LETTER

Efficient Windowing Scheme for MDCT-Based TCX in AMR-WB+

Jae-seong LEE[†a)], ***Nonmember***, **Young-cheol PARK**[††], ***Member***, **Dae-hee YOUN**[†], ***and*** **Kyung-ok KANG**[†††], ***Nonmembers***

SUMMARY Although the AMR-WB+ coder provides excellent quality for speech signal, its coding model for music signals is not as optimal as the HE-AAC v2. The main causes of the poor quality of the AMR-WB+ TCX are the non-critical sampling and block artifacts. The new TCX windowing scheme proposed in this paper uses an MDCT with a 50% frame overlap, so that the problems of non-critical sampling and blocking artifacts are significantly mitigated. Due to long overlaps, the proposed scheme involves an additional codec delay. It is, however, moderate for audio services. The results of objective and subjective tests indicate that the proposed scheme achieves noticeable quality improvements for music signals over the previous TCX schemes.

key words: *AMR-WB+, TCX, MDCT*

1. Introduction

The Extended Adaptive Multi-Rate - Wideband (AMR-WB+) audio coder, standardized in 2004 by the 3GPP [1], uses a hybrid coding model that switches automatically, depending on the characteristics of the input signal, between an Algebraic Code-Excited Linear Prediction (ACELP) and a Transform Coded eXcitation (TCX) coding model. The coder using a multi-mode encoding model can switch frame by frame between time and frequency-domain encoding. In time-domain mode, the input signal is encoded using an Adaptive Multi-Rate - Wideband(AMR-WB) speech coding standard [2]. In frequency-domain mode, a weighted version of the input signal is encoded in the discrete Fourier transform (DFT) domain using TCX [3], [4].

Although the AMR-WB+ performs well for speech signals at low bit rates, its quality for music signals is much lower than other audio coders such as the High Efficiency Advanced Audio Coding (HE-AAC) v2. On the other hand, HE-AAC does not perform well for speech signals, since it can not use a small bit budget as efficiently as linear predictive (LP) coders when encoding speech [5], [6]. At 16~20 kbps, the music quality of the AMR-WB+ is significantly worse than that of the HE-AAC v2 [6]. One of the major reasons is overhead information, particularly during the core-coding transitions, due to non-critical sampling with a low-frequency resolution. To solve this problem, modified discrete cosine transform (MDCT)-based TCX coder has been proposed [8]. By employing the MDCT, the overhead information during the transitions could be reduced, which resulted in significant quality improvement. The other important reason for the poor quality of the original TCX in the AMR-WB+ is a block artifact being caused by the short overlap between frames. The windows in [8] allowed longer frame overlaps than the original. However, it was still not enough to reduce the block artifacts. In essence, analysis windows with short overlap cannot achieve accurate frequency resolution.

This paper proposes a new TCX windowing scheme that can improves the quality of AMR-WB+ at low bit rates. The proposed windowing scheme utilizes MDCT instead of DFT, with a longer overlap (50%) than that of the original (12.5%). Due to long overlaps, the proposed scheme involves an additional codec delay. However, it is still moderate for audio services. By implementing Time-domain Aliasing Cancellation (TDAC) with windows having low sidelobes, the proposed scheme can reduce both the overhead information and block artifacts, which directly results in quality improvement, especially for music signals. This paper is organized as follows. Section 2 briefly describes the AMR-WB+ TCX scheme, and the new TCX windowing scheme is presented in Sect. 3. Finally, the results of objective and subjective quality evaluations are presented in Sect. 4, with the conclusions in Sect. 5.

2. AMR-WB+ TCX

In TCX of AMR-WB+, the input signal is first filtered through a time-varying weighting filter to obtain a weighted signal. Then the weighted signal is transformed by using the DFT, and the DFT coefficients are quantized using a Lattice Vector Quantizer. In total, there are 20-ms TCX (TCX20), 40-ms TCX (TCX40), and 80-ms TCX (TCX80) coding modes.

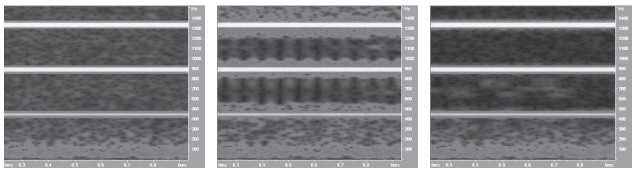
The TCX mode in AMR-WB+ uses overlapping. The window shape simultaneously improves the transform coding performance and allows a smooth transition from ACELP frame to TCX frame and between two consecutive TCX frames.

