A MODULATED COMPLEX LAPPED TRANSFORM AND ITS APPLICATIONS TO AUDIO PROCESSING

Henrique Malvar

Microsoft Research
One Microsoft Way
Redmond, Washington 98052, USA

ABSTRACT

This paper introduces a new structure for a modulated complex lapped transform (MCLT), which is a complex extension of the modulated lapped transform (MLT). The MCLT is a particular kind of a 2x oversampled generalized DFT filter bank, whose real part corresponds to the MLT. That property can be used for efficient implementation of joint echo cancellation, noise reduction, and coding, for example. Fast algorithms for the MCLT are presented, as well as examples that show the good performance of the MCLT in noise reduction and echo cancellation.

1. INTRODUCTION

The modulated lapped transform (MLT) is an efficient tool for localized frequency decomposition of signals and transform/subband signal processing [1]. It is a particular form of a cosine-modulated filter bank [2] that allows for perfect reconstruction, has no blocking artifacts, and has almost optimal performance for transform coding of a wide variety of signals. Because of those properties, the MLT is being used in most modern audio coding systems, such as Dolby AC-3, MPEG-2 Layer III, and others [3].

One disadvantage of the MLT for some applications is that its transform coefficients are real, and so they do not explicitly carry phase information. In audio processing, two applications that need complex subbands are noise reduction via spectral subtraction [4], and acoustic echo cancellation [5].

In [6] a complex lapped transform (CLT) was introduced, with the purpose of using its phase information for motion estimation in video coding. The CLT was developed in [6] as a complex extension of the lapped orthogonal transform (LOT) [1]. The LOT of a signal segment can be approximately computed from the real and imaginary parts of the CLT.

We present here a new construction for a complex lapped transform, the modulated complex lapped transform (MCLT), as a very simple extension to the MLT. The MCLT has many interesting properties, which we review in Section 2. In Section 3 we present two alternatives for fast MCLT computation, and in Section 4 we present experimental results of the application of the MCLT to noise reduction and echo cancellation.

2. THE MODULATED COMPLEX LAPPED TRANSFORM

The MLT is based on the oddly-stacked time-domain aliasing cancellation (TDAC) filter bank introduced by Princen, Johnson, and Bradley [7]. Its basis functions can be obtained by cosine modulation of smooth windows, in the form [1], [8]

$$
\begin{aligned}
p_a(n,k) &\equiv h_a(n)\sqrt{\frac{2}{M}}\cos\left[\left(n+\frac{M+1}{2}\right)\left(k+\frac{1}{2}\right)\frac{\pi}{M}\right] \\
p_s(n,k) &\equiv h_s(n)\sqrt{\frac{2}{M}}\cos\left[\left(n+\frac{M+1}{2}\right)\left(k+\frac{1}{2}\right)\frac{\pi}{M}\right]
\end{aligned}
\tag{1}
$$

where $p_a(n,k)$ and $p_s(n,k)$ are the basis functions for the direct (analysis) and inverse (synthesis) transforms, and $h_a(n)$ and $h_s(n)$ are the analysis and synthesis windows, respectively. The time index n varies from 0 to $2M-1$ and the frequency index k varies from 0 to $M-1$, where M is the block size. The MLT is the TDAC for which the windows generate a lapped transform with maximum DC concentration, that is [1]

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

The direct transform matrix $\mathbf{P}_a$ is the one whose entry in the n-th row and k-th column is $p_a(n,k)$. Similarly, the inverse transform matrix $\mathbf{P}_s$ is the one with entries $p_s(n,k)$. For a block $\mathbf{x}$ of $2M$ input samples of a signal $x(n)$, its corresponding vector $\mathbf{X}$ of transform coefficients is computed by $\mathbf{X} = \mathbf{P}_a^T\mathbf{x}$. For a vector $\mathbf{Y}$ of processed transform coefficients, the reconstructed $2M$-sample vector $\mathbf{y}$ is given by $\mathbf{y} = \mathbf{P}_s\mathbf{Y}$. Reconstructed $\mathbf{y}$ vectors are superimposed with M-sample overlap [1], generating the reconstructed signal $y(n)$.

