Lapped Transforms for Efficient Transform/Subband Coding

HENRIQUE S. MALVAR, MEMBER, IEEE

Abstract—Two new lapped transforms for subband/transform coding of signals are introduced: a new version of the lapped orthogonal transform (LOT), which can be efficiently computed for any transform length; and the modulated lapped transform (MLT), which is based on a modulated quadrature mirror filter (QMF) bank. The MLT can also be efficiently computed by means of a type-IV discrete sine transform (DST-IV). The LOT and the MLT are both asymptotically optimal lapped transforms for coding an AR(1) signal with a high intersample correlation coefficient. The coding gains of the LOT and MLT of length M are higher than that of the discrete cosine transform (DCT) of the same length; they are actually close to the coding gains obtained with a DCT of length $2M$. An MLT-based adaptive transform coder (ATC) for speech signals has been simulated; the coder is essentially free from frame rate noise and has a better spectral resolution than DCT-based ATC systems.

I. Introduction

TRANSFORM coding and subband coding are well-known approaches to the efficient waveform representation at medium and low bit rates [1], [2]. In fact, transform coding is a special case of subband coding, in which the impulse responses of the synthesis filters are the transform basis functions, the impulse responses of the analysis filters are equal to the time-reversed basis functions, and the decimation factor in each band is equal to the transform length [1], [3], [4]. The design of efficient transform or subband coders has two major aspects [2]–[5]: first, the choice of the transform or the filter bank; and second, the design of adaptive quantizers for the transform coefficients or the outputs of the analysis filter bank. This paper will focus on a class of solutions to the first problem.

In transform coding, the discrete cosine transform (DCT) [1] is almost always employed, because it is a good approximation to the statistically optimal Karhunen–Loève transform (KLT), for a wide class of signals [2]. In subband coding, the design of the band-splitting filters depends on the application. For speech, FIR filters with impulse response lengths of up to six times the number of bands are commonly used [5], mainly if the number of bands is small. In image, coding, shorter filters are preferred, in order to avoid ringing effects around sharp edges [4]. In any case, the cascade of the analysis and synthesis filter banks should lead to an overall impulse response that is as close as possible to a delayed impulse, so that reconstruction errors will be due only to quantization noise.

Manuscript received March 22, 1989; revised August 3, 1989.
The author is with the Department of Electrical Engineering, Universidade de Brasília, 70910 Brasília, DF, Brazil.
IEEE Log Number 9034981.

Since the outputs of a critically sampled filter bank with M bands are decimated by a factor of M, aliasing will occur. Thus, a necessary condition for nearly perfect reconstruction is that the synthesis filter bank should be able to cancel out that aliasing. Filter banks with such a property are generally referred to as quadrature mirror filters (QMF's) [7]. A rigorous analysis of the conditions for alias-free reconstruction, and also for perfect reconstruction, appears in [6]. In [8], a general synthesis procedure is presented for QMF structures with perfect signal reconstruction based on FIR filters.

In this paper we will present a family of perfect-reconstruction QMF filter banks that are characterized by three properties: 1) the synthesis filters are FIR, with lengths equal to twice the number of bands; 2) the analysis filters are identical, within time reversal, to the synthesis filters, which implies that the analysis and synthesis banks have identical magnitude frequency responses; and 3) the outputs of the filter banks can be obtained by means of a fast transform of length M and little additional computation. The perfect reconstruction property is guaranteed by means of an orthogonality condition that is in fact equivalent to the losslessness property defined in [8].

The general structure of such filter banks is shown in Fig. 1, where $\boldsymbol{P}$ is a $2M \times M$ matrix whose columns are the impulse responses of the synthesis filters, and $\boldsymbol{P}'$ (the transpose of $\boldsymbol{P}$) defines the analysis filters. The system is critically sampled, so that the outputs of the analysis filter bank are decimated by a factor of M, which is equal to the number of bands, and the decimated subband signals are $X_0(m), X_1(m), \cdots, X_{M-1}(m)$. The coded subband signals $Y_0(m), Y_1(m), \cdots, Y_{M-1}(m)$ are upsampled by a factor of M, and fed to the synthesis filter bank defined by $\boldsymbol{P}$. In terms of practical implementation, the outputs of the analysis filter bank are computed only once for every M samples that are shifted in on the left of Fig. 1. Also, the outputs of the synthesis filter bank are computed once for every M samples that are shifted out on the right of Fig. 1, in an overlap-add operation [3].

Because of their relatively short lengths, these filter banks can also be viewed as block transforms, in which the basis functions overlap the adjacent blocks by 50%.

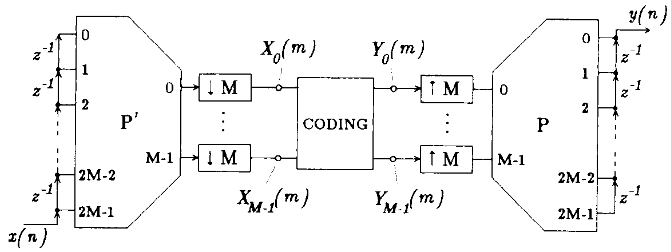

Fig. 1. Signal coding with a Lapped Transform P.

Thus, we will refer to these filter banks as *lapped transforms*. Within this class, the lapped orthogonal transfrom (LOT) [9]–[13] will be considered in Section II, where a new general formulation for the LOT will be presented. In Section III, we will introduce the modulated lapped transform, which is a particular case of a family of filter banks suggested in [15]. In Section IV, we will show that the LOT and MLT are asymptotically optimal lapped transforms. Finally, in Section V, an example of adaptive transform coding (ATC) of speech demonstrates that the MLT has two advantages over the DCT: no frame rate noise and better spectral resolution.

II. Lapped Orthogonal Transforms

The LOT, introduced in [9], was developed with the aim of reducing the blocking effects (discontinuities in the reconstructed signal at the block boundaries) in image coding. In order to keep the direct and inverse transform matrices as the transpose of each other (so that the direct transform flowgraph is just the transpose of the inverse flowgraph), each LOT basis function must be orthogonal not only to the other functions in the same block, but also to the functions in the two adjacent blocks. In terms of the matrix P of Fig. 1, the following conditions [10], [11], [13] must hold:

$$P'P = I \tag{1}$$

and

$$P'WP = 0 \tag{2}$$

where

$$W \equiv \begin{pmatrix} 0 & I \\ 0 & 0 \end{pmatrix}. \tag{3}$$

The I and 0 matrices above have dimensions $M \times M$.

