

## Research Article

# Synthesis of an Optimal Wavelet Based on Auditory Perception Criterion

Abhijit Karmakar,<sup>1</sup> Arun Kumar,<sup>2</sup> and R. K. Patney<sup>3</sup>

<sup>1</sup>Integrated Circuit Design Group, Central Electronics Engineering Research Institute/Council of Scientific and Industrial Research, Pilani 333031, India

<sup>2</sup>Centre for Applied Research in Electronics, Indian Institute of Technology Delhi, New Delhi 110016, India

<sup>3</sup>Department of Electrical Engineering, Indian Institute of Technology Delhi, New Delhi 110016, India

Correspondence should be addressed to Abhijit Karmakar, abhijit.karmakar@gmail.com

Received 2 July 2010; Revised 3 November 2010; Accepted 4 February 2011

Academic Editor: Antonio Napolitano

Copyright © 2011 Abhijit Karmakar et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

A method is proposed for synthesizing an optimal wavelet based on auditory perception criterion for dyadic filter bank implementation. The design method of this perceptually optimized wavelet is based on the critical band (CB) structure and the temporal resolution of human auditory system (HAS). The construction of this compactly supported wavelet is done by designing the corresponding optimal FIR quadrature mirror filter (QMF). At first, the wavelet packet (WP) tree is obtained that matches optimally with the CB structure of HAS. The error in passband energy of the CB channel filters is minimized with respect to the ideal QMF. The optimization problem is formulated in the lattice QMF domain and solved using bounded value global optimization technique. The corresponding wavelet is obtained using the cascade algorithm with the support being decided by the temporal resolution of HAS. The synthesized wavelet is maximally frequency selective in the critical bands with temporal resolution closely matching with that of the human ear. The design procedure is illustrated with examples, and the performance of the synthesized wavelet is analyzed.

## 1. Introduction

Wavelet transform is an important signal processing tool to analyze nonstationary signals with frequent transients, as in the case of speech and audio signals. It divides a signal into different frequency components, and each component can be analyzed with a resolution matched to its scale. Its major advantage over the short-time Fourier transform (STFT) is that it is possible to construct orthonormal wavelet bases that are well localized over both time and frequency.

It has also long been recognized that human auditory perception plays a crucial role in various speech and audio applications. Some of the many applications where models of auditory perception have been exploited are in speech and audio coding, speech enhancement, and audio watermarking [1–3]. Wavelet-based time-frequency transforms have also been applied in these applications, and models of auditory perception such as critical band (CB) structure and auditory masking have been incorporated [4, 5]. In many wavelet-

based speech and audio processing applications such as in [4, 5], the input signal is decomposed in accordance with the perceptual frequency scale of human auditory system. Thus, perceptually motivated wavelet packet (WP) transform is a popular method for dividing the signal into auditory inspired frequency components, before processing them with a resolution matched to their scale.

The next important thing in these WP-based speech and audio applications is the choice of suitable wavelet and its synthesis. A systematic framework for obtaining orthogonal wavelet was developed by Mallat [6]. Daubechies gave a construction technique for obtaining compactly supported wavelets with arbitrarily high regularity [7]. The requirement of regularity of a wavelet is an important consideration for some applications but their importance is unknown for many other applications [8]. It is evident that appropriate design of wavelet based on the perceptual frequency scale and temporal resolution of the human auditory system is of interest. In the literature, we do find methods of designing mother

wavelet based on the perceptual frequency scale of human auditory system, such as in [9, 10] for continuous wavelet transform. These methods do not provide the requisite filter bank structure for the dyadic multiresolution analysis.

In this paper, we have proposed a design method for synthesizing an optimal mother wavelet for auditory perception-based dyadic filter bank implementation. The design method optimally exploits the CB structure and temporal resolution of human auditory system. The proposed method for synthesizing this compactly supported wavelet is by designing the corresponding optimal wavelet-generating FIR quadrature mirror filter (QMF). The approach followed for the construction of this wavelet is to first obtain the WP tree which closely mimics the CB structure of the human auditory system. This is followed by obtaining the error in passband energy of the CB channel filter responses with respect to the case where the QMFs in the WP tree are replaced by the ideal brick-wall QMFs. These error components are suitably weighted to obtain the performance measure of optimization. The optimization problem is formulated as a single objective unconstrained optimization problem in lattice QMF domain, and the solution is obtained by bounded value global optimization technique. Then, the corresponding wavelet is derived using the cascade algorithm. The support of the wavelet is decided by the temporal resolution of the human auditory system. The synthesized optimal wavelet is found to be maximally frequency selective in the critical bands with temporal resolution matched with that of the human ear. The wavelet design procedure is elaborated with an example, and the performance is compared with respect to other important wavelets such as the Daubechies wavelet, Symlet, and Coiflet.

The rest of the paper is organized as follows. Section 2 describes the broad framework of the design of the proposed perceptually optimized wavelet. In Section 3, the design criterion is elaborated for obtaining the optimal wavelet packet tree. Section 4 deals with the details of the design procedure of the optimal wavelet. Section 5 presents the results. Finally, the paper is concluded with Section 6.

## 2. Synthesis Framework of the Perceptually Optimized Wavelet

The method of designing the perceptually motivated wavelet starts with the design of the optimal WP tree that closely matches the CB structure of the human auditory system. The widely used Zwicker's model of CB structure is used for this purpose which gives a mapping from the physical frequency scale to the critical band rate scale, as given by [11, 12]

$$\begin{aligned} z &= F(f) \\ &= 13 \arctan\left(\frac{0.76 \times f}{10^3}\right) \\ &\quad + 3.5 \arctan\left[\left(\frac{10^{-3} f}{7.5}\right)^2\right], \end{aligned} \quad (1)$$

$$B(f) = 25 + 75(1 + 1.4 \times 10^{-6} f^2)^{0.69}. \quad (2)$$

Here,  $z$  is the critical band rate scale in Bark,  $f$  is the frequency in Hz, and  $B(f)$  is the critical bandwidth in Hz at frequency  $f$ . The criterion for obtaining the optimal wavelet packet tree based on Zwicker's model can be found in [13]. The perceptual criterion minimizes the cost function and allocates an optimal set of terminating nodes at each decomposition depth of the WP tree so that the error in quantizing  $B(f)$  in (2) is minimal in the Bark domain.

In the present paper, using the optimal WP tree obtained from [13], we construct a wavelet which produces maximally frequency selective filter response in each of the CB channels for the corresponding filter bank implementation. Further, the support length of the wavelet is determined by the temporal resolution of the human auditory system.

From the optimal WP tree, the nontree filter structure is obtained which represents the equivalent filtering followed by the combined decimator for each of the CB channels. Using the equivalent nontree filter structure, the error in energy of each of the CB filter impulse responses with respect to the ideal brickwall QMF is obtained. These error components in each channel are minimized with respect to the constraints of QMF. The multiple-objective constrained global optimization problem is converted into a single-objective constrained global optimization problem by taking a suitably weighted average of the energy error terms, denoted as the performance measure of optimization.

The optimization problem is reformulated into an unconstrained optimization problem by converting the QMF constraints in lattice QMF domain [14, 15]. In lattice QMF domain, the performance measure is expressed in terms of Givens rotations [14, 15] which absorb the QMF constraints of the optimization problem. Using the  $2\pi$ -periodicity of Givens rotations, the problem is converted into a bounded value optimization problem [16]. The solution of the global optimization problem is obtained using multilevel coordinate search (MCS) [17]. Using the cascade algorithm (also known as the successive approximation algorithm) [18], the desired wavelet is synthesized. The support of the wavelet is selected in accordance with the temporal resolution of the human ear [19]. This is done by choosing the support of the wavelet so that its time duration is less than the temporal resolution of human auditory system.

Thus, the wavelet synthesized as above is optimal with respect to the critical band structure and temporal resolution of the human auditory system. The design process of the wavelet is elaborated for the case of sampling frequency of  $f_s = 16$  kHz.

## 3. The Optimal WP Tree Based on CB Structure

In [13], a criterion is given to obtain an optimal wavelet packet (WP) tree based on the CB structure of human auditory system for time-frequency decomposition of speech and audio signals. Here, we refer to certain relevant parts from [13]. We first give a brief contextual review of the wavelet packet transform followed by a brief description of the design of the optimal WP tree and an example.

