ON THE RELATIONSHIP BETWEEN MDCT, SDFT AND DFT

Ye Wang[1], Leonid Yaroslavsky[2], Miikka Vilermo[1]

[1]Nokia Research Center, Speech and Audio Systems Lab, FIN-33720 Tampere, Finland

Email: ye.wang@nokia.com, miikka.vilermo@nokia.com

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

Email: yaro@eng.tau.ac.il

ABSTRACT

Modified Discrete Cosine Transform (MDCT) has emerged as a dominant time-frequency decomposition method in high quality audio compression. The MDCT is a special case of the Lapped Transforms (LTs) with 50% overlap. This paper establishes the relationship between the MDCT and Shifted Discrete Fourier Transform (SDFT). The analysis provides insight into the following issues: (1) the relationship between MDCT, Shifted DFT (SDFT) and DFT, (2) characteristics of the MDCT in the time and frequency domain, (3) the concept of Time Domain Aliasing Cancellation (TDAC).

1. INTRODUCTION

High quality audio compression has been widely used in various applications, such as VCD, DVD, DAB, HDTV, and Internet music distribution, etc. High quality audio will play an increasingly important role in future wireless communication systems, such as WAP and BLUETOOTH. The objective in these applications is to achieve a low bit rate in the digital representation of an input signal with minimum perceived loss of signal quality. In order to achieve this objective, three basic coding tools are used: time-frequency decomposition (or transform), a psychoacoustic model that also assumes time-frequency decomposition, and a quantizer. The use of transforms in audio compression is determined by their energy compaction property.

Among many different transforms, the modified discrete cosine transform (MDCT) [1][2] has become dominant in practically all state-of-the-art audio codecs [3]. The MDCT method is a block transform method, in which the input signal is partitioned into smaller blocks. Each block is MDCT transformed and the transform coefficients are quantized prior to code assignment. The principal reason for using the MDCT rather than other known transform methods is that, in addition to the effective decorrelation/energy compaction similar to DCT, MDCT implies 50% time-domain window overlap, thus greatly reducing the block effects, while maintaining critical sampling. The time domain alias introduced by the MDCT and inverse MDCT is independent for each half of the window. This leads to the realization of adaptive window switching systems [4]. Perfect reconstruction (PR) of the signal in the overlapped region can be achieved by the overlap-add (OA) procedure.

In an MDCT-based audio encoder, the quantization of the MDCT coefficients is the only source of quality degradation. The objective of the quantization process is to reduce the bit rate without compromising the perceptual audio quality. Psychoacoustics and physiology have revealed that the human auditory system performs frequency analysis in the basilar membrane [5][6]. Its sensitivity is highly frequency dependent. In terms of simultaneous masking properties, psychoacoustic experiments have often been conducted in the Fourier transform domain [5][6]. Since MDCT coefficients in an audio encoder are quantized according to a human auditory model, which is based on signal Fourier analysis implemented by means of DFT [7], a better understanding of the natural frequency distribution in the MDCT domain is very important for designing efficient audio encoders. The purpose of our analysis in this paper is to introduce SDFT as a bridge between MDCT and DFT.

MDCT was initially introduced in terms of the filterbank ideology [1][8][9]. This approach, however, does not display clearly why the MDCT possesses an energy compaction property, and what its relationship with Fourier analysis is, which is fundamental in audio coding. We present our analysis as follows: Section 2 suggests a relationship between the MDCT, SDFT and the DFT in such a way as to allow us to gain insight into the frequency distribution properties in the MDCT domain. Furthermore, the section analyzes the symmetric properties of MDCT and illustrates the TDAC concept in a very intuitive way based on our theoretical analysis, Section 3 concludes the paper.

2. THE RELATIONSHIP BETWEEN MDCT, SDFT AND DFT

In order to investigate the frequency characteristics of the MDCT, we derive in this section the relationship between the MDCT, SDFT and DFT.

The MDCT of a signal sequence a_k of *2N* samples is defined as [1][2]:

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

where h_k is a window function. We assume an identical analysis-synthesis time window. The constraints of perfect reconstruction are [3][9]:

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

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

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 be applied as well [3]. The sine window is defined as:

$$h_k = \sin[\pi(k+1/2)/2N], \quad (4)$$

In the following we will prove that the MDCT is equivalent to a Shifted Discrete Fourier Transform (SDFT) [10][11] of a modified input signal. SDFT is a generalization of DFT that allows an arbitrary shift in position of the samples in the time and frequency domain with respect to the signal and its spectrum coordinate system.

