Xie, S. C. Chan and T. I. Yuk, "On the theory and design of a class of perfect-reconstruction nonuniform cosine-modulated filter-banks", *Proc. ISCAS*, vol. 5, pp. 285-288, May 2002.

[15] J. Li, T. Q. Nguyen and S. Tantaratana, "A simple design method for nonuniform multirate filter banks", *Proc. 28th Asilomar Conference on Circuits, Systems and Computers*, vol. 2, no. 31, pp. 1015-1019, Oct. 1994.

[16] F. Argenti and E. Del Re, "Non-uniform filter banks based on a multi-prototype cosine modulation", *Proc. ICASSP*, pp. 1511-1514, May 1996.

[17] K. Nayebi, T. P. Barnwell III and M. J. T. Smith, "Nonuniform filter banks: a reconstruction and design theory", *IEEE Trans. Signal Processing*, vol. 41, no. 3, pp. 1114-1127, March 1993.

[18] T. Q. Nguyen, "Digital filter bank design quadratic-constrained formulation", *IEEE Trans. Signal Processing*, vol. 43, no. 9, pp. 2103-2108, Sep. 1995.

[19] T. Q. Nguyen, "Near-perfect-reconstruction pseudo-QMF banks", *IEEE Trans. Signal Processing*, vol. 42, no. 1, pp. 65-76, Jan. 1994.