Equation (1) forces orthogonality of the basis functions within the same block, whereas (2) forces orthogonality of the overlapping portions of the basis functions of adjacent blocks.

In [10]–[12] it was shown that the orthogonality conditions were satisfied by the following construction:

$$P = \frac{1}{2}\begin{pmatrix} D_e - D_o & D_e - D_o \\ J(D_e - D_o) & -J(D_e - D_o) \end{pmatrix} Z \tag{4}$$

where Z is an orthogonal matrix of order M, J is the "counteridentity"

$$J = \begin{pmatrix} 0 & \cdots & 0 & 1 \\ 0 & \cdots 0 & 1 & 0 \\ \vdots & & & \vdots \\ 1 & 0 & \cdots & 0 \end{pmatrix} \tag{5}$$

and D_e and D_o are the $M \times M/2$ matrices containing the even and odd DCT functions, respectively. Calling $[A]_{nk}$ the element of a matrix A in the nth row and kth column, we have

$$[D_e]_{nk} = c(k)\sqrt{\frac{2}{M}}\cos\left(\frac{\pi}{M}2k\left(n + \frac{1}{2}\right)\right) \tag{6}$$

and

$$[D_o]_{nk} = \sqrt{\frac{2}{M}}\cos\left(\frac{\pi}{M}(2k+1)\left(n + \frac{1}{2}\right)\right) \tag{7}$$

for $n = 0, 1, \cdots, M-1$, $k = 0, 1, \cdots, M/2 - 1$, where

$$c(k) = \begin{cases} 1/\sqrt{2}, & k = 0, \\ 1, & \text{otherwise.} \end{cases} \tag{8}$$

The definition in (4) is somewhat restrictive, since not all matrices P that satisfy (1) and (2) can be written as in (4). Nevertheless, efficient LOT's can be generated by appropriate choices for the matrix Z, and two criteria have been considered for the design of Z. In [10]–[12], Z was obtained as the orthogonal matrix that maximizes the coding gain, by means of the solution of an eigenvector problem, and in [13] Z was obtained, through a QR decomposition, as the orthogonal matrix that leads to good stopband attentuations for the frequency responses of all basis functions. It is interesting that, in both cases, the optimal Z can be closely approximated by

$$Z = \begin{pmatrix} I & 0 \\ 0 & \tilde{Z} \end{pmatrix} \tag{9}$$

where $\tilde{Z}$ is a cascade of $M/2 - 1$ plane rotations [11], [12]

$$\tilde{Z} = T_1 T_2 \cdots T_{M/2-1} \tag{10}$$

where each plane rotation is defined by

$$T_i = \begin{pmatrix} I & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & Y(\theta_i) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & I \end{pmatrix}. \tag{11}$$

The matrix $Y(\theta_i)$ is a 2×2 butterfly

$$Y(\theta_i) = \begin{pmatrix} \cos\theta_i & \sin\theta_i \\ -\sin\theta_i & \cos\theta_i \end{pmatrix} \tag{12}$$

where θ_i is the rotation angle, and the top-left identity factor in (11) is of order $i - 1$. With Z as in (9), the first $M/2$ columns of P will have even symmetry, and the last $M/2$ columns will have odd symmetry.

The LOT defined in (4), (9), and (10) can be computed efficiently. We need M butterflies with factors equal to $+1$ or -1 and $M/2 - 1$ nontrivial butterflies for $\tilde{Z}$, besides a length-M DCT [11], [12]. Therefore, the computational overhead involved in moving from DCT-based coding to LOT-based coding is small enough that in most systems there will be enough computing power to accommodate the LOT. However, the approximation in (10) is not good when M is greater then 16, because the first and second odd basis functions will show some discontinuities in their central values for any choice of the angles θ_i. In some applications, like adaptive transform coding of speech [1], [16], [17], M is usually much larger, ranging typically from 64 to 256. For such applications, we need another formulation of the fast LOT.

We introduce now a new definition for the LOT that is fast-computable for any length. The basic point is to use another approximation for Z, which changes (4) into

$$P = \frac{1}{2}\begin{pmatrix} D_e - D_o & D_e - D_o \\ J(D_e - D_o) & -J(D_e - D_o) \end{pmatrix} \cdot \begin{pmatrix} I & \mathbf{0} \\ \mathbf{0} & C^{\mathrm{II}}_{M/2} S^{\mathrm{IV}}_{M/2} \end{pmatrix} R \tag{13}$$

where $C^{\mathrm{II}}_{M/2}$ and $S^{\mathrm{IV}}_{M/2}$ are the DCT-II and DST-IV matrices [19], [20], defined by

$$[C^{\mathrm{II}}_K]_{kr} = c(k)\sqrt{\frac{2}{K}}\cos\left(\frac{\pi}{K}k\left(r + \frac{1}{2}\right)\right) \tag{14}$$

with $c(k)$ as in (8), and

$$[S^{\mathrm{IV}}_K]_{kr} = \sqrt{\frac{2}{K}}\sin\left(\frac{\pi}{K}\left(k + \frac{1}{2}\right)\left(r + \frac{1}{2}\right)\right). \tag{15}$$

We should note that the DCT-II is actually an inverse DCT. The factor R in (13) is a permutation matrix given by

$$R = \begin{bmatrix} 1 & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & \vdots & 1 & \vdots & & \vdots & \vdots \\ \vdots & & 0 & & & & \\ & & \vdots & & & \vdots & \\ & \vdots & & & & 0 & \\ & 0 & & \vdots & & 1 & \\ & 1 & & 0 & & 0 & \\ & 0 & & 1 & & \vdots & \\ & \vdots & & 0 & & & \vdots \\ \vdots & & \vdots & \vdots & & \vdots & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 1 \end{bmatrix} \tag{16}$$

which serves to order the LOT basis functions in terms of increasing center frequencies of the associated frequency responses.

It is easy to verify that the LOT basis functions defined in (13) satisfy the following symmetry condition:

$$[P]_{nk} = (-1)^k [P]_{2M-1-n,k} \tag{17}$$

which guarantees that all the filters in the analysis and synthesis filter banks have linear phase. In this way, the delay from input to the output of any analysis filter is the same, and is frequency independent. This property may be advantageous in applications where features are detected from the subband signals.

