Some Peculiar Properties of the MDCT

Ye Wang[1], Leonid Yaroslavsky[2], Miikka Vilermo[1] and Mauri Väänänen[1]

***Abstract*—Having established the interconnection between MDCT, SDFT (Shifted Discrete Fourier Transform) and DFT recently, we have successfully applied the results in audio encoder design. This paper presents some new observations and analyses: 1) MDCT is not an orthogonal transform and this has an impact on audio coding, 2) analysis of Time Domain Alias Cancellation (TDAC) during the window switching in the case of MPEG-2 AAC. Finally, we report some experimental results on the energy compaction properties of MDCT.**

I. INTRODUCTION

Signal representation in the Modified Discrete Cosine Transform (MDCT) domain has emerged as a dominant tool in high quality audio coding because it combines critical sampling, reduction of block effect and flexible window switching. However, its mismatch with the Fourier transform domain based psychoacoustic model led us to study the characteristics of MDCT in the time and frequency domains. We have established the interconnection between MDCT, SDFT and DFT and applied the results in audio encoder design.

The purpose of this paper is to provide some new results on MDCT characteristics in the time and frequency domain and their impact on audio coding performance. We show that MDCT is not an orthogonal transform and does not fulfill Parseval's theorem, in contrast with orthogonal transforms. In general, performing MDCT and then IMDCT with one single frame of time domain samples, the original time samples cannot be perfectly reconstructed, instead the reconstructed samples are normally an alias-embeded version [1]. MDCT itself is a lossy process. This paper gives the conditions in which MDCT coefficients become zero with non-zero time domain samples, and the conditions in which original time domain samples can be perfectly reconstructed by performing direct and inverse MDCT even without overlap-add (OA) procedure. Using the relationship between MDCT and SDFT, the TDAC concept is illustrated during window switching process. Finally, we have examined the energy compaction properties of DFT, SDFT, DCT and MDCT experimentally with real life music samples.

The corresponding author is with Nokia Research Center, Visiokatu 1, FIN-33720 Tampere, Finland (e-mail: ye.wang@nokia.com).

[1]Nokia Research Center, P.O.Box 100, FIN-33721 Tampere, Finland

[2]Department of Interdisciplinary Studies, Tel Aviv University, Ramat Aviv 69978, Israel

II. PRELIMINARIES

The direct and inverse MDCT are defined as [2][3]:

$$\alpha_r = \sum_{k=0}^{2N-1} \tilde{a}_k \cos\left[\pi \frac{(k+(N+1)/2)(r+1/2)}{N}\right], \quad (1)$$

$r = 0,..., N-1$

$$\hat{a}_k = \frac{1}{N}\sum_{r=0}^{N-1} \alpha_r \cos\left[\pi \frac{(k+(N+1)/2)(r+1/2)}{N}\right], \quad (2)$$

$k = 0,...,2N-1$

where $\tilde{a}_k = h_k a_k$ is the windowed input signal, a_k is the input signal of $2N$ samples. h_k is a window function. We assume an identical analysis-synthesis time window. The constraints of perfect reconstruction are [4][5]:

$$h_k = h_{2N-1-k} \quad (3)$$

$$h_k^2 + h_{k+N}^2 = 1 \quad (4)$$

A sine window is widely used in audio coding because it offers good stop-band attenuation, provides good attenuation of the block edge effect and allows perfect reconstruction. Other optimized windows can also be applied [4]. The sine window is defined as:

$$h_k = \sin[\pi (k+1/2)/2N], \quad (5)$$

$\hat{a}_k$ in (2) contain time domain aliasing [1]:

$$\hat{a}_k = \begin{cases} \frac{1}{2}\tilde{a}_k - \frac{1}{2}\tilde{a}_{N-1-k}, & k = 0,..., N-1 \\ \frac{1}{2}\tilde{a}_k + \frac{1}{2}\tilde{a}_{3N-1-k}, & k = N,...,2N-1 \end{cases} \quad (6)$$