**3.1. Wavelet Packet Transform.** In discrete wavelet transform (DWT) a signal,  $s(t)$  in  $L^2(R)$ , limited to a scale  $J$  can be represented as

$$s(t) = \sum_{k=-\infty}^{\infty} c_0[k] \phi_{0,k}(t) + \sum_{k=-\infty}^{\infty} \sum_{j=0}^{J-1} d_j[k] \psi_{j,k}(t), \quad (3)$$

where  $\phi_{j,k}(t)$  and  $\psi_{j,k}(t)$  are the two-dimensional families of functions generated from the scaling function  $\phi(t)$  and the wavelet  $\psi(t)$  as

$$\begin{aligned} \phi_{j,k}(t) &= 2^{j/2} \phi(2^j t - k), \\ \psi_{j,k}(t) &= 2^{j/2} \psi(2^j t - k). \end{aligned} \quad (4)$$

Here  $j$  denotes the scale, and  $k$  denotes the integer translates of the scaling function and wavelet as defined below. Also,  $c_j[k]$  and  $d_j[k]$  are the approximation and detail coefficients of the DWT at scale  $j$ .

The scaling function and the wavelet are recursively defined as:

$$\phi(t) = \sqrt{2} \sum_{k=-\infty}^{\infty} h[k] \phi(2t - k), \quad (5)$$

$$\psi(t) = \sqrt{2} \sum_{k=-\infty}^{\infty} g[k] \phi(2t - k), \quad (6)$$

where,  $h[k]$  is the lowpass scaling filter and  $g[k]$  is the highpass wavelet filter [18]. The functions  $\phi_{j,k}(t)$ ,  $\psi_{j,k}(t)$ , form an orthonormal basis in  $L^2(R)$ . The DWT can be implemented by the pyramid algorithm [6] where the approximation coefficients  $c_j[k]$  and the detail coefficients  $d_j[k]$  in (3) can be obtained by passing through the approximation coefficients of the next higher scale,  $c_{j+1}[k]$ , to the filters  $h[k]$  and  $g[k]$  and downsampled by a factor of two for  $j = 0, 1, 2, \dots, J-1$ . The filters  $h[k]$  and  $g[k]$  form a quadrature mirror filter (QMF) pair [18]. In the Fourier transform domain they are related by

$$\left| H(e^{j\omega}) \right|^2 + \left| G(e^{j\omega}) \right|^2 = 2, \quad (7)$$

and filters  $h[k]$  and  $g[k]$  are related by [18]

$$g[k] = (-1)^k h[1 - k]. \quad (8)$$

For a finite even length filter of order  $K$ , (8) can be written as [18]

$$g[k] = \pm (-1)^k h[K - k]. \quad (9)$$

After the signal is processed by the tree-structured analysis filter bank, the inverse process of interpolation and filtering can be used to reconstruct the signal. The perfect reconstruction of a signal can be achieved using a realizable orthogonal filter bank [14, 20]. The perfect reconstruction lowpass and highpass synthesis QMF pair,  $h_1[k]$  and  $g_1[k]$ , is related to the analysis filters by [20]

$$\begin{aligned} h_1[k] &= h[K - k], \\ g_1[k] &= g[K - k]. \end{aligned} \quad (10)$$

The wavelet packet transform (WPT) is an extension of DWT, where both the approximation and detail coefficients are decomposed. A sequence of functions,  $\{\nu_n(t)\}_{n=0}^{\infty}$ , can be defined as

$$\begin{aligned} \nu_{2n}(t) &= \sqrt{2} \sum_{k=-\infty}^{\infty} h[k] \nu_n(2t - k), \\ \nu_{2n+1}(t) &= \sqrt{2} \sum_{k=-\infty}^{\infty} g[k] \nu_n(2t - k), \end{aligned} \quad (11)$$

where  $\nu_0(t) = \phi(t)$ , that is, the scaling function, and  $\nu_1(t) = \psi(t)$ , that is, the wavelet [21]. The collection of functions  $\nu_n(t - k)$ , as defined in (11), forms an orthonormal basis of  $L^2(R)$ . The library of wavelet packet bases is the collection of orthonormal basis functions composed of functions of the form [21]

$$\nu_{n,j,k}(t) = 2^{j/2} \nu_n(2^j t - k). \quad (12)$$

Denoting the space formed by the basis  $\nu_{n,j,k}(t)$  by  $W_{n,j}$ , the signal  $s(t)$  limited to a scale  $J$ , that is,  $s(t) \in W_{0,J}$ , can be decomposed in a manner similar to (3), as follows:

$$s(t) = \sum_{k=-\infty}^{\infty} \sum_{j=0}^{J-1} \sum_{n \in I_j} d_{n,j}[k] \nu_{n,j,k}(t), \quad (13)$$

where  $I_j = \{0, 1, 2, \dots, 2^{j-1} - 1\}$  [22]. Here,  $d_{n,j}[k]$  are the WPT coefficients. Further,  $j$  denotes the scale, and  $n$  gives their position in the wavelet packet tree. The WPT can be implemented using an extension of the pyramid algorithm where both the approximation and detail coefficients are decomposed in a tree-structured QMF bank.

**3.2. Criterion for Obtaining Optimal WP Tree Based on Bark Scale.** The criterion minimizes a cost function and allocates an optimal set of number of terminating nodes at each level of decomposition so that the error in quantizing  $B(f)$  is minimal in the Bark domain. Here, we seek to identify the segments of  $B(f)$ , which correspond to dyadically related critical bandwidths, and the number of nodes in each segment so that the error in Bark domain as defined below is minimum.

Let us assume that a signal is limited to a scale  $J$  and  $j$  is the variable of scale as given in (13). We define the variable  $p$  as the depth of decomposition given by  $p = J - j$ . The input signal sampled at Nyquist rate is taken as the scaling coefficients at the  $J$ th scale. As the signal is decomposed through all the levels, the depth of decomposition varies from  $p = 0$  to  $J$ . The bandwidth available at decomposition depth  $p$  is given by

$$\Delta f_{WP}(p) = \frac{f_s}{2^{p+1}}, \quad (14)$$

where  $f_s$  is the sampling frequency of the input signal. For a dyadic WP tree with maximum depth of decomposition  $L$ , minimum depth of decomposition  $M$ , and  $n_p$  number of terminal nodes at decomposition depth  $p$ ,

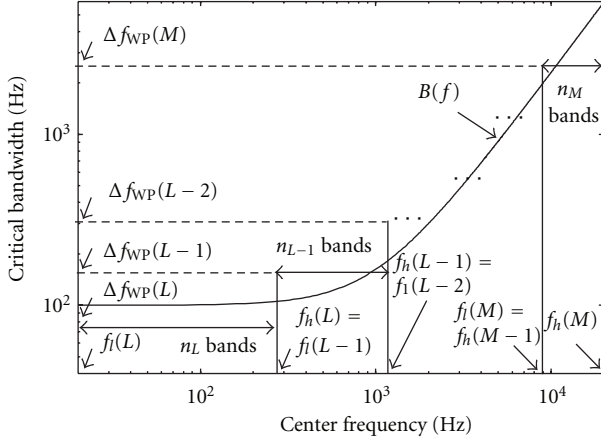


FIGURE 1: Illustration of WP bandwidths and number of terminating nodes at various decomposition depths with respect to  $B(f)$ .

$M \leq p \leq L$ , the terms  $L$ ,  $M$ ,  $n_p$ , and  $\Delta f_{WP}(p)$  are related by  $\sum_{p=M}^L n_p \Delta f_{WP}(p) = f_s/2$  which can alternatively be written as

$$\sum_{p=M}^L \frac{n_p}{2^p} = 1. \quad (15)$$

The critical bandwidth in (2) is a monotonically increasing function of the frequency. So the lower frequency bands are progressively decomposed to a deeper depth compared to the higher frequency bands. The frequency range covered by the  $p$ th depth of decomposition is  $f_l(p) \leq f \leq f_h(p)$ , where  $f_h(p) = \sum_{m=p}^L n_m \Delta f_{WP}(m)$ ,  $f_l(p) = \sum_{m=p+1}^{L+1} n_m \Delta f_{WP}(m)$ ,  $n_{L+1} = 0$ , and  $M \leq p \leq L$ . Here,  $f_l(p)$ , and  $f_h(p)$  are respectively, the lower and higher limits of the frequency range covered by the  $p$ th depth of decomposition of the WP tree. In Figure 1, the terms  $L$ ,  $M$ ,  $n_p$ ,  $f_l(p)$ ,  $f_h(p)$  and  $\Delta f_{WP}(p)$  are illustrated with respect to  $B(f)$ , for the complete auditory range of 20 Hz–20 kHz.