However, previous research [5], [6] indicated that the coding gain of AMR-WB+, especially for music signals, is much lower than HE-AAC v2. The major cause of a

Manuscript received November 8, 2010.
Manuscript revised February 9, 2011.

[†]The authors are with the Dept. of Electrical & Electronic Eng. Yonsei University, Seoul, Korea.
[††]The authors is with the Dept. of Computer & Telecommunications Eng. Yonsei University, Wonju, Korea.
[†††]The authors is with the Electronics and Telecommunications Research Institute, Daejun, Korea.
a) E-mail: dream7070@dsp.yonsei.ac.kr

DOI: 10.1587/transinf.E94.D.1341



(a) (b) (c)

Fig. 1 Example of block artifacts: (a) original, (b) AMR-WB+ and (c) HE-AAC v2.

poor coding gain is that the conventional DFT-based TCX is not based on critical sampling, which causes low-frequency resolution and overhead data during the core-coding transitions. Another problem associated with AMR-WB+ TCX is the block artifact, which is caused by the short overlap between TCX frames. Although the windowing scheme of AMR-WB+ TCX is designed to minimize the blocking artifacts, experimental results indicate that artifacts caused by the frame transition still exist between frames, and these artifacts become more serious as bit rate decreases. Figure 1 shows spectrograms of synthesized music signals for an input music of pitch pipe sound ("phi7") encoded at 20 kbps mono. Since this music signal contains fairly stationary tones, AMR-WB+ runs almost in TCX80 mode. In the figure, the artifacts are visible between frames of the AMR-WB+, whereas the HE-AAC shows no visible artifacts. These artifacts are usually masked when the signal contents are dynamically varied as in frames containing vocal music. However for stationary tone signals such as the ones shown in Fig. 1, the block artifacts are easily audible.

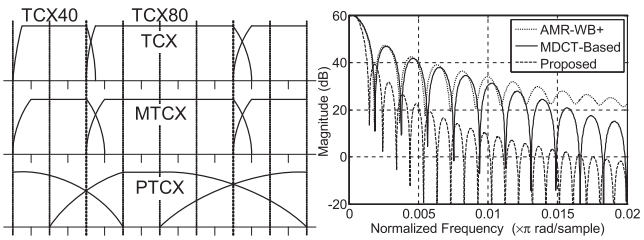
3. New TCX Windowing Scheme

In AMR-WB+ TCX, the weighted and windowed input signal is mapped to the frequency domain using a DFT [1]. The proposed windowing scheme is based on MDCT and it implements TDAC with a 50% overlap between adjacent frames. The TCX windows in [8] allow a 128-point overlap between all TCX frames, which is longer than for the original TCX, however insufficient to mitigate the block artifact problem. As a result, artifacts are still perceptible. In the proposed algorithm, frame overlaps depend on frame length, but the 50% frame overlap is always guaranteed, except in the case of non-homogeneous window connections. More details of the proposed scheme are described below.

ACELP/TCX of the AMR-WB+ uses 1024-sample super-frames, in which one of 256, 512 or 1024 sample frames is chosen based on the closed loop operation. That is, all possible modes within the super-frame are tried in the encoder and it selects the best mode combination based on segmental SNR values. In the new windowing scheme, the mode selection algorithm of the AMR-WB+ is maintained. DFTs for sub-frame data are replaced by MDCTs with appropriate windows designed to satisfy TDAC. In detail, a 2048-sample MDCT is used for TCX80 frame, and a 1024/512-sample MDCT is used for TCX40/20 frames. To deal with the 1024-sample super-frame used in the AMR-WB+, we defined a 2048-sample super-frame for the MDCT processing by adding 1024 samples (previous 512 samples and next 512 samples) to the current 1024-sample frame.

Table 1 Constants l_1, l_2, l_3 and l_4 for the proposed window.

Neighbor Frame \ Current Frame		TCX80	TCX40	TCX20
Previous (l_1, l_2)	TCX80	1024, 0	512, 0	256, 0
	TCX40	512, 256	512, 0	256, 0
	TCX20	256, 384	256, 128	256, 0
	ACELP	0, 512	0, 256	0, 128
Next (l_3, l_4)	TCX80	0, 1024	0, 512	0, 256
	TCX40	256, 512	0, 512	0, 256
	TCX20	384, 256	128, 256	0, 256
	ACELP	512, 0	256, 0	128, 0



(a) (b)

Fig. 2 (a) Window switching diagrams of the AMR-WB+ (top), modified TCX in [8] (middle) and the proposed TCX(bottom) TCX80 windows, in a case in which the previous frame is TCX40 and the next frame is TCX80. (b) Frequency responses of AMR-WB+ (dotted line), modified TCX (dashed line) and proposed TCX (solid line) windows.

