HIGH-QUALITY AUDIO-CODING AT LESS THAN 64 KBIT/S BY USING TRANSFORM-DOMAIN WEIGHTED INTERLEAVE VECTOR QUANTIZATION (TWINVQ)

Naoki Iwakami, Takehiro Moriya, and Satoshi Miki

NTT Human Interface Labs., 3-9-11 Midori-cho, Musashino-shi, Tokyo, 180 Japan

ABSTRACT

A new audio-coding method is proposed. This method is called transform-domain weighted interleave vector quantization (TwinVQ) and achieves high-quality reproduction at less than 64 kbit/s. The method is a transform coding using modified discrete cosine transform (MDCT). There are three novel techniques in this method: flattening of the MDCT coefficients by the spectrum of linear predictive coding (LPC) coefficients; interframe backward prediction for flattening the MDCT coefficients; and weighted interleave vector quantization. Subjective evaluation tests showed that the quality of the reproduction of TwinVQ exceeded that of an MPEG Layer II coder at the same bitrate.

1. INTRODUCTION

Digital audio applications, such as broadcasting and storage, require audio signal compression, in particular to less than 64 kbit/s, which is the bitrate of N-ISDN. This will enable us to send and store music easily at low cost.

A popular technique for audio-coding is to apply scalar quantization to sub-band or transformed signals. The quantizers are optimized for music signals, which have dynamic frequency characteristics, by using adaptive bit allocation. MPEG Layer I, II, and III coders [1], various adaptive transform codings (ATC) [2]-[4], and adaptive spectral perceptual entropy coding (ASPEC) [5] are based on this technique.

Using vector quantization as an alternative to scalar quantization [6] is more efficient, but the computational complexity is usually greater. TC-WVQ [7] is a coding method that applies vector quantization to the discrete cosine transform (DCT) coefficients, but its computational complexity is reasonable because it uses an interleaving technique. This method is designed for speech signals, so some problems occur when it is used to code audio signals. Low pitch signals generate interframe noise due to the non-contiguous quantization error at the frame boundary of inverse DCT. Moreover, the pitch filter in the coder does not work effectively for a signal with irregular harmonics, because it is designed for a speech signal, which has regular harmonics. So we propose an audio-coding version of TC-WVQ, called transform-domain weighted interleave vector quantization (TwinVQ).

In this paper, first we introduce the architecture of TwinVQ. Then, we describe three major techniques used in the method. Finally, we report results of subjective evaluation tests.

2. BASIC STRUCTURE

Figure 1 shows the basic structure of TwinVQ. In this method, the input signal is first transformed into frequency domain coefficients. To avoid interframe noise, we use MDCT [8] as an alternative to DCT. Then the coefficients are flattened at the following two stages, and flattened coefficients are normalized by their power. The coder quantizes the normalized flattened MDCT coefficients, power normalization factor, and spectral envelope parameters.

There are three novel techniques in this method: flattening of the MDCT coefficients by the LPC spectrum; interframe backward prediction for flattening the MDCT coefficients; and weighted interleave vector quantization of flattened MDCT coefficients.

3. FLATTENING OF MDCT COEFFICIENTS BY LPC SPECTRUM

At the first stage of TwinVQ, input samples are divided into two paths. In one path they are transformed into frequency domain coefficients using MDCT. In the other path, input samples are analyzed using LPC, and LPC coefficients are produced. Next, the spectrum of the LPC coefficients is calculated. Finally, each MDCT coefficient is divided by its corresponding spectrum element and first-stage residual coefficients are produced.

The frequency characteristics of the LPC synthesis filter are an envelope of the input-signal spectrum, so the MDCT coefficients are flattened by using this procedure.

