A Low-Delay MDCT/IMDCT

Sangkil Lee and Insung Lee

This letter presents an algorithm for selecting a low delay for the modified discrete cosine transform (MDCT) and inverse MDCT (IMDCT). The implementation of conventional MDCT and IMDCT requires a 50% overlap-add (OLA) for a perfect reconstruction. In the OLA process, an algorithmic delay in the frame length is employed. A reduced overlap window and MDCT/IMDCT phase shifting is used to reduce the algorithmic delay. The performance of the proposed algorithm is evaluated by applying the low-delay MDCT to the G.729.1 speech codec.

Keywords: MDCT, IMDCT, low delay, aliasing cancellation.

I. Introduction

The modified discrete cosine transform (MDCT), based on the type-IV discrete cosine transform (DCT-IV), is used widely in a range of speech codecs and audio codecs, such as G.729.1 for speech and MPEG and AAC for audio. Unlike a Fourier transform, the inverse MDCT (IMDCT) results are not reconstructed perfectly, due to aliasing. Therefore, MDCT and IMDCT use the time domain aliasing cancellation technique to obtain a perfect reconstruction [1]. On the other hand, an algorithmic delay occurs during this process. The algorithmic delay of the MDCT-based codec is equal to the frame length and is an important factor in the codec because it can cause quality problems in the communication system. The MDCT-based codec, however, requires the algorithmic delay of the frame length.

This letter presents an algorithm called the low-delay MDCT (LD-MDCT). The proposed algorithm can select the algorithmic delay for the MDCT or IMDCT using a reduced overlap window and phase shifting. In addition, details of the LD-MDCT algorithm along with an evaluation of its performance are presented.

II. LD-MDCT and LD-IMDCT

As shown in Fig. 1, the non-overlapping transforms, such as DCT-IV, cause aliasing in lossy compression coding. To eliminate this aliasing, the MDCT based on the DCT-IV is developed by an overlapping transform. MDCT requires a 50% overlap-add (OLA), and an algorithmic delay occurs. The LD-MDCT is developed to reduce the delay needed for the lookahead. The 2*N*-point MDCT and *N*-point IMDCT are respectively defined as

$$X(k)=\sqrt{\frac{2}{N}}\sum_{n=0}^{2N-1} w(k)x(n)\cos\frac{(2n+1+D)(2k+1)\pi}{4N}, \tag{1}$$

$$\hat{x}(n)=w(n)\sqrt{\frac{2}{N}}\sum_{k=0}^{N-1} X(k)\cos\frac{(2n+1+D)(2k+1)\pi}{4N}, \tag{2}$$

where *N* is the frame length, *x*(*k*) is the original signal, $\hat{x}(n)$ is the reconstruction signal, *w*(*n*) is the reduced overlap window, and *D* is the phase for aliasing cancellation.

D in the conventional MDCT is fixed to *N*. On the other hand, *D* in the LD-MDCT changes according to the desired delay. *D* is equal to the lookahead length and is defined as:

$$D=\frac{N}{2^{d-1}},\quad 2^{d-1}<N, \tag{3}$$

where *d* is a positive number. The result of the MDCT or IMDCT is changed by the value of *D* and has the following symmetry properties [2]:

$$\begin{aligned}&\hat{x}(n+(2N-\frac{D}{2}))=\hat{x}((2N-\frac{D}{2})-1-n), n=0,\ldots,\frac{D}{2}-1,\\&\hat{x}(n-\frac{D}{2})=-\hat{x}((2N-\frac{D}{2})-1-n),\qquad n=\frac{D}{2},\ldots,D.\end{aligned} \tag{4}$$

Manuscript received Dec. 13, 2012; revised Mar. 18, 2013; accepted Mar. 21, 2013.

This study was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (No. 2012 0004611).

