Low-Delay Nonuniform Pseudo-QMF Banks With Application to Speech Enhancement

Ying Deng, *Member, IEEE*, V. John Mathews, *Fellow, IEEE*, and Behrouz Farhang-Boroujeny, *Senior Member, IEEE*

***Abstract*—This paper presents a method for designing low-delay nonuniform pseudo quadrature mirror filter (QMF) banks. This method is motivated by the work of Li, Nguyen, and Tantaratana, in which the nonuniform filter bank is realized by combining an appropriate number of adjacent sub-bands of a uniform pseudo-QMF bank. In prior work, the prototype filter of the uniform pseudo-QMF bank was constrained to have linear phase and the overall delay associated with the filter bank was often unacceptably large for filter banks with a large number of sub-bands. This paper proposes a pseudo-QMF filter bank design technique that significantly reduces the delay by relaxing the linear phase constraints. An example in which an oversampled critical-band nonuniform filter bank is designed and applied to a two-state modeling speech enhancement system is presented in this paper. Comparison of the performance of this system to competing methods employing tree-structured, linear phase multiresolution analysis indicates that the approach described in this paper strikes a good balance between system performance and low delay.**

***Index Terms*—Low-delay nonuniform filter bank design, multiresolution analysis, speech enhancement.**

I. Introduction

Speech enhancement systems employing uniform decompositions such as discrete Fourier transform are commonly used as front-end processing elements of robust speech communication systems [1]–[5]. Subjective tests have shown a preference for systems employing filter banks with nonuniform bandwidths [6]–[8]. Wavelet denoising by thresholding the wavelet coefficients was proposed by Donoho and Johnstone [9] and was shown to be successful in a number of situations where nonwavelet-based methods had problems. As a result, wavelet/wavelet packet transform has been widely studied as a multiresolution analysis tool in speech enhancement systems [6]–[8], [10]–[14]. Wavelet/wavelet packet transform can be conveniently performed using tree-structured filter banks [15]. However, nonuniform filter banks created by cascading wavelet bases often result in sub-band filters with poor frequency localization. Speech enhancement algorithms usually process different sub-bands with different spectral gains. When decompositions with poor frequency localization are employed, aliasing distortion will result in the "enhanced" speech signals [8], [16].

One way of alleviating this problem is to use longer wavelet bases that have high stopband attenuation. For example, after comparing the frequency response of different Daubechies filter pairs (D_6, D_{12}, D_{18}) [17] with Vaidyanathan filter pair (V_{24}) [18], Gülzow [7] chose the Vaidyanathan filter of length 24 to get a better separation of the adjacent bands. Cohen [8] chose the discrete Meyer wavelet filters of length 62.

In an attempt to resolve the poor frequency localization problem, Cvetković and Johnston [16] proposed a method of designing nonuniform oversampled filter banks for audio signal processing by combining sections of different uniform filter banks using transition filters. Oversampled filter banks are used to minimize aliasing by reducing the number of significant aliasing terms [8], [16]. Cvetković and Johnston [16] employed nearly perfect reconstruction oversampled complex modulated uniform filter banks. The prototype filters of the complex modulated uniform filter banks were designed to have a sharp transition band and high stopband attenuation. In this method, a separate uniform filter bank must be designed for every sub-band that has a different bandwidth from the other bands. A set of transition filters are then designed to combine filters from the uniform filter banks. This design approach is somewhat complicated. Li *et al.* [19] proposed a simple method for designing nonuniform filter banks by combining an appropriate number of sub-bands of a uniform pseudo-quadrature mirror filter (QMF) bank [20]. This design approach is simpler and more efficient. Li *et al.* [19] did not address the issue of near perfect reconstruction in the oversampled case.

Besides frequency localization, the delay introduced into the signal by a decomposition method is also a crucial component for speech communication applications. The total delay for a maximally decimated tree-structured filter bank has a duration of $(2^M - 1)D$ samples of the filter bank input [21], where M denotes the number of levels in the decomposition and D is the delay associated with a single analysis and synthesis pair. For the critical-band wavelet packet decomposition with six levels and a discrete Meyer wavelet basis of length 62 $(D = 61)$ the total delay is 3843 samples. A delay of 3843 samples corresponds to more than 240 ms for a sampling rate of 16 kHz. This is unacceptable for many speech applications.

Heller *et al.* [22] and Schuller *et al.* [23] proposed two different approaches for designing perfect reconstruction cosine-modulated filter banks with arbitrary delay. The authors of these two papers derived the necessary and sufficient conditions for low-delay perfect reconstruction biorthogonal cosine-modulated filter banks (different prototype filters for analysis and synthesis filter banks) using polyphase structure. Kliewer *et al.* [24] derived the perfect reconstruction conditions for oversampled cosine-modulated filter banks with arbitrary delay. It was shown that by taking advantage of oversampling, the number of perfect reconstruction constraints are reduced

Manuscript received September 15, 2005. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Antonia Papandreou-Suppappola.

The authors are with the Department of Electrical and Computer Engineering, University of Utah, Salt Lake City, UT 84112 USA (e-mail: yd4@utah.edu).

Digital Object Identifier 10.1109/TSP.2007.892707

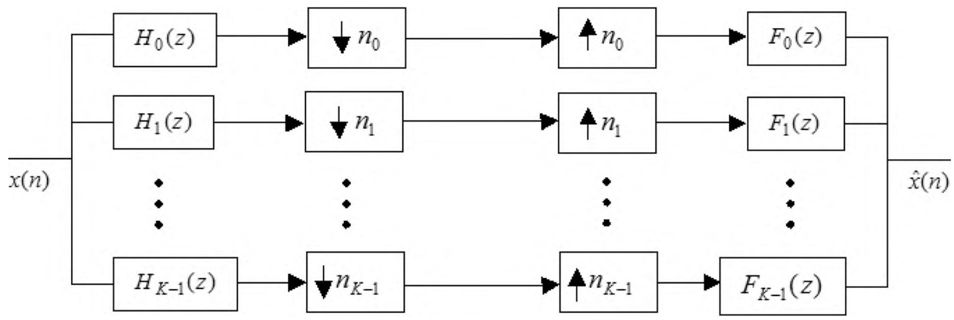


Fig. 1. M-channel uniform filter bank.

as compared to the critically sampled case. However, perfect reconstruction is overly restrictive in many practical applications. More importantly, perfect reconstruction filter banks usually cannot achieve as high stopband attenuation as nearly perfect reconstruction filter banks can [20]. High stopband attenuation is important to minimize the distortion caused by combining uniform bands to achieve nonuniform bands [19]. Furthermore, we will show later in this paper that generating nonuniform filter banks as in [19] requires the analysis and synthesis prototype filters to satisfy a special relation, which precludes the general biorthogonal filter banks.

In this paper, we present a method for designing low-delay nonuniform pseudo-QMF banks that achieves sharp transition bands and high stopband attenuation. Sharp transition bands and high stopband attenuation allow unequal processing of the subbands without introducing significant aliasing distortion into the reconstructed signals. The low-delay nonuniform pseudo-QMF bank is constructed by combining an appropriate number of bands of a low-delay uniform pseudo-QMF bank. The combining process is the same as that in [19]. However, we generalize the theory to include filter banks with different analysis and synthesis prototype filters. We then prove that the synthesis prototype has to be a scaled version of the analysis prototype filters in order for significant aliasing cancellation to occur in low-delay uniform pseudo-QMF banks and to achieve significant distortion cancelation in low-delay nonuniform filter banks. Our design assumes an oversampled filter bank. A version of this paper based on the same theory, but a different design approach, was presented in [25]. The design method in [25] can achieve flatter passband and higher stopband attenuation at the cost of higher amplitude distortion. The design methodology presented in this paper exhibits smaller overall amplitude distortion than the method of [25].

This paper is organized as follows. In Section II, we briefly review the theoretic background of pseudo-QMF banks [15] and the work by Nguyen [20] and Li *et al.* [19]. We present the new scheme for designing low-delay, oversampled nonuniform pseudo-QMF banks in Section III. Section IV provides a design example and the design of a critical-band nonuniform filter bank that simulates the front-end of the human auditory system. We also present an application of this filter bank in a two-state modeling speech enhancement scheme [8] in this section. Comparison of the performance of this system to competing methods employing tree-structured, linear phase multiresolution analysis indicates that the approach described in this paper strikes a good balance between system performance and low delay. Finally, we make our concluding remarks in Section V.