Assuming symmetrical analysis and synthesis windows, i.e. $h_a(n) = h_a(2M-1-n)$ and $h_s(n) = h_s(2M-1-n)$, it is easy to verify that perfect reconstruction is obtained with

$$
h_a(n) = \frac{h_s(n)}{h_s^2(n) + h_s^2(M-1-n)} \tag{3}
$$

It is interesting to considering the product window $h_p(n) = h_a(n)h_s(n)$. From (3), it follows that

$$
h_p(n) + h_p(n+M) = h_p(n) + h_p(M-1-n) = 1 \tag{4}
$$

With either the MLT window in (2) or the biorthogonal windows in [8], the product window satisfies

$$
h_p(n) = \sin^2\left[\left(n+\frac{1}{2}\right)\frac{\pi}{2M}\right] = \frac{1}{2} - \frac{1}{2}\cos\left[\left(n+\frac{1}{2}\right)\frac{\pi}{M}\right] \tag{5}
$$

We now introduce the MCLT. Its basis functions are defined by cosine and sine modulation of the analysis and synthesis windows, in the form:

$$
\begin{aligned}
p_a(n,k) &= p_a^c(n,k) - j\, p_a^s(n,k) \\
p_a^c(n,k) &= h_a(n)\sqrt{\frac{2}{M}}\cos\left[\left(n+\frac{M+1}{2}\right)\left(k+\frac{1}{2}\right)\frac{\pi}{M}\right] \\
p_a^s(n,k) &= h_a(n)\sqrt{\frac{2}{M}}\sin\left[\left(n+\frac{M+1}{2}\right)\left(k+\frac{1}{2}\right)\frac{\pi}{M}\right]
\end{aligned}
\tag{6}
$$

with $j \equiv \sqrt{-1}$, and

$$
\begin{aligned}
p_s(n,k) &= \frac{1}{2}\left[p_s^c(n,k) + j\, p_s^s(n,k)\right] \\
p_s^c(n,k) &= h_s(n)\sqrt{\frac{2}{M}}\cos\left[\left(n+\frac{M+1}{2}\right)\left(k+\frac{1}{2}\right)\frac{\pi}{M}\right] \\
p_s^s(n,k) &= h_s(n)\sqrt{\frac{2}{M}}\sin\left[\left(n+\frac{M+1}{2}\right)\left(k+\frac{1}{2}\right)\frac{\pi}{M}\right]
\end{aligned}
\tag{7}
$$

The MCLT transform coefficients $X(k)$ are computed from the input signal block $x(n)$ by $\mathbf{X} = \mathbf{P}_a^T \mathbf{x}$, or

$$
X(k) = \sum_{n=0}^{2M-1} x(n)\, p_a(n,k) \tag{8}
$$

Comparing (1) and (6), it is clear that the MLT of a signal is given by the real part of its MCLT.

There are two main interpretations of the MCLT construction above. First, we can view the additional sine-modulated functions as a 2x oversampling in the frequency domain, because for every new M real-valued input samples the MCLT computes M complex frequency components. We could also say that the MCLT functions above form an overcomplete basis. The second view is that the MCLT is in fact a 2x oversampled DFT filter bank (using a doubly-odd DFT [1] instead of the traditional DFT), in which the DFT length size is $2M$ and the frame (block) size is M. Note that, unlike in DFT filter banks, the lowest-frequency subband (the "DC" subband) is complex-valued.

With the MCLT, if we cascade the direct and inverse transform for a block, without modifying the transform coefficients, we obtain

$$
\mathbf{X} = \mathbf{P}_a^T \mathbf{x}, \quad \mathbf{Y} = \mathbf{X}, \quad \mathbf{y} = \mathbf{P}_s \mathbf{Y} \quad \Rightarrow \quad \mathbf{y} = \mathbf{P}_s \mathbf{P}_a^T \mathbf{x} \tag{9}
$$

with

$$
\mathbf{P}_s \mathbf{P}_a^T = \operatorname{diag}\left\{h_p(n)\right\} \tag{10}
$$

Thus, we note that the mapping from the input block $\mathbf{x}$ to the reconstructed block $\mathbf{y}$ is done via a diagonal matrix of order $2M$. This is quite different from the MLT, for which the product $\mathbf{P}_s \mathbf{P}_a^T$ is not diagonal. In fact, the off-diagonal terms of $\mathbf{P}_s \mathbf{P}_a^T$ for the MLT are the time-domain aliasing terms, which are cancelled when the overlapped blocks are superimposed. When the subband signals are processed such that $\mathbf{Y} \neq \mathbf{X}$, then the time-domain aliasing terms will not cancel exactly, producing artifacts. The MCLT, because of its 2x oversampling, does not rely on time-domain aliasing cancellation.

