INCORPORATION OF BIORTHOGONALITY INTO LAPPED TRANSFORMS FOR AUDIO COMPRESSION

Shiufun Cheung and Jae S. Lim

Research Laboratory of Electronics
Massachusetts Institute of Technology
Cambridge, Massachusetts, U.S.A.

ABSTRACT

Acoustic signal representations used in current audio coding algorithms can be improved by the incorporation of biorthogonality into Malvar's Extended Lapped Transform (ELT). Biorthogonality allows more flexibility in the design of the analysis and synthesis windows by increasing the number of degrees of freedom. This paper examines this increase for two special cases and demonstrates the importance of the additional flexibility to the proper implementation of psychoacoustic modeling, a feature central to all modern audio compression schemes.

1. INTRODUCTION

Among the most important issues in the design of an audio compression scheme is the selection of an appropriate and efficient acoustic signal representation. In conventional audio coders, this representation is derived from a short-time spectral decomposition which serves to recast the audio signal in a domain that is not only amenable to perceptual modeling but also conducive to achieving transform coding gain. The signal decomposition is commonly achieved by a multirate filter bank or, equivalently, a lapped transform. In recent years, much attention has been devoted to the study of these filter banks. Many different designs have emerged in the context of audio compression, including schemes that use nonuniform subbands [1] and time-varying structures [2]. Present throughout these developments are two criteria that have consistently been regarded as important in the construction of these filter banks.

1. *Critical Sampling* This means that the aggregate rate of the subband channels is the same as the input sample rate. Critically sampled filter banks are also called maximally decimated.
2. *Perfect Reconstruction* This refers to signal decompositions from which the original signal can be exactly recovered in the absence of quantization distortion.

Although neither of the two features is essential to audio compression, they are nevertheless desirable. Critical sampling ensures that subsequent stages of the audio coder are not required to operate at a higher aggregate rate than the input sample rate. Perfect reconstruction allows us to isolate the introduction of signal distortion in the quantization-and-coding module, thereby simplifying the system design process.

Malvar's lapped transforms are popular implementations of critically sampled perfect-reconstruction filter banks [3]. Although lapped transforms are intended only to be uniform filter banks, they are fairly versatile. For example, the transforms can be cascaded in a hierarchical structure to yield a composite filter bank with nonuniform subbands [1]. In this paper, we consider improving one particular realization of lapped transforms by the incorporation of biorthogonality. One previous work in this direction is described in [4] in which the Lapped Orthogonal Transform (LOT) is generalized to become the BiOrthonormal Lapped Transform (BOLT). We, on the other hand, are primarily interested in improving the Extended Lapped Transform (ELT).

2. BIORTHOGONALITY IN LAPPED TRANSFORMS

2.1. The Extended Lapped Transform

Figure 1 shows the general structure of a system based on a lapped transform. The implementation in the figure is that of an M-channel filter bank with an overlapping factor of K. The duration of each transform frame or, equivalently, the length of each analysis filter is $2KM$ samples. Note that the decimation factor for each channel is M, which is equal to the total number of channels, thereby guaranteeing that the filter bank is critically sampled.

In its most general form, a lapped transform can be shown to be completely equivalent to a uniform paraunitary multirate filter bank. For the purposes of this paper, however, we are mainly interested in the particular realization that implements a cosine-modulated filter bank. For this type of filter bank, Malvar has coined the terms, Extended Lapped Transform (ELT) and Modulated Lapped Transform (MLT); the MLT corresponds to a special case ($K = 1$) of the ELT.

In matrix notation, the forward ELT is represented by the $2KM \times M$ matrix $\mathbf{P}$ where the (n,k)th element, p_{nk}, is given by

$$p_{nk} = h[n]\sqrt{\frac{2}{M}}\cos\left[\left(n+\frac{M+1}{2}\right)\left(k+\frac{1}{2}\right)\frac{\pi}{M}\right] \quad (1)$$