Sangkil Lee (phone: +82 10 2568 6319, ggeguli@nate.com) and Insung Lee (inslee@chungbuk.ac.kr) are with the Department of Radio Engineering, Chungbuk National University, Cehongju, Rep. of Korea.

http://dx.doi.org/10.4218/etrij.13.0212.0559

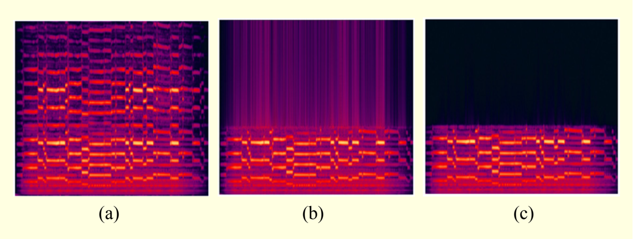


Fig. 1. Spectrogram of DCT-IV and MDCT in lossy compression coding: (a) spectrogram of original signal, (b) spectrogram of DCT-IV without OLA operation, and (c) spectrogram of MDCT with OLA operation.

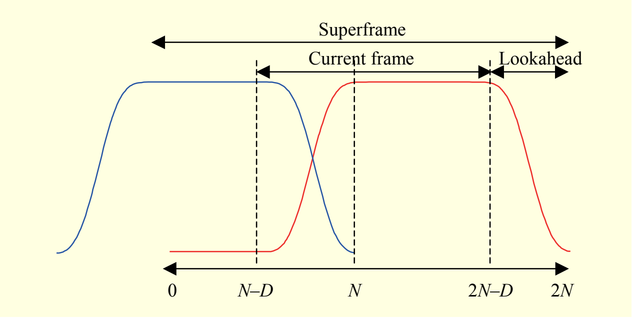


Fig. 2. Reduced overlap window used in LD-MDCT.

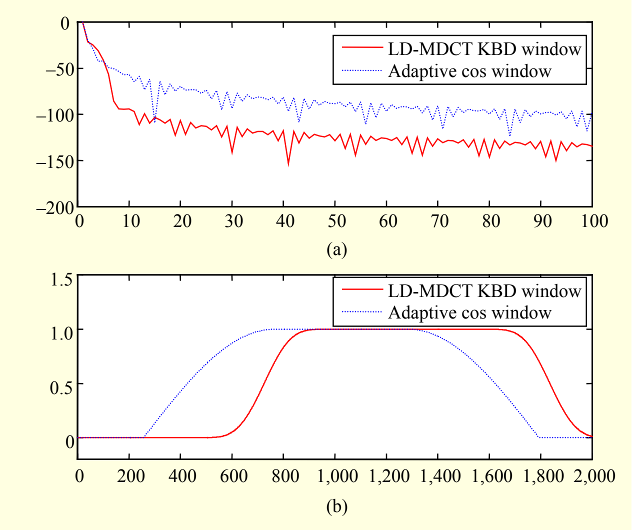


Fig. 3. Comparison of LD-MDCT window (α=4) and adaptive window: (a) frequency responses and (b) window shapes.

Using these properties, the OLA length of the LD-MDCT can be changed. The modified Kaiser-Bessel derived (KBD) window is used for the LD-MDCT, which can be expressed as

$$w(n) = \begin{cases} 0, & 0 \le n < N-D, \\ w_{\mathrm{KBD}}(n-N+D), & N-D \le n < N, \\ 1, & N \le n < 2N-D, \\ w_{\mathrm{KBD}}(n-2N+2D), & 2N-D \le n < 2N, \end{cases} \tag{5}$$

where $w_{\mathrm{KBD}}(n)$ is the KBD window. The KBD window function is defined in terms of the Kaiser window $w_{\mathrm{K}}(n)$ and is defined as

$$w_{\mathrm{K}}(n) = \frac{I_0\left(\pi\alpha\sqrt{1-(\frac{2n}{D}-1)^2}\right)}{I_0(\pi\alpha)}, \quad 0 \le n < D, \tag{6}$$