The direct and inverse Shifted Fourier transforms are defined as [10][11]:

$$\alpha_r^{u,v} = \sum_{k=0}^{2N-1} a_k \exp[i2\pi(k+u)(r+v)/2N], \tag{5}$$

$$a_k^{u,v} = 1/(2N)\sum_{r=0}^{2N-1}\alpha_r^{u,v}\exp[-i2\pi(k+u)(r+v)/2N], \tag{6}$$

where u and v represent arbitrary time and frequency domain shifts respectively.

We will now prove that the MDCT of a windowed signal $\tilde{a}_k$ of 2N samples is a $SDFT_{(u,v)}$ of an alias-embedded signal $\hat{a}_k$ with $u = (N+1)/2, v = 1/2$.

We define

$$\beta(k,r,N) = \frac{(k+(N+1)/2)(r+1/2)}{N}. \tag{7}$$

Denote $\tilde{a}_k = h_k a_k$ as the windowed input signal. Then the MDCT coefficients are:

$$\alpha_r = \sum_{k=0}^{2N-1}\tilde{a}_k \cos[\pi\beta(k,r,N)], \tag{8}$$

Represent the cosine term via complex exponents and split the summation in (8) into four parts:

$$\alpha_r = \frac{1}{2}\sum_{k=0}^{2N-1}\tilde{a}_k\{\exp[i\pi\beta(k,r,N)] + \exp[-i\pi\beta(k,r,N)]\} =$$

$$\frac{1}{2}\sum_{k=0}^{N-1}\tilde{a}_k\exp[i\pi\beta(k,r,N)] + \frac{1}{2}\sum_{k=0}^{N-1}\tilde{a}_k\exp[-i\pi\beta(k,r,N)] +$$

$$\frac{1}{2}\sum_{k=N}^{2N-1}\tilde{a}_k\exp[i\pi\beta(k,r,N)] + \frac{1}{2}\sum_{k=N}^{2N-1}\tilde{a}_k\exp[-i\pi\beta(k,r,N)], \tag{9}$$

Replace the summation index k in the second and fourth terms of (9) with $N-1-k$ and $3N-1-k$ respectively. This results in:

$$\alpha_r = \frac{1}{2}\sum_{k=0}^{N-1}\tilde{a}_k\exp[i\pi\beta(k,r,N)] + \frac{1}{2}\sum_{k=0}^{N-1}\tilde{a}_{N-1-k}\exp[-i\pi(2-\beta(k,r,N))] +$$

$$\frac{1}{2}\sum_{k=N}^{2N-1}\tilde{a}_k\exp[i\pi\beta(k,r,N)] + \frac{1}{2}\sum_{k=N}^{2N-1}\tilde{a}_{3N-1-k}\exp[-i\pi(4-\beta(k,r,N))], \tag{10}$$

or, described in another way:

$$\alpha_r = \frac{1}{2}\sum_{k=0}^{N-1}\tilde{a}_k\exp[i\pi\beta(k,r,N)] +$$

$$\frac{1}{2}\sum_{k=0}^{N-1}\tilde{a}_{N-1-k}\exp[-i2\pi(r+1/2)]\exp[i\pi\beta(k,r,N)] + \frac{1}{2}\sum_{k=N}^{2N-1}\tilde{a}_k\exp[i\pi\beta(k,r,N)] +$$

$$\frac{1}{2}\sum_{k=N}^{2N-1}\tilde{a}_{3N-1-k}\exp[-i4\pi(r+1/2)]\exp[i\pi\beta(k,r,N)], \tag{11}$$

Eq. (11) can be further simplified by substitution,

$$\exp[-i2\pi(r+1/2)] = -1, \tag{12}$$

$$\exp[-i4\pi(r+1/2)] = 1 \tag{13}$$

Introduce time domain aliasing:

$$\hat{a}_k = \begin{cases}\tilde{a}_k/2 - \tilde{a}_{N-1-k}/2, & k = 0,\dots,N-1\\ \tilde{a}_k/2 + \tilde{a}_{3N-1-k}/2, & k = N,\dots,2N-1\end{cases} \tag{14}$$

and finally obtain:

$$\alpha_r = \sum_{k=0}^{2N-1}\hat{a}_k\exp\left[i2\pi\frac{(k+(N+1)/2)(r+1/2)}{2N}\right] \tag{15}$$