and the inverse transform is similarly represented by $\mathbf{Q}$

0-7803-2431-5/95 $4.00

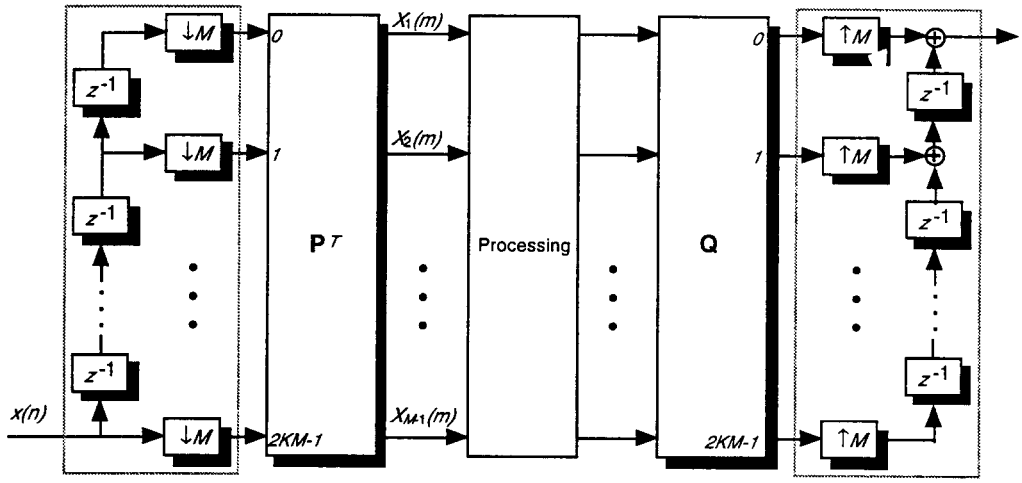


Figure 1: General structure of a system based on a lapped transform.

where q_{nk} is given by

$$q_{nk} = f[n]\sqrt{\frac{2}{M}}\cos\left[\left(n+\frac{M+1}{2}\right)\left(k+\frac{1}{2}\right)\frac{\pi}{M}\right]. \quad (2)$$

In the above notation, it is conventional to refer to $h[n]$ and $f[n]$ as the analysis window and the synthesis window, respectively.

2.2. Perfect Reconstruction—Orthogonal Case

The ELT originally conceived by Malvar is an orthogonal transform in which

$$\mathbf{P} = \mathbf{Q} \Longleftrightarrow h[n] = f[n]. \quad (3)$$

Aside from equating the analysis window and the synthesis window, Malvar further stipulates that the windows used are symmetric, such that

$$h[2KM-1-n] = h[n]. \quad (4)$$

Under these assumptions, time-domain analysis shows that perfect reconstruction is achieved if both the analysis and synthesis windows satisfy the following constraint:

$$\sum_{m=0}^{2K-1-2s} h[mM+n]h[(m+2s)M+n] = \delta(s) \quad (5)$$

for $s = 0,\ldots,K-1$ and $n = 0,\ldots,\frac{M}{2}-1$. The set of nonlinear equations represents $KM/2$ independent conditions on the KM different coefficients in the analysis window. This leaves $KM/2$ degrees of freedom for design purposes.

2.3. Perfect Reconstruction—Biorthogonal Case

Relaxation of the requirement in (3), i.e. the analysis window equals the synthesis window, leads to the loss of orthogonality in the transform. Nevertheless, perfect reconstruction can still be achieved. The resulting filter bank is known as biorthogonal. For this biorthogonal case, an equivalent time-domain analysis can be performed to show that the perfect-reconstruction requirement leads to the following constraints on the coefficients of the analysis and synthesis windows:

$$\sum_{m=0}^{2K-1-2s} f[mM+n]h[(m+2s)M+n] = \delta(s) \quad (6)$$

$$\sum_{m=0}^{2K-1-2s} (-1)^m f[mM+n] \cdot h[(m+2s)M+(M-n-1)] = 0 \quad (7)$$

for $s = 0,\ldots,K-1$ and $n = 0,\ldots,M-1$. The $2KM$ nonlinear equations shown here are not independent. Fortunately, practical experience has revealed patterns suggesting that these equations can be reduced to $(3K-1)M/2$ independent conditions. Given that there is a combined total of $2KM$ different coefficients in the analysis and synthesis windows, this leads to $(K+1)M/2$ degrees of freedom. If proven, the result would represent an increase of $M/2$ degrees of freedom over the orthogonal formulation. In what follows, we shall subtantiate this claim for two special cases.

2.4. Special Case I — K = 1

When the overlapping factor is one, only adjacent transform frames overlap. For this special case, the constraints in (6) and (7) reduce to

$$f[n]h[n] + f[n+M]h[n+M] = 1 \quad (8)$$

$$f[n]h[n+M] - f[n+M]h[n] = 0 \quad (9)$$

for $n = 0,\ldots,\frac{M}{2}-1$. This set of M equations allows M degrees of freedom in choosing the combined total of $2M$ coefficients in the analysis and synthesis windows.

One method of satisfying the above conditions is to choose freely a symmetric window $h[n]$, and then to solve for $f[n]$ by using

$$f[n] = \frac{h[n]}{h^2[n] + h^2[n+M]}. \quad (10)$$

An equivalent result has been observed in the context of time-domain aliasing cancellation [5].

