ISCA Archive
http://www.isca-speech.org/archive

First International Conference on
Spoken Language Processing (ICSLP 90)
Kobe, Japan
November 18-22, 1990

ADAPTIVE TRANSFORM CODING WITH AN ADAPTIVE BLOCK SIZE USING A MODIFIED DCT

Fumie Hazu, Akihiko Sugiyama, Masahiro Iwadare and Takao Nishitani

C&C Systems Research Laboratories, NEC Corporation

Miyazaki 4-1-1, Miyamae-ku, Kawasaki-shi, KANAGAWA, 213 JAPAN

Abstract

A Modified Discrete Cosine Transform (MDCT), which is effective for reducing block boundary noise, is applied to Adaptive Transform Coding with an Adaptive Block Size (ATC-ABS) reported by authors. As original MDCT is defined to have a fixed block size for realizing time domain aliasing cancellation, a new overlap window set is introduced to enable variable block size existence, in ATC-ABS. In addition to MDCT employment, a new adaptive bit assignment algorithm based on the auditory characteristic is also introduced to prevent distortion by deficiency of assigned quantization bit amount.

In the results of paired comparison tests, the original signals and the reproduced signals by the proposed algorithm are preferred by more than 40% of the subjects, while the reproduced signal by ATC-ABS is preferred by less than 10% of the subjects. The proposed method is promising for digital recording media and for digital communications.

1. Introduction

For the purpose of the application to Integrated Services Digital Network (ISDN) and a variety of recording media, the coding of high-quality audio at low bit rates has been widely studied[1]-[4]. An international standard for high-quality audio recording in compact discs with moving pictures has been under study in ISO/MPEG (Moving Picture Expert Group)[4],[5].

Adaptive Transform Coding (ATC) is one of the promising high-quality audio coding algorithms. The coding quality of ATC is deteriorated by pre-echoes, which are noise components preceding front-end of an individual sound. ATC with an Adaptive Block Size (ATC-ABS) has been reported which adaptively changes the transform block size according to input signal characteristics to reduce pre-echoes[6]. ATC-ABS coding quality is improved for many kinds of sources, such as castanets and drums. However, in some sources, including a bassoon and a piano, block boundary noise is perceived. Moreover, in the source of cabasa which has widely spreaded spectrum, coding quality is not improved.

ATC-ABS has room for improvement in DCT with overlapping and in a quantizer based on the Zelinski and Noll algorithm[7]. Adjacent blocks in a conventional DCT are overlapped to each other to reduce the block boundary noise[8]. However, there is a trade-off between the overlap length and the coding efficiency. A large overlap is more effective for reducing block boundary noise. On the other hand, such an overlap results in less efficient data compression since the overlapped portion is transmitted twice. As an efficient algorithm to reduce the block boundary noise, Princen et al. proposed an MDCT based on TDAC algorithm[9]. However, the MDCT cannot be directly applied to ATC-ABS as it is. MDCT assumes that adjacent block sizes are the same and it never holds for ATC-ABS. Introduction of a special window set could expands the MDCT operation form to having a different block size in adjacent blocks.

ATC-ABS is equipped with Zelinski-Noll quantizer which first assigns limited quantization bit amount to each transformed coefficient according to its power, and then quantizes each transformed coefficient with the assigned bits. Quantization bit amount is effectively assigned by taking advantage of power concentration in the frequency domain. When transformed coefficient power is distributed over a broad band, the quality of the quantizer is deteriorated. Therefore, a new bit assignment algorithm which assigns more quantization bit amount to lower band based on the auditory characteristic should be employed.

This paper proposes a new ATC-ABS algorithm whose auditory quality is improved by using the MDCT with a new window set and a new bit assignment algorithm. Section 2 describes the MDCT introduction. A new bit assignment algorithm is described in section 3. Section 4 shows the results of subjective tests.