The transforms involved in the new LOT, namely the DCT, the DCT-II, and the DST-IV, can all be computed by means of fast algorithms [19]–[22]. Since the overall computational complexity of the DCT-II, the DST-IV (which are of length $M/2$), and the butterflies with factors equal to $+1$ and -1 in (13) is about the same as that of a length-M DCT, we see that the new LOT leads to a computational overhead of about 100% over the DCT. In a speech coding application, for example, the computation of the direct and inverse LOT's for $M = 256$ would require fewer than 250 operations (multiplies and adds) per sample. With a sampling frequency of 8 kHz, a signal processing chip like the TMS320C25 or DSP56001 can perform about 1000 operations per sample, so that less than 25% of the chip's computing power would be used for the filter banks.

The block diagram of the fast LOT is shown in Fig. 2. Comparing it to Fig. 1, we note that the input signal passes only through M unity-delay (z^{-1}) branches, and that $M/2$ branches with a delay of M samples now appear after the DCT. We recall that the outputs are only computed for every M samples that are shifted in, because of the decimators at the right of Fig. 2. The flowgraph of the inverse LOT is just the transpose of that of Fig. 2, but with the z^{-M} delays moved to the branches marked with asterisks.

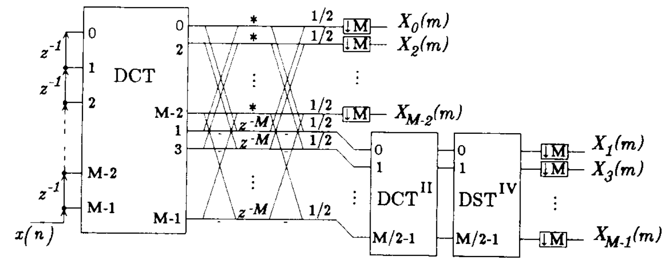

Fig. 2. Flowgraph of the Lapped Orthogonal Transform.

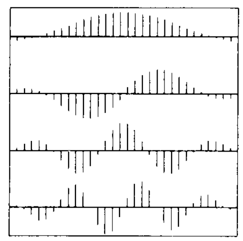

Fig. 3. The first four basis functions of the LOT, for $M = 16$.

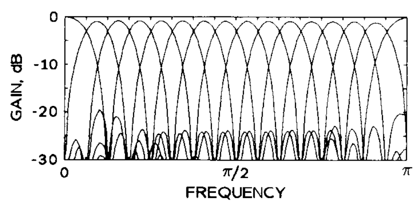

Fig. 4. Magnitude frequency responses of the LOT filter bank, for $M = 16$.

In terms of practical implementation, the decimators can be put before the DCT, so that the z^{-M} delays will become z^{-1} delays.

The first four basis functions (synthesis filter impulse responses) of the new LOT are shown in Fig. 3, for $M = 16$, and the magnitude frequency responses of all filters of the synthesis bank are shown in Fig. 4. We recall that the magnitude responses of the analysis filter bank are identical to those of the synthesis bank, by the definition of lapped transforms. All filters have a linear phase response, because of the symmetry in (17). The center frequency of the kth basis function is $(2k + 1)\pi/2M$, and the -3 dB bandwidth is π/M. We note that the mainlobes of all but the first and last functions have approximately the same shape, and most sidelobes are at a relatively low level of -24 dB, although some of the sidelobes are a little higher (at -20 dB).

An important property of the LOT filter bank is that all filter responses have zero gain at $\omega = 0$, except for the first function. Thus, a constant input signal can be represented only by the output of the first filter. This is useful in image coding, because flat background areas can be coded with a minimum number of bits. In fact, in the Appendix we will show that an optimal lapped transform should have this property.

III. Modulated Lapped Transforms

Another way to arrive at a good set of basis functions for a lapped transform is to define an appropriate low-pass filter prototype in a modulated filter bank structure [23], [24]. If the length of the low-pass prototype is chosen to be equal to $2M$, it is possible to achieve not only aliasing cancellation, but also perfect reconstruction with identical analysis and synthesis filters, as noted in [25].

Referring to Fig. 1, and calling $p_k(n) = [\boldsymbol{P}]_{nk}$ the impulse response of the kth synthesis filter, the modulated filter bank is based on the construction

$$p_k(n) = h(n)\sqrt{\frac{2}{M}}\cos\left[\frac{\pi}{M}\left(k + \frac{1}{2}\right)\cdot\left(n - \frac{L-1}{2}\right) + \frac{\pi}{4}\right] \tag{18}$$

for k even, and

$$p_k(n) = h(n)\sqrt{\frac{2}{M}}\sin\left[\frac{\pi}{M}\left(k + \frac{1}{2}\right)\cdot\left(n - \frac{L-1}{2}\right) + \frac{\pi}{4}\right] \tag{19}$$

for k odd, where L is the length of the low-pass prototype $h(n)$. The factor $\sqrt{2/M}$ is necessary to generate an orthogonal transform implementation, as we will see later.

The above construction was suggested in [23], and it was analyzed in more detail in [14], [15], and [24], without any restrictions on the length L of the low-pass prototype. In [23] and [24], it was shown that the aliasing terms between neighboring bands are cancelled, but perfect reconstruction is not necessarily achieved. When $M \leq L \leq 2M$, however, Princen and Bradley [14], [15] have shown that perfect reconstruction is possible, if the low-pass prototype (also referred to as the window function) satisfies the constraints

$$h(2M - 1 - n) = h(n) \tag{20}$$

and

$$h^2(n) + h^2(n + M) = a \tag{21}$$

where a is any constant. In order to keep orthogonality of the $\boldsymbol{P}$ matrix generated by (18) and (19), we must choose $a = 1$.

We will refer to the filter bank defined in (18)–(21) with $L = 2M$ as the Modulated Lapped Transform (MLT), since it belongs to the class of lapped orthogonal transforms, i.e., its $\boldsymbol{P}$ matrix satisfies (1) and (2), and its basis functions come from a modulated filter bank. We will reserve the term MLT, however, to be used with a particular choice for the window $h(n)$ to be introduced now, namely

$$h(n) = \sin\left[\frac{\pi}{2M}\left(n + \frac{1}{2}\right)\right] \tag{22}$$

with $n = 0, 1, \cdots, 2M - 1$, because other terms have been used for the general case: "polyphase quadrature filters" [23], "parallel QMF banks" [24], "pseudo-QMF's" [25], and "time domain aliasing cancellation" [14]–[16]. The choice for $h(n)$ in (22) will be justified in Section IV, where we will study the coding efficiency of the LOT and the MLT. It is importrant to note that the MLT is not a particular case of the fast LOT in (4), since there is no matrix $\boldsymbol{Z}$ that can generate the MLT as defined in (18)–(22). In fact, the MLT is a particular case of the oddly stacked time-domain aliasing cancellation filter bank [15].