To obtain the perceptual cost function in the Bark domain, we define  $\tilde{B}(z)$  as an expression relating the critical bandwidth in Hz as a function of center frequency in Bark, that is,

$$\tilde{B}(z) = \tilde{B}(F(f)) = B(F^{-1}(z)) = B(f). \quad (16)$$

At the  $p$ th decomposition depth, the integral squared error in critical bandwidth in the Bark domain can be obtained as

$$q_e(p) = \int_{z_l(p)}^{z_h(p)} (\tilde{B}(z) - n_p \Delta f_{WP}(p))^2 dz, \quad (17)$$

where  $z_h(p) = F(f_h(p))$  and  $z_l(p) = F(f_l(p))$ . The total error  $Q_E$ , in quantizing  $\tilde{B}(z)$ , for the complete frequency range  $0 \leq f \leq f_s/2$ , can be given by  $Q_E = \sum_{p=M}^L q_e(p)$ . Substituting the expression of  $q_e(p)$  and replacing  $z$  by  $F(f)$  in the expression of  $Q_E$ , we obtain

$$Q_E = \sum_{p=M}^L \int_{f_l(p)}^{f_h(p)} \left( B(f) - n_p \left( \frac{f_s}{2^{p+1}} \right) \right)^2 F'(f) df, \quad (18)$$

where  $F'(f)$  is obtained by differentiating (1). The perceptual criterion for obtaining the optimal WP tree is to minimize the cost function  $Q_E$ , that is,

$$\begin{aligned} & (L^{\text{opt}}, M^{\text{opt}}, n_M^{\text{opt}}, n_{M+1}^{\text{opt}}, \dots, n_L^{\text{opt}}) \\ & = \arg \min_{(L, M, n_M, n_{M+1}, \dots, n_L)} \{Q_E\}, \end{aligned} \quad (19)$$

subject to the constraint given in (15). One can exhaustively search the possible candidate trees using (15) and obtain the optimum value of  $L$ ,  $M$ , and the number of terminal nodes at different decomposition depths, that is,  $n_M, n_{M+1}, \dots, n_L$  by evaluating (18).

**3.3. Optimal WP Tree for  $f_s = 16$  kHz and Auditory Band Indexed WP Bases.** The above design is explained for the case where the input signal is sampled at  $f_s = 16$  kHz. The solution of (19) gives the optimum values  $L = 6$ ,  $M = 3$ ,  $n_3 = 4$ ,  $n_4 = 3$ ,  $n_5 = 6$ , and  $n_6 = 8$ . Thus, the minimum depth of decomposition is  $M = 3$ , the maximum depth of decomposition is  $L = 6$ , and the total number of decomposition depths is  $L - M + 1 = 4$ . Figure 2 compares the WP tree with Zwicker's critical band structure. For this case, the signal can be decomposed as in (13) as follows:

$$\begin{aligned} & \int |s(t)|^2 dt \\ & = \sum_{k=0,1,3,2,6,7,5,4} |d_{n,0}[k]|^2 + \sum_{k=6,7,5,4,12,13} |d_{n,1}[k]|^2 \\ & + \sum_{k=7,5,4} |d_{n,2}[k]|^2 + \sum_{k=6,7,5,4} |d_{n,3}[k]|^2. \end{aligned} \quad (20)$$

In (20),  $n$  denotes the position of the WPT coefficients in the WP tree and assumes the appropriate values at the various scales such that the frequency bands are ordered in an ascending manner for the WPT. It is noticed that  $n$  is not in ascending order with respect to the band-ordered WPT coefficients at the various scales. This is because of the fact that, in a dyadic filter bank implementation, when a highpass region is decomposed by a QMF bank, the highpass and lowpass frequency regions swap with each other [22].

## 4. Design Procedure of the Perceptually Motivated Optimal Wavelet

**4.1. Auditory Band Indexed Optimal WP Tree and Its Filter Bank Implementation.** The solution as obtained from the previous section is restated in terms of the CB-indexed WP tree where the indexing is done with increasing center frequencies of the CBs. As given in the previous section, let the optimal WP tree be obtained with maximum depth of decomposition  $L$ , minimum depth of decomposition  $M$ , and number of branches  $n_p$  in  $p$ th depth of decomposition. The CBs numbering  $i = 1$  to  $N$  span  $L - M + 1$  sets where  $i = 1$  to  $n_L$ , for  $m = 1$ ;  $i = n_L + 1$  to  $n_L + n_{L-1}$  for  $m = 2$ ;  $\dots$ ;  $i = n_L + n_{L-1} + \dots + n_{M-1} + 1$  to  $N$  for  $m = L - M + 1$ . Here,  $m$  is



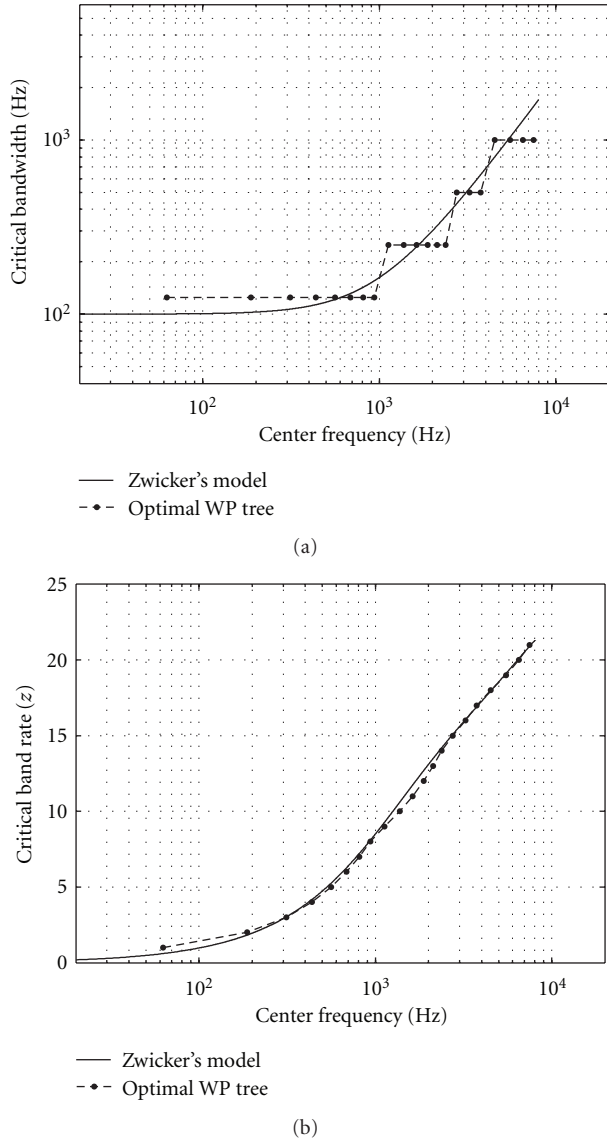


FIGURE 2: Comparison of optimal WP tree ( $f_s = 16$  kHz) with Zwicker's critical band structure: (a) critical bandwidth as a function of center frequency, (b) critical band rate as a function of center frequency.

the index of sets having same bandwidth,  $m = 1$  to  $L - M + 1$ , and  $N$  is the total number of CBs. For the example case of  $f_s = 16$  kHz, the first set of critical bands, that is,  $m = 1$ , are numbered from  $i = 1$  to 8, the second set ( $m = 2$ ) from 9 to 14, the third set ( $m = 3$ ) from 15 to 17 and the fourth set ( $m = 4$ ) from  $i = 18$  to 21, and  $N = 21$ .

The filter bank implementation for this case is shown in Figure 3. These sets of critical bands and the corresponding CB-indexing can be observed from this figure. Note that, each set of critical band is associated with its time resolution and frequency bandwidth. This CB-indexing will be used for obtaining the CB filter impulse responses for the different channels.

Now, the binary tree structure of the optimal WP tree is converted into an equivalent nontree filter structure

using the noble identity for a downsampler as shown in Figure 4 [14]. As can be seen from the figure, a filter  $A(z)$  following a decimator  $M$  is equivalent to  $A(z^M)$  preceding the same decimator. Using the noble identity, the nontree filter structure is obtained for the optimal WP tree. As an illustration, the nontree filter structure corresponding to Figure 3 is shown in Figure 5. In this figure,  $H_i(z)$  represents the equivalent filtering at the  $i$ th critical band, and the combined decimators follow them. Note that  $i$  in  $H_i(z)$  denotes the critical band numbers in ascending order of center frequencies of the respective bands.