For the design of windows, Sine window is utilized. To deal with all possible transitions between different-sized frames, we defined a total of 29 window types. The windows are formed by concatenating four sub-windows defined as

$$\begin{aligned} w_1(n) &= \sin(\pi n/(2l_1)) && for\ n = 0, \ldots, l_1 - 1, \\ w_2(n) &= 1 && n = 0, \ldots, l_2 - 1, \\ w_3(n) &= 1 && for\ n = 0, \ldots, l_3 - 1, \\ w_4(n) &= \sin(\pi(n/(2l_4) + 0.5)) && n = 0, \ldots, l_4 - 1. \end{aligned} \tag{1}$$

The constants l_1, l_2, l_3 and l_4 respectively control the shape of the left, left-middle, right-middle and right part of the window. The constants l_1, l_2, l_3 and l_4 for the proposed window are summarized in Table 1. For example, when the current frame is TCX80, the previous frame is TCX20 and the next frame is TCX80, the constants are $l_1 = 256$, $l_2 = 384$, $l_3 = 0$ and $l_4 = 1024$. The window shape at the end of the super-frame depends on the mode of the next frame. For example, when the current mode uses a 2048 point MDCT and the mode of the next frame starts with a 512 point MDCT, the current window shape is determined by $l_4 = 512$. In the case of transitions from TCX to ACELP, the l_4 is always zero, so the right overlap is discarded. Figure 2 (a) shows an example of a new window for the case in which the previous frame is TCX40 and the next frame is TCX80. The new window has a longer overlap region than the original TCX window in the head and tail parts. The proposed windowing scheme always satisfies the 50% frame overlap, except in

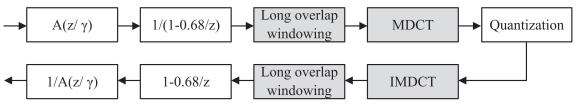


Fig. 3 Structure of the proposed windowing scheme in the modified TCX [8].

the cases of transitions between TCX to ACELP. The new windowing scheme allows a slow transition in the head tail part, so that the effect of spectral leakage can be alleviated. Figure 2 (b) compares the frequency responses of the original TCX80 window and the new TCX80 window. The new window has narrower main-lobe and lower side-lobe levels than the original TCX window. On the other hand, the window used in [8] has similar frequency characteristics to that of AMR-WB+. The long overlap can introduce additional codec delay. The algorithmic codec delay of AMR-WB+ for the typical Internal Sampling Frequency (ISF) of 25.6 kHz is 113 ms [10]. The maximum additional delay by using our method is 40 ms for TCX80 mode. And the minimum additional delay is 10 ms for TCX20 mode. Considering inherent delay of AMR-WB+, and in a situation that AMR-WB+ can be used for audio services, the additional delay would not be critical. Finally, Fig. 3 shows the structure of the proposed TCX windowing scheme. The overall structure is the same as the modified TCX in [8].

4. Experimental Results

4.1 Objective Performance Measure

We first compared the spectrograms of the output signals; these are shown in Fig. 4. For ease of comparison, we chose the same "phi7" mono signal that was used in Fig. 1 at 20 kbps. The figure shows that the proposed algorithm noticeably reduces block artifacts compared to the original AMR-WB+ TCX and the TCX in [8]. The results in Fig. 1 confirm that the proposed TCX windowing scheme can mitigate the block artifact problem by allowing 50% overlap between adjacent MDCT frames. However, a long overlap is disadvantageous in terms of preserving the details of spectral envelops; thus, we next evaluated the negative effects of the long frame overlap. Since the frame overlap averages the time features of adjacent frames, the temporal envelope is likely to be blurred. It should be mentioned that although the frame overlap of the proposed algorithm is longer than the AMR-WB+ and the modified TCX in [8], the time resolution is the same because the frame size is unchanged.

To assess the average effect of overlap, we measured the Perceptual Evaluation of Audio Quality (PEAQ) [7]. Figure 5 shows the measured PEAQ for 12 test items at 20 kbps mono. The audio items for the tests covered three categories: speech, music, and mixed speech/music signals. Twelve audio items were used, with four items in each of the three categories. These items have been commonly used in the process of Unified Speech and Audio Coding (USAC) standardization [9].