II. Review of Previous Works

Fig. 1 illustrates an M-channel uniform filter bank. For the maximally decimated case, i.e., $K = M$, the reconstructed signal $\hat{X}(z)$ can be expressed as [15]

$$\hat{X}(z) = \sum_{l=0}^{M-1} X\left(zW_M^l\right) T_l(z) \tag{1}$$

where $W_M = e^{-j(2\pi/M)}$ and

$$T_l(z) = \frac{1}{M} \sum_{k=0}^{M-1} H_k\left(zW_M^l\right) F_k(z). \tag{2}$$

For perfect reconstruction of $X(z)$, we need to have $T_0(z) = z^{-\Delta}$ and $T_l(z) = 0,\ 1 \le l \le M-1$, where Δ is a positive integer corresponding to the delay of the filter bank. The system function $T_l(z)$ contributes to the aliasing associated with $X(zW_M^l)$.

Modulated filter banks are popular due to their simplicity. In such systems, the M analysis and M synthesis filters can be generated by modulating a single prototype filter. The system developed in this paper is a pseudo-QMF bank, which is a nearly perfect reconstruction cosine modulated filter bank. Nearly perfect reconstruction means that only "adjacent channel aliasing" (significant aliasing) is canceled and the distortion function $T_0(z)$ is only approximately a delay [15]. When the stopband attenuation of the prototype filter is high, the aliasing due to nonadjacent channels will be small.

A. Uniform Pseudo-QMF Bank

For uniform pseudo-QMF banks, the real valued coefficients of the analysis and synthesis filters are generated using [15], [20]

$$H_k(z) = a_k U_k(z) + a_k^* V_k(z) \tag{3}$$

$$F_k(z) = b_k U_k(z) + b_k^* V_k(z) \tag{4}$$

where $U_k(z)$ and $V_k(z)$ are defined as

$$U_k(z) = c_k H\left(zW_{2M}^{(k+0.5)}\right) \tag{5}$$

and

$$V_k(z) = c_k^* H\left(zW_{2M}^{-(k+0.5)}\right) \tag{6}$$

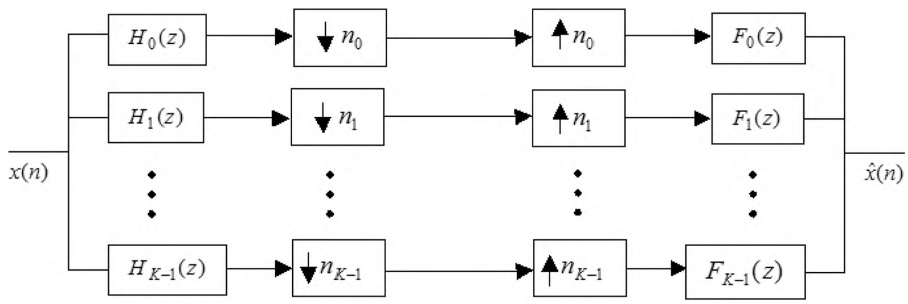


Fig. 2. K-channel nonuniform filter bank.

respectively. Here, $W_{2M} = e^{-j(\pi/M)}$, $H(z)$ is the prototype low-pass filter, and the coefficients a_k, b_k, and c_k are constants with unit magnitude. $*$ denotes complex conjugation. To cancel the significant aliasing components, the following relationship should hold [15]:

$$a_k b_k^* = -a_{k-1} b_{k-1}^*, \qquad 1 \le k \le M-1. \tag{7}$$

The prototype filter $H(z)$ was restricted to be a linear phase filter in [15] and [20]. It was shown in [15] that if we further choose $a_k = e^{j(-1)^k(\pi/4)}$, $b_k = a_k^*$, and $c_k = W_{2M}^{(N-1)(k+0.5)/2}$, the distortion function reduces to

$$T_0(e^{j\omega}) = e^{-j\omega(N-1)} \frac{1}{M} \sum_{k=0}^{M-1} \left|H_k(e^{j\omega})\right|^2 \tag{8}$$

where N is the length of the prototype filter.

Furthermore, it is proven in [20] that if we further constrain the prototype filter $H(z)$ to be a linear phase spectral factor of a $2M$th band filter, the overall distortion function $T_0(z)$ is a pure delay. In other words, if $H^2(z) = z^{-(N-1)}H(z)H(z^{-1})$ is a $2M$th band filter, the uniform pseudo-QMF bank will have no amplitude distortion. Thus, $T_0(z) = (1/M)z^{-(N-1)}$.

B. Nonuniform Pseudo-QMF Bank

Li, *et al.* [19] proposed a nonuniform pseudo-QMF bank design approach by simply combining the neighboring bands of a uniform pseudo-QMF bank from [20]. In order to understand the method, we will first introduce the concept of a feasible partition filter bank [26]. Consider the nonuniform filter bank in Fig. 2. Let M be the least common multiple of $n_0, n_1, \ldots, n_{K-1}$ and let $k_p = M/n_p$, $p = 0, 1, \ldots, K-1$. If for each $p \in \{0, 1, \ldots, K-1\}$, $\sum_{i=0}^{p-1} k_i = k_p v_p$ for some integer v_p, we call the filter bank a feasible partition filter bank. It is shown in [26] that it is possible to achieve perfect reconstruction with a feasible partition filter bank when the analysis and synthesis filters are properly designed. However, if a filter bank does not have a feasible partition, no matter how well the analysis and synthesis filters are designed, aliasing cannot be completely canceled and perfect reconstruction is impossible. We, thus, limit our attention to feasible partition filter banks in this paper.

It is possible to design a feasible partition nonuniform pseudo-QMF bank by combining an appropriate number of the adjacent bands of a uniform pseudo-QMF bank [19]. A uniform M-channel pseudo-QMF bank is first designed using the technique in [20], where M is equal to the least common multiple of $n_0, n_1, \ldots, n_{K-1}$. The pth sub-band of the nonuniform filter bank is then formed by combining M/n_p uniform sub-bands starting at $k_p = \sum_{i=0}^{p-1}(M/n_i) + 1$ so that

$$H_p^{NU}(z) = \frac{1}{\sqrt{M/n_p}} \sum_{j=0}^{\frac{M}{n_p}-1} H_{k_p+j}(z) \tag{9}$$

and

$$F_p^{NU}(z) = \frac{1}{\sqrt{M/n_p}} \sum_{j=0}^{\frac{M}{n_p}-1} F_{k_p+j}(z). \tag{10}$$

The distortion resulting from the combining operation may be expressed as $T_0^{NU}(z) = (1/M)(z^{-(N-1)} + D(z))$, where the distortion $D(z)$ is very small when the stopband attenuation of the prototype filter is sufficiently high [19].

III. Design of Low-Delay Nonuniform Pseudo-QMF Banks

In this section, we discuss the design of low-delay, maximally decimated, and oversampled uniform pseudo-QMF banks. The nonuniform filter bank is then constructed in the same way as in [19].

A. Theoretic Statements

Consider the general case of different analysis and synthesis prototype filters with impulse responses $h(n)$ and $f(n)$, respectively. With DCT-IV modulation and arbitrary integer Δ (the meaning of Δ will be illustrated in lemma 1), the analysis and synthesis filters are given by

$$h_k(n) = 2h(n)\cos\left(\frac{\pi}{M}(k+0.5)\left(n - \frac{\Delta}{2}\right) + (-1)^k \frac{\pi}{4}\right)$$
$$f_k(n) = 2f(n)\cos\left(\frac{\pi}{M}(k+0.5)\left(n - \frac{\Delta}{2}\right) - (-1)^k \frac{\pi}{4}\right)$$

or equivalently in the z domain by

$$H_k(z) = a_k c_k H\left(zW_{2M}^{(k+0.5)}\right) + a_k^* c_k^* H\left(zW_{2M}^{-(k+0.5)}\right) \tag{11}$$
$$F_k(z) = b_k c_k F\left(zW_{2M}^{(k+0.5)}\right) + b_k^* c_k^* F\left(zW_{2M}^{-(k+0.5)}\right) \tag{12}$$

with $a_k = e^{j(-1)^k(\pi/4)}$, $b_k = a_k^*$, and $c_k = W_{2M}^{\Delta(k+0.5)/2}$. This generation of the analysis and synthesis filter banks is the same as the general DCT-IV cosine-modulated biorthogonal filter banks in [22]. Instead of deriving the constraints on the polyphase components of the analysis and synthesis prototype filters to satisfy perfect reconstruction [22], we consider pseudo-QMF banks.

We start with the following two lemmas for uniform pseudo-QMF banks. Lemma 1 extends the theory of pseudo-QMF bank, which in the past has been limited to identical analysis and synthesis prototype filters with linear phase, and shows that the same theory is applicable to analysis and synthesis prototype filters with certain relations and nonlinear phase as well. A proof of this lemma is given in Appendix A.