The lower and upper passband edges of  $H_i(e^{j\omega})$  are denoted as  $\omega l_i$  and  $\omega h_i$ , respectively, and can be expressed as

$$\omega l_i = \frac{2\pi}{f_s} \left[ f_i(L+1-m) + \frac{f_h(L+1-m) - f_l(L+1-m)}{n_{L+1-m}} i \right] \quad (21)$$

$$\omega h_i = \frac{2\pi}{f_s} \left[ f_h(L+1-m) + \frac{f_h(L+1-m) - f_l(L+1-m)}{n_{L+1-m}} i \right], \quad (22)$$

where  $i$  is the index of critical bands in ascending order as explained previously. In (21) and (22),  $f_l(L+1-m)$  and  $f_h(L+1-m)$  are defined as in between (15) and (16). Further,  $f_s$  denotes the sampling frequency. For the ideal brickwall QMF pair [14],  $H_i^{\text{Ideal}}(e^{j\omega})$  and  $G_i^{\text{Ideal}}(e^{j\omega})$ , the magnitude squared frequency response of the individual channels is shown in Figure 6, where  $H_i^{\text{Ideal}}(e^{j\omega})$  is the frequency response of the equivalent nontree filter structure of the  $i$ th critical band in the nontree filter structure. This figure also shows the passband edges of the CB filters for the particular example being considered.

#### 4.2. CB Channel Filter Errors and the Optimization Problem.

The integral squared error in passband energy of the individual CB channel filters with respect to the ideal case can be expressed as

$$E_i = 2 \int_{\omega l_i}^{\omega h_i} \left( |H_i^{\text{Ideal}}(e^{j\omega})|^2 - |H_i(e^{j\omega})|^2 \right) \frac{d\omega}{2\pi}, \quad i = 1, \dots, N, \quad (23)$$

where  $\omega l_i$  and  $\omega h_i$  are the low and high band edges of the  $i$ th critical band and are given by (21) and (22), respectively. Note that, the term in the integral is always a nonnegative quantity, as

$$2 \int_{\omega l_i}^{\omega h_i} |H_i^{\text{Ideal}}(e^{j\omega})|^2 \frac{d\omega}{2\pi} = 1, \quad (24)$$

$$2 \int_{\omega l_i}^{\omega h_i} |H_i(e^{j\omega})|^2 \frac{d\omega}{2\pi} \leq 1.$$

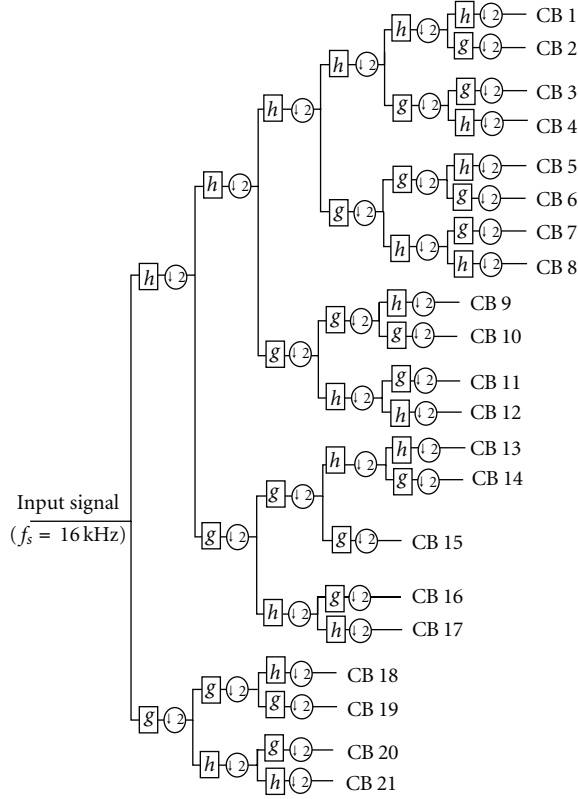


FIGURE 3: The filter bank implementation of the WP tree for  $f_s = 16$  kHz.

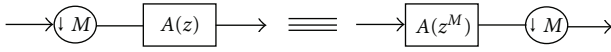


FIGURE 4: The noble identity for downsampler.

Hence,  $E_i$  can alternatively be expressed as

$$E_i = 1 - 2 \int_{\omega_i}^{\omega_{hi}} |H_i(e^{j\omega})|^2 \frac{d\omega}{2\pi}. \quad (25)$$

Also,  $E_i \geq 0$ . The optimal lowpass QMF is defined as the solution of the following multi-objective function:

$$h^{\text{opt}}[n] = \arg \min_{\{h[n]\}} \{E_1, E_2, E_3, \dots, E_N\}, \quad (26)$$

where  $h[n]$  represents all possible wavelet-defining, lowpass QMFs.

This multi-objective optimization problem can be simplified however, if we convert it into a conventional single-objective optimization problem using the average of suitably weighted objective functions as the performance measure of optimization. We have used the weighting function of the outer and middle ear (OME) for this purpose. The OME function weights the CB energy errors in the passband such that it is smaller in the mid-frequency regions compared to the low and high frequency regions. Thus, the expression of the single objective performance measure (28), as obtained

below, gives more importance to the perceptually significant mid-frequency region due to OME weighting. A well-known model of the OME transfer function  $W_{\text{dB}}(f)$  is given by

$$W_{\text{dB}}(f) = -0.6 \times 3.64(10^{-3}f)^{-0.8} + 6.5 \\ \times \exp\left(-0.6 \times (10^{-3}f - 3.3)^2\right) - 10^{-3}(10^{-3}f)^4, \quad (27)$$

where  $W_{\text{dB}}(f)$  is the weighting in dB scale as a function of frequency  $f$  in Hz [23]. The OME weighting function  $W_{\text{dB}}(f)$  is shown as a function of the frequency  $f$  in Figure 7.

Now, the single objective performance measure is obtained as

$$P = \frac{1}{N} \sum_{i=1}^N E_i w(i), \quad (28)$$

where

$$w(i) = 10^{W_{\text{dB}}(f_i)} \quad (29)$$

with  $f_i$  as the center frequency of the  $i$ th critical band. Substituting (25) to (28), the performance measure  $P$  can alternately be written as

$$P = \frac{1}{N} \sum_{i=1}^N \left( 1 - 2 \int_{\omega_i}^{\omega_{hi}} |H_i(e^{j\omega})|^2 \frac{d\omega}{2\pi} \right) w(i), \quad (30)$$

and the optimization problem can be restated as

$$h^{\text{opt}}[n] = \arg \min_{h[n]} \{P\}. \quad (31)$$

Here,  $P$  is the performance measure of optimization, and  $h^{\text{opt}}[n]$  is the perceptually optimized wavelet-defining QMF. This is a global optimization problem in which  $h[n]$  is constrained to satisfy the QMF condition given in (7).

**4.3. Lattice QMF Representation and the Unconstrained Optimization Problem.** By utilizing the lattice representation of QMF bank [14, 15], the constrained optimization problem (31) is converted into an unconstrained one. The use of lattice QMF representation for converting the constrained optimization problem to an unconstrained one can be found in [24], where the authors have used this technique for designing a minimum duration orthonormal wavelet.

It is well known that any FIR two-channel paraunitary QMF bank can be represented by the so-called paraunitary QMF lattice as shown in Figure 8, where the filter pair  $H(z)$  and  $G(z)$  is written in a matrix form as [14]

$$\begin{bmatrix} H(z) \\ G(z) \end{bmatrix} = R_J \Lambda(z^2) R_{J-1} \Lambda(z^2) \cdots \Lambda(z^2) R_0 \begin{bmatrix} 1 \\ z^{-1} \end{bmatrix}. \quad (32)$$