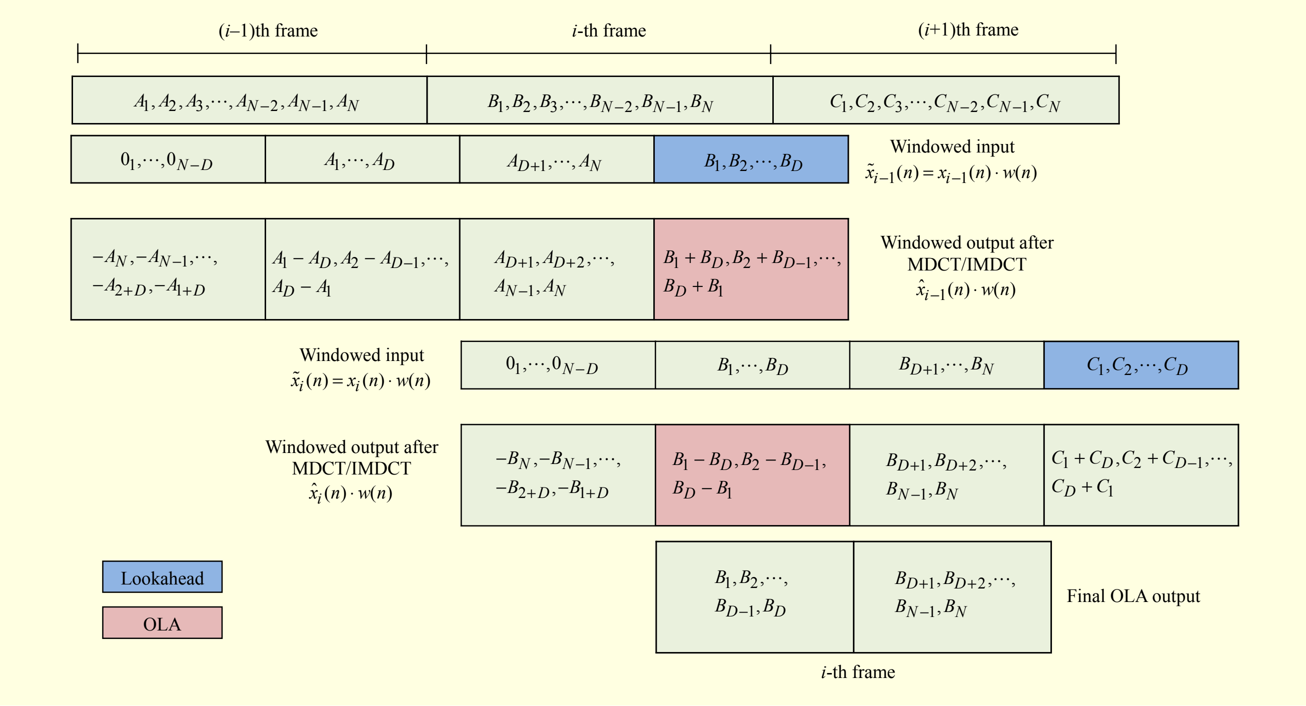
$$w_{\mathrm{KBD}}(n) = \begin{cases} \sqrt{\dfrac{\sum_{j=0}^{n} w_K(j)}{\sum_{j=0}^{D} w_K(j)}}, & 0 \le n < D, \\ \sqrt{\dfrac{\sum_{j=0}^{2D-1-n} w_K(j)}{\sum_{j=0}^{D} w_K(j)}}, & D \le n < 2D, \end{cases} \tag{7}$$

where I_0 is the zeroth-order modified Bessel function of the first type and α is an arbitrary real number that determines the shape of the window. In the frequency domain, it determines the tradeoff between the main lobe width and side lobe level. Figure 2 gives an illustration of (5) [3], [4]. The proposed window shape is similar to the adaptive window shape proposed in [5]. As shown in Fig. 3(a), however, the proposed KDB window is better than the adaptive window in terms of the analysis tools for the frequency response [6].