The relationship between MDCT and DFT can be established via Shifted Discrete Fourier Transforms (SDFT, [6]). The direct and inverse SDFTs are defined as:

$$\alpha_r^{u,v} = \sum_{k=0}^{2N-1} a_k \exp[i2\pi(k+u)(r+v)/2N], \quad (7)$$

$$a_k^{u,v} = 1/(2N)\sum_{r=0}^{2N-1} \alpha_r^{u,v} \exp[-i2\pi(k+u)(r+v)/2N], \quad (8)$$

where u and v represent arbitrary time and frequency domain shifts respectively. SDFT is a generalization of DFT allowing arbitrary shifts in the position of the samples in the time and frequency domain with respect to the signal and its spectrum coordinate system.

We have proven that the MDCT is equivalent to an SDFT of a modified input signal in (6):

$$\alpha_r = \sum_{k=0}^{2N-1} \hat{a}_k \exp\left[i2\pi \frac{(k+(N+1)/2)(r+1/2)}{2N}\right] \quad (9)$$

The right side of (9) is $SDFT_{(N+1)/2,1/2} = \left(\alpha_r^{(N+1)/2,1/2}\right)$ of the signal $\hat{a}_k$ formed from the initial windowed signal $\tilde{a}_k$ according to (6). The physical interpretation of (6) and (9) is straightforward. MDCT coefficients can be obtained by adding the $SDFT_{(N+1)/2,1/2}$ coefficients of the initial windowed signal and the alias.

III. PROPERTIES OF MDCT

MDCT differs somewhat from orthogonal transforms used for signal coding. The main peculiar properties of MDCT are:

- MDCT is not an orthogonal transform. Perfect signal reconstruction can be achieved in the overlap-add (OA) process. For the overlap-add window of 2N samples, first N and last N samples of the signal will remain modified according to (6). One can easily see this from the fact that performing MDCT and IMDCT of an arbitrary signal $\tilde{a}_k$ reconstructs signal $\hat{a}_k$ defined in (6).
- If a signal exhibits local symmetry such that
$$\begin{cases} \tilde{a}_k = \tilde{a}_{N-k-1}, & k = 0,...,N-1 \\ \tilde{a}_k = -\tilde{a}_{3N-k-1}, & k = N,...,2N-1 \end{cases}, \qquad (10)$$
its MDCT degenerates to zero: $\alpha_r = 0$ for $r = 0,...,N-1$. This property follows from (6). This is a good example that MDCT does not fulfill Parseval's theorem, i.e. the time domain energy is not equal to the frequency domain energy (see Figure 1).
- If a signal exhibits local symmetry such that
$$\begin{cases} \tilde{a}_k = -\tilde{a}_{N-k-1}, & k = 0,...,N-1 \\ \tilde{a}_k = \tilde{a}_{3N-k-1}, & k = N,...,2N-1 \end{cases}, \qquad (11)$$
MDCT and IMDCT will perfectly reconstruct the original time domain samples. This property also follows from (6).
- Nevertheless, on average, MDCT, similar to such orthogonal transforms as DFT, DCT, DST, etc, possesses energy compaction capability and acceptable Fourier spectrum analysis.

Prove: $MDCT(\tilde{a}_k) = 0 \Leftrightarrow \tilde{a}_k$ fulfills (10).

Proof:

"$\Rightarrow$"

If $\tilde{a}_k$ fulfills (10) $\Rightarrow$ the aliased signal

$\hat{a}_k = 0 \Rightarrow MDCT(\tilde{a}_k) = 0$

"$\Leftarrow$"

We prove the negation of the claim: If $\tilde{a}_k$ does not fulfill (10) $\Rightarrow$ $MDCT(\tilde{a}_k) \neq 0$ for some $k = 0,...,2N$.

If $\tilde{a}_k$ does not fulfill (10) $\Rightarrow$ the aliased signal

$\hat{a}_k \neq 0$ for some $k = 0,...,2N$ $\Rightarrow$ $SDFT_{\frac{N+1}{2},\frac{1}{2}}(\hat{a}_k) \neq 0$,

because SDFT fulfills Parseval's theorem $\Rightarrow$

$MDCT(\tilde{a}_k) = SDFT_{\frac{N+1}{2},\frac{1}{2}}(\hat{a}_k) \neq 0$ [1], which concludes the proof.