2. The MDCT introduction

One problem in the conventional ATC-ABS is overlapped DCT processing. An overlapped DCT is effective for reducing block boundary noise. However, it is not always possible to overlap as much as desired, because a larger overlap requires more repeated transmission as is mentioned in the previous section. As one of the solutions to this problem, an MDCT based on TDAC algorithm[9] has been proposed which allows 50% overlap processing. The following equations define the MDCT with a block size M. In the equation, $x(n)$, $y(n)$, and $x'(n)$ is defined input signal, MDCT transformed coefficient, and inverse MDCT transformed coefficient, respectively.

$$y(k)=\sum_{n=0}^{M-1} x(n)\cdot h(M-1-n)\cdot\cos\left(\frac{2\pi}{M}\cdot(k+\frac{1}{2})\cdot(n+n_0)\right), \quad (1)$$

$$x'(n)=\frac{1}{M}\cdot\sum_{k=0}^{M-1} y(k)\cdot h(n)\cdot\cos\left(\frac{2\pi}{M}\cdot(k+\frac{1}{2})\cdot(n+n_0)\right), \quad (2)$$

$$k=0,1,\cdots,M-1, \quad n=0,1,\cdots,M-1, \quad n_0=\frac{M}{4}+\frac{1}{2},$$

where $h(n)$ defines the window function which needs to comply with the following conditions.

$$\begin{cases} h(n)=h(M-n-1) \\ h^2(n)+h^2(n+\frac{M}{2})=1 \end{cases} \quad (3)$$

The reason why MDCT is effective is that MDCT with a block size M needs to transmit just M/2 coefficients, thanks to the symmetry property of $y(k)=-y(M-k)$. Inverse MDCT is performed by using M transformed coefficients unfolded from M/2 coefficients obtained by the forward transform.

However, the MDCT could not be directly applied to ATC-ABS which selects a block size from some candidate blocks. This is because the MDCT assumes the same block size for the adjacent blocks. Moreover, ATC-ABS may have some regions where more than one blocks are overlapped.

This problem is overcomed by introducing a set of windows ,{ $h_i(n) \mid i=1,\cdots,l$ }, which corresponds to different block sizes ,{ $M_i \mid i=1,\cdots,l$ }. The end shape of the window is defined so that the substantial overlapped portions from the left and the right are symmetric independent of the adjacent block size. Therefore, new MDCT operations can be summarized in the following equation. In the equation, $y_i(n)$ defines input signal MDCT transformed coefficient. Subscript i ($i=1,\cdots,l$) defines the candidate block number.

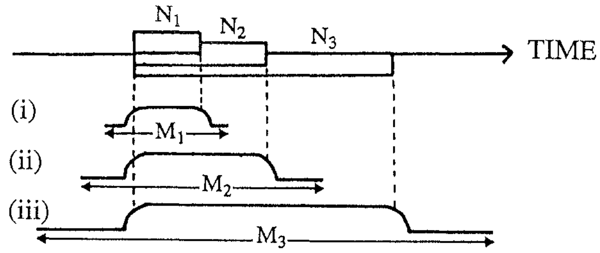
$$y_i(k)=\sum_{n=0}^{M_i-1} x(n)\cdot h_i(M_i-1-n)\cdot\cos\left(\frac{2\pi}{M_i}\cdot(k+\frac{1}{2})\cdot(n+n_0)\right), \quad (1)'$$

$$x'(n)=\frac{1}{M_i}\cdot\sum_{k=0}^{M_i-1} y_i(k)\cdot h_i(n)\cdot\cos\left(\frac{2\pi}{M_i}\cdot(k+\frac{1}{2})\cdot(n+n_0)\right), \quad (2)'$$

$$k=0,1,\cdots,M_i-1, \quad n=0,1,\cdots,M_i-1, \quad n_0=\frac{M_i}{4}+\frac{1}{2},$$