Unlike a conventional MDCT, the delay of the LD-MDCT can be reduced. For example, a conventional MDCT with a frame size of 320 must have a lookahead of 320 samples, making its superframe size 640 samples, resulting in a 320-sample delay. On the other hand, only a 160-sample delay occurs if the D value of the LD-MDCT is 160 (N/2). The superframe size of the LD-MDCT is equal to that of the conventional MDCT. The front 160 samples of the current frame are reconstructed by an OLA using the prior frame's lookahead region. The remaining 160 samples of the current frame are reconstructed without overlapping. This allows the delay to be selected according to the changes in the D value. Finally, Fig. 4 shows the entire process of the LD-MDCT algorithm. In Fig. 4, $\tilde{x}_i(n)$ is the windowed input signal, i is the frame index, and $\hat{x}_i(n)$ is the LD-IMDCT coefficient of $X_i(k)$, which contains time domain aliasing:

$$\hat{x}_i(n) = \begin{cases} \tilde{x}_i(n) - \tilde{x}_i(2N-D-1-n), & n = 0,\ldots,2N-D-1, \\ \tilde{x}_i(n) + \tilde{x}_i(4N-D-1-n), & n = 2N-D,\ldots,2N-1. \end{cases} \tag{8}$$

This signal is not the same as the input signal. A perfect reconstruction is achieved by removing the aliasing signal using the OLA. The OLA length is reduced with the zeros on



(i–1)th frame
i-th frame
(i+1)th frame
$A_1, A_2, A_3, \cdots, A_{N-2}, A_{N-1}, A_N$
$B_1, B_2, B_3, \cdots, B_{N-2}, B_{N-1}, B_N$
$C_1, C_2, C_3, \cdots, C_{N-2}, C_{N-1}, C_N$
$0_1, \cdots, 0_{N-D}$
$A_1, \cdots, A_D$
$A_{D+1}, \cdots, A_N$
$B_1, B_2, \cdots, B_D$
Windowed input
$\tilde{x}_{i-1}(n) = x_{i-1}(n) \cdot w(n)$
$-A_N, -A_{N-1}, \cdots, -A_{2+D}, -A_{1+D}$
$A_1 - A_D, A_2 - A_{D-1}, \cdots, A_D - A_1$
$A_{D+1}, A_{D+2}, \cdots, A_{N-1}, A_N$
$B_1 + B_D, B_2 + B_{D-1}, \cdots, B_D + B_1$
Windowed output after MDCT/IMDCT
$\hat{x}_{i-1}(n) \cdot w(n)$
Windowed input
$\tilde{x}_i(n) = x_i(n) \cdot w(n)$
$0_1, \cdots, 0_{N-D}$
$B_1, \cdots, B_D$
$B_{D+1}, \cdots, B_N$
$C_1, C_2, \cdots, C_D$
Windowed output after MDCT/IMDCT
$\hat{x}_i(n) \cdot w(n)$
$-B_N, -B_{N-1}, \cdots, -B_{2+D}, -B_{1+D}$
$B_1 - B_D, B_2 - B_{D-1}, B_D - B_1$
$B_{D+1}, B_{D+2}, \cdots, B_{N-1}, B_N$
$C_1 + C_D, C_2 + C_{D-1}, \cdots, C_D + C_1$
Lookahead
OLA
$B_1, B_2, \cdots, B_{D-1}, B_D$
$B_{D+1}, B_{D+2}, \cdots, B_{N-1}, B_N$
Final OLA output
i-th frame

Fig. 4. Whole process of LD MDCT/IMDCT.