2.5. Special Case II — K = 2

This case is slightly more complicated than the previous one. After some algebraic manipulation, however, the constraints in (6) and (7) can be reduced to the following set of equations:

$$\begin{aligned} f[n]h[n] + f[M+n]h[M+n] \\ +f[2M+n]h[2M+n] \\ + f[3M+n]h[3M+n] &= 1 \quad (11) \\ f[n]h[3M+n] - f[M+n]h[2M+n] \\ +f[2M+n]h[M+n] \\ - f[3M+n]h[n] &= 0 \quad (12) \\ f[n]h[2M+n] + f[M+n]h[3M+n] &= 0 \quad (13) \\ h[n]h[2M+n] + h[M+n]h[3M+n] &= 0 \quad (14) \end{aligned}$$

for $n = 0, \ldots, \frac{M}{2} - 1$ in (11), (12) and (14), and $n = 0, \ldots, M-1$ in (13). This represents $5M/2$ independent conditions on a total of $4M$ window coefficients leaving $3M/2$ degrees of freedom. The increase of $M/2$ over the orthogonal case is as predicted.

One way of designing the windows such that the above conditions are satisfied is to construct first the analysis window $h[n]$ under the constraint in (14). Once $h[n]$ is chosen, (11), (12) and (13) form a set of linear equations that can be solved to yield $f[n]$, the synthesis window.

2.6. Summary

Table I summarizes the results of the two special cases discussed above. Similar results can be derived for some higher values of K. It is worth noting that because the increase in the number of degrees of freedom appears to be $M/2$ for all K, the improvement is much more pronounced for the lower values.

Table I: Summary of the results in sections 2.4 and 2.5.

	K	Orthogonal	Biorthogonal
Number of Design Variables	1	M	$2M$
	2	$2M$	$4M$
	general	KM	$2KM$
Number of Independent Conditions	1	$M/2$	M
	2	M	$5M/2$
	general	$KM/2$	$(3K-1)M/2$
Number of Degrees of Freedom	1	$M/2$	M
	2	M	$3M/2$
	general	$KM/2$	$(K+1)M/2$
Note that the general case is not yet proven.			

3. SIGNIFICANCE OF USING BIORTHOGONALITY

The flexibility gained by incorporating biorthogonality into the lapped transform can be very useful in the design of audio compression algorithms. To understand fully these benefits requires some knowledge of the use of psychoacoustic modeling.

3.1. Psychoacoustic Modeling

Much of the recent progress in audio coding can be attributed to successful application of psychoacoustics to the coding process [6]. Unlike speech coding, and to some extent image coding, it is difficult to find effective models for the diverse sources from which sounds are generated. On the other hand, the audio receiver, namely the human auditory system, has been extensively studied and some of the results are applicable to audio coding. Of particular importance is interband masking, an effect in which one signal is rendered less audible by another signal in close spectral or temporal proximity. By properly shaping the quantization noise which is introduced by the coding process, it is possible to code at a lower rate for equivalent subjective fidelity.

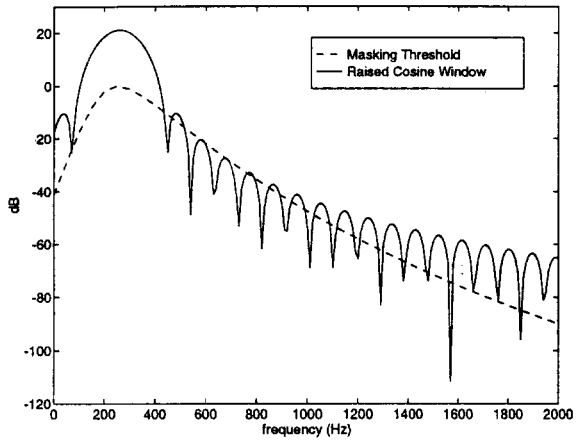
3.2. Masking Thresholds and Window Design

The application of quantization-noise shaping requires the computation of a signal-adaptive masking curve. This overall masking curve is formed by taking a weighted sum of the masking thresholds, each of which is based on the energy in its corresponding critical band. These masking thresholds, therefore, determine the necessary stop-band attenuation for the analysis filters in each subband channel. We know from [3] that the magnitude frequency response of the analysis filters can be approximately given by the magnitude frequency response of the analysis window modulated to the appropriate frequency. It follows that shaping the stop-band attenuation for the analysis filters is equivalent to shaping the stop-band attenuation of the analysis windows.