Modulated filter banks can generally be implemented by means of a fast sine or cosine transform of length $2M$ [15], [16], [23]–[26]. The MLT, however, is more efficient because it can be implemented by means of a fast transform of length M, as we will see in the following.

With the synthesis filters of the MLT defined in (18)–(22), the outputs of the analysis filter bank in Fig. 1 can be written as

$$X_k(m) = \sum_{n=0}^{2M-1} x(mM + n)\, p_k(2M - 1 - n) \tag{23}$$

because the impulse responses of the analysis filters are equal to the time-reversed impulse responses of the synthesis filters. With the symmetry constraint for $h(n)$ in (20), (23) is equivalent to

$$X_k(m) = \sum_{n=0}^{2M-1} x_m(h)\, h(n) \sqrt{\frac{2}{M}} \cos\left[\frac{\pi}{M}\left(k + \frac{1}{2}\right) \cdot \left(M - n - \frac{1}{2}\right) + \frac{\pi}{4}\right] \tag{24}$$

for k even, and

$$X_k(m) = \sum_{n=0}^{2M-1} x_m(h)\, h(n) \sqrt{\frac{2}{M}} \sin\left[\frac{\pi}{M}\left(k + \frac{1}{2}\right) \cdot \left(M - n - \frac{1}{2}\right) + \frac{\pi}{4}\right] \tag{25}$$

for k odd. We note that m is the block index, i.e., the time index after decimation by M, so that $x_m(n) \equiv x(mM + n)$, and $m = 0, 1, \cdots, M - 1$.

It is possible to put (24) anad (25) in a common form, after some manipulations. The result is

$$X_k(m) = \gamma(k) \sqrt{\frac{2}{M}} \sum_{n=0}^{2M-1} x_m(n)\, h(n) \cdot \cos\left[\frac{\pi}{M}\left(k + \frac{1}{2}\right)\left(n + \frac{M+1}{2}\right)\right] \tag{26}$$

where

$$\gamma(k) = \begin{cases} (-1)^{(2+k)/2}, & k \text{ even} \\ (-1)^{(k-1)/2}, & k \text{ odd.} \end{cases} \tag{27}$$

The key to an efficient implementation is to define a new sequence $y_m(n)$ by

$$y_m(n) = \begin{cases} x_m(n + M/2)\, h(n + M/2) \\ \quad - x_m(M/2 - n - 1) \\ \quad \cdot h(M/2 - n - 1), \\ \qquad n = 0, \cdots, M/2 - 1 \\ x_m(n + M/2)\, h(n + M/2) \\ \quad + x_m(5M/2 - n - 1) \\ \quad \cdot h(5M/2 - n - 1), \\ \qquad n = M/2, \cdots, M - 1. \end{cases} \tag{28}$$

Using the above equation, it is easy to show that the filter bank outputs are given by

$$X_k(m) = \sqrt{\frac{2}{M}} \sum_{n=0}^{M-1} y_m(n) \sin\left[\frac{\pi}{M}\left(k + \frac{1}{2}\right)\left(n + \frac{1}{2}\right)\right] \tag{29}$$

where we have dropped $\gamma(k)$, since it only affects the relative signs of the filter bank outputs. We recognize in the above equation that $X_k(m)$ can be obtained as the orthogonal DST-IV of $y_m(n)$.

From (28) and (29), we see that the MLT analysis filter bank can be implemented in two steps: first, we compute the butterflies in (28); and second, we calculate the DST-IV of the sequence $y_m(n)$. We recall that the computational complexity of the DST-IV is the same as that of the DCT [19], [20]. The flowgraph of the direct MLT (analysis) is shown in Fig. 5(a), where $c^r \equiv \cos(\pi r/4M)$ and $s^r \equiv \sin(\pi r/4M)$. The inverse MLT (synthesis) flowgraph is shown in Fig. 5(b). With $M = 64$, for example, the number of operations per sample necessary to compute the direct and inverse MLT is about 30, which is a 40% savings over the LOT.

The first four basis functions (synthesis filter impulse responses) of the MLT are shown in Fig. 6, for $M = 16$, and the magnitude frequency responses of all filters are shown in Fig. 7. As for the LOT, the center frequency of

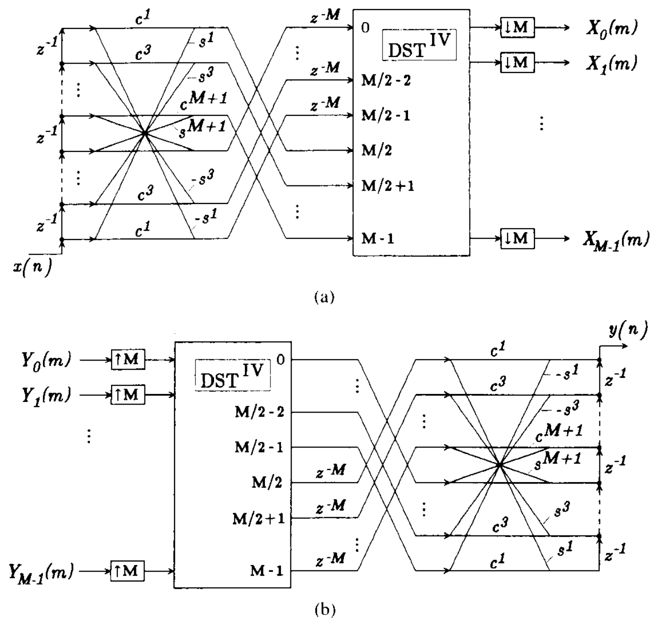

Fig. 5. Flowgraph of the Modulated Lapped Transform. The butterfly transmittances are $c^r \equiv \cos(\pi r/4M)$ and $s^r \equiv \sin(\pi r/4M)$.

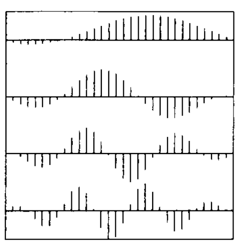

Fig. 6. The first four basis functions of the MLT, for $M = 16$.

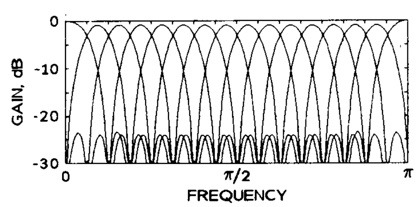

Fig. 7. Magnitude frequency responses of the MLT filter bank, for $M = 16$.

the kth basis function is $(2k + 1)\pi/2M$, and the -3 dB bandwidth is π/M. The sidelobes are at virtually the same level, -24 dB, which is slightly better than the sidelobe levels of the LOT. The MLT filter bank, like the LOT, has the property that all filter responses have infinite attenuation at $\omega = 0$, except for the first function. Thus, DC is captured only by the first function.