Another interesting property of the MCLT is that the reconstruction formula

$$
y(n) = \sum_{k=0}^{M-1} Y(k)\, p_s(n,k) \tag{11}
$$

is not unique. Perfect reconstruction (with $X(k) = Y(k)$, of course) can also be achieved with the choices

$$
y_c(n) = \sum_{k=0}^{M-1} \operatorname{Re}\left\{Y(k)\right\} p_s^c(n,k) \tag{12}
$$

or

$$
y_s(n) = \sum_{k=0}^{M-1} \operatorname{Im}\left\{Y(k)\right\} p_s^s(n,k) \tag{13}
$$

We recognize in (12) an inverse MLT. Although $y(n)$, $y_c(n)$, and $y_s(n)$ in (11)–(13) are not block-by-block identical, they build exactly the same reconstructed signal after overlapping.

The magnitude frequency responses of the MCLT filter bank are the same as those of the MLT [1], [8]. For each frequency $\omega_k = (k+1/2)\pi/M$ there are two subbands with the same magnitude frequency response but $\pi/2$ radians out of phase. There is significant overlap among the frequency responses of neighboring subbands, and the stopband attenuation is around –22 dB with the sine window in (2).

3. FAST COMPUTATION

As with the MLT, the MCLT can be computed via the type-IV discrete cosine transform (DCT-IV). For a signal $u(n)$, its length-M orthogonal DCT-IV is defined by [9]

$$
U(k) \equiv \sqrt{\frac{2}{M}} \sum_{n=0}^{M-1} u(n) \cos\left[\left(n+\frac{1}{2}\right)\left(k+\frac{1}{2}\right)\frac{\pi}{M}\right] \tag{14}
$$

We see that the frequencies of the cosine functions that form the DCT-IV basis are $(k+1/2)\pi/M$, the same as those of the MLT and MCLT. The type-IV discrete sine transform (DST-IV) of a signal $v(n)$ is defined by [9]

$$
V(k) \equiv \sqrt{\frac{2}{M}} \sum_{n=0}^{M-1} v(n) \sin\left[\left(n+\frac{1}{2}\right)\left(k+\frac{1}{2}\right)\frac{\pi}{M}\right] \tag{15}
$$

For a signal $x(n)$ with MCLT coefficients $X(k)$ determined by (8), it is easy to show that $\operatorname{Re}\{X(k)\} = U(k)$ and $\operatorname{Im}\{X(k)\} = V(k)$, if $u(n)$ in (14) is related to $x(n)$, for $n = 0,1,\ldots, M/2-1$, by [1]

$$
\begin{aligned}
u(n+M/2) &= \Delta_M\{x(M-1-n)h_a(M-1-n) - x(n)h_a(n)\} \\
u(M/2-1-n) &= x(M-1-n)h_a(n) + x(n)h_a(M-1-n)
\end{aligned}
$$

and $v(n)$ in (15) is related to $x(n)$ by

$$
\begin{aligned}
v(n+M/2) &= \Delta_M\{x(M-1-n)h_a(M-1-n) + x(n)h_a(n)\} \\
v(M/2-1-n) &= -x(M-1-n)h_a(n) + x(n)h_a(M-1-n)
\end{aligned}
$$

where $\Delta_M\{\cdot\}$ is the M-sample (one block) delay operator.

Thus, the MCLT can be computed from two window operators, a length-M DCT-IV, and a length-M DST-IV, as shown in the simplified flowgraph in Figure 1.

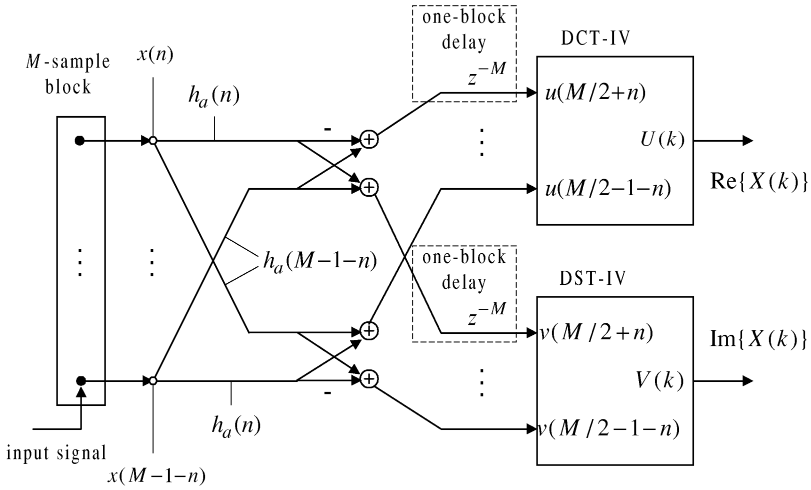

Figure 1. Flowgraph for the fast direct MCLT, $n = 0, 1, \ldots, M/2{-}1$, $k = 0, 1, \ldots, M/2{-}1$. The DCT-IV and DST-IV can be implemented with the fast algorithms in [1]. The inverse MCLT can be computed by simply transposing the flowgraph, moving the delays to the bottom half ouputs of the DCT-IV and DST-IV [1], replacing the coefficients $h_a(n)$ by $h_s(n)$, and multiplying the contents of the final buffer by 1/2.