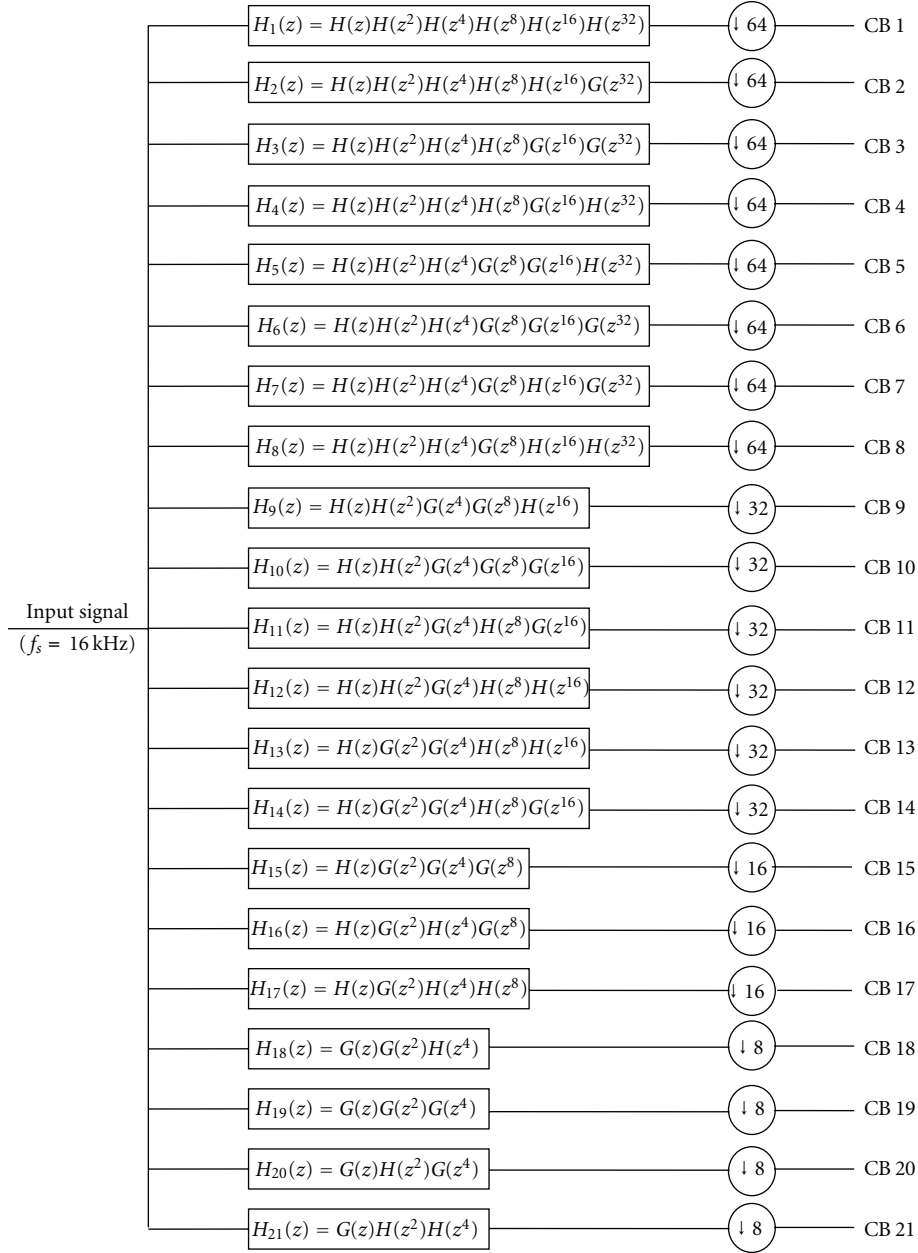


FIGURE 5: Equivalent nontree filter structure of Figure 3.

In (32),  $J$  relates to the QMF length  $M$  via  $M = 2J + 2$  and  $R_m$ ,  $0 \leq m \leq J$ , is a  $2 \times 2$  unitary matrix (i.e.,  $R_m^T R_m = I$ ) and is expressed as

$$R_m = \begin{bmatrix} \cos \theta_m & \sin \theta_m \\ -\sin \theta_m & \cos \theta_m \end{bmatrix}. \quad (33)$$

The matrix  $R_m$  is known as Givens rotation with  $\theta_m$  as the angle. In Figure 9, the details of the unitary matrix  $R_m$  are shown. Also, in (32),  $\Lambda(z^2)$  is given by

$$\Lambda(z^2) = \begin{bmatrix} 1 & 0 \\ 0 & z^{-2} \end{bmatrix}. \quad (34)$$

The filter pair  $H(z)$  and  $G(z)$  must also satisfy the additional constraint that  $H(z)$  is lowpass or, equivalently,  $G(z)$  is highpass

$$H(1) = 1. \quad (35)$$

The constraint of (35) on  $h[n]$  can be transformed to Givens rotations by evaluating (32) for  $z = 1$ , that is,  $\omega = 0$ , and obtained as

$$\theta_J + \theta_{J-1} + \dots + \theta_0 = -\frac{\pi}{4}. \quad (36)$$

Thus, any one of the  $\theta_m$ ,  $0 \leq m \leq J$ , can be expressed as a function of others. The performance measure  $P$  of (30)

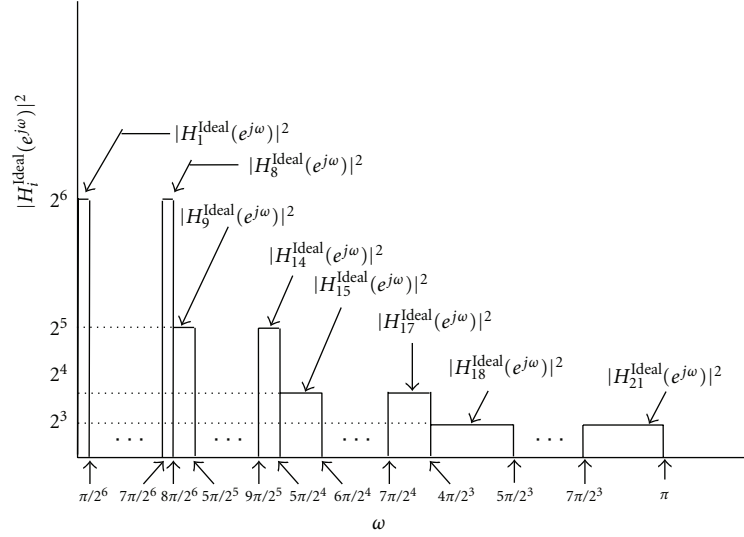


FIGURE 6: Magnitude squared frequency response and passband edges of the CB channel filters for the ideal case.

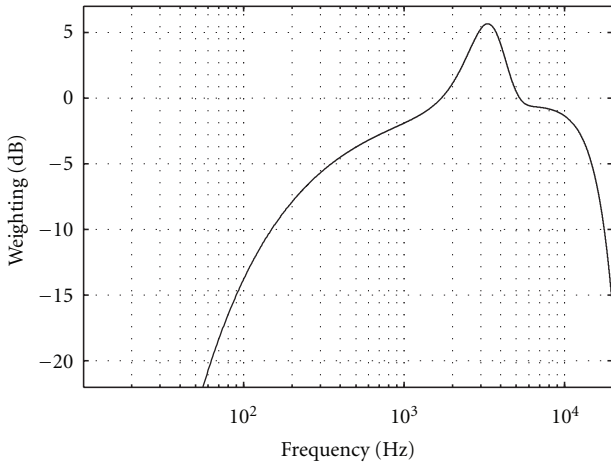


FIGURE 7: Outer and middle ear transfer function  $W_{dB}(f)$ .

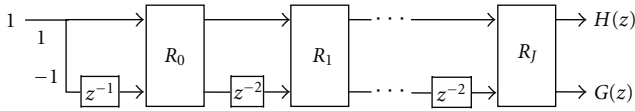


FIGURE 8: The QMF lattice structure of  $H(z)$  and  $G(z)$ .

can now be expressed as a function of  $\theta_0$  to  $\theta_{J-1}$ , and the optimization problem can be posed as an unconstrained global optimization problem as

$$\begin{aligned} & (\theta_0^{\text{opt}}, \theta_1^{\text{opt}}, \dots, \theta_{J-1}^{\text{opt}}) \\ & = \arg \min_{\theta_0, \theta_1, \dots, \theta_{J-1}} P \left( \theta_0, \theta_1, \dots, \theta_{J-1}, -\frac{\pi}{4} - \sum_{m=0}^{J-1} \theta_m \right). \end{aligned} \quad (37)$$

This optimization problem can be solved by using standard unconstrained global optimization program.

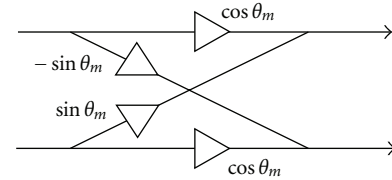


FIGURE 9: Details of the unitary matrix  $R_m$ .

**4.4. Bound Constrained Global Optimization.** We give a simple approach to solve the optimization problem (37). The search space for the optimal values of  $\theta_m$ ,  $m = 0, \dots, J-1$ , is reduced by exploiting the periodic property of  $R_m$ . It is observed from (33) that Givens rotations are periodic functions in  $2\pi$ , that is,

$$R_m(\theta_m) = R_m(\theta_m + 2\pi). \quad (38)$$

Thus, instead of searching for a globally optimized solution, the optimal solution of (37) can be obtained by using a bounded value global optimization program where the bounds on  $\theta_m$  are  $0 \leq \theta_m \leq 2\pi$ ,  $0 \leq m \leq J-1$ .