The small price to be paid for the lower computational complexity of the MLT is that the filter impulse responses do not have even/odd symmetry, like the LOT. Therefore, their frequency responses do not have linear phase. Nevertheless, since the analysis filters are the equal to the time-reversed synthesis filters, the overall impulse response of any channel has even symmetry, and so the overall group delay is equal to $2M - 1$, for all channels. Linear phase is not generally a required property, since in most applications (including transmultiplexing) the analysis or synthesis filters are not used alone, i.e., if a signal is processed by the kth analysis filter, it will also be processed by the kth synthesis filter. Thus, only their cascade connection is required to have linear phase.

IV. Coding Efficiency

A major issue in transform and subband coding is the efficiency of the transform or filter bank employed. There are many meaningful measures of transform efficiency [2], but perhaps the single most important measure is the coding gain G_{TC}, defined by [1]

$$G_{TC} = \frac{\frac{1}{M}\sum_{i=1}^{M} \sigma_i^2}{\left(\prod_{i=1}^{M} \sigma_i^2\right)^{1/M}} \tag{30}$$

where σ_i^2 is the variance of the output of the ith analysis filter, i.e., the ith diagonal entry of the matrix

$$\boldsymbol{R}_0 = \boldsymbol{P}'\boldsymbol{R}_{xx}\boldsymbol{P}. \tag{31}$$

In the above equation, $\boldsymbol{P}$ is the lapped transform matrix, and $\boldsymbol{R}_{xx}$ is the covariance matrix of a block of $2M$ samples of $x(n)$. The importance of the G_{TC} measure is that it indicates the factor by which the mean-square reconstruction error is reduced, compared to direct quantization of the signal (PCM).

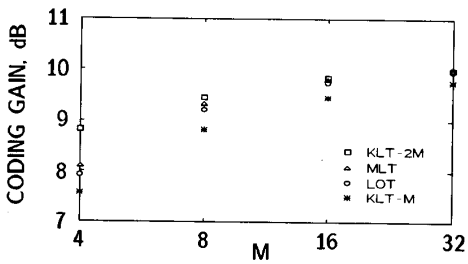

Fig. 8. Coding gain of the LOT and the MLT, for an AR(1) signal with $\rho = 0.95$. The curves labeled KLT-$2M$ and KLT-M correspond to KLT's of length $2M$ and M, respectively.

The relative efficiency of transforms is generally evaluated for a first-order autoregressive process, for which

$$[R_{xx}]_{ij} = \rho^{|i-j|}, \tag{32}$$

where ρ is the intersample correlation coefficient. In fact, in [27] it was shown that the DCT is an asymptotically optimal transform, because it is the limit of the KLT as ρ approaches unity.

The maximum G_{TC} that can be achieved by a lapped transform is upper-bounded by that of a KLT of length $2M$, as discussed in the Appendix, since the KLT maximizes G_{TC}, by definition [1]. This upper bound is not tight, however, since a lapped transform $\boldsymbol{P}$ must be restricted to the space of matrices that satisfy (1) and (2). In this section we will compare the fast-computable LOT of (13) to the MLT and the KLT.

In Fig. 8, the coding gain is plotted as a function of the number of bands (or the transform size) M, for $\rho = 0.95$. We see that the coding gains of the LOT and the MLT rapidly approach that of a KLT of length $2M$. For $M = 8$, a common transform size in image coding, the LOT has a coding gain that is 0.4 dB above that of a KLT of length 8, and the coding gain of the MLT is 0.11 dB higher than that of the LOT. Also, the coding gain of the MLT is only 0.13 dB below the maximum value of 9.46, which is the coding gain of a KLT of length 16.

The plot in Fig. 8 shows that the coding gains of the LOT and the MLT approach that of the KLT of length $2M$ as M grows. In fact, in the Appendix, we show that both the LOT and the MLT are asymptotically optimal, in the sense of satisfying the necessary conditions for optimality of a lapped transform, as $\rho \to 1$. This is an important property in adaptive signal coding, because it means that if the signal gets too close to an AR(1) process with $\rho \simeq 1$, e.g., in flat background areas in images, the transform will lead to a maximum coding gain.

The good performance of the LOT for image coding with small M was reported in [10]–[12]. In [29], it was shown that transform domain filtering of speech with the LOT is free from blocking effects (also referred to as frame rate noise [17]). In the next section, we will consider an example of the application of the LOT and the MLT to adaptive speech coding.

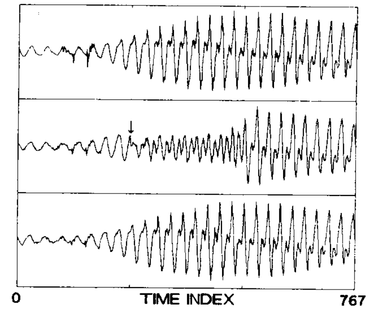

Fig. 9. Top: original speech segment, containing three frames of 256 samples each. Middle: coded speech at 10 kbps, using the DCT; note the discontinuity indicated by the arrow. Bottom: coded speech at 10 kbps, using the MLT.

V. Application to Speech Coding

In order to evaluate experimentally the performance of the new LOT and the MLT, we have simulated on a general-purpose computer of the ATC coder described in [17], which is a variation of the ATC coder of [18]. The transform coefficients (i.e., the outputs of the analysis bank) were scalar quantized, with an adaptive bit allocation based on an estimate of the power spectrum derived from an windowed "pseudocepstrum" [17]. Pitch information was added to the windowed pseudocepstrum exactly as described in [17]. The differences between our ATC implementation and that of [17] were: first, we used an inverse DCT to compute the pseudospectrum, instead of the symmetrical discrete Fourier transform (SDFT), since the results are virtually the same, but the SDFT requires an odd block length; second, instead of using the SDFT filter bank, we used the DCT, the LOT, and the MLT.

In Fig. 9, we have, in the top trace, a segment of 768 samples of speech that is the onset of the word "wait" spoken by a female speaker. The speech was low-pass filtered at 3.3 kHz, and sampled at 8 kHz. In the middle trace, we have the reconstructed signal using the DCT, for a total bit rate of 10 kbps (with 20% used for side information), with $M = 256$. We note the strong discontinuity at the beginning of the second frame (indicated by the arrow). We note also that the harmonics were not well represented in the second frame, which was poorly reconstructed. In the bottom trace, we have the same segment processed by the same coder, but with the DCT replaced by the MLT. With the MLT, there are no discontinuities, and the second frame is well reconstructed.