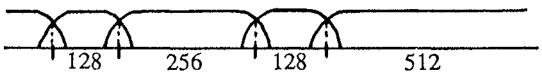
where the window $\{h_i(n) \mid i=1,\cdots,l\}$ is defined by the following equation. In the equation, L defines the actual overlap length : $L=\min\{M_i\}$.

$$h_i(n)=\begin{cases} 0 & (n=0,\cdots,\frac{M_i}{4}-\frac{L}{2}-1) \\ & (n=3\cdot\frac{M_i}{4}+\frac{L}{2},\cdots,M_i-1), \\ \sin\left[\frac{\pi\cdot(n+0.5)}{2\cdot L}\right] & (n=\frac{M_i}{4}-\frac{L}{2},\cdots,\frac{M_i}{4}+\frac{L}{2}-1) \\ & (n=3\cdot\frac{M_i}{4}-\frac{L}{2},\cdots,3\cdot\frac{M_i}{4}+\frac{L}{2}-1), \\ 1 & (n=\frac{M_i}{4}+\frac{L}{2},\cdots,3\cdot\frac{M_i}{4}-\frac{L}{2}-1). \end{cases} \quad (4)$$



(N_1=128, N_2=256, N_3=512,)
(M_1=256, M_2=512, M_3=1024)

(a) An Independent Window for Each Block Size



(b) Block-Size Sequence

Fig. 1. The MDCT Application to ATC-ABS

Figure 1 shows newly developed windows for the MDCT. Each block size has one specific window which is independent of the others. Windows (i), (ii), (iii) in Fig. 1 (a) have different length M_1, M_2 and M_3 , respectively. However, the shape of their shoulders or the transition regions are the same. A block size M_i is defined twice as large as corresponding DCT block size, N_i ; that is $M_i=2\cdot N_i$ (i=1,2,3). Figure 1. (b) illustrates an example of the resulting block-size sequence. As is clear from the figure, the substantial overlapping regions are all the same. Each window shape is not affected by the adjacent block size. Thus, the window for a block is uniquely defined once the block-size has been selected.

3. A New Bit Assignment Algorithm

The Zelinski-Noll quantizer assigns limited total quantization bit amount to each transformed coefficient according to its power, and quantizes each transformed coefficient with the assigned bits. When transformed coefficient power is distributed over broad band in the frequency domain, enough quantization bit amount is not assigned to each transformed coefficient. Figures 2 (a) and (c) show transformed coefficient power in the frequency domain, and (b) and (d) show quantization bit amount to be assigned to each transformed coefficient. Transformed coefficient power is distributed over broad band, for a cabasa as in Fig. 2 (c), while each transformed coefficient power is concentrated on a narrow band, for a piano as in Fig. 2 (a). Therefore, for a piano, the maximum quantization bits per sample (8-bit) is assigned to around the peak of transformed coefficient power as in Fig. 2 (b). However, as Fig. 2 (d) indicates, assigned bits to each coefficient for a cabasa is near the average quantization bits per sample (3-bit). If enough quantization bit amount is not assigned to each transformed coefficient, such as for a cabasa, coding

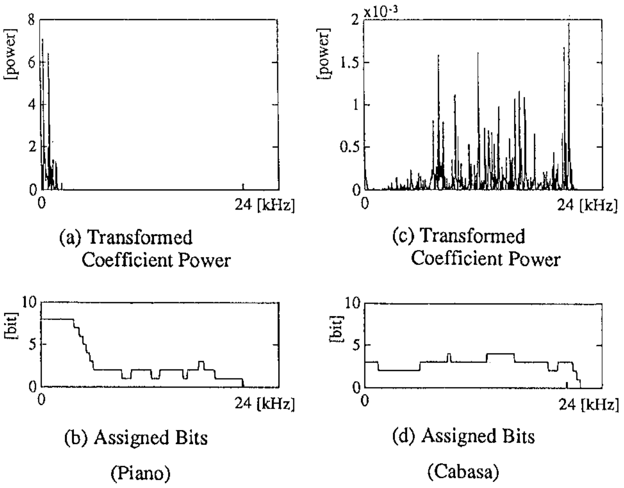


Fig. 2. Transformed Coefficient Power and Assigned Quantization Bit Amount

quality is deteriorated.