Lemma 1: Consider the general case as in (11) and (12). The significant aliasing components, i.e., the "adjacent channel aliasing" terms are completely canceled if and only if

$$F(z) = cH(z) \tag{13}$$

with c being an arbitrary constant. Furthermore, if $G(z) = H^2(z) = \sum_{n=0}^{2(N-1)} g(n)z^{-n}$ is a $2M$th band filter, i.e., for some arbitrary delay Δ and integer values p

$$g(\Delta + p2M) = \begin{cases} 1/2, & p = 0 \\ 0, & \text{otherwise.} \end{cases} \tag{14}$$

The pseudo-QMF bank exhibits no amplitude and phase distortion with $T_0(z) = cz^{-\Delta}$. That is, a pseudo-QMF bank defined by (11) and (12) and satisfying (13) and (14) exhibits only aliasing distortion caused by nonadjacent bands.

From this Lemma, we know that in order for "adjacent channel aliasing" cancellation, $H(z)$ and $F(z)$ must satisfy (13). It is also clear from this lemma that we can design a low-delay maximally decimated uniform pseudo-QMF bank by relaxing the linear phase constraint on the prototype filter. We note that as formulated here, Δ is the overall system delay and is a free design parameter that we can select.

All the pseudo-QMF banks mentioned before assume maximal decimation. In general, the pseudo-QMF banks lose their aliasing cancellation property for the oversampled case. However, for a special subclass of oversampling factors, we can design oversampled filter banks by simply changing the sampling factor associated with a maximally-decimated filter bank. This is formally stated in the next lemma.

Lemma 2: Let R be an integer such that $K = M/R$ is an integer. Then an oversampled pseudo-QMF bank with oversampling factor R can be designed from a maximally decimated pseudo-QMF bank such that there is no amplitude and phase distortion and the significant aliasing terms are completely canceled. As in lemma 1, the only source of reconstruction error in such systems is aliasing caused by nonadjacent channels.

Proof: A maximally decimated pseudo-QMF bank satisfies the following perfect reconstruction constraints approximately:

$$\begin{pmatrix} H_0(z) & H_1(z) & \cdots & H_{M-1}(z) \\ H_0(zW_M) & H_1(zW_M) & \cdots & H_{M-1}(zW_M) \\ \vdots & \vdots & \ddots & \vdots \\ H_0\left(zW_M^{M-1}\right) & H_1\left(zW_M^{M-1}\right) & \cdots & H_{M-1}\left(zW_M^{M-1}\right) \end{pmatrix} \times \begin{pmatrix} F_0(z) \\ F_1(z) \\ \vdots \\ F_{M-1}(z) \end{pmatrix} = \begin{pmatrix} z^{-\Delta} \\ 0 \\ \vdots \\ 0 \end{pmatrix}. \tag{15}$$

For an oversampled filter bank with sampling factor $K = M/R$, the perfect reconstruction constraint is

$$\begin{pmatrix} H_0(z) & H_1(z) & \cdots & H_{M-1}(z) \\ H_0(zW_K) & H_1(zW_K) & \cdots & H_{M-1}(zW_K) \\ \vdots & \vdots & \ddots & \vdots \\ H_0\left(zW_K^{K-1}\right) & H_1\left(zW_K^{K-1}\right) & \cdots & H_{M-1}\left(zW_K^{K-1}\right) \end{pmatrix} \times \begin{pmatrix} F_0(z) \\ F_1(z) \\ \vdots \\ F_{M-1}(z) \end{pmatrix} = \begin{pmatrix} z^{-\Delta} \\ 0 \\ \vdots \\ 0 \end{pmatrix}. \tag{16}$$

Since $W_K = W_{M/R} = W_M^R$ and R is an integer, the set of constraints in (16) is a subset of the constraints in (15). This indicates that if we design a maximally decimated filter bank that approximately satisfies (15), an oversampled filter bank obtained by simply increasing the sampling rate of the maximally decimated filter bank R times will satisfy (16) approximately.■

Using the previous lemmas, one may argue that the method of [19] can be applied to construct a low-delay nonuniform oversampled filter bank from a low-delay uniform oversampled filter bank. We prove in Appendix B that significant aliasing cancellation (13) is also a sufficient condition for the cancellation of significant distortion caused by combining uniform sub-bands to achieve a nonuniform filter bank.

B. Design Procedure

From Lemma 2, we know it is simple to construct an oversampled pseudo-QMF bank from a maximally decimated pseudo-QMF bank if the oversampling factor is an integer factor of the maximal decimation rate. Consequently, we only discuss the design procedure of a maximally decimated pseudo-QMF bank and assume $c = 1$ in what follows.

Let $\mathbf{h} = [h(0)\ h(1)\ \cdots\ h(N-1)]^T$ and $\mathbf{e}(z) = [1\ z^{-1}\ \cdots\ z^{-(N-1)}]^T$, where the superscript T denotes transposition. Then, the transfer function of the prototype filter is

$$H(z) = \mathbf{h}^T\mathbf{e}(z). \tag{17}$$

Now

$$\begin{aligned} G(z) &= \sum_{n=0}^{2(N-1)} g(n)z^{-n} \\ &= \mathbf{h}^T\mathbf{e}(z)\mathbf{e}^T(z)\mathbf{h} \\ &= \mathbf{h}^T\mathbf{R}(z)\mathbf{h} \end{aligned} \tag{18}$$

where

$$\begin{aligned} \mathbf{R}(z) &= \mathbf{e}(z)\mathbf{e}^T(z) \\ &= \begin{pmatrix} 1 \\ z^{-1} \\ \vdots \\ z^{-(N-1)} \end{pmatrix} \begin{pmatrix} 1 & z^{-1} & \cdots & z^{-(N-1)} \end{pmatrix} \\ &= \sum_{n=0}^{2(N-1)} z^{-n}\mathbf{S}_n \end{aligned} \tag{19}$$

and $\mathbf{S}_n$ are constant matrices whose elements take values from 0 or 1 according to [20]

$$[\mathbf{S}_n]_{k,l} = \begin{cases} 1, & k+l=n \\ 0, & \text{otherwise.} \end{cases} \quad (20)$$

Substituting (19) into (18) gives

$$G(z) = \sum_{n=0}^{2(N-1)} (\mathbf{h}^T\mathbf{S}_n\mathbf{h})z^{-n}. \quad (21)$$

From Lemma 1, we know that we can design a pseudo-QMF bank by designing a prototype filter $H(z)$ with high stopband attenuation and satisfying $G(z) = H^2(z)$ as in (14) with $p \in [-\lfloor\Delta/2M\rfloor, \lfloor 2(N-1)-\Delta/2M\rfloor]$. Comparing (21) and (14), we can now constrain the filter coefficients vector $\mathbf{h}$ so that whenever $n = \Delta + p2M$

$$\mathbf{h}^T\mathbf{S}_n\mathbf{h} = \begin{cases} 1/2, & p=0 \\ 0, & \text{otherwise.} \end{cases} \quad (22)$$

In addition to the previous constraints, $\mathbf{h}$ should also yield a prototype filter with good stopband attenuation and a reasonable pass-band response. Accordingly, the prototype filter is designed to minimize a weighted sum of the stopband energy and the difference between the ideal low-pass filter and the designed filter in the pass-band subject to the constraints in (22). This leads to a filter design procedure that aims at minimizing the objective function

$$\begin{aligned} E &= (1-\alpha)\frac{1}{2\pi}\int_{\omega_s}^{2\pi-\omega_s} \left|H(e^{j\omega})\right|^2 d\omega \\ &\quad + \alpha\frac{1}{2\pi}\int_{-\omega_p}^{\omega_p} \left(\left|H(e^{j\omega})\right|^2 - \left|H(e^{j0})\right|^2\right) d\omega \\ &= (1-\alpha)E_s + \alpha E_p \end{aligned} \quad (23)$$

subject to the constraints in (22). In (23), α is a tradeoff parameter. The parameter α is initially set to 0 during the iterative optimization process. If the deviation of the pass-band response of the resulting prototype filter from a flat response is unacceptably large, the optimization process is repeated using a small positive value of α. The value of α is increased gradually with iterations untill an acceptable pass-band response is obtained. In our paper, the largest value of α we have employed is 0.00025. To solve the constrained optimization problem, we note from (17) that $H(e^{j\omega}) = \mathbf{h}^T\mathbf{e}(e^{j\omega})$ and, thus

$$E_s = \mathbf{h}^T\mathbf{\Phi}_s\mathbf{h} \quad (24)$$

where the (i,j)th element of the matrix $\mathbf{\Phi}_s$ is obtained as

$$\begin{aligned} \mathbf{\Phi}_s(i,j) &= \frac{1}{2\pi}\int_{\omega_s}^{2\pi-\omega_s} e^{-j\omega(i-j)}d\omega \\ &= \begin{cases} 1-\frac{\omega_s}{\pi}, & i=j \\ -\frac{\sin[\omega_s(i-j)]}{\pi(i-j)}, & i\neq j. \end{cases} \end{aligned} \quad (25)$$

Similarly, we express E_p as

$$E_p = \mathbf{h}^T\mathbf{\Phi}_p\mathbf{h} \quad (26)$$

where

$$\begin{aligned} \mathbf{\Phi}_p(i,j) &= \frac{1}{2\pi}\int_{-\omega_p}^{\omega_p} (e^{-j\omega(i-j)}-1)d\omega \\ &= \begin{cases} 0, & i=j \\ \frac{1}{2\pi}\left(\frac{\sin[\omega_p(i-j)]}{(i-j)} - \omega_p\right), & i\neq j. \end{cases} \end{aligned} \quad (27)$$

Substituting (24) and (26) into (23), we obtain

$$E = \mathbf{h}^T\mathbf{\Phi}\mathbf{h} \quad (28)$$

with $\mathbf{\Phi} = (1-\alpha)\mathbf{\Phi}_s + \alpha\mathbf{\Phi}_p$.