Prove:

$IMDCT(MDCT(\tilde{a}_k)) = \tilde{a}_k \Leftrightarrow \tilde{a}_k$ fulfills (11).

Proof:

$MDCT(\tilde{a}_k) = SDFT_{\frac{N+1}{2},\frac{1}{2}}(\hat{a}_k)$,

$\Rightarrow ISDFT_{\frac{N+1}{2},\frac{1}{2}}(MDCT(\tilde{a}_k)) = \hat{a}_k$

$\Rightarrow IMDCT(MDCT(\tilde{a}_k)) = \hat{a}_k$, because

$IMDCT = ISDFT_{\frac{N+1}{2},\frac{1}{2}}$ [7]

$\Rightarrow \tilde{a}_k = \hat{a}_k$, i.e. perfect reconstruction can be achieved without overlap-add procedure.

q.e.d.

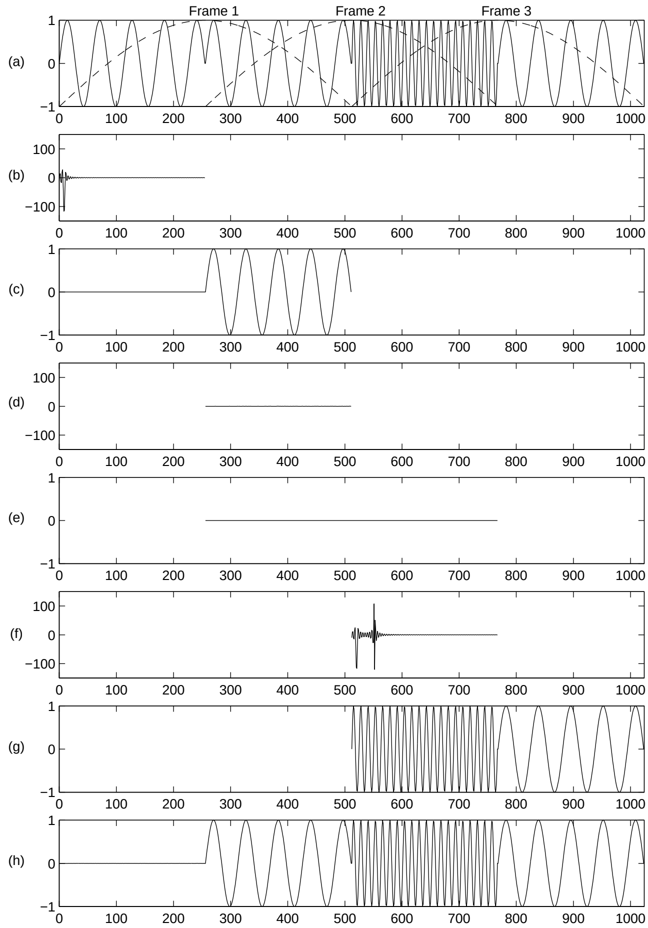

Figure 1. Illustration of signal analysis/synthesis with MDCT, overlap-add procedure and perfect reconstruction of time domain samples. (a) a phase/frequency-modulated time signal; (b)(d)(f) MDCT spectra in different time slots, indicated as frames 1, 2, 3 in (a); (c)(e)(g) reconstructed time domain samples (with IMDCT) of frames 1, 2, 3 respectively; (h) the reconstructed time samples after the overlap-add procedure.