In listening tests, the MLT-coded speech is completely free from the periodic "clicks" that are generated by these discontinuities. Although this frame rate noise can be reduced in a DCT-based coder by frame overlapping [17], the low-frequency components of the frame rate noise are still perceivable, unless an overlapping of about 10% is used, which slightly increases the bit rate.

The better performance of the MLT can also be verified in Fig. 10. In the top trace, we have the log magnitude DCT spectrum (solid line), superimposed to its cepstrally based estimate (dashed line), for the second frame of Fig.

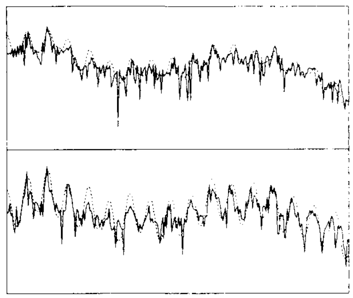

Fig. 10. Top: log magnitude spectra for the second frame of Fig. 9; solid line: DCT spectrum, dashed line: estimate based on the windowed pseudocepstrum. Bottom: log magnitude spectra for the same speech segment, but using the MLT; solid line: MLT spectrum, dashed line: estimate.

9. We note the poor resolution of the harmonics, due to the rectangular window that is intrinsic to the DCT. This is the cause of the inaccurate reconstruction in Fig. 9.

In the bottom trace, we have the log magnitude MLT spectrum, superimposed to its cepstrally based estimate. The harmonics are well resolved because of two facts. First, the basis functions of the MLT have twice the length of those of the DCT, and so each analysis filter of the MLT bank has a narrower bandwidth than those of the DCT. Second, the sine window (22) that is applied to all basis functions makes the stopband attenuations of the MLT filter bank much higher than those of the DCT filter bank. We should note that the same spectral resolution could not be obtained by using a DCT of length $M = 512$, because of the rectangular window of the DCT. With the better harmonic resolution of the MLT, the reconstructed speech sounds less resonant than the one coded with the DCT, mainly at low bit rates.

With the LOT, results similar to those for the MLT are obtained. However, with the LOT there is still some slight frame rate noise (at a much lower level than that from the DCT) that can be heard during sustained voiced segments. The reason for the incomplete removal of frame noise is that the second basis function of the LOT in eqrefeq:newlot (the one that corresponds to the filter with the second lowest center frequency) has a small jump in its values near the center.

It should be noted that the distortions in frame two of the DCT-processed speech of Fig. 9 are not typical; in most frames, the distortions are at lower levels, but still quite noticeable in listening tests. The segmental SNR of the MLT-coded speech was, in most cases, only between 0.5 and 1.5 dB higher than that for DCT-coded speech. Thus, the main advantage of the MLT is the elimination of the frame rate noise.

VI. Conclusion

We have presented a new general formulation for the Lapped Orthogonal Transform (LOT) [(13)], and we have introduced the Modulated Lapped Transform (MLT) [(18), (19), and (22)]. They are called lapped transforms because they are realizations of the general filter bank in Fig. 1 with identical analysis and synthesis filters, i.e., they satisfy (1) and (2). We have shown in Figs. 2 and 5 that the LOT and the MLT can be implemented using fast transforms of length M, whereas previous approaches toward the realization of filter banks for the system in Fig. 1 required a transform of length $2M$. We have also shown that the coding gain (over PCM) of the LOT and MLT is actually close to the maximum value for a length-2 M transform, and that they both asymptotically optimal lapped transforms for coding an AR(1) signal with a high intersample correlation coefficient.

The MLT leads to a higher coding gain than the LOT, especially for small transform sizes. The MLT also has a lower computational complexity than that of the LOT, in general. However, the MLT analysis and synthesis filters do not have a linear phase response; this is actually a small penalty, since the cascaded responses of all filters are linear phase, with a delay of $2M - 1$ samples, for all channels. Thus, it seems that the MLT may be a better choice for most signal coding applications, except when the number of bands M is small (e.g., $M = 8$), where previous algorithms for the LOT [12], [13] are more efficient.

Finally, it should be noted that lapped transforms have a better performance than nonlapped ones (like the DCT), not only because they have higher coding gains, but also because they lead to a strong reduction in discontinuities in the reconstructed signal at the block boundaries (the "blocking effects" in image coding [11], [12] or "clicking" sounds in speech coding [17], [29]). Lapped transforms are also more efficient for adaptive speech coding, as we have demonstrated in Section V, since their better spectral resolution leads to bit assignments that can more accurately reproduce most harmonics of voiced sounds.

Appendix

The purpose of this Appendix is to derive the necessary conditions for optimality of a lapped transform, and to show that the LOT and MLT are at least locally optimal solutions to the problem of maximizing the coding gain with a lapped transform.

Let us consider first the problem of optimizing the first basis function. Calling it $\boldsymbol{p}_0 = [\boldsymbol{P}]_{n0}$, the problem is

$$\begin{aligned} \text{maximize} \quad & \xi = \boldsymbol{p}_0' \boldsymbol{R}_{xx} \boldsymbol{p}_0 \\ \text{subject to} \quad & \boldsymbol{p}_0' \boldsymbol{W} \boldsymbol{p}_0 = 0 \\ \text{and} \quad & \boldsymbol{p}_0' \boldsymbol{p}_0 = 1 \end{aligned} \tag{A.1}$$

where $\boldsymbol{W}$ is defined in (3). This is the problem originally considered in [9]; it states that the first basis function should capture most of the signal energy, i.e., the first transform coefficient should have the maximum variance, given the orthogonality restrictions in (1) and (2).

In [9], an augmented Lagrangian method was used to solve the problem, but a restriction of even symmetry was imposed on $\boldsymbol{p}_0$. With this restriction, the optimal $\boldsymbol{p}_0$ was essentially identical to the first basis function of the LOT definition in (13). However, a solution to (A.1) may be neither symmetric nor unique, and nonsymmetrical solutions actually do exist, as we will see later.

A necessary condition for a vector p_0 to be a solution to the problem in (A.1) is that it must be a zero of the associated Lagrangian [28], i.e., there must exist two Lagrange multipliers μ and λ, associated to the restrictions in (A.1), such that

$$2R_{xx}p_0 + \mu(W + W')p_0 + 2\lambda p_0 = 0. \quad (A.2)$$

Premultiplying (A.2) by p_0', we get

$$p_0'R_{xx}p_0 + \lambda p_0'p_0 = 0 \quad (A.3)$$

from which we get $\lambda = -p_0'R_{xx}p_0$.