In summary, the optimization problem is to find a prototype filter $\mathbf{h}$ that minimizes $\mathbf{h}^T\mathbf{\Phi}\mathbf{h}$ and at the same time satisfies (22). We used the MATLAB optimization routine *fmincon* to solve this problem. The *fmincon* routine uses the sequential quadratic programming method to find the constrained minimum multivariate function. For our case, the inputs to *fmincon* are the cost function $E = \mathbf{h}^T\mathbf{\Phi}\mathbf{h}$ and the constraint (22). We also have the option of adding the first derivative of E, i.e., $\nabla_{\mathbf{h}}E = 2\mathbf{\Phi}\mathbf{h}$, as an additional input. This reduces the CPU time of the design. The design procedure we employed is as follows.

1) Given M, the middle of the transition band of the prototype filter is located at $\pi/2M$ [15], [20]. Choose ω_s in the range of $((\pi/2M), (\pi/M))$. Set α to 0 and calculate $\mathbf{\Phi} = \mathbf{\Phi}_s$ as in (25). Compute the quadratic objective function $\mathbf{h}^T\mathbf{\Phi}\mathbf{h}$ and its first order derivative to be used in the MATLAB routine *fmincon*.
2) Given N and Δ, calculate $\mathbf{S}_n$ as in (20) with $n = \Delta + p2M$ and $p \in [-\lfloor\Delta/2M\rfloor, \lfloor(2(N-1)-\Delta)/2M\rfloor]$. Compute the quadratic constraints (22) and their first-order derivative to be used in *fmincon*.
3) Design a low-pass filter $h_0(n)$ approximating the frequency specifications of the prototype filter using the MATLAB routine *fir1*. Use this filter to initialize the iterations in *fmincon*.
4) Call MATLAB routine *fmincon* to solve the constrained quadratic optimization problem. The *TolCon* option was set to 10^{-13} in our design.
5) If the deviation from flat magnitude response is unacceptably large for the design of the prototype filter (this usually happens when Δ approaches $N/2$), choose ω_p in the range of $(0, (\pi/2M))$ and calculate $\mathbf{\Phi}_p$ as in (27). Choose a small value for α (usually in the order of 0.0001) and recalculate $\mathbf{\Phi} = (1-\alpha)\mathbf{\Phi}_s + \alpha\mathbf{\Phi}_p$. Compute the new objective function $\mathbf{h}^T\mathbf{\Phi}\mathbf{h}$ and its first-order derivative. Repeat step 4) with the new changes. If the new result is unacceptable, try increasing α until a satisfying pass-band response is achieved.

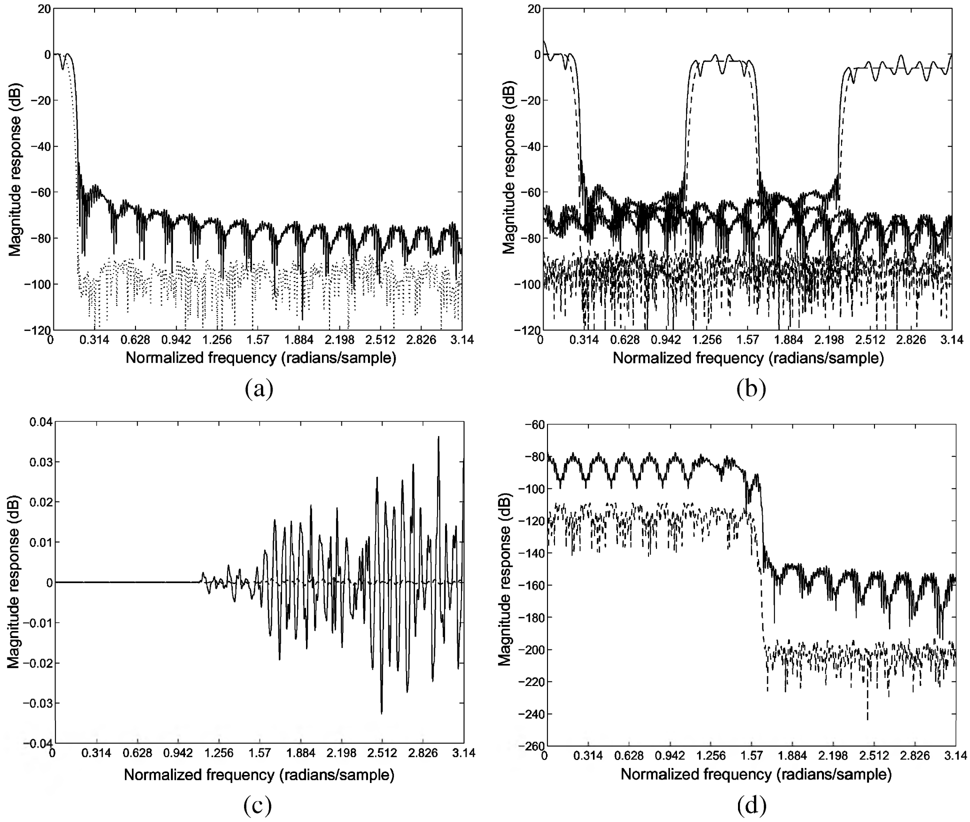


Fig. 3. Comparison of the two filter bank designs in the example: Solid line: $\Delta = 192$, Dashed line: linear phase design with $\Delta = 383$. (a) Magnitude response of the prototype filter $H(z)$. (b) Magnitude response of the nonuniform analysis filters $H_k(z), k = 1, 7, 9$. (c) Magnitude response of the overall distortion $T_0(z)$. (d) Magnitude response plots for the aliasing transfer functions $T_l(z), l = 1$.

IV. Design Example and an Application to Speech Enhancement

A. Design of a Maximally Decimated Pseudo-QMF Bank

We present the results of designing a nine-channel nonuniform filter bank that was obtained from combining sub-bands of a 16-channel pseudo-QMF bank. The first six channels of the nonuniform filter bank were the same as those in the uniform filter bank. The seventh channel of the nonuniform filter bank was created by combining the seventh and eighth channels, the eighth by combining the 9th–12th channels and the ninth channel by combining the 13th–16th channels of the uniform filter bank. The parameters of the prototype filter were $N = 384$ and $\omega_s = 0.059\pi$. Fig. 3 presents the design results of a filter bank with a delay of 192 samples, $\omega_p = 0.03\pi$ and $\alpha = 0.00025$ (solid line) and a linear phase design according to [20] with a delay of 383 samples (dashed line). The plots presented in Fig. 3(a)–(d) are the magnitude responses of the prototype filter $H(z)$, the analysis filters $H_k(z)$, $k = 1$, 7, and 9, the overall distortion function $T_0(z)$, and the aliasing transfer functions $T_l(z)$, $l = 1$, respectively. Comparing the magnitude responses in Fig. 3(a), we can see that the lower delay was obtained at the cost of lower stopband attenuation. Moreover, the reduced delay design caused a magnitude boost in the transition band, and larger pass-band ripples. However, the pass-band ripples are completely compensated for by the synthesis filter bank and will cause little amplitude distortion as can be seen from Fig. 3(c) (where combining is done, some small amplitude distortion is observed). As the delay gets smaller, the amplitude distortion and aliasing gets larger. The maximum amplitude distortion in this example is less than 0.04 dB and the aliasing distortion is below -72 dB for all cases. Such distortions are negligible and, thus, acceptable in a variety of applications. When we processed speech signals with the analysis and synthesis filter banks of this example, there were no audible differences between input and output signals.