In order to cope with this difficulty, frequency component manupilation is carried out, only when transformed coefficient power is distributed in a wide range of frequency. This is because the target is HiFi audio signals. Therefore, degradation to reproduced audio waveforms should be minimized.

The employed bit assignment algorithm is based on the human auditory characteristic. In general, human auditory sensitivity is higher in low band. It means lower frequency components of the transformed coefficients are more important than those in the high band. Therefore, when transformed coefficient power is distributed over broad band, assigned bits to the number of the higher band should be reduced and squeezed bits are used for the lower band, to improve the auditory quality.

A block diagram of this algorithm is shown in Fig. 3, where the enclosed area by dotted line is newly developed to achieve the new bit assignment algorithm. Distribution of each transformed coefficient power is first calculated and then compared with a given threshold in DETECTOR. WEIGHTING weights to each transformed coefficient based on the auditory characteristic. If the distribution exceeds the threshold, SELECTOR selects the signal from WEIGHTING, otherwise, the input to SELECTOR is selected instead. The weighting function is determined that the weighted values in the lower band is larger than those in the higher band. When DETECTOR recognize that transformed coefficient power is distributed over broad band, weighted transformed coefficients are fed into ADAPTIVE BIT ASSIGNMENT through SELECTOR.

4. Simulation and Subjective Tests

The newly developed MDCT and the bit assignment approaches are introduced to ATC-ABS. This new ATC-ABS block diagram is shown in Fig. 4. It consists of

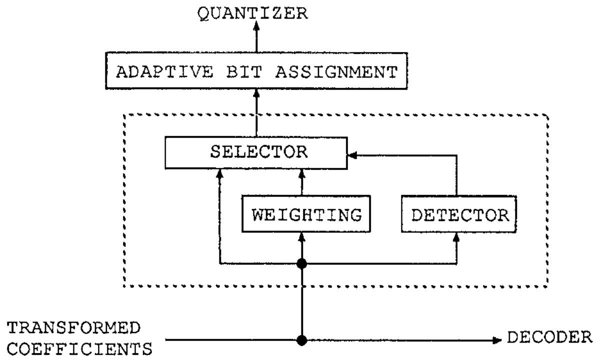


Fig. 3. A Block Diagram of the New Bit Assignment Algorithm

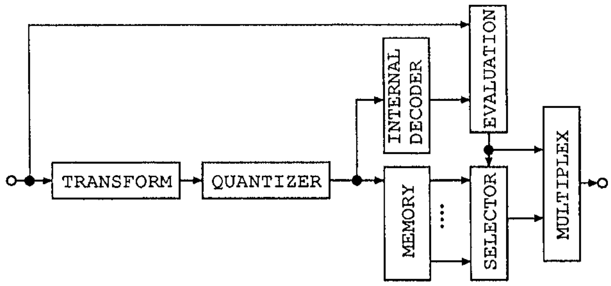


Fig. 4.A Block Diagram of the New ATC-ABS

TRANSFORM using the MDCT, QUANTIZER using the new bit assignment algorithm, INTERNAL DECODER for IMDCT, and EVALUATION for block-size selection. EVALUATION first calculates SNR using input signal and reproduced signal for each candidate block size, and then the block size with the maximum SNR is chosen one after another. This system compresses a 48kHz sampled HiFi audio signal into 128 kbit/s.

Reproduced signal by the new ATC-ABS is evaluated based on a paired comparison test with the original signal and the reproduced signal by conventional ATC-ABS. The test is executed by 20 people using loudspeakers in a listening room. Sources are castanets, a piano, a bassoon, and a cabasa, with the most of which, block boundary noise and quantization distortion noise is perceived by ATC-ABS.