The solution of the above optimization problem as in (37) is achieved by using multilevel coordinate search (MCS), a “bound constrained global optimization” program [16, 17]. Bound constrained global optimization problem can be formalized as

$$\begin{aligned} & \min f(x) \\ & \text{s.t. } x \in [u, v] \end{aligned} \quad (39)$$

with finite or infinite bounds, where the interval notation is used for rectangular boxes,  $[u, v] := \{x \in \mathbb{R}^n \mid u_i \leq x_i \leq v_i, i = 1, \dots, n\}$  with  $u$  and  $v$  being  $n$ -dimensional vectors with components in  $\bar{\mathbb{R}} := \mathbb{R} \cup \{-\infty, \infty\}$  and  $u_i < v_i$  for  $i = 1, 2, \dots, n$ . To handle the bounded constraint optimization problem, we use a global optimization algorithm called



TABLE 1: Filter coefficients of the optimal decomposition and reconstruction QMF pair for length  $M = 6$  and  $M = 8$ .

$n$	QMF of length = 6				QMF of length = 8			
	Decomposition		Reconstruction		Decomposition		Reconstruction	
	$h[n]$	$g[n]$	$h_1[n]$	$g_1[n]$	$h[n]$	$g[n]$	$h_1[n]$	$g_1[n]$
0	0.49985	0.11923	0.11923	-0.49985	0.4083	-0.0501	-0.0501	-0.4083
1	0.73350	0.17497	-0.17497	0.73350	0.7608	-0.0934	0.0934	0.7608
2	0.38225	-0.14563	-0.14563	-0.38225	0.4296	0.0631	0.0631	-0.4296
3	-0.14563	-0.38225	0.38225	-0.14563	-0.0668	0.2241	-0.2241	-0.0668
4	-0.17497	0.73350	0.73350	0.17497	-0.2241	-0.0668	-0.0668	0.2241
5	0.11923	-0.49985	0.49985	0.11923	0.0631	-0.4296	0.4296	0.0631
6					0.0934	0.7608	0.7608	-0.0934
7					-0.0501	-0.4083	0.4083	-0.0501

Multiple Coordinate Search (MCS) algorithm as proposed in [16, 17]. The MCS method combines both global search and local search into one unified framework via multilevel coordinate search. It is guaranteed to converge if the function is continuous.

The multilevel coordinate search balances between global and local search. The local search is done via sequential quadratic programming. The search in MCS is not exhaustive, and thus, the global minimum may be missed. However, in comparison with other global optimization algorithms, MCS shows excellent performance in many cases, especially for smaller dimensions [16].

**4.5. Construction of the Optimal Wavelet.** After obtaining the optimal solution of (37), the corresponding lowpass QMF is derived from (32). Subsequently, the optimal highpass component of the QMF pair  $g[n]$  is obtained from (9). The cascade algorithm is used to solve the basic recursion equation of (5), known as the two-scale equation for the scaling function. This is an iterative algorithm that generates successive approximations to  $\phi(t)$ . The iterations are defined by

$$\phi^{k+1}(t) = \sum_{n=0}^{M-1} h[n] \sqrt{2} \phi^k(2t - n), \quad (40)$$

where  $M$  is the support of the scaling function,  $k$  is the iteration number, and  $\phi^{k+1}(t)$  denotes the  $k$ th iteration of the scaling function with  $\phi^0(t)$  being the initial value of iteration. From the scaling function, the wavelet  $\psi(t)$  is obtained by using the two-scale equation for the wavelet as given in (6).

The order of the FIR filter and in turn, the support of the wavelet, is taken into consideration from the temporal resolution of the human auditory system. The time duration of a wavelet  $\psi(t)$  is defined by [25]

$$\Delta t = \frac{\int_{-\infty}^{\infty} (t - t_0)^2 |\psi(t)|^2 dt}{\int_{-\infty}^{\infty} |\psi(t)|^2 dt}, \quad (41)$$

where

$$t_0 = \frac{\int_{-\infty}^{\infty} t |\psi(t)|^2 dt}{\int_{-\infty}^{\infty} |\psi(t)|^2 dt}. \quad (42)$$

Here,  $t_0$  is the first moment of the wavelet and provides the measure of where  $\psi(t)$  is centered along the time axis. The time duration of wavelet  $\Delta t$  (41) is the root mean square (RMS) measure of duration and gives the spread of wavelet in time. This definition of time-duration gives a measure of time localization of the wavelet [25].

Now, the above definition of time duration is used for selecting the support of the wavelet. The support of the optimal wavelet is chosen depending on the temporal resolution of human ear. Temporal resolution of the ear refers to its ability to detect changes in stimuli over time [19]. It is usually characterized by the ability to detect a brief gap between two stimuli or to detect the amplitude modulation of a sound [19]. Temporal resolution measured for the detection of gaps in broadband noise is typically 2-3 ms. Further, temporal resolution measured by the discrimination of stimuli with identical magnitude spectra is in the range of 2-6 ms [19]. For our wavelet construction algorithm, we have taken temporal resolution of the human auditory system to be less than 4 ms.

We choose the support of the wavelet so that its time-duration (41) is less than the temporal resolution of the ear, that is, 4 ms. Thus, the proposed wavelet can detect the short duration acoustic stimuli that the ear can perceive. The higher support length of the wavelet will give better frequency selectivity at the critical band channels. The enhanced support length is also associated with increased time-duration of the wavelet.

## 5. Results

The perceptually optimized wavelet has been obtained using the bound-constrained global optimization program known as multilevel coordinate search (MCS) [16]. The algorithm for construction of the optimal wavelet is implemented in MATLAB. The MATLAB code for MCS algorithm is available at [17]. After obtaining the  $\theta_m$  values,  $0 \leq \theta_m \leq 2\pi$ ,  $0 \leq m \leq J - 1$ ,  $J = (M - 2)/2$  for the desired support length of  $M$ , the coefficients of the QMF pair are obtained using (32). The perfect reconstruction QMF pair is obtained from (10). The filter coefficients of decomposition and synthesis QMF bank are shown in Table 1 for filter lengths  $M = 6$  and 8. In Figures 10(a) and 11(a), the magnitude squared frequency responses

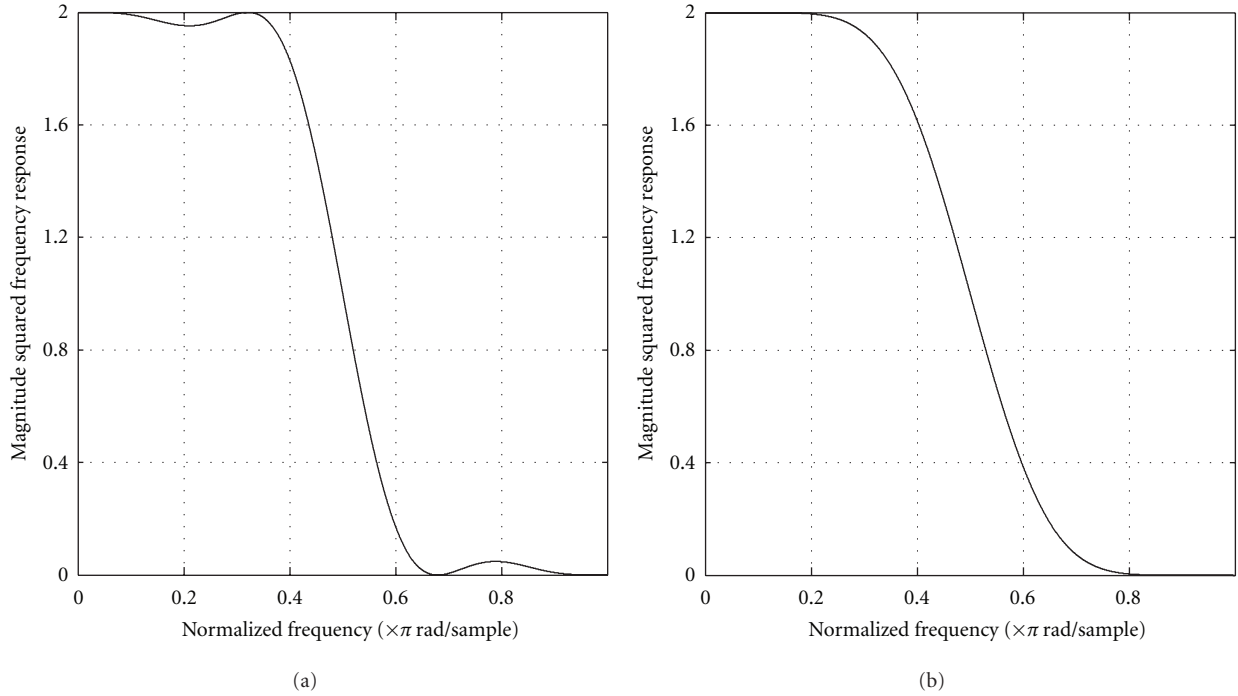


FIGURE 10: Magnitude squared frequency response of the perceptually optimized QMF and Daubechies QMF of length  $M = 8$ ; (a) the perceptually optimized lowpass QMF, (b) Daubechies lowpass QMF.