Also, premultiplying (A.2) by $p_0'W$ and $p_0'W'$, and using $W^2 = \mathbf{0}$, we obtain

$$2p_0'WR_{xx}p_0 + \mu p_0'WW'p_0 = 0 \quad (A.4)$$

and

$$2p_0'W'R_{xx}p_0 + \mu p_0'W'Wp_0 = 0. \quad (A.5)$$

Adding the two conditions above and using $WW' + W'W = I$ leads to

$$\mu = -2p_0'(W + W')R_{xx}p_0. \quad (A.6)$$

With R_{xx} given by (32) with $\rho \to 1$, all elements of R_{xx} will have a unity value, in the limit. Thus, we will have $(W + W')R_{xx} = R_{xx}$, and (A.2) will assume the form

$$[R_{xx} - (p_0'R_{xx}p_0)(W' + W + I)]p_0 = \mathbf{0}. \quad (A.7)$$

An optimal p_0 must satisfy not only the above equation, but also the restrictions in (A.1). Thus, (A.7) implies either

$$p_0(n) + p_0(n + M) = \left[\sum_{k=0}^{2M-1} p_0(k)\right]^{-1} \quad (A.8)$$

for $n = 0, 1, \cdots, M - 1$, or

$$\sum_{k=0}^{2M-1} p_0(k) = 0. \quad (A.9)$$

The condition in (A.8) states that if we superimpose the first basis functions of all blocks, the result should be a constant value. This is generally considered as a reasonable property in image coding, because it leads to the most efficient representation of flat background areas, as we have discussed before. What we have done above was to show that it is in fact a necessary condition for asymptotic optimality as $\rho \to 1$.

If the first basis function p_0 satisfies (A.8), the objective function ξ will have the value, as $\rho \to 1$

$$\xi = \left[\sum_{n=0}^{2M-1} p_1(n)\right]^2 = M. \quad (A.10)$$

In the second equality, we have assumed that the optimal p_0 does not satisfy (A.9), because if so, we would have $\xi = 0$ in the limit.

Thus, all feasible solutions that satisfy the necessary condition for optimality in (A.8) lead to the same value of the objective function, and so they are all globally optimal. It is easy to verify that both the first function of the LOT in (13) and the first function of the MLT in (18), (19), and (22) satisfy the necessary condition in (A.8). Thus, the first basis functions of the LOT and the MLT are both globally optimal, asymptotically.

Now we turn to the problem of optimizing all basis functions. The problem can be formulated sequentially, that is, once an optimal p_0 is found, we optimize p_1, then p_2, and so on, as suggested originally in [9]. Then, the optimization problem for the ith basis function is

$$\begin{aligned} \text{maximize} \quad & \xi_i = p_i'R_{xx}p_i \\ \text{subject to} \quad & p_i'Wp_i = 0 \\ & p_i'p_i = 1 \\ & p_i'p_j = 0, \qquad j = 0, 1, \cdots, i-1 \\ & p_i'Wp_j = 0, \qquad j = 0, 1, \cdots, i-1 \\ \text{and} \quad & p_i'W'p_j = 0, \qquad j = 0, 1, \cdots, i-1. \end{aligned} \quad (A.11)$$

The necessary condition for optimality in this case is that there must exist Lagrange multipliers μ, λ, α_j, β_j, and γ_j, for $j = 0, 1, \cdots, i - 1$, such that

$$2R_{xx}p_i + \mu(W + W')p_i + 2\lambda p_i + \sum_{j=0}^{i-1}(\alpha_j I + \beta_j W + \gamma_j W')p_j = 0. \quad (A.12)$$

Premultiplying (A.12) by p_i' and by p_r', with $0 \le r \le i - 1$, and after some manipulations, we get

$$\lambda = -p_i'R_{xx}p_i \quad (A.13)$$

and

$$\alpha_r = -2p_r'R_{xx}p_i. \quad (A.14)$$

Now, premultiplying (A.12) by $p_i'W$ and $p_i'W'$, and adding the results, we arrive at

$$2p_i'(W + W')R_{xx}p_i + \mu + \sum_{j=0}^{i-1} p_i'(\beta_j W'W + \gamma_j WW')p_j = 0. \quad (A.15)$$

Also, premultiplying (A.12) by $p_r'W$ and $p_r'W'$, with $0 \le r \le i - 1$, and adding the results, we obtain

$$2p_r'(W + W')R_{xx}p_i + \mu + \sum_{j=0}^{i-1} p_r'(\beta_j W'W + \gamma_j WW')p_j = 0. \quad (A.16)$$

Using again the fact that, in the limit as $\rho \to 1$, we have $(W + W')R_{xx} = R_{xx}$, we conclude from (A.15) that $\mu = 2\lambda$. Furthermore, since (A.16) must hold for all $r < i$, we get $\gamma_i = \beta_j = 0$, for $j = 0, 1, \cdots, i - 1$, i.e., the last two constraints in (A.11) are not binding at an optimal solution. Then, (A.12) will be equivalent to

$$[R_{xx} - (p_i'R_{xx}p_i)(W' + W + I)]p_i = \mathbf{0}. \quad (A.17)$$

We note that the above equation is identical to the necessary condition for p_0 in (A.7). However, since p_i must be orthogonal to p_0, for $i > 0$, it follows

$$\sum_{k=0}^{2M-1} p_i(k) = 0. \quad (A.18)$$

It is easy to verify that the basis functions of the LOT and the MLT satisfy the necessary conditions in (A.8) and (A.18), as well as the constraints in (A.1) and (A.11). Thus, the LOT and the MLT are at least locally optimal solutions to the problem of maximizing the coding gain. We recall that their first basis functions are globally optimal.

Due to the highly nonlinear nature of the problem in (A.11), deriving sufficient conditions for global optimality is virtually impossible. However, we must recall that if we remove the restriction in (2), i.e., if we do not require orthogonality of the overlapping portions of the basis functions of neighboring blocks, the problem becomes trivial, and the maximum coding gain is that obtained by the Karhunen–Loève transform of length $2M$. As we saw in Fig. 8, the coding gain of the MLT is so close to that of a length-$2M$ KLT that, for all practical purposes, we can consider the MLT as the optimal lapped transform. The LOT is also optimal, since its coding gain is close to that of the MLT.