B. Application to Speech Enhancement

Cohen [8] proposed a bark-scaled wavelet packet decomposition with discrete Meyer wavelet filters for speech enhancement. The bark-scaled wavelet packet decomposition is obtained by cascading each sub-band of the 21 bands of the critical-band wavelet packet tree with two more levels of nondecimated filtering as shown in Fig. 4(a). The two levels of oversampled decomposition exist for the outputs of all maximally decimated filter bank outputs even though only one set of such decompositions are shown in the figure. The last two levels of the decomposition increase the frequency resolution by four without reducing the time resolution at the output. It was shown that this decomposition outperforms a critical-band wavelet packet decomposition and a uniform decomposition when applied to a modified version of Ephraim and Malah's speech enhancement algorithm [2]. We used the same speech enhancement algorithm in this example. Details of the algorithm are available in [8]. In [8], the noise spectrum was assumed to be known. In our paper,

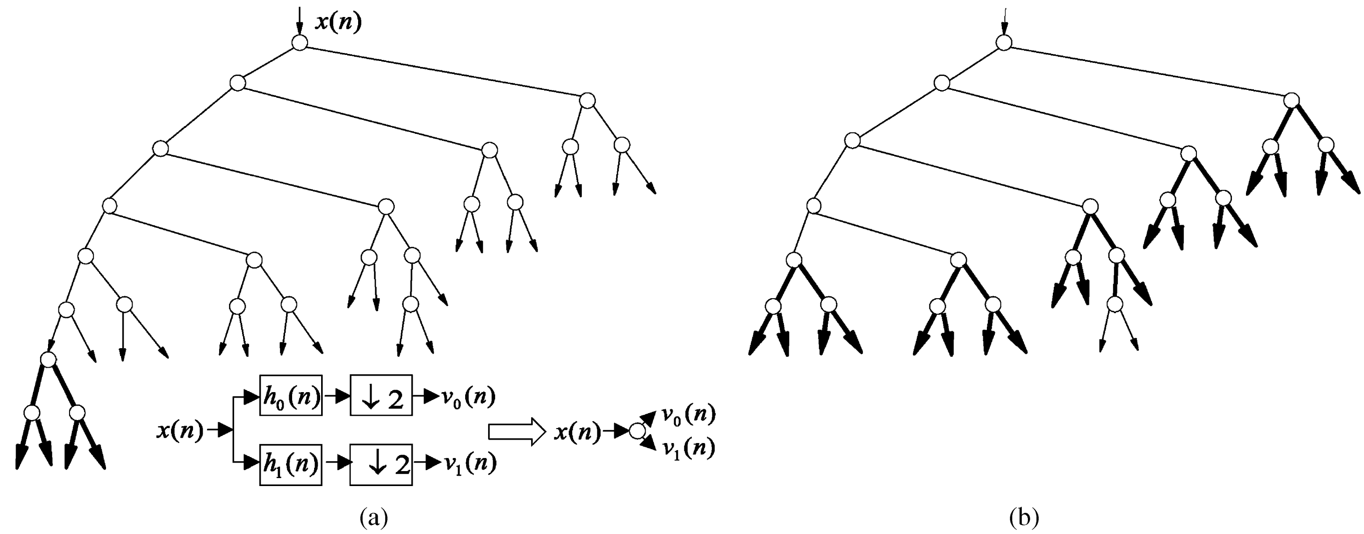


Fig. 4. (a) Bark scaled decomposition with 84 bands and oversampling by four. (a) The two levels of oversampled decomposition exist for the outputs of all maximally decimated filter bank outputs even though only one set of such decompositions are shown. (b) Critical band decomposition with 21 bands and oversampling by four.

the noise spectrum was estimated using the minimum controlled recursive averaging method [5].

To implement the enhancement system, we designed a low-delay oversampled nonuniform filter bank that achieves critical-band decomposition. In order to achieve the frequency resolution of a critical-band decomposition, we first designed a 64-band uniform filter bank. The design parameters of the prototype filter were $N = 896$ and $\omega_s = 0.0147\pi$. In order to achieve high enough stopband attenuation, we chose a delay of 512 samples which is slightly larger than half the filter length. For the design, we choose $\alpha = 0$. The overall magnitude distortion of this filter bank was less than 0.15 dB. The maximum aliasing distortion was less than -70 dB. In order to compare the capabilities of our system with tree-structured filter banks, we constructed two speech enhancement systems using the oversampled critical-band wavelet packet decomposition in Fig. 4(b) with the Daubechies wavelet basis of length 12 and the discrete Meyer wavelet filters of length 62, respectively. We now compare the performance of the speech enhancement system using the design in this paper to those using the critical-band wavelet packet decomposition and the bark-scaled wavelet packet decomposition [8]. The oversampling ratio was four for all the systems.

The evaluation consists of an objective quality measure based on the segmental signal-to-noise ratio (SNR) improvement and informal subjective listening tests. The segmental SNR is a widely used objective measure for speech enhancement systems and is defined as [12]

$$\mathrm{SegSNR} = \frac{1}{M}\sum_{m=0}^{M-1} 10\log_{10} \times \left(\frac{\sum\limits_{n=0}^{L-1} |s(n+mL)|^2}{\sum\limits_{n=0}^{L-1} |s(n+mL)-\hat{s}(n+mL)|^2} \right) \tag{29}$$

where $s(n)$ and $\hat{s}(n)$ denote the clean speech and the enhanced speech, respectively. Here, M is the number of frames in the speech segment and L is the number of samples per frame. Four speech signals, two male and two female, were used in our tests. Four different types of noise, a white Gaussian noise, an F16 cockpit noise, a factory noise, and a traffic noise, were added to the clean speech signals at different segmental SNRs ranging from -5 to 10 dB. The segmental SNR improvement was estimated by subtracting the SegSNR of the enhanced speech from the SegSNR associated with the noisy speech. The results shown in Fig. 5 are averages of the SegSNRs obtained for the four speech signals contaminated with each type of noise. All the speech and noise signals were sampled at 16 kHz.

From Fig. 5, we can see that the critical-band wavelet packet decomposition using the discrete Meyer basis achieved a better speech enhancement performance than the system equipped with the shorter Daubechies wavelet basis. This is because of the higher stopband attenuation and the correspondingly lower aliasing distortion provided by the Meyer basis filter. The delay associated with the oversampled critical-band wavelet packet transform with Daubechies wavelet basis with 12 coefficients and the discrete Meyer wavelet basis with 62 coefficients were 517 samples and 2867 samples, respectively. The method of Cohen [8] performed better than the other three systems in low SNR cases and worse in high SNR cases. In [8], it was shown that when the noise spectrum is known, the bark-scaled wavelet packet decomposition performs better than the critical-band wavelet packet decomposition in both high and low SNR cases due to its higher frequency resolution. In practice, the noise spectrum is unknown and the minimum controlled recursive averaging or another approach is used to estimate the noise spectrum [12]. In high SNR cases, higher frequency resolution decompositions tend to over estimate the noise spectrum and result in speech distortion. This causes the lower segmental SNR improvement for the bark-scaled wavelet packet decomposition in high SNR cases. The delay associated with the

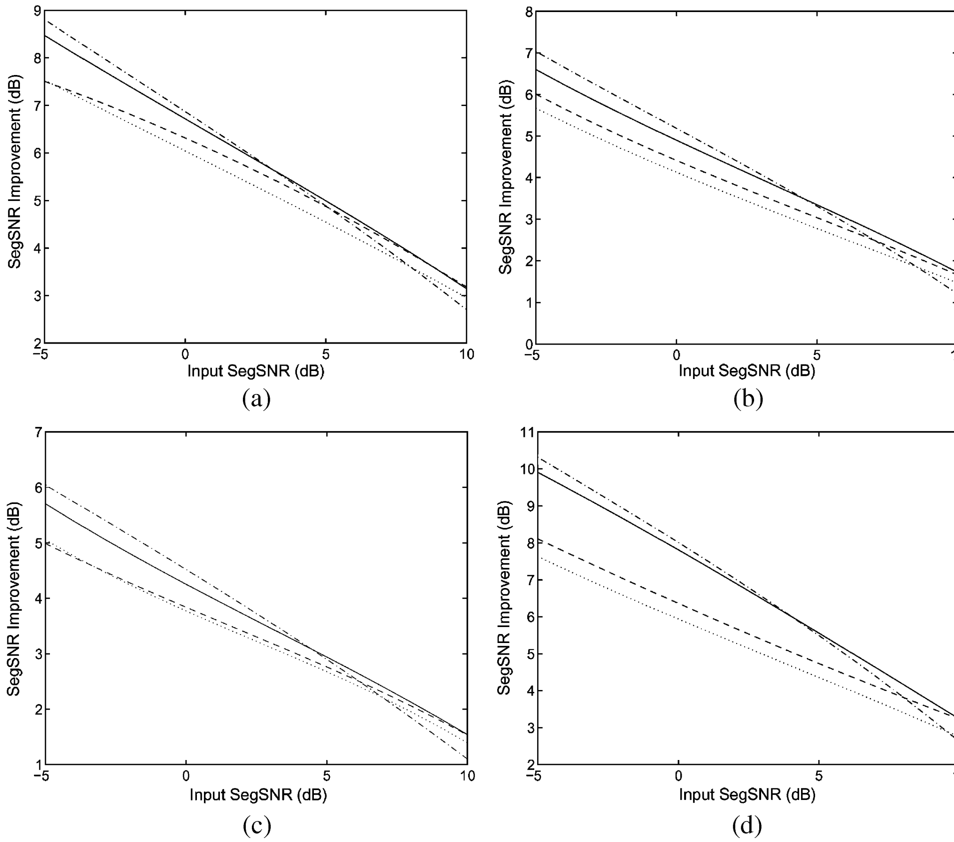


Fig. 5. Average segmental SNR improvement. (a) White Gaussian noise. (b) F16 cockpit noise. (c) Factory noise. (d) Traffic noise. Solid line: method of this paper. Dash-dotted line: bark-scaled wavelet packet decomposition with discrete Meyer basis. Dotted line: critical-band wavelet packet with Daubechies wavelet basis of length 12. Dashed line: critical-band wavelet packet with discrete Meyer wavelet basis.