In order to illustrate the special characteristics of the MDCT and their impact on audio coding in an intuitive way, we have designed a phase/frequency-modulated time signal in Figure 1 (a), which has two different frequency elements with the duration of half of the frame size (= 512 samples). Dashed lines in Figure 1 (a) illustrate the 50% window overlap. However, MDCT spectra of different time slots in Figures 1 (b)(d)(f) are calculated with rectangular windows for simplicity. The IMDCT time domain samples of frame 1, 2, 3 are shown in Figures 1 (c)(e)(g) respectively. The reconstructed time domain samples after overlap-add (OA) procedure is shown in Figure 1 (h). With frame 2 the condition in (10) holds, and the MDCT coefficients are all zero! Nevertheless, the time domain samples in frame 2 can still be perfectly reconstructed after the overlap-add procedure. With frame 3 the condition (11) holds, and the original time samples are perfectly reconstructed even without overlap-add procedure. These are, of course, very special occurrences, which are rare in real life audio signals. If the signal is close to the condition in (10), however, MDCT spectrum will be very unstable in comparison with DFT spectrum. In this case, using the output of the DFT based psychoacoustic model to quantise MDCT coefficients will not be logical. This is an important limitation of MDCT.

Figure 2 shows the fluctuation of MDCT spectrum in comparison with DFT and $SDFT_{(N+1)/2,1/2}$ spectra. With a frequency-modulated time signal in Figure 2 (a), the DFT power spectrum is very stable despite a moving window. Conversely, the MDCT spectrum is very unstable. The $SDFT_{(N+1)/2,1/2}$ spectrum is in between. This is at least one evidence that the $SDFT_{(N+1)/2,1/2}$ can be used as a bridge to connect MDCT and DFT in audio coding applications.

In order to illustrate TDAC concept during the window switching in MPEG-2 AAC, we define two overlapping windows with window functions h_k and g_k. The conditions for perfect reconstruction are [8]:

$$h_{N+k} \cdot h_{2N-1-k} = g_k \cdot g_{N-1-k} \tag{12}$$

$$h_{N+k}^2 + g_k^2 = 1 \tag{13}$$

Using (6) one can easily see one of the important properties of MDCT: the time domain alias in each half of the window is independent, which allows adaptive window switching [8]. The TDAC concept during window switching in MPEG-2 AAC is illustrated in Figure 3.

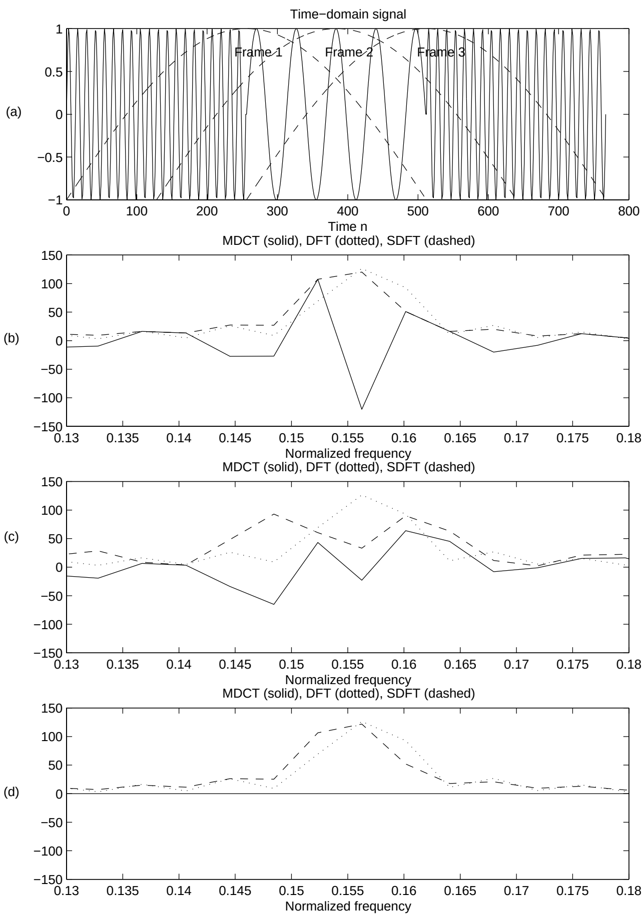

Figure 2 Comparison of DFT, $SDFT_{(N+1)/2,1/2}$ and MDCT spectra in different time slots. (a) a frequency-modulated time signal (solid line) with a moving window, (b)(c)(d) DFT (dotted lines), $SDFT_{(N+1)/2,1/2}$ (dashed lines) and MDCT (solid lines) spectra of Frames 1, 2, 3.