Acknowledgment

The author would like to thank the following people, whose comments on his work on LOT's have contributed to this paper: Prof. D. H. Staelin (Massachusetts Institute of Technology), Dr. A. Tabatabai (Bell Communications Research), Dr. M. Vetterli (Columbia University), Prof. K. R. Rao (University of Texas at Arlington), Dr. P. P. Vaidyanathan (California Institute of Technology), and Dr. T. Lookabaugh (Compression Labs). The detailed comments of the reviewers are gratefully acknowledged.

References

[1] N. S. Jayant and P. Noll, *Digital Coding of Waveforms.* Englewood Cliffs, NJ: Prentice-Hall, 1984.

[2] R. J. Clarke, *Transform Coding of Images.* London: Academic, 1985.

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

[4] J. W. Woods and S. D. O'Neal, "Subband coding of images," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-34, pp. 1278–1288, Oct. 1986.

[5] J. H. Derby and J. W. Woods, "Multirate subband coding applied to digital speech interpolation," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-35, pp. 1684–1698, Dec. 1987.

[6] M. Vetterli, "A theory of multirate filter banks," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-35, pp. 356–372, Mar. 1987.

[7] P. P. Vaidyanatha, "Quadrature mirror filter banks, *M*-band extensions, and perfect reconstruction techniques," *IEEE ASSP Mag.*, vol. 4, pp. 4–20, July 1987.

[8] Z. Doğanata, P. P. Vaidyanathan, and T. Q. Nguyen, "General synthesis procedures for FIR lossless transfer matrices, for perfect-reconstruction multirate filter bank applications," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 36, pp. 1561–1574, Oct. 1988.

[9] P. Cassereau, "A new class of optimal unitary transforms for image processing," S. M. thesis, Dep. Elec. Eng. Compute Sci., Mass. Inst. Technol., Cambridge, May 1985.

[10] H. S. Malvar, "Optimal pre- and post-filters in noisy sampled-data systems," Tech. Rep. 519, Res. Lab. Electron., Mass. Inst. Technol., Aug. 1986.

[11] H. S. Malvar and D. H. Staelin, "Reduction of blocking effects in image coding with a lapped orthogonal transform," in *Proc. ICASSP 88*, New York, Apr. 1988, pp. 781–784.

[12] —, "The LOT: Transform coding without blocking effects," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, pp. 553–559, Apr. 1989.

[13] H. S. Malvar, "The LOT: A link between block transform coding and multirate filter banks," in *Proc. ISCAS 88*, Espoo, Finland, June 1988, pp. 835–838.

[14] J. P. Princen and A. B. Bradley, "Analysis/synthesis filter bank design based on time domain aliasing cancellation," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-34, pp. 1153–1161, Oct. 1986.

[15] J. P. Princen, A. W. Johnson, and A. B. Bradley, "Subband/transform coding using filter bank designs based on time domain aliasing cancellation," in *Proc. ICASSP 87*, Dallas, TX, Apr. 1987, pp. 2161–2164.

[16] A. W. Johnson and A. B. Bradley, "Adaptive transform coding incorporating time domain aliasing cancellation," *Speech Commun.*, vol. 6, pp. 299–308, Dec. 1987.

[17] R. V. Cox and R. E. Crochiere, "Real-time simulation of adaptive transform coding," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-29, pp. 147–154, Apr. 1981.

[18] J. M. Tribolet and R. E. Crochiere, "Frequency domain coding of speech," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-27, pp. 512–530, Oct. 1979.

[19] Z. Wang, "Fast algorithms for the discrete *W* transform and for the discrete Fourier transform," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-32, pp. 803–816, Aug. 1984.

[20] P. Yip and K. R. Rao, "Fast decimation-in-time algorithms for a family of discrete sine and cosine transforms," *Circuits, Syst., Signal Processing*, vol. 3, pp. 387–408, 1984.

[21] H. S. Malvar, "Fast computation of the discrete cosine transform through fast Hartley transform," *Electron Lett.*, vol. 22, pp. 352–353, Mar. 1986.

[22] Z. Wang, "On computing the discrete Fourier and cosine transforms," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-33, pp. 1341–1344, Oct. 1985.

[23] J. H. Rothweiler, "Polyphase quadrature filters—A new subband coding technique," in *Proc. ICASSP 83*, Boston, Mar. 1983, pp. 1280–1283.

[24] P. L. Chu, "Quadrature mirror filter design for an arbitrary number of equal bandwidth channels," *IEEE Trans. Acoust., Signal Processing*, vol. ASSP-33, pp. 203–218, Feb. 1985.

[25] M. Vetterli and D. Le Gall, "Perfect reconstruction FIR filter banks: Lapped transforms, pseudo-QMF's and paraunitary matrices," in *Proc. ISCAS 88*, Espoo, Finland, June 1988, pp. 2249–2253.

[26] H. Nussbaumer and M. Vetterli, "Computationally efficient QMF filter banks," in *Proc. ICASSP 84*, San Diego, CA, Mar. 1984, pp. 11.3.1–11.3.4.

[27] R. J. Clarke, "Relation between the Karhunen–Loève and cosine transforms," *IEE Proc., Part F*, vol. 128, pp. 359–360, Nov. 1981.

[28] D. M. Bertsekas, *Constrained Optimization and Lagrange Multiplier Methods.* New York: Academic, 1982.

[29] H. S. Malvar and R. Duarte, "Transform/subband coding of speech with the lapped orthogonal transform," in *Proc. ISCAS 89*, Portland, OR, May 1989, pp. 1268–1271.

Henrique S. Malvar (M'79–S'82–M'86) was born in Rio de Janeiro, Brazil, in 1957. He received the B.S. degree in 1977 from Universidade de Brasilia, Brazil, the M.S. degree in 1979 from Universidade Federal do Rio de Janeiro, Brazil, and the Ph.D. degree in 1986 from the Massachusetts Institute of Technology, Cambridge, all in electrical engineering.

Since 1979 he has been on the Faculty of the Universidade de Brasilia, Brazil, where he is now Associate Professor of Electrical Engineering. From 1982 to 1987 he was on leave at the Massachusetts Institute of Technology, where he had been a doctoral student from 1982 to 1986, and a Visiting Assistant Professor of Electrical Engineering for the 1986–1987 academic year. From 1985 to 1987 he had been a consultant for PictureTel Corporation, Peabody, MA, in the area of algorithms for image compression. His research interests include analog and digital filter synthesis, digital signal processing, and low-bit-rate coding of video and speech.

Dr. Malvar is a member of Sigma Xi. He was the recipient of the Young Scientist Award from the Marconi International Fellowship in 1981.