bark-scaled wavelet packet transform in Cohen's work is 11651 samples, which is more than 728 ms for a sampling rate of 16 kHz. The system of this paper achieved better performance than the two critical-band wavelet packet decomposition with the same frequency resolution, especially in low SNR cases. Our method performed slightly poorer than the bark-scaled wavelet packet decomposition with the performance difference being less than 0.4 dB in low SNR cases. However, our method achieves a much smaller delay of 512 samples (32 ms). On the basis of these comparisons, we believe that the system of this paper achieved the best compromise between segmental SNR improvement and processing delay among all the methods tested.

In order to see the performance loss due to even lower delay, we designed low-delay critical-band filter banks with delays of 256 and 128 samples, respectively. The design parameters of the prototype filters were $\omega_s = 0.0147\pi$ and $\omega_p = 0.0075\pi$. For delay values of 256 and 128 samples, we choose $N = 896$, $\alpha = 0.005$ and $N = 256$, $\alpha = 0.001$, respectively. For a comparison to the linear phase design [19], [20], we also designed the linear phase critical-band filter banks of delays 512, 256, and 128 samples. We then performed the same speech enhancement experiments for these critical-band decompositions as before. We averaged the SegSNRs improvements obtained for the 16 possible combinations of the four speech signals contaminated with the four types of noise. The result is shown in Fig. 6. In Fig. 6, LD represents low-delay designs and LP represents linear phase designs. From the plots, we can see that for low-delay designs reducing the delay from 512 to 256 samples, the maximum SegSNR degradation is less than 0.5 dB. A maximum SegSNR degradation of less than 1 dB is observed when the delay is further reduced from 256 to 128 samples. The low-delay designs consistently achieve higher SegSNR improvements than the linear phase designs. The differences between the low-delay and linear phase designs are not significant for large delay values. They become pronounced at lower delay values as can be seen from Fig. 6. As delay gets lower, the linear phase design has shorter filter and lower stopband attenuation. These cause higher amplitude and aliasing distortions in the combining processing of the nonuniform filter bank. Simply processing the clean speech signals with the linear phase critical-band filter bank of delay 128 samples resulted in audible distortions in the output signals.

To further compare the capabilities of the new design technique with the tree-structured filter banks, an informal listening test based on mean opinion scores (MOS) was performed with twelve subjects. A female speech signal, "A lathe is a big tool; Grab every dish of sugar,"contaminated with the previous four types of noise and then enhanced using different decomposition methods were used for the test. Three input SegSNRs of 0, 5, and 10 dB were selected to present the results. Table I shows the informal mean opinion scores obtained by averaging the results provided by the twelve subjects. Here, 5 represents the determination (by the subject) that the speech signal con-

TABLE I
INFORMAL MOS COMPARISON

Noise type	Input SegSNR (dB)	Informal MOS Scores						
		Noisy Signal	Enhanced Signal					
			Cohen [8] delay 11651	Meyer 62 delay 2867	Daubechies 12 delay 517	Low Delay delay 512	Low Delay delay 256	Low Delay delay 128
White Gaussian Noise	0	1.79	2.46	2.29	1.67	2.42	2.33	2.17
	5	1.92	3.63	3.42	2.63	3.46	3.33	3.33
	10	2.13	3.50	3.96	3.42	4.04	4.00	3.50
F16 Cockpit Noise	0	1.63	2.00	1.96	1.42	1.96	1.92	1.88
	5	1.88	2.96	2.42	2.29	2.83	2.71	2.33
	10	2.25	3.71	3.63	3.08	3.96	3.54	3.13
Factory Noise	0	1.58	2.29	1.63	1.50	2.13	1.79	1.63
	5	1.83	3.33	2.38	2.33	2.92	2.71	2.54
	10	2.54	3.75	3.71	3.21	3.83	3.71	3.63
Traffic Noise	0	1.67	3.13	2.75	1.63	2.83	2.50	2.08
	5	2.08	4.29	2.92	2.33	4.13	3.50	3.25
	10	2.83	4.79	3.67	2.96	4.88	4.38	4.25

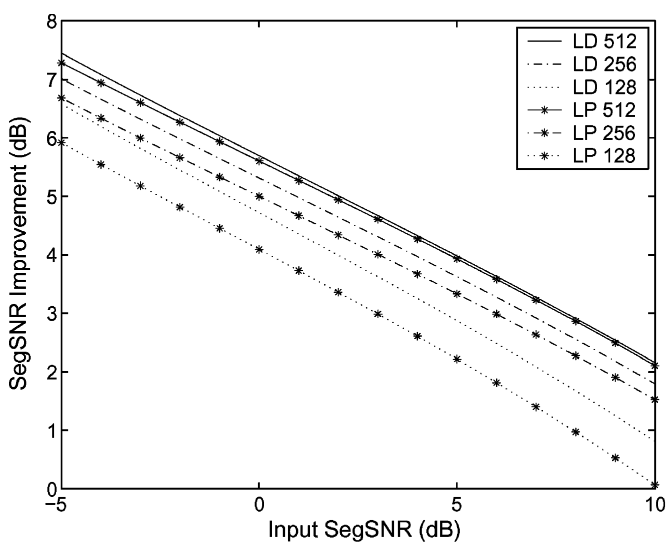


Fig. 6. Average segmental SNR improvement comparison for different delays.

tains no audible distortions and it is virtually identical to the clean speech segment. A score of 1 indicates that the distortions and residual noise in the speech signal are very high and the segment sounds unacceptably annoying. From the informal MOS scores, we can see that the three low-delay designs consistently provided better speech quality when compared with the critical-band wavelet packet decomposition using Daubechies wavelet basis of length 12. We note that the delay associated with the Daubechies wavelet basis was the smallest among the competing methods, but still at least slightly larger than the three low-delay designs. The speech enhancement system employing the decomposition method using Daubechies wavelet basis of length 12 created high aliasing distortion. Even though it provides an SegSNR improvement, the informal MOS scores show that the aliasing presented in the enhanced signals sounds more annoying than the input noisy speech signals at 0 dB SegSNR cases. As compared to the speech enhancement system based on the critical-band decomposition using Meyer wavelet basis, the low-delay design of delay 512 samples showed superior performance and the low-delay designs of delays 256 and 128 samples gave comparable performance. Our method exhibited slightly higher residual noise in low SegSNR cases as compared to Cohen's bark-scaled wavelet packet decomposition. However, as discussed before, the higher frequency resolution of Cohen's method [8] tends to over estimate the noise spectrum in high SNR cases and results in speech distortion. This can also be seen from the comparison of the MOS scores of Cohen's method with the low-delay design of delay 512 samples at input SegSNR of 10 dB cases.

V. CONCLUSION

This paper presented an approach for designing low-delay nonuniform filter banks. The low delay is achieved by relaxing the linear phase constraints of traditional pseudo-QMF banks. A low delay oversampled critical-band division filter bank was designed and applied to a two-state modeling speech enhancement system. The performance of the proposed system was compared to that of using tree-structured filter banks employing wavelet bases. According to the objective SegSNR comparison and the subjective MOS test, we can see that this paper's method strikes a good balance between delay and system performance in this application.