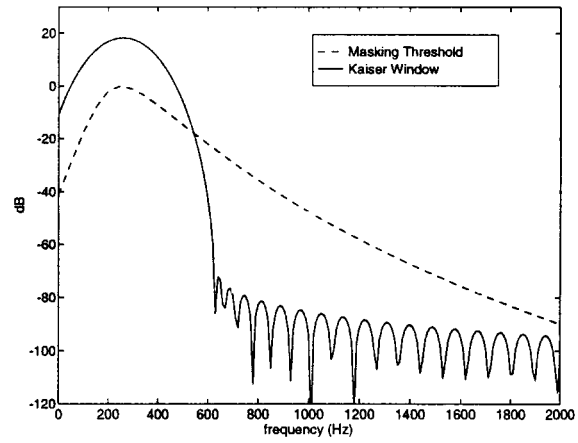
One method of achieving the stop-band attenuation necessary for the masking-curve computation is to increase the number of degrees of freedom by lengthening the analysis window. If M, the number of subband channels, is held constant, this approach increases K, the overlapping factor, a side effect that would be acceptable in the absence of quantization noise. In the presence of quantization noise, however, audio coding schemes often suffer from pre-echo artifacts when the signal contains sharp transients. This artifact is exacerbated if K is large. It is, therefore, to our advantage to keep the value of K relatively small. In light of this, we propose that biorthogonality be used to provide enough degrees of freedom to achieve the necessary stop-band attenuation.

3.3. Orthogonal vs. Biorthogonal for K = 1

Consider the design of the analysis window for the case of $K = 1$. In the orthogonal case, the design is subject to the $M/2$ constraints given in (5). This greatly restricts the



(a) The Orthogonal Case—Raised Cosine Window



(b) The Biorthogonal Case—Kaiser Window, $\beta = 12$

Figure 2: Comparison of a window used for the orthogonal lapped transform and a window used for the biorthogonal lapped transform in relation to the masking threshold of a signal at 250 Hz.

ability of the designer to choose an appropriate window. One popular choice that satisfies the perfect-reconstruction requirement is the raised-cosine or Hanning window. In figure 2(a), we have shown the magnitude frequency response of the window modulated to 250 Hz in relation to a masking threshold derived from a signal at the same frequency. For this, we have assumed a standard sampling rate of 48 kHz so the 512-point window has a duration of 10.7 msec. The masking threshold is computed based on [7]. Note that in the bands close to the signal frequency the sidelobes of the window are sufficiently high to invalidate subsequent masking-curve calculations. Unfortunately, the nonlinear nature of the equations in (5) makes it difficult for designers to improve upon the analysis window by performing the usual mainlobe-width-versus-sidelobe-attenuation tradeoff.

This situation is rectified in the biorthogonal case. The increase in the number of degrees of freedom to M means that the designer can choose any appropriate symmetric window for the analysis filter. For example, it is possible to use the Kaiser window in which the tradeoff between mainlobe width and sidelobe attenuation is controlled by a single parameter β. The relation between the magnitude frequency response of the modulated Kaiser window with $\beta = 12$ and the corresponding masking threshold is shown in figure 2(b).

4. CONCLUSION

In this paper, we have demonstrated, for two special cases, that the incorporation of biorthogonality into the ELT provides an increase of $M/2$ degrees of freedom in our choice of windows. The increase, while modest, is nevertheless significant, especially for lower values of K. In particular, the added flexibility is important for the proper implementation of psychoacoustic modeling in audio compression schemes.

5. REFERENCES

[1] P. A. Monta and S. Cheung, "Low rate audio coder with hierarchical filterbanks and lattice vector quantization," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. II, pp. 209–212, April 1994.

[2] D. Sinha and A. H. Tewfik, "Low bit rate transparent audio compression using adapted wavelets," *IEEE Transactions on Signal Processing*, vol. 41, pp. 3463–3479, December 1993.

[3] H. S. Malvar, *Signal Processing with Lapped Transforms*. Artech House, 1991.

[4] P. P. Vaidyanathan, "Causal FIR matrices with anticausal FIR inverses, and application in characterization of biorthonormal filter banks," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. III, pp. 177–180, April 1994.

[5] G. Smart and A. B. Bradley, "Filter bank design based on time domain aliasing cancellation with non-identical windows," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. III, pp. 181–184, April 1994.

[6] N. Jayant, J. Johnston, and R. Safranek, "Signal compression based on models of human perception," *Proceedings of the IEEE*, vol. 81, pp. 1385–1422, October 1993.

[7] M. R. Schroeder, B. S. Atal, and J. L. Hall, "Optimizing digital speech coders by exploiting masking properties of the human ear," *Journal of the Acoustical Society of America*, vol. 66, pp. 1647–1652, December 1979.