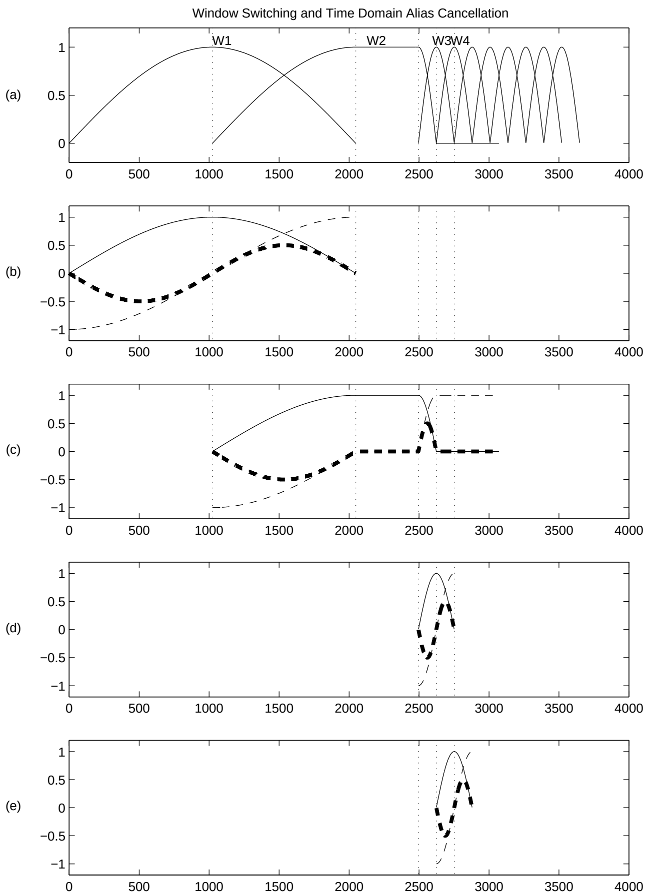

Figure 3. TDAC in the case of window switching. (a) three types of window shape in MPEG-2 AAC indicated with W1, W2, W3. (b) time domain alias in the long window, thick dashed line indicates the time domain alias after weighting with window function, (c) time domain alias in a transition window, (d)(e) time domain alias in short windows.

IV. EXPERIMENTAL RESULTS

Experiments were performed to compare the cumulative spectra of DFT, DCT, $SDFT_{(N+1)/2,1/2}$ and MDCT with a large number of test samples. We have gained insights into the energy compaction properties of different transforms experimentally. With these experiments we have observed the following:

1) 90% energy is concentrated within 10% of the normalized frequency scale for most of the test signals for all transforms concerned. The energy compaction property of different transforms becomes more unified with increasing window length.

2) Window shape has an impact on MDCT energy compaction property. Tests were performed with rectangular and Hanning windows. In the case of a rectangular window, DCT has always the best energy compaction property (see Figure 4), because DCT corresponds to an even extension of the signal [6].