which is $SDFT_{((N+1)/2,1/2)}$ of the signal $\hat{a}_k$ formed from the initial windowed signal $\tilde{a}_k$ according to (14). For real-valued signals $SDFT_{(N+1)/2,1/2}$ has the following property [11]:

if

$$\hat{a}_k = -\hat{a}_{N-1-k}, k = 0,\dots,2N-1 \tag{16}$$

then

$$\alpha_k = (\alpha_k)^*, k = 0,\dots,N-1, \tag{17}$$

where * stands for the complex conjugate. Since $\hat{a}_k$ fulfills (14), (16) holds for $k = 0,\dots,N-1$. For $k = N,\dots,2N-1$ the generalized cyclicity property of the SDFT [11] can be employed:

$$\hat{a}_{k+2N} = -\hat{a}_k \tag{18}$$

Therefore (16) can be written as

$$\hat{a}_k = \hat{a}_{2N-1-k}, k = N,\dots,2N-1 \tag{19}$$

(19) is true when $\hat{a}_k$ fulfills (14).

This proves that the right side of (15) is real-valued if $\hat{a}_k$ fulfills (14), which is always the case for real-valued $\tilde{a}_k$.

Physical interpretation of (14) is straightforward. MDCT coefficients can be obtained by adding the SDFT coefficients of the initial windowed signal and the alias. In other words, we can rewrite (15) as:

$$MDCT(signal) = SDFT_{(N+1)/2,1/2}(signal) + SDFT_{(N+1)/2,1/2}(alias) \tag{20}$$

With reference to (14) and Figure 1(c), the alias is added to the original signal in such a way that the first half of the window 1 (the signal portion between points A and B) is mirrored in the time domain and then inverted before being subsequently added to the original signal. The second half of the window 1 (signal portion between points B and C) is also mirrored in the time domain and added to the original signal.

The $SDFT_{(N+1)/2,1/2}$ can be expressed by means of the conventional DFT as:

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

$$= \left\{\sum_{k=0}^{2N-1}\left[\hat{a}_k\exp\left(i2\pi\frac{k}{4N}\right)\right]\exp(i2\pi\frac{kr}{2N})\right\}\exp\left(i2\pi\frac{(N+1)r}{4N}\right)\exp\left(i\pi\frac{N+1}{4N}\right) \tag{21}$$

To the right side of (21), the first exponential function corresponds to a modulation of $\hat{a}_k$ that result in a signal spectrum shift in frequency domain by ½ of the frequency-sampling interval. The second exponential function corresponds to the conventional DFT. The third exponential function modulates the signal spectrum that is equivalent to a signal shift by (N+1)/2 of the sampling interval in the time domain [11]. Therefore, $SDFT_{(N+1)/2,1/2}$ is the conventional DFT of this signal shifted in time domain by (N+1)/2 of the sampling interval and evaluated with the shift of ½ of the frequency-sampling interval.

MDCT transform coefficients exhibit symmetric properties:

$$\alpha_{2N-r-1} = (-1)^{N+1}\alpha_r \tag{22}$$

To show this, replace in (1) r with $2N-r-1$ to obtain

$$\alpha_{2N-1-r} = \sum_{k=0}^{2N-1} \tilde{a}_k \cos[\pi\beta(k, 2N-r-1, N)] \tag{23}$$

Rearrange terms:

$$= \sum_{k=0}^{2N-1} \tilde{a}_k \cos[2\pi(k + (N+1)/2) - \pi\beta(k, r, N)] \tag{24}$$

For integers k and n,

$$\cos(2\pi k + \pi n + x) = (-1)^n \cos(x) \tag{25}$$

Therefore,

$$\alpha_{2N-1-r} = (-1)^{N+1} \sum_{k=0}^{2N-1} \tilde{a}_k \cos[\pi\beta(k, r, N)] = (-1)^{N+1}\alpha_r \tag{26}$$

Apparently, the MDCT coefficients are odd symmetric, only if N is even, which is often true in audio coding applications. However, they are even symmetric if N is odd. This new conclusion is more general in comparison with [12]. Using this newly derived property we can now easily derive the Inverse MDCT (IMDCT). From (6) it follows that

$$\hat{a}_k = \frac{1}{2N} \sum_{r=0}^{2N-1} \alpha_r \exp[-i2\pi\beta(k, r, N)/2] \tag{27}$$

Divide the summation into two parts:

$$\hat{a}_k = \frac{1}{2N} \sum_{r=0}^{N-1} \alpha_r \exp[-i\pi\beta(k, r, N)] + \frac{1}{2N} \sum_{r=N}^{2N-1} \alpha_r \exp[-i\pi\beta(k, r, N)] \tag{28}$$