Table 1. shows scores that each signal was perceived to be closer to the original signal in quality, of all paired comparison. The original signal and the reproduced signal by the new ATC-ABS got higher scores than the reproduced signal by ATC-ABS for a piano, a bassoon and a cabasa. For example, the original signal and the reproduced signal by the new ATC-ABS scored 46.8% and 45.2%, respectively, while the reproduced signal by the ATC-ABS scored 8.0%, for a piano. In the source of castanets, all of the three signals are scored around 30%. This result prove that there are almost same auditory quality because, for a castanets, noise is not perceived in reproduced signal even by ATC-ABS.

Table 1. Subjective Test Results [%]

	PIANO	BASSOON	CABASA	CASTANETS
ORIGINAL	46.8	43.2	54.8	39.2
ATC-ABS	8.0	8.8	0.8	34.6
NEW ATC-ABS	45.2	48.0	44.4	26.2

It is confirmed by these subjective tests that the new ATC-ABS using the MDCT and the new bit assignment algorithm provides coding quality comparable to the source signals.

5. Conclusion

The new ATC-ABS using the MDCT and the new adaptive bit assignment algorithm has been proposed. In order to apply the MDCT to ATC-ABS, a new window set, which is determined independent of the adjacent block sizes, is developed. The end shape of the window is defined so that the substantial overlapped portions from left and right are symmetric independent of the adjacent block sizes. When transformed coefficient power is distributed over broad band in the frequency domain, quantization bit amount is assigned to each transformed coefficient through weighting based on the auditory characteristic.

Distortions caused by Block operations are greatly reduced by MDCT introduction without degrading pre-echo suppression capability. A wide-band signal encoding capability is also improved by auditory weighting introduction in the bit assignment algorithm. The reproduced signals at the bit rate of 128 kbit/s are almost the same quality of the original signals.

Acknowledgment

The authors thank to Tastuo Ishiguro, a vice President of NEC corporation, and Botaro Hirosaki, the manager of Network Research Laboratory, for their encouragement and guidance on this work. The authors also thank Koji Okazaki, Fujitu Laboratories Ltd., for discussions.

References

[1] J. D. Johnston, "Perceptual Transform Coding of Wideband Stereo Signals," Proc. ICASSP'90, pp. 1993 - 1996, May 1989.

[2] Y. Mahieux et al., "Transform Coding of Audio Signals Using Correlation Between Successive Transform Block," Proc. ICASSP '89, pp. 2021 - 2024, May 1989.

[3] M. Iwadare, "A Robast 384KBIT/S Stereo Hifi Audio Codec for ISDN Application. ," Proc. IEEE GCOM '89 pp. 54. 4. 1. - 54. 4. 5. Nov. 1989.

[4] L. Chiarigrione, "Documents of Kurihama / Osaka Meeting, Part C," Oct. 1989.

[5] T. Nishitani, "Coding Scheme for Audio Information based on Adaptive Block Size Assignment," Contribution to ISO-IEC / JTC1 / SC2 / WG8, Jul. 1989.

[6] A. Sugiyama et al., "Adaptive Transform Coding with an Adaptive Block Size (ATC-ABS)," Proc. ICASSP '90, pp. 1093 - 1096, Apl. 1990.

[7] P. Zelinski and P. Noll., "Adaptive Transform Coding of Speech Signals," IEEE Trans. ASSP - 25, pp. 299 - 309, Aug. 1977.

[8] P. Zelinski and P. Noll et al., "Approaches to Adaptive Transform Coding at Low Bit Rates," IEEE Trans. ASSP - 27, pp. 89 - 95, Feb. 1979.

[9] J. P. Princen et al., "Analysis / Synthesis Filter Bank Design Based on Time Domain Aliasing Cancellation," IEEE Trans. ASSP - 34, pp. 1153 - 1161, Oct. 1986.