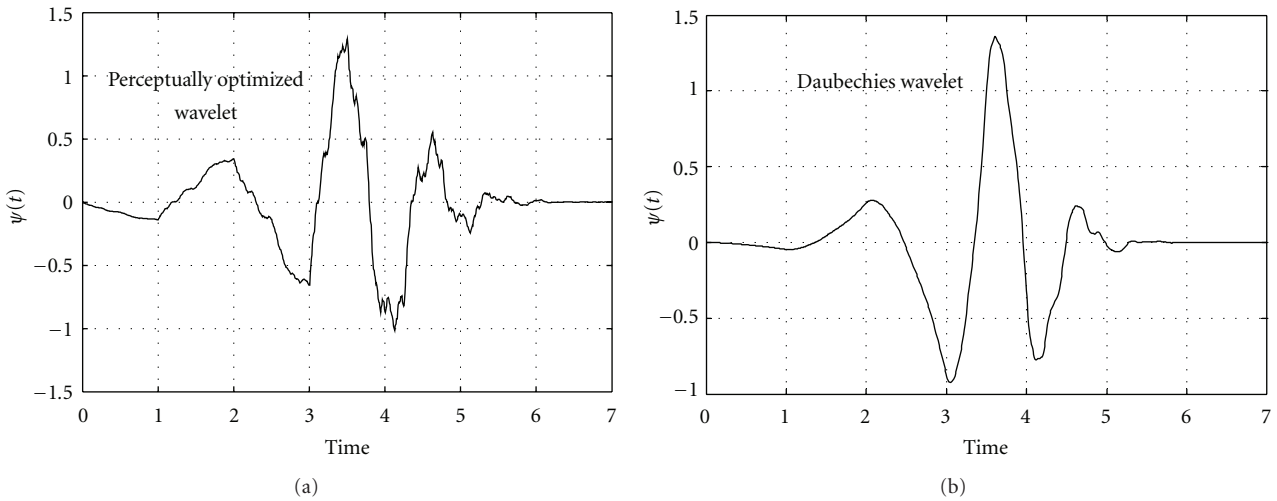


FIGURE 11: The perceptually optimized wavelet and Daubechies wavelet of length  $M = 8$ ; (a) the perceptually optimized wavelet, (b) Daubechies wavelet.

of the optimal lowpass QMF and the corresponding mother wavelet are shown for the case of  $M = 8$ . In Figures 10(b) and 11(b), the magnitude squared frequency responses of the Daubechies QMF and the corresponding wavelet are shown for the same value of  $M$  for comparison.

The magnitude squared frequency responses  $|H_i(e^{i\omega})|^2$  of the CB channel filters are shown in Figure 12 for the optimal QMF with length,  $M = 8$ . The responses are

grouped in Figures 12(a) to 12(d) according to the depth of decomposition of the CB channel filters. The RMS time-duration of this optimal wavelet is found to be 2.8 ms.

The perceptually optimized wavelet is compared with Daubechies wavelet, Symlet, and Coiflet in terms of the energy error in CB channel impulse response. In Figure 13, we show the energy error  $E_i$  in the critical bands as computed

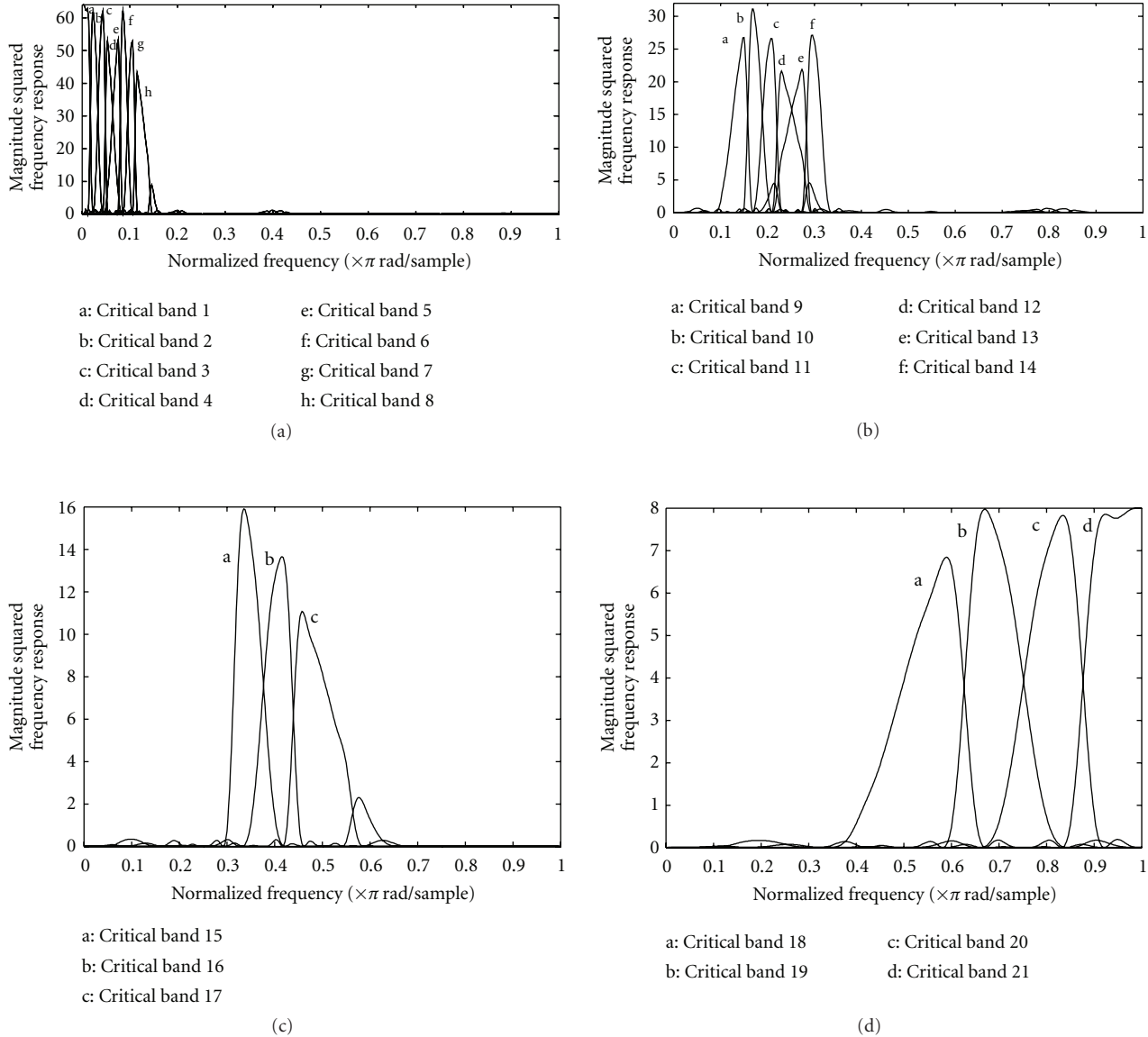


FIGURE 12: Magnitude squared frequency response of the CB channel filters with perceptually optimized QMF of length  $M = 8$ ; (a) critical bands 1 to 8, (b) critical bands 9 to 14, (c) critical bands 15 to 17, (d) critical bands 18 to 21.

from (25) by using Daubechies QMF, Symlet QMF, Coiflet QMF, and the perceptually motivated QMF of length  $M = 6$ . Figure 14 shows the energy error  $E_i$  for Daubechies QMF, Symlet and the perceptually motivated QMF for  $M = 8$ . The advantage of the perceptually optimized wavelet is clearly visible from the Figures 13 and 14 in terms of the reduction in energy error in the CB channel filter impulse responses in comparison to the other wavelets.