Replace r with $2N-r-1$ and change the summation order in the second term:

$$\hat{a}_k = \frac{1}{2N} \sum_{r=0}^{N-1} \alpha_r \exp[-i\pi\beta(k, r, N)] +$$

$$\frac{1}{2N} \sum_{r=0}^{N-1} \alpha_{2N-r-1} \exp[-i\pi\beta(k, 2N-r-1, N)]\cdot \tag{29}$$

Rearrange terms in the last exponent:

$$\hat{a}_k = \frac{1}{2N} \sum_{r=0}^{N-1} \alpha_r \exp[-i\pi\beta(k, r, N)] +$$

$$\frac{1}{2N} \sum_{r=0}^{N-1} \alpha_{2N-r-1} \exp[-i\pi\beta(k, 1/2, N)2N] \exp[i\pi\beta(k, r, N)] \tag{30}$$

Because $\exp(i2\pi n + x) = \exp(x)$ for integer n, we get

$$\hat{a}_k = \frac{1}{2N} \sum_{r=0}^{N-1} \alpha_r \exp[-i\pi\beta(k, r, N)] +$$

$$\frac{1}{2N} \sum_{r=0}^{N-1} \alpha_{2N-r-1} \exp[-i\pi(N+1)] \exp[i\pi\beta(k, r, N)] \tag{31}$$

Because $\exp[-i\pi(N+1)] = (-1)^{N+1}$, and using the symmetry of α_r in (22) we get:

$$\hat{a}_k = \frac{1}{2N} \sum_{r=0}^{N-1} \alpha_r \exp[-i\pi\beta(k, r, N)] +$$

$$\frac{1}{2N} \sum_{r=0}^{N-1} \alpha_r (-1)^{N+1} (-1)^{N+1} \exp[i\pi\beta(k, r, N)] \tag{32}$$

Finally add the summations together:

$$\hat{a}_k = \frac{1}{N} \sum_{r=0}^{N-1} \alpha_r \cos\left[\pi \frac{(k + (N+1)/2)(r + 1/2)}{N}\right] \tag{33}$$

This proves that IMDCT is equivalent to the $ISDFT_{(N+1)/2, 1/2}$. From (33) we can see that, in comparison with conventional orthogonal transforms, MDCT has a special property: the input signal cannot be perfectly reconstructed from the MDCT coefficients even without quantization. MDCT itself is a lossy process (therefore not an orthogonal transform). That is, the imaginary coefficients of the $SDFT_{(N+1)/2, 1/2}$ are lost in the MDCT transform. However, the lost information can be recovered using the redundancy of the 50% overlap of neighboring frames to gain perfect reconstruction. Applying MDCT and IMDCT converts the input signal into one that contains certain twofold symmetric alias (see (14) and Figure 1(c)). The introduced alias will be cancelled in the overlap-add process (see Figure 1).