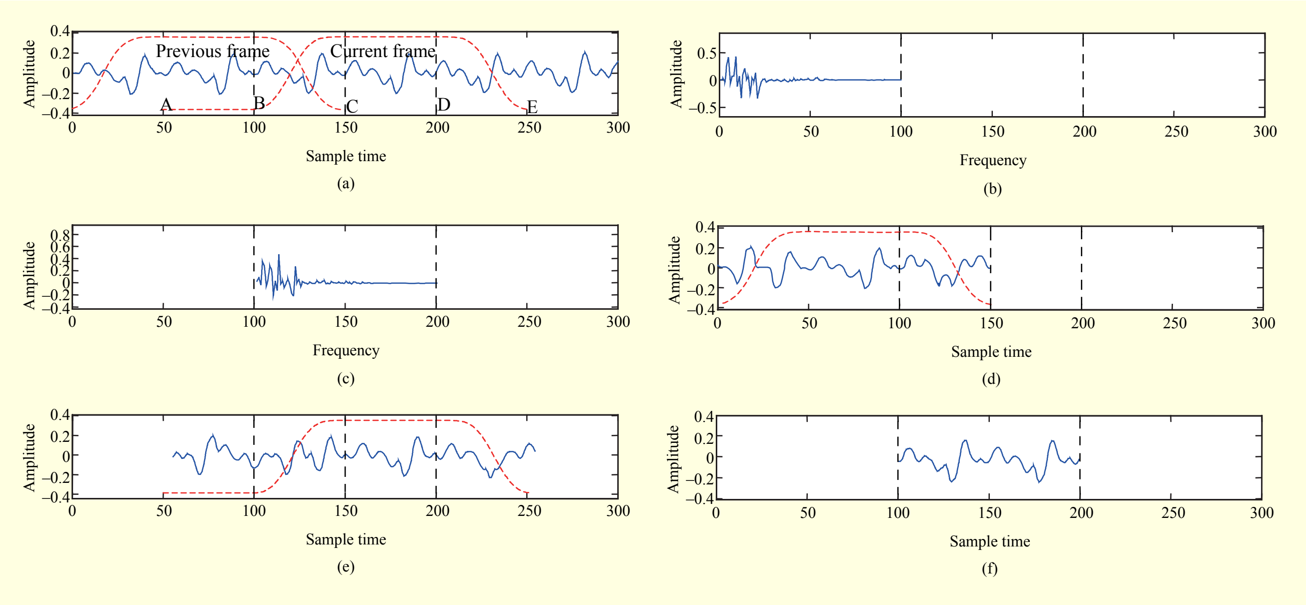


Fig. 5. Illustration of LD-MDCT(N/2 delay), OLA procedure, and perfect reconstruction: (a) speech input signal, dashed lines reduced-overlap window; (b) LD-MDCT coefficients of signal in previous frame; (c) LD-MDCT coefficients of signal in current frame; (d) LD-IMDCT coefficients of signal in (b); (e) LD-IMDCT coefficients of signal in (c); and (f) reconstructed time domain signal after OLA procedure.

the front (N–D samples) and achieves a perfect reconstruction:

$$\hat{x}_i^{\mathrm{PR}}(n)=\begin{cases}\hat{x}_{i-1}(2N-D+n)+\hat{x}_i(N-D+n), & n=0,\dots,D-1,\\ \hat{x}_i(N+n), & n=D,\dots N-1,\end{cases} \qquad (9)$$

where $\hat{x}_i^{\mathrm{PR}}(n)$ is the perfect reconstruction signal.

III. Performance Evaluation

The LD-MDCT is applied to the G.729.1 speech codec

(16 kbps) to evaluate the performance of the LD-MDCT [7]. The input speech signals consist of 20 sentences spoken by males and 20 sentences spoken by females. G.729.1 has an algorithmic delay of 48.9375 ms (20 ms for the input superframe, 20 ms for the MDCT lookahead, 5 ms for the LPC lookahead, and 3.9375 ms for the QMF analysis-synthesis filter bank). The MDCT lookahead delay is reduced using the LD-MDCT. If D is $N/16$, the MDCT lookahead delay becomes 1.25 ms. The result of the perceptual evaluation of the speech quality (PESQ) test for the G.729.1 output (3.8642) is the same whether using the conventional MDCT or the LD-MDCT. Despite the reduced delay, the LD-MDCT indicates the same performance as that of the conventional MDCT.