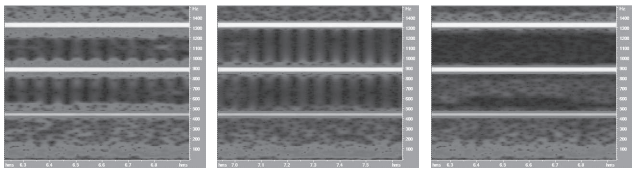


Fig. 4 Example spectrograms obtained using (a) AMR-WB+, (b) modified TCX in [8] and (c) proposed TCX windowing scheme.

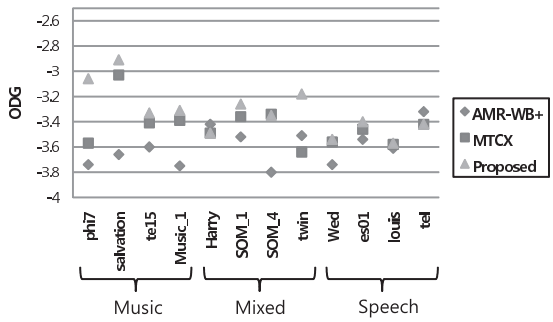


Fig. 5 Objective Difference Grade of proposed TCX(Proposed), modified TCX(MTCX) [8] and AMR-WB+.

In the figure, Objective Difference Grade(ODG) values are shown. For speech items, the proposed algorithm shows a similar level of ODG to the AMR-WB+ and modified TCX. For music items, however, the proposed algorithm shows a slightly higher level of ODG than AMR-WB+ and the modified TCX. One exceptional case is"phi7" which consists of fairly stationary temporal envelopes. For this particular music, the proposed algorithm outperforms both the original AMR-WB+ and the modified TCX without overlap. Based on these results, it can be said that the distortion due to the long overlap is not particularly problematic in the proposed algorithm. Furthermore, a more faithful preservation of decoded signal is expected for stationary inputs than in the other TCX methods.

The proposed scheme was also evaluated by measuring the segSNR, since it is the criterion on which the closed-loop decision of AMR-WB+ relies. For comparison, we also measured the segSNR of the original TCX and the TCX in [8]. The measured segSNR values are plotted in Fig. 6 for coding bit rate 20 kbps. We can see that the proposed TCX windowing scheme provides similar segSNR for all categories considered. However, a significant improvement was achieved for "phi7", which confirms effectiveness of the proposed windowing scheme for stationary audio contents.

4.2 Subjective Evaluation

Next, we performed formal listening tests using the same audio items. An AMR-WB+ coder was used as a main frame, and only the TCX module was switched. The tests were carried out according to the MUSHRA procedure [11]. The test items were same as those used in Fig. 5 and 15 experienced listeners participated in the test. The listening tests were

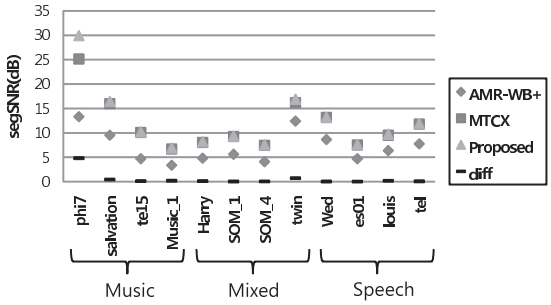
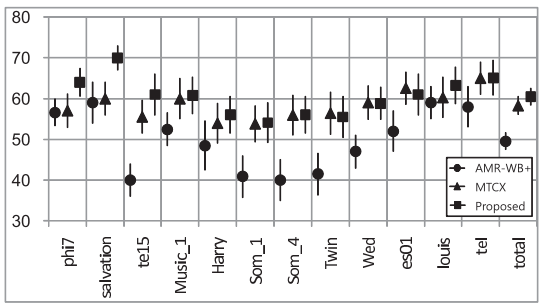
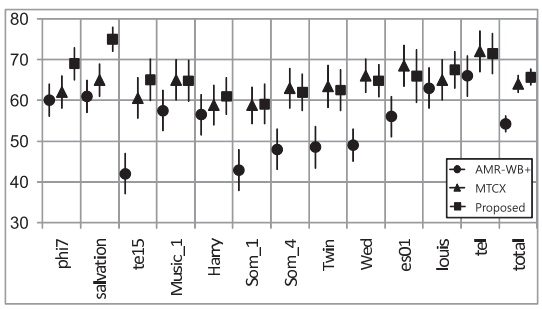


Fig. 6 Segmental SNR Performance of AMR-WB+ core coding using different flavors of TCX for music, mixed and speech.