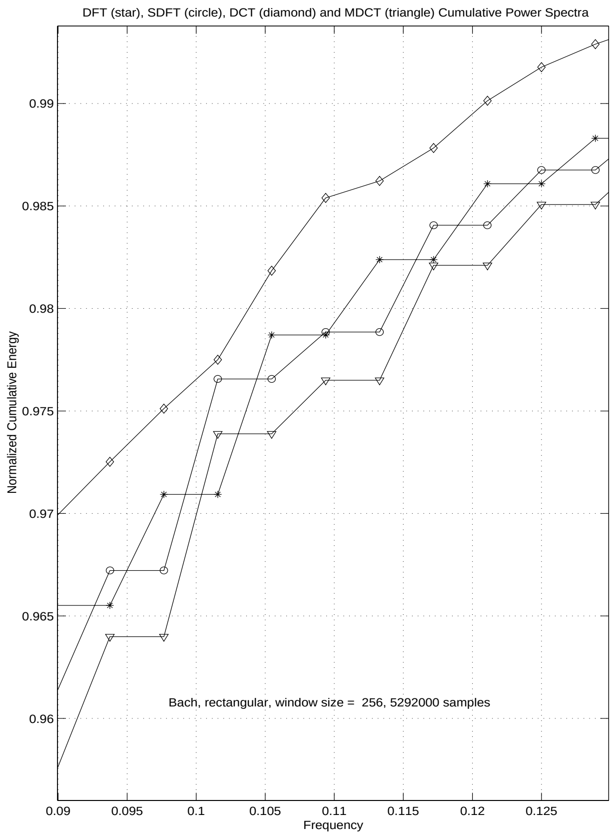

Figure 4. Cumulative power spectra of DCT, $SDFT_{(N+1)/2,1/2}$, DFT and MDCT with a fraction of classical music with rectangular windows. The window size is 256. The length of the test sequence is 5292000 PCM samples.

V. CONCLUSION AND FUTURE WORK

In this paper we have shown that MDCT exhibits some peculiar properties which distinguish it from orthogonal transforms. We then examined its energy compaction properties experimentally with real-life music samples.

After analyzing MDCT both theoretically and experimentally, we can conclude that MDCT is a very useful concept with its Time Domain Alias Cancellation (TDAC) characteristics. However, its special features described in this paper and its mismatch with the DFT domain based psychoacoustic model must be kept in mind when developing a MDCT based audio codec with its full potential in terms of coding performance.

In terms of energy compaction property, MDCT does not have any advantage in comparison to DFT and DCT as indicated in Figure 4. Apparently, the distinct advantage of MDCT lies in its critical sampling property, reduction of block effect and the possibility of adaptive window switching.

We also believe that the disappointing performance of wavelet based audio codecs may be caused by the mismatch between the two fundamental tools of audio coding -- audio signal representation and the auditory system perceptual model. Therefore, the interconnection between discrete wavelet and Fourier transform will be our next focus in the hope of making some progress; possibly even a breakthrough in wavelet domain based audio coding algorithms.

VI. ACKNOWLEDGEMENT

Ye Wang wishes to thank Mr. James Johnston (AT&T Research Laboratories) and Prof. Anibal Ferreira (University of Porto) for very inspiring discussions during the 17th AES International Conference on High Quality Audio Coding in Florence, Italy, September 2nd – 5th, 1999. Those discussions inspired us to complete this paper.

VII. REFERENCES

[1] Wang, Y., Yaroslavsky, L., Vilermo, M., Väänänen, M. “Restructured Audio Encoder for Improved Computational Efficiency”, AES 108th International Convention, February 19-22, 2000, Paris, France.

[2] Princen, J. P., Bradley, A. B., “Analysis/Synthesis Filter Bank Design Based on Time Domain Aliasing Cancellation”, IEEE Transactions on Acoustics, Speech, and Signal Processing, Vol. ASSP-34, No. 5, October 1986.

[3] Princen, J. P., Johnson, A. W., Bradley, A. B., “Subband/Transform Coding Using Filter Bank Designs Based on Time Domain Aliasing Cancellation”, IEEE International Conference on Acoustics, Speech, and Signal Processing, 1987, Dallas, USA, pp. 2161-2164.

[4] Ferreira, A., “Spectral Coding and Post-Processing of High Quality Audio”, Ph.D. thesis http://telecom.inescn.pt/doc/phd_en.html.

[5] Malvar, H., “Signal Processing with Lapped Transforms”, Artech House, Inc., 1992.

[6] Yaroslavsky, L., Eden, M., “Fundamentals of Digital Optics”, Birkhauser, Boston, 1996.

[7] Wang, Y., Yaroslavsky, L., Vilermo, M., “On The Relationship Between MDCT, SDFT and DFT”, submitted to the 5th International Conference on Signal Processing (ICSP2000), August 21-25, 2000, Beijing, China.

[8] Edler, B., “Coding of Audio Signals with Overlapping Block Transform and Adaptive Window Functions”, (in German), *Frequenz*, vol.43, pp.252-256, 1989.