The fast MCLT flowgraph in Figure 1 does not assume identical analysis and synthesis windows. Therefore, it can be used to compute a biorthogonal MCLT, as long as the windows satisfy the perfect reconstruction condition in (3). Good biorthogonal designs are presented in [8].

4. APPLICATIONS TO AUDIO PROCESSING

In most applications, the MCLT can be used in place of a DFT filter bank. Thus, the MCLT is well suited for noise suppression and echo cancellation.

We have implemented a noise reducer using the MCLT with spectral subtraction [4]. The original 8-second speech signal was captured at 16 kHz sampling rate, with the microphone near a very noisy personal computer (PC), whose noise spectrum is approximately pink. We adjusted the depth of subtraction for a noise reduction of about 15 dB. The results are shown in Figure 2, where the signal-to-noise ratio (SNR) was successfully increased from 15 dB to 30 dB. More importantly, the processed file has very fewer artifacts than the results obtained using a commercial product that uses standard DFT filter banks for spectral subtraction [10].

Another practical application of filter banks is in acoustic echo cancellation (AEC) for real-time teleconferencing. AEC uses an adaptive filter that estimates the feedback transfer function from the loudspeaker to the microphone. The estimated echo return is then subtracted from the microphone signal [5]. Simple FIR filters are not recommended, because of the length of the impulse response necessary to obtain a reasonable amount of echo reduction (for a 16 kHz sampling rate and an echo window of 100 ms, a 1,600-point impulse response is needed). With subband adaptive filtering, the long FIR full-band filter is replaced by a collection of short FIR filters, one for each subband. A critically sampled filter bank such as the MLT can be used for adaptive filtering, but the uncancelled aliasing due to subband processing limits the amount of echo reduction to 10 dB or less [5]. Performance can be improved by using cross-filters among neighboring subbands [11], but the extra degrees of freedom in such adaptive cross-filters usually slows down convergence significantly.

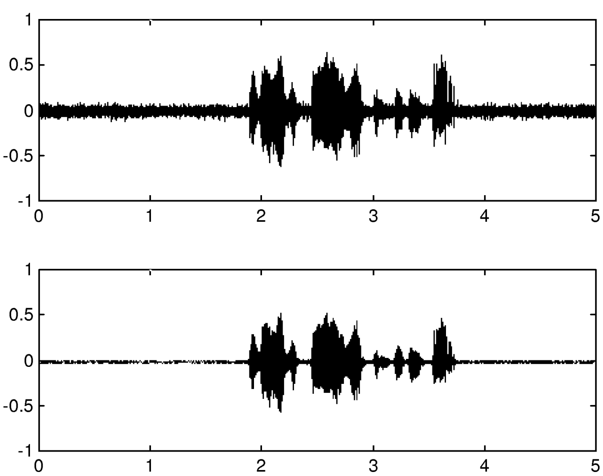

Figure 2. Noise reduction with the CMLT. Top: original speech, corrupted by PC noise, SNR ≈ 15 dB. Bottom: processed speech, SNR ≈ 30 dB.

With the MLCT we can perform subband AEC without cross-filters. We can process each subband by a short FIR filter with complex taps. With a large number of subbands, the subband signals are essentially white, and so each adaptive filter can be adjusted via the normalized LMS algorithm [5].

Results of an AEC simulation are shown in Figure 3. The original signal is an actual echo return recorded at 16 kHz sampling from a microphone located at about 20” from the loudspeaker (using a 4” driver). The signals in Figure 3 show the cancelled echo after convergence of each AEC (which takes a few seconds in all cases). The MLT and MCLT AECs used M = 512 subbands and a four-tap adaptive filter in each band (corresponding to an echo window of about 128 ms). It is clear that the MLT-based AEC does not work well without cross-filters, since its echo attenuation is only about 5 dB. With the MCLT, however, the echo attenuation is about 20 dB, which is adequate for many practical teleconferencing applications.

It is easy to combine AEC and spectral subtraction using a single MLCT decomposition. For example, we can apply spectral subtraction to the subband signals immediately after the AEC adaptive filters. If the resulting signal is to be encoded by an MLT-based codec such as in [12], then the MLT coefficients for the audio codec can be obtained by simply taking the real part of the outputs of the spectral subtraction. Therefore, only a single transformation step with the MCLT is necessary to perform simultaneous signal enhancement and coding.