Further, it is noticed from Figures 13 and 14 that the energy errors in the CB channel filters for  $M = 8$  are less than those for  $M = 6$ . This is because of the better frequency selectivity of the CB filters for the case with higher support length of the wavelet. The higher support length of the optimal wavelet is also associated with higher time-duration of the wavelet.

## 6. Conclusions

We have presented a design method for synthesizing a perceptually optimized wavelet that is optimally frequency selective at the critical bands with temporal resolution closely matching with that of human auditory system. The performance measure of the optimization criterion takes into account the critical band structure, the frequency-dependent weighting due to outer and middle ear, and the temporal resolution of human auditory system. The design method is elaborated with an example case of sampling frequency of 16 kHz. The energy error of the CB filter impulse responses,  $E_i$ , is found to be less for the case of the synthesized optimal wavelet in comparison to other important wavelets such as Daubechies wavelet, Symlet, and Coiflet.

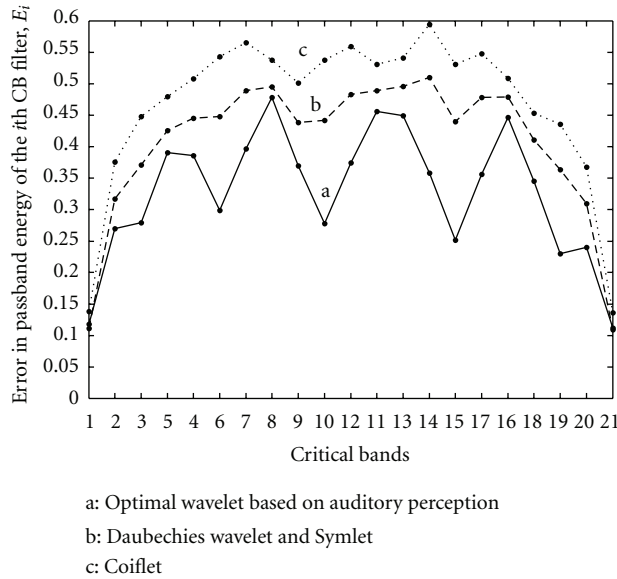


FIGURE 13: Comparison of CB channel filter passband energy error,  $E_i$ , in the critical bands with the perceptually optimized wavelet, Daubechies wavelet, Symlet and Coiflet of support length  $M = 6$ .

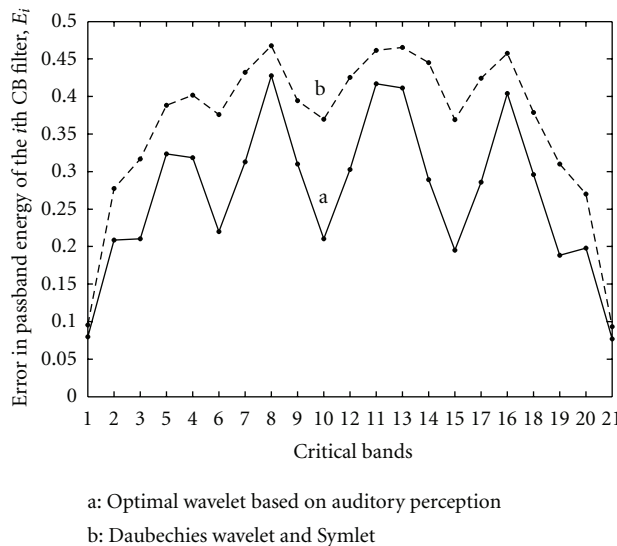


FIGURE 14: Comparison of CB channel filter passband energy error,  $E_i$ , in the critical bands with the perceptually optimized wavelet, Daubechies wavelet and Symlet of support length  $M = 8$ .

## Acknowledgments

The authors wish to thank the anonymous reviewers for their constructive comments and suggestions that have improved the paper. The authors would also like to thank the Director, Central Electronics Engineering Research Institute (CEERI)/CSIR, for providing the requisite support for conducting this research work.

## References

- [1] N. Jayant, J. Johnston, and R. Safranek, "Signal compression based on models of human perception," *Proceedings of the IEEE*, vol. 81, no. 10, pp. 1385–1422, 1993.
- [2] N. Virag, "Single channel speech enhancement based on masking properties of the human auditory system," *IEEE Transactions on Speech and Audio Processing*, vol. 7, no. 2, pp. 126–137, 1999.
- [3] M. D. Swanson, B. Zhu, A. H. Tewfik, and L. Boney, "Robust audio watermarking using perceptual masking," *Signal Processing*, vol. 66, no. 3, pp. 337–355, 1998.
- [4] D. Sinha and A. H. Tewfik, "Low bit rate transparent audio compression using adapted wavelets," *IEEE Transactions on Signal Processing*, vol. 41, no. 12, pp. 3463–3479, 1993.
- [5] B. Carnero and A. Drygajlo, "Perceptual speech coding and enhancement using frame-synchronized fast wavelet packet transform algorithms," *IEEE Transactions on Signal Processing*, vol. 47, no. 6, pp. 1622–1635, 1999.
- [6] S. G. Mallat, "A theory for multiresolution signal decomposition: the wavelet representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, no. 7, pp. 674–693, 1989.
- [7] I. Daubechies, "Orthonormal bases of compactly supported wavelets," *Communications on Pure and Applied Mathematics*, vol. 41, no. 7, pp. 909–996, 1988.
- [8] O. Rioul and M. Vetterli, "Wavelets and signal processing," *IEEE Signal Processing Magazine*, vol. 8, no. 4, pp. 14–38, 1991.
- [9] I. Pinter, "Perceptual wavelet-representation of speech signals and its application to speech enhancement," *Computer Speech and Language*, vol. 10, no. 1, pp. 1–22, 1996.
- [10] J. Yao and Y. T. Zhang, "Bionic wavelet transform: a new time-frequency method based on an auditory model," *IEEE Transactions on Biomedical Engineering*, vol. 48, no. 8, pp. 856–863, 2001.
- [11] E. Zwicker and H. Fastl, *Psychoacoustics: Facts and Models*, Springer, New York, NY, USA, 2nd edition, 1999.
- [12] E. Zwicker and E. Terhardt, "Analytical expression for critical-band rate and critical bandwidth as a function of frequency," *Journal of the Acoustical Society of America*, vol. 68, no. 5, pp. 1523–1525, 1980.
- [13] A. Karmakar, A. Kumar, and R. K. Patney, "Design of optimal wavelet packet trees based on auditory perception criterion," *IEEE Signal Processing Letters*, vol. 14, no. 4, pp. 240–243, 2007.
- [14] P. P. Vaidyanathan, *Multirate Systems and Filter Banks*, Prentice-Hall, Englewood Cliffs, NJ, USA, 1993.
- [15] P. P. Vaidyanathan and P. Q. Hoang, "Lattice structures for optimal design and robust implementation of two-channel perfect reconstruction QMF banks," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 36, no. 1, pp. 81–94, 1988.
- [16] W. Huyer and A. Neumaier, "Global optimization by multi-level coordinate search," *Journal of Global Optimization*, vol. 14, no. 4, pp. 331–355, 1999.
- [17] "MCS: Global optimization by multilevel coordinate search," <http://www.mat.univie.ac.at/~neum/software/mcs/>.
- [18] C. S. Burrus, R. A. Gopinath, and H. Guo, *Introduction to Wavelets and Wavelet Transforms*, Prentice-Hall, Englewood Cliffs, NJ, USA, 1995.
- [19] B. C. J. Moore, *An Introduction to the Psychology of Hearing*, Academic Press, 4th edition, 1997.
- [20] G. Strang and T. Nguyen, *Wavelets and Filter-Banks*, Wellesley-Cambridge Press, Wellesley, Mass, USA, 1996.

- [21] R. R. Coifman and M. V. Wickerhauser, "Entropy-based algorithms for best basis selection," *IEEE Transactions on Information Theory*, vol. 38, no. 2, pp. 713–718, 1992.
- [22] D. B. Percival and A. T. Walden, *Wavelet Methods for Time Series Analysis*, Cambridge University Press, Cambridge, UK, 2000.
- [23] ITU-R Recommendation BS.1387, "Method for objective measurements of perceived audio quality," 1998.
- [24] J. M. Morris and V. Akunuri, "Minimum duration orthonormal wavelets," *Optical Engineering*, vol. 35, no. 7, pp. 2079–2087, 1996.
- [25] R. M. Rao and A. S. Bopardikar, *Wavelet Transforms: Introduction to Theory and Applications*, Addison-Wesley, Reading, Mass, USA, 1998.