Figure 5 shows the process of a perfect reconstruction in the LD-MDCT ($N/2$ delay, where N is the frame length [100 samples]). Figure 5(a) shows a speech input signal of 300 samples. Figure 5(b) presents the LD-MDCT coefficients of the signal in the previous frame. The alias is introduced due to the 50% decimation in the LD-MDCT (from $2N$ time domain samples to N independent frequency domain coefficients) [8]. Because of this aliasing, the LD-MDCT introduces redundancy (from N frequency domain coefficients in Fig. 5(b) to the $2N$ time domain samples in Fig. 5(d)). Figure 5(c) presents the LD-MDCT coefficients of the signal in the current frame. Figure 5(e) shows the corresponding LD-IMDCT time domain signal. If the OLA procedure is performed with the aspects shown in Figs. 5(d) and 5(e) (between points B and C), a perfect reconstruction of the original signal in the OLA part (between points B and C) can be achieved. In addition, the non-OLA part (between points C and D) can achieve the perfect reconstruction without an OLA.

IV. Conclusion

This letter presented an algorithm for LD-MDCT and LD-IMDCT. The LD-MDCT algorithm can select the lookahead length and reduce the delay using a reduced overlap window and phase shifting. Despite the reduced delay, the LD-MDCT still achieves a perfect reconstruction. The result of the PESQ test using the LD-MDCT in the G.729.1 codec showed the same performance as the conventional MDCT.

References

[1] J.P. Princen and A.B. Bradley, “Analysis/Synthesis Filter Bank Design Based on Time Domain Aliasing Cancellation,” *IEEE Trans. Acoustics*, *Speech*, *Signal Process.*, vol. 34, no. 5, Oct. 1986, pp. 1153-1161.

[2] H.S. Malvar, “Lapped Transforms for Efficient Transform/Subband Coding,” *IEEE Trans. Acoustics*, *Speech*, *Signal Process.*, vol. 38, no. 6, June 1990, pp. 969-978.

[3] Xiph.Org Foundation, “Vorbis I Specification,” Feb. 2012. http://xiph.org/vorbis/doc/Vorbis_I_spec.html

[4] M. Iwadare et al., “A 128 kb/s Hi-Fi Audio CODEC Based on Adaptive Transform Coding with Adaptive Block Size MDCT,” *IEEE Trans. Sel. Areas Commun.*, vol. 10, no. 1, Jan. 1992, pp. 138-144.

[5] J.-M. Valin et al., “A High-Quality Speech and Audio Codec with Less Than 10 ms Delay,” *IEEE Trans. Audio*, *Speech*, *Language Process.*, vol. 18, no. 1, Jan. 2010, pp. 58-67.

[6] T. Lee et al., “Adaptive TCX Windowing Technology for Unified Structure MPEG-D USAC,” *ETRI J.*, vol. 34, no.3, June 2012, pp. 474-477.

[7] S. Ragot et al., “ITU-T G,729.1: An 8-32 kbit/s Scalable Wideband Coder Bitstream Interoperable with G.729 for Wideband Telephony and Voice Over IP,” *IEEE Int. Conf. Acoustics*, *Speech*, *Signal Process.*, Honolulu, HI, USA, Apr. 2007, pp. IV:529-IV:532.

[8] J.P. Princen, A.W. Johnson, and A.B. Bradley, “Subband/Transform Coding Using Filter Bank Designs Based on Time Domain Aliasing Cancellation,” *IEEE Int. Conf. Acoustics*, *Speech*, *Signal Process.*, vol. 12, 1987, pp. 2161-2164.