APPENDIX A

Let us first prove that $F(z) = cH(z)$ is a necessary and sufficient condition for significant aliasing cancellation. Define $U_k(z) = c_k H(zW_{2M}^{(k+0.5)})$, $V_k(z) = c_k^* H(zW_{2M}^{-(k+0.5)})$, and $U_k'(z) = c_k F(zW_{2M}^{(k+0.5)})$, $V_k'(z) = c_k^* F(zW_{2M}^{-(k+0.5)})$. (11) and (12) can be rewritten as

$$H_k(z) = a_k U_k(z) + a_k^* V_k(z) \quad (30)$$

$$F_k(z) = b_k U_k'(z) + b_k^* V_k'(z). \quad (31)$$

Using the same analysis in [15], the significant aliasing term is canceled if and only if

$$a_k b_k^* U_k\left(zW_M^{-k}\right) V_k'(z) + a_{k-1} b_{k-1}^* U_{k-1}\left(zW_M^{-k}\right) V_{k-1}'(z) = 0, \quad 1 \le k \le M-1. \quad (32)$$

Notice that $U_k(zW_M^{-k}) = c_k c_{k-1} V_{k-1}(z)$ and $U_{k-1}(zW_M^{-k}) = c_k c_{k-1} V_k(z)$, (32) can be simplified as

$$V_{k-1}(z)V_k'(z) - V_k(z)V_{k-1}'(z) = 0, \quad 1 \le k \le M-1 \quad (33)$$

or equivalently as

$$\frac{F\left(zW_{2M}^{-(k+0.5)}\right)}{H\left(zW_{2M}^{-(k+0.5)}\right)} = \frac{F\left(zW_{2M}^{-(k-0.5)}\right)}{H\left(zW_{2M}^{-(k-0.5)}\right)}, \quad 1 \le k \le M-1. \quad (34)$$

(34) is satisfied only if $F(z) = cH(z)$ with c being some arbitrary constant.

We now show that $T_0(z) = cz^{-\Delta}$. Using the facts that $a_k b_k^* + a_k^* b_k = 0$ and $a_k b_k = a_k^* b_k^* = 1$ for the choice of the parameters, it is relatively straightforward to show that

$$\begin{aligned} T_0(z) &= \frac{1}{M}\sum_{k=0}^{M-1} H_k(z)F_k(z) \\ &= c\frac{1}{M}\sum_{k=0}^{M-1}\left[U_k^2(z) + V_k^2(z)\right] \\ &= c\frac{1}{M}\sum_{k=0}^{M-1}\left[c_k^2 H^2\left(zW_{2M}^{(k+0.5)}\right)\right. \\ &\qquad \left. + \left(c_k^2\right)^* H^2\left(zW_{2M}^{-(k+0.5)}\right)\right]. \end{aligned} \quad (35)$$

Since $c_k = W_{2M}^{\Delta(k+0.5)/2}$, we can write

$$\begin{aligned} T_0(z) &= c\frac{1}{M}\sum_{k=0}^{M-1}\left[W_{2M}^{\Delta(k+0.5)} H^2\left(zW_{2M}^{(k+0.5)}\right)\right. \\ &\qquad \left. + W_{2M}^{-\Delta(k+0.5)} H^2\left(zW_{2M}^{-(k+0.5)}\right)\right] \\ &= c\frac{1}{M}\sum_{k=0}^{M-1}\left[W_{2M}^{\Delta(k+0.5)} H^2\left(zW_{2M}^{(k+0.5)}\right)\right. \\ &\qquad + W_{2M}^{\Delta(2M-1-k+0.5)} H^2 \\ &\qquad \left. \times \left(zW_{2M}^{(2M-1-k+0.5)}\right)\right] \\ &= c\frac{1}{M}\sum_{k=0}^{2M-1} W_{2M}^{\Delta(k+0.5)} H^2\left(zW_{2M}^{(k+0.5)}\right). \end{aligned} \quad (36)$$

Substituting $G(z) = H^2(z) = \sum_{n=0}^{2(N-1)} g(n)z^{-n}$ in (36), we get

$$\begin{aligned} T_0(z) &= c\frac{1}{M}\sum_{k=0}^{2M-1}\sum_{n=0}^{2(N-1)} W_{2M}^{(\Delta-n)(k+0.5)} g(n)z^{-n} \\ &= c\frac{1}{M}\sum_{n=0}^{2(N-1)} g(n)z^{-n}\sum_{k=0}^{2M-1} W_{2M}^{(\Delta-n)(k+0.5)}. \end{aligned} \quad (37)$$

Since

$$\sum_{k=0}^{2M-1} W_{2M}^{(\Delta-n)(k+0.5)} = \begin{cases} (-1)^p 2M, & n = \Delta + p2M \\ 0, & \text{otherwise} \end{cases} \quad (38)$$

and

$$g(\Delta + p2M) = \begin{cases} 1/2, & p = 0 \\ 0, & \text{otherwise} \end{cases} \quad (39)$$

(37) reduces to

$$T_0(z) = cz^{-\Delta}. \quad (40)$$

Appendix B

In this appendix, we prove that the significant aliasing cancellation condition (33) is also a sufficient condition for significant distortion cancellation in nonuniform filter bank. To simplify the proof, we assume there is only one nonuniform channel, the ith channel, generated from combining the ith untill the $i + k' - 1$ channels of an M-channel uniform low-delay pseudo-QMF bank. For feasible partition filter bank, we know $i = k'\nu$ for some integer ν. The proof can be extended to more than one nonuniform band in a straightforward manner. The proof starts by first converting the ith branch to an equivalent k'-branch uniform filter bank using the technique originally described in [27] and also presented in [19]. The nonuniform filter bank distortion function $T_0^{NU}(z)$ can be written as

$$\begin{aligned} T_0^{NU}(z) &= \frac{1}{M}\sum_{\substack{k=0 \\ k\notin\{i,\cdots,i+k'-1\}}}^{M-1} H_k(z)F_k(z) \\ &\quad + \frac{1}{M}\sum_{k=i}^{i+k'-1}\left[\left(\frac{1}{\sqrt{k'}}\sum_{j=0}^{k'-1} H_{i+j}(z)\right) z^{-\frac{M}{k'}(k-i)}\right] \\ &\quad \times \left[\left(\frac{1}{\sqrt{k'}}\sum_{j=0}^{k'-1} F_{i+j}(z)\right) z^{\frac{M}{k'}(k-i)}\right] \\ &= \frac{1}{M}\sum_{k=0}^{M-1} H_k(z)F_k(z) \\ &\quad + \frac{1}{M}\sum_{j=0}^{k'-1}\sum_{\substack{m=0 \\ m\neq j}}^{k'-1} H_{i+j}(z)F_{i+m}(z). \end{aligned} \quad (41)$$

For the low-delay uniform pseudo-QMF bank, we know from Appendix A that $1/M\sum_{k=0}^{M-1} H_k(z)F_k(z) = cz^{-\Delta}$. Therefore, the distortion associated with the nonuniform filter bank is caused by

$$D(z) = \frac{1}{M}\sum_{j=0}^{k'-1}\sum_{\substack{m=0 \\ m\neq j}}^{k'-1} H_{i+j}(z)F_{i+m}(z). \quad (42)$$

The significant distortion is caused by adjacent bands, i.e., $|m - j| = 1$. When $|m - j| > 1$, $H_{i+j}(z)$, and $F_{i+m}(z)$ fall in the stopband of each other. If the stopband attenuation of the prototype filter is high, the distortion caused by nonadjacent band will be small. The distortion caused by adjacent bands in (42) can then be expressed as

$$D_a(z) = \frac{1}{M}\sum_{j=0}^{k'-2}\left[H_{i+j}(z)F_{i+j+1}(z) + H_{i+j+1}(z)F_{i+j}(z)\right]. \tag{43}$$

Using the definitions of $H_l(z)$ and $F_m(z)$ from (30) and (31) and employing the values of a_k and b_k from Section III-A in (43) and further suppressing the cross terms that fall in the stop-band of each other, such as $U_{i+j}(z)V'_{i+j+1}(z)$, the significant distortion can be expressed as

$$\begin{aligned} D_s(z) =& \frac{1}{M}\sum_{j=0}^{k'-2}(-1)^{i+j}j \\ &\times \left[U_{i+j}(z)U'_{i+j+1}(z) - U_{i+j+1}(z)U'_{i+j}(z)\right] \\ &- \frac{1}{M}\sum_{j=0}^{k'-2}(-1)^{i+j}j \\ &\times \left[V_{i+j}(z)V'_{i+j+1}(z) - V_{i+j+1}(z)V'_{i+j}(z)\right]. \end{aligned} \tag{44}$$

According to (33), $U_k(z) = V_k^*(z)$ and $U_k(z) = V_k'^*(z)$, it is straightforward to get $D_s(z) = 0$.

Acknowledgment

The authors would like to thank R. Malvar for his careful reading of an earlier draft and his many suggestions for improvement.

References

[1] S. Boll, "Suppression of acoustic noise in speech using spectral subtraction," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-27, no. 2, pp. 113–120, Apr. 1979.