(a)



(b)

Fig. 7 MUSHRA tests results for (a) 16 kbps, (b) 20 kbps.

carried out for mono signals at 16 and 20 kbps. The test results are summarized in Figs. 7 (a) and (b). The results show that, especially for music and mixed signals, the proposed scheme outperforms the AMR-WB+. For speech signals, the quality is also better than the original AMR-WB+ TCX.

Compared with the modified TCX in [8], the proposed algorithm shows a slight improvement in total score. However, the improvement is significant for music items such as "phi7", "Salvation", and "te15". The reason for this is that those music items mainly consist of consecutive long frames. The percentages of the selected frame modes are summarized in Fig. 8. As the figure depicts, the music items are encoded mostly in the TCX80 mode so that the long overlap scheme proposed in this paper effectively reduces the block artifacts. One exception is "Music_1." Although it consists of 80% TCX80 frames, there were occasional switches between TCX80 and ACELP the during non-harmonic attack transitions, and these offset the positive

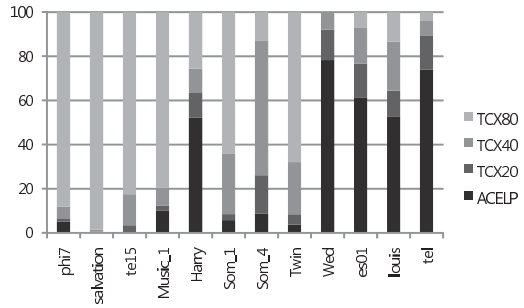


Fig. 8 Percentages of selected frame modes.

effect of the long overlap.

5. Conclusion

We have presented a new windowing scheme for TCX in AMR-WB+. The proposed scheme is based on MDCT with 50% overlapping frames implementing TDAC. Long overlap is accompanied by additional coding delay, but it is moderate for audio services. Objective evaluations confirmed the effectiveness of the proposed windowing scheme operating at low bitrates. Listening tests also showed that the proposed scheme could provide better quality than the conventional TCX schemes for music signals and maintain similar quality for speech and mixed signals.

References

[1] R. Salami, R. Lefebvre, A. Lakaniemi, K. Kontola, S. Bruhn, and A. Taleb, "Extended AMR-WB for high-quality audio on mobile devices," IEEE Commun. Mag., vol.44, no.5, pp.90–97, May 2006.
[2] 3GPP TS 26.190, "AMR wideband speech codec; Transcoding functions," 2009.
[3] R. Lefebvre, R. Salami, C. Laflamme, and C. Laflamme, "8 kbps coding of speech with 6 ms frame-length," Proc. ICASSP'93, pp.II-612–615, Minnesota, 1993.
[4] S. Ragot, B. Bessette, and R. Lefebvre, "Low complexity multi-rate lattice vector quantization with application to wideband TCX speech coding at 32 kbit/s," IEEE Conf. ASSP, Canada, 2004.
[5] ISO/IEC 14496-3, "Coding of audio-visual objects : Audio," 2005.
[6] M. Neuendorf, P. Gournay, M. Multrus, J. Lecomte, B. Bessette, R. Geiger, S. Bayer, G. Fuchs, J. Hilpert, N. Rettelbach, R. Salami, G. Schuller, R. Lefebvre, and B. Grill, "Unified speech and audio coding scheme for high quality at low bitrates," ICASSP'09, Taipei, 2009.
[7] ITU-R. Recommendation BS-1387, "Method for objective measurements of perceived audio quality," ITU, Switzerland, 1998.
[8] G. Fuchs, M. Multrus, M. Neuendorf, and R. Geiger, "MDCT based coder for highly adaptive speech and audio coding," 17th EUSIPCO Scotland, 2009.
[9] ISO/IEC JTC1/SC29/WG11, "Call for proposals on unified speech and audio coding.," MPEG2007/N9519, China, 2007.
[10] 3GPP TR 26.936, "Performance characterization of audio codecs," 2009.
[11] ITU Radiocommunication Bureau, "Method for the subjective assessment of intermediate quality level of coding systems," ITU-R Recommendations, Supplement 1 to vol.2000 - BS-BT Series, Recommendation ITU-R BS.1534, June 2001.