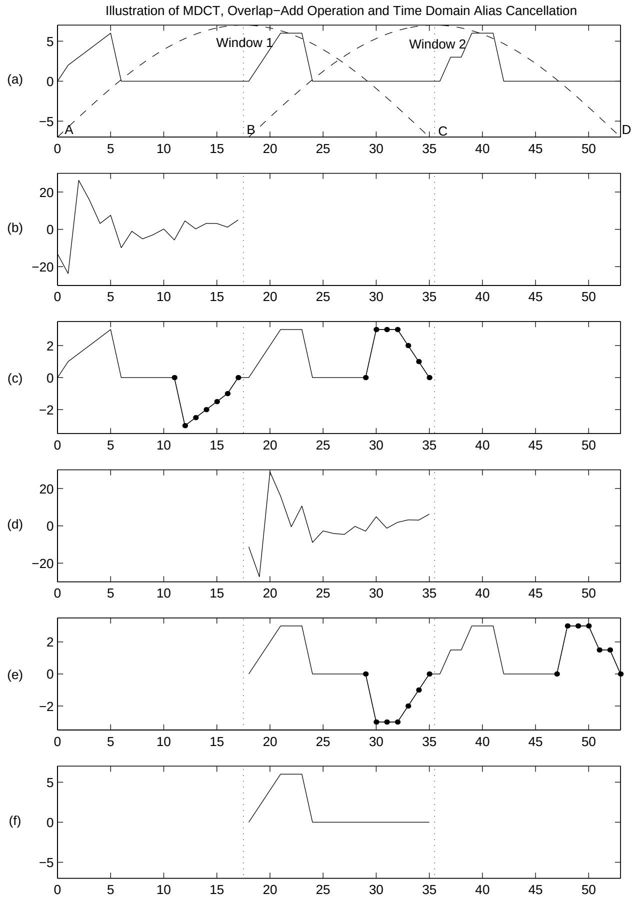

Figure 1. Illustration of the MDCT, overlap-add (OA) procedure and the concept of the Time Domain alias cancellation (TDAC). (a) An artificial time signal, dashed lines indicating the 50% overlapped windows; (b) MDCT coefficients of the signal in Window 1; (c) IMDCT coefficients of the signal in (b), the alias is shown by markers on the line; (d) The MDCT coefficients of the signal in Window 2; (e) IMDCT coefficients of the signal in (d), the alias is shown by markers on the line; (f) The reconstructed time domain signal after the overlap-add (OA) procedure. The original signal in the overlapped part (between points B and C) is perfectly reconstructed.

Based on our theoretical analysis, we have designed an artificial time domain signal to illustrate the Time Domain Aliasing Cancellation (TDAC) concept in a very intuitive way. The artificial signal of 54 samples is shown in Figure 1(a). The MDCT coefficients of the signal in Window 1 are shown in Figure 1(b). Obviously the coefficients are subsampled by 50% in MDCT (from 2N time domain samples to N independent frequency domain coefficients), and the alias is introduced as well. The IMDCT coefficients of the signal in Figure 1(b) are illustrated in Figure 1(c). This step introduces redundancy (from N frequency domain coefficients to 2N time domain samples). The MDCT coefficients of the signal in Window 2 are presented in Figure 1(d). The corresponding IMDCT time domain signal is shown in Figure 1(e). If the overlap-add procedure is performed with Figure 1(c) and (e), perfect reconstruction (PR) of the original signal in the overlapped part (between points B and C) can be achieved.

It is clear that one cannot achieve perfect reconstruction (PR) for the first half of the first window and the second half of the last window as indicated in Figure 1.

3. CONCLUSION AND FUTURE WORK

The analysis presented in this paper demonstrates the interconnection between MDCT, SDFT and DFT, and vividly explains the Time Domain Alias Cancellation (TDAC) concept of MDCT. Essentially MDCT is the DFT of a signal modified in a certain way.

The presented analysis provides explicit relationships between the DFT coefficients and the MDCT coefficients for the same input samples. Therefore, many existing useful results in the DFT domain can be mapped to the MDCT domain.

4. ACKNOWLEDGEMENT

The authors wish to thank Prof. Jaakko Astola and Dr. Karen Egiazarian (Tampere University of Technology), Dr. Deepa Kundur (University of Toronto), and Prof. Petri Haavisto and Dr. Jilei Tian (Nokia Research Center) for reading of the manuscript and for constructive suggestions.

5. REFERENCES

[1] 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.

[2] 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.

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

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

[5] Zwicker, E., Fastl, H., "Psychoacoustics, Facts and Models", Springer-Verlag, Berlin Heidelberg, Germany, 1990.

[6] Moore, B. C. J., "An Introduction to the Psychology of Hearing", 4. Edition, Academic Press, 1997.

[7] ISO/IEC JTC1/SC29/WG11, "Coding of moving pictures and audio - MPEG-2 Advanced Audio Coding AAC", ISO/IEC 13818-7 International Standard, 1997.

[8] Jayant, N., Johnston, J., Safranek, R., "Signal Compression Based on Models of Human Perception", Proceedings of the IEEE, vol. 81, no. 10, 1993.

[9] Malvar, H., "Signal Processing with Lapped Transforms", Artech House, Inc., 1992.

[10] Yaroslavsky, L., "Shifted Discrete Fourier Transforms", Digital Signal Processing, Ed. by V. Cappellini and A. G. Constantinides, Academic Press, London, 1980.

[11] Yaroslavsky, L., Eden, M., "Fundamentals of Digital Optics", Birkhauser, Boston, 1996.

[12] Bosi, M., Brandenburg, K., Quackenbush, S., Fielder, L., Akagiri, K., Fuchs, H., Dietz, M., Herre, J., Davidson, G., Oikawa, Y., "ISO/IEC MPEG-2 Advanced Audio Coding", Journal of Audio Engineering Society, vol. 45, no. 10, 1997.