[2] Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean square error short time spectral amplitude estimator," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-32, no. 6, pp. 1109–1121, Dec. 1984.

[3] ——, "Speech enhancement using a minimum mean square error log spectral amplitude estimator," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-33, no. 2, pp. 443–445, Apr. 1985.

[4] N. Virag, "Single channel speech enhancement based on masking properties of the human auditory system," *IEEE Trans. Speech Audio Process.*, vol. 7, no. 2, pp. 126–137, Mar. 1999.

[5] I. Cohen and B. Berdugo, "Speech enhancement for non-stationary noise environments," *Signal Process.*, vol. 81, no. 11, pp. 2403–2418, Nov. 2001.

[6] T. Gülzow, A. Engelsberg, and U. Heute, "Comparison of a discrete wavelet transformation and a nonuniform polyphase filterbank applied to spectral-subtraction speech enhancement," *Signal Process.*, vol. 64, no. 1, pp. 5–19, Jan. 1998.

[7] T. Gülzow, T. Ludwig, and U. Heute, "Spectral-subtraction speech enhancement in multirate systems with and without non-uniform and adaptive bandwidths," *Signal Process.*, vol. 83, no. 8, pp. 1613–1631, Aug. 2003.

[8] I. Cohen, "Enhancement of speech using bark-scaled wavelet packet decomposition," in *Proc. 7th Eur. Conf. Speech, Commun. Technol.*, 2001, pp. 1933–1936.

[9] D. Donoho and I. Johnstone, "Ideal spatial adaptation via wavelet shrinkage," *Biometrika*, vol. 81, pp. 425–455, Sep. 1994.

[10] B. Camero and A. Drygajlo, "Perceptual speech coding and enhancement using frame-synchronized fast wavelet packet transform algorithms," *IEEE Trans. Signal Process.*, vol. 47, no. 6, pp. 1622–1635, Jun. 1999.

[11] M. Bahoura and J. Rouat, "Wavelet speech enhancement based on the teager energy operator," *IEEE Signal Process. Lett.*, vol. 8, no. 1, pp. 10–12, Jan. 2001.

[12] C. Lu and H. Wang, "Enhancement of single channel speech based on masking property and wavelet transform," *Speech Commun.*, vol. 41, no. 2–3, pp. 409–427, Oct. 2003.

[13] S. Chen and J. Wang, "Speech enhancement using perceptual wavelet packet decomposition and teager energy operator," *J. VLSI Signal Process. Syst. Signal, Image, Video Technol.*, vol. 36, no. 2–3, pp. 125–139, Feb.–Mar. 2004.

[14] J. Yang and S. Krishnan, "Wavelet packets-based speech enhancement for hearing aids application," *Can. Acoust.*, vol. 33, no. 3, pp. 66–67, Sep. 2005.

[15] P. Vaidyananthan, *Multirate Systems and Filter Banks*. Upper Saddle River, NJ: Prentice-Hall, 1993.

[16] Z. Cvetkovic and J. Johnston, "Nonuniform oversampled filler banks for audio signal processing," *IEEE Trans. Speech Audio Process.*, vol. 11, no. 5, pp. 393–399, Sep. 2003.

[17] I. Daubechies, *Ten Lectures on Wavelets*. Philadelphia, PA: SIAM, 1992.

[18] M. Wickerhauser, *Adapted Wavelet Analysis From Theory to Software*. Wellesley, MA: A.K. Peters, 1994.

[19] J. Li, T. Nguyen, and S. Tantaratana, "A simple design method for near-perfect-reconstruction nonuniform filter banks," *IEEE Trans. Signal Process.*, vol. 45, no. 8, pp. 2105–2109, Aug. 1997.

[20] T. Nguyen, "Near-perfect-reconstruction pseudo-qmf banks," *IEEE Trans. Signal Process.*, vol. 42, no. 1, pp. 65–76, Jan. 1994.

[21] N. Chong, I. Burnett, and J. Chicharo, "A new waveform interpolation coding scheme based on pitch synchronous wavelet transform decomposition," *IEEE Trans. Speech Audio Process.*, vol. 8, no. 3, pp. 345–348, May 2000.

[22] P. Heller, T. Karp, and T. Nguyen, "A general formulation of modulated filler banks," *IEEE Trans. Signal Process.*, vol. 47, no. 4, pp. 986–1002, Apr. 1999.

[23] G. Schuller and T. Karp, "Modulated filter banks with arbitrary system delay: Efficient implementations and the time varying case," *IEEE Trans. Signal Process.*, vol. 48, no. 3, pp. 737–748, Mar. 2000.

[24] J. Kliewer and A. Mertins, "Oversampled cosine-modulated filter banks with arbitrary system delay," *IEEE Trans. Signal Process.*, vol. 46, no. 4, pp. 941–955, Apr. 1998.

[25] Y. Deng, V. J. Mathews, and B. Farhang-Boroujeny, "The design of low-delay nonuniform pseudo QMF banks," presented at the Eur. Signal Process. Conf., Florence, Italy, 2006.

[26] J. Kovacevic and M. Vetterli, "Perfect reconstruction filter banks with rational sampling factors," *IEEE Trans. Signal Process.*, vol. 41, no. 6, pp. 2047–2066, Jun. 1993.

[27] P. Hoang and P. Vaidyanathan *et al.*, "Non-uniform multirate filter banks: Theory and design," in *Proc. ISCAS*, 1989, pp. 371–374.

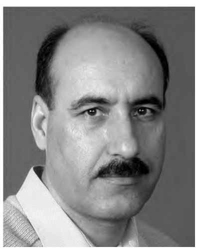
Ying Deng (S'06–M'07) received the M.S. degree from Zhejiang University, Hangzhou, China, in 2000, and the Ph.D. degree, both in electrical engineering, from the University of Utah, Salt Lake City, in 2006.

Her research interests include the general area of speech processing, especially in the area of developing nonuniform multirate filter banks and particle filtering algorithms to speech enhancement applications.

V. John Mathews (S'82–M'84–SM'90–F'02) received the B.E. (honors) degree in electronics and communication engineering from the University of Madras, Madras, India, in 1980, and the M.S. and Ph.D. degrees in electrical and computer engineering from the University of Iowa, Iowa City, in 1981 and 1984, respectively.

At the University of Iowa, he was a Teaching/Research Fellow from 1980 to 1984 and a Visiting Assistant Professor in the Department of Electrical and Computer Engineering during the 1984-1985 academic year. He joined the University of Utah in 1985, where he is a Professor in the Department of Electrical and Computer Engineering. He served as the chairman of the department from 1999 to 2003. His research interests include adaptive and nonlinear filtering and application of signal processing techniques in communication, audio, aerospace, and biomedical engineering. He is the author of the book *Polynomial Signal Processing* (Wiley) and coauthored with Prof. G. L. Sicuranza, University of Trieste, Italy. He has published more than 100 technical papers.

Dr. Mathews served as the Vice President-Finance of the IEEE Signal Processing Society during 2003-2005, and has served on the Publications Board and the Conference Board of the Society. He is a past Associate Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING, and the IEEE SIGNAL PROCESSING LETTERS. He serves on the editorial boards of the *IEEE Signal Processing Magazine* and the IEEE JOURNAL ON SPECIAL TOPICS IN SIGNAL PROCESSING. He was a member of the Signal Processing Theory and Methods and Education Technical Committees and served as the General Chairman of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP) 2001.



Behrouz Farhang-Boroujeny (M'84–SM'90) received the B.Sc. degree in electrical engineering from Teheran University, Iran, in 1976, the M.Eng. degree from University of Wales Institute of Science and Technology, U.K., in 1977, and the Ph.D. degree from Imperial College, University of London, U.K., in 1981.

From 1981 to 1989, he was with the Isfahan University of Technology, Isfahan, Iran. From 1989 to 2000, he was with the National University of Singapore. Since August 2000, he has been with the University of Utah, Utah City, where he is now a Professor and Associate Chair of the department. He is an expert in the general area of signal processing. His current scientific interests include adaptive filters, multicarrier communications, detection techniques for space-time coded systems, cognitive radio, and signal processing applications to optical devices. In the past, he has worked and has made significant contribution to areas of adaptive filters theory, acoustic echo cancellation, magnetic/optical recoding, and digital subscriber line technologies. He is the author of the book *Adaptive Filters: Theory and Applications* (Wiley, 1998).

Dr. Farhang-Boroujeny was a recipient of the UNESCO Regional Office of Science and Technology for South and Central Asia Young Scientists Award in 1987. He served as Associate Editor of IEEE TRANSACTIONS ON SIGNAL PROCESSING from July 2002 to July 2005. He has also been involved in various IEEE activities, including the chairmanship of the Signal Processing/Communications chapter of IEEE of Utah in 2004 and 2005.