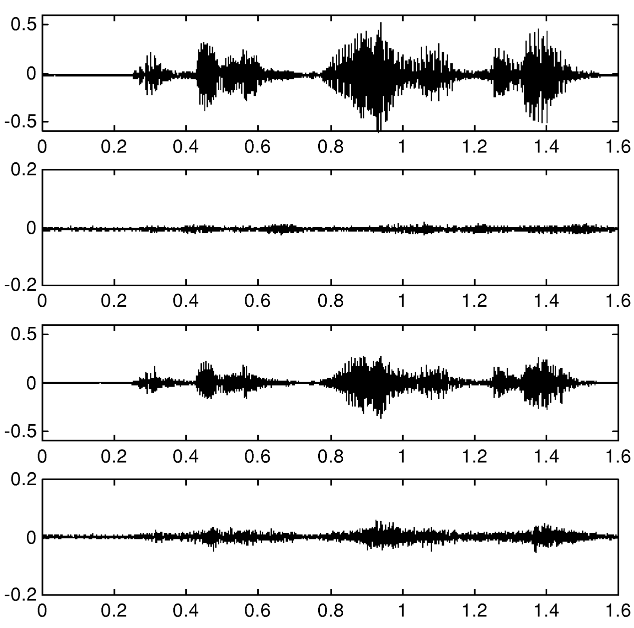

Figure 3.Acoustic echo cancellation experiments. Top: recorded echo return. Second trace: output of a full-band AEC, echo reduction ratio (ERR) ≈ 26 dB (note the different vertical scale). Third trace: output of a 512-band MLT AEC without cross filters, ERR ≈ 5 dB. Bottom: output of a 512-band MCLT AEC without cross filters, ERR ≈ 20 dB.

5. CONCLUSION

We have introduced a new filter bank structure with complex-valued subbands, which we referred to as the MCLT – modulated complex lapped transform. The MLCT is a simple extension of the MLT, obtained by using both cosine and sine modulations. That generates a subband structure that is 2x oversampled in the frequency domain, because for every new M real-valued input samples the MCLT computes M complex-valued frequency components. Using the DCT-IV and DST-IV, we presented fast algorithms for the MCLT. They show that the MCLT has about twice the complexity of the MLT.

The MCLT performs well in applications that usually require a complex filter bank, such as noise reduction and acoustic echo cancellation. We have shown that the MLCT can be used to efficiently implement those operations with good performance in practice. Furthermore, if the MLT of the enhanced signal is necessary for coding, it can be obtained simply by taking the real part of the processed MLCT coefficients.

REFERENCES

[1] H. S. Malvar, *Signal Processing with Lapped Transforms.* Norwood, MA: Artech House, 1992.

[2] P. P. Vaidyanathan, *Multirate Systems and Filter Banks.* Englewood Cliffs, NJ: Prentice Hall, 1993.

[3] S. Shlien, “The modulated lapped transform, its time-varying forms, and applications to audio coding,” *IEEE Trans. Speech Audio Processing,* vol. 5, pp. 359–366, July 1997.

[4] S. Godsill, P. Rayner, and O. Cappé, “Digital audio restoration,” in *Applications of Digital Signal Processing to Audio and Acoustics,* M. Kahrs and K. Brandenburg, Eds. Boston, MA: Kluwer, 1998.

[5] P. L. De León and D. M. Etter, “Acoustic echo cancellation using subband adaptive filtering,” in *Subband and Wavelet Transforms,* A. N. Akansu and M. J. T. Smith, Eds. Boston, MA: Kluwer, 1996.

[6] R. W. Young and N. Kingsbury, “Frequency domain motion estimation using a complex lapped transform,” *IEEE Trans. Image Processing,* vol. 2, pp. 2–17, Jan. 1993.

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

[8] H. S. Malvar, “Biorthogonal and nonuniform lapped transforms for transform coding with reduced blocking and ringing artifacts,” *IEEE Trans. Signal Processing,* vol. 46, pp. 1043–1053, Apr. 1998.

[9] K. R. Rao and P. Yip, *Discrete Cosine Transform: Algorithms, Advantages, and Applications.* New York: Academic Press, 1990.

[10] See http://www.syntrillium.com/cooledit/.

[11] A. Gilloire and M. Vetterli, “Adaptive filtering in subbands with critical sampling: analysis, experiments, and applications to acoustic echo cancellation,” *IEEE Trans. Signal Processing,* vol. 40, pp. 1862–1875, Aug. 1992.

[12] H. S. Malvar, “Enhancing the performance of subband audio coders for speech signals,” *Proc. IEEE ISCAS,* Monterey, CA, June 1988.