0-7803-2431-5/95 $4.00

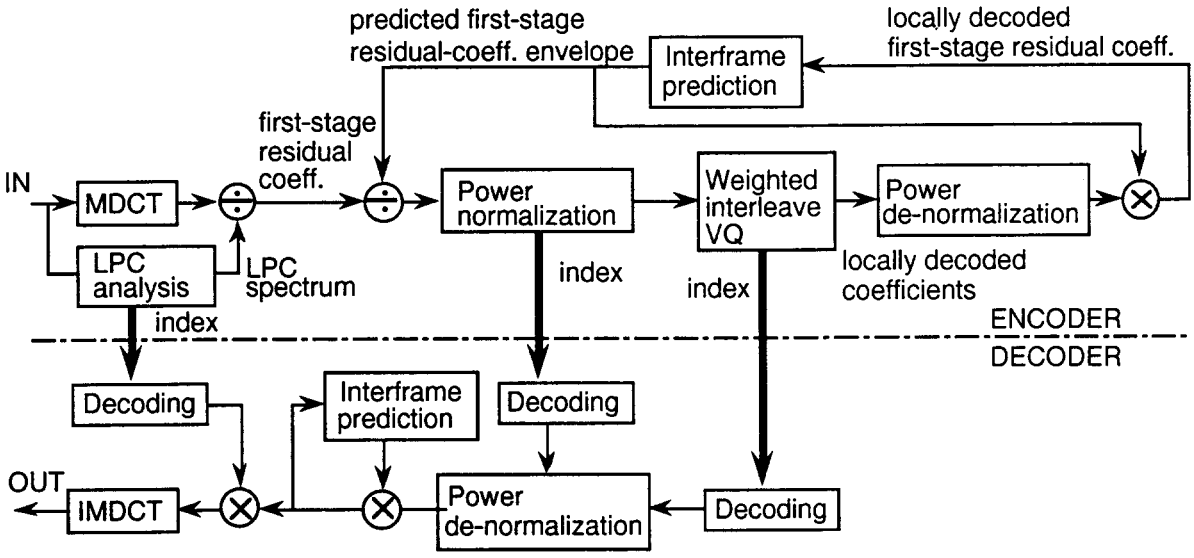

Fig. 1. Basic structure of TwinVQ.

The main advantage of this flattening method is that fewer bits are required to normalize the frequency characteristics of input signals. And the main advantage of flattening in the frequency domain instead of time domain filtering is that it can be applied to MDCT, which has time domain aliasing, so interframe noise is less than when using DCT.

4. INTERFRAME BACKWARD PREDICTION

An LPC spectrum does not have fine structures such as harmonics, which are present in the input-signal spectrum. So there remains a fine structure in the first-stage residual coefficients. Such coefficients have a large dynamic range, so vector quantization is less efficient.

To reduce this fine structure, TwinVQ also uses another flattening method. In this method, the first-stage residual coefficients are divided by a fine-structure envelope predicted from previous frames, and second-stage residual coefficients are produced.

Figure 2 shows the procedure of interframe backward prediction. A first-stage residual-coefficients envelope vector is first calculated. Next, the envelope vector is successively delayed by one frame. Then, the whole of the delayed envelope vector is combined using predictive coefficients ω_i. The predictive coefficients are determined so that the combiner produces the nearest vector of the current-frame envelope vector of the first-stage residual-coefficients, and the determined values are used in the next frame.

Exactly the same procedure can be done in the decoder as in the encoder, because this procedure needs only the decoded first-stage residual coefficients. So there is no need to transmit any side information for this procedure.

The segmental SNRs of audio signals synthesized by TwinVQ with and without interframe prediction are shown in Fig. 3. This procedure improves the segmental SNR by 1 - 10 dB.

5. WEIGHTED INTERLEAVE VECTOR QUANTIZATION

TwinVQ applies weighted interleave vector quantization, which is shown in Fig. 4, to second-stage residual coefficients. Inputs of the quantizer are the second-stage residual coefficient vector and the LPC spectrum vector. The second-stage residual coefficient vector is interleaved and the rearranged vector is split into subvectors. The LPC spectrum vector is divided into subvectors in the same way. Each residual coefficient subvector is quantized using a weighted distortion measure. The corresponding LPC spectrum subvector is used as a weight. For perceptual control, elements of the LPC spectrum subvector are applied using a non-linear function. Weighted vector quantization fits the dynamic frequency characteristics of audio signals as effectively as adaptive bit allocation, even if we use a constant bit allocation. Moreover, there is little loss in performance due to division into subvectors since the averaged power of the envelope for each subvector can be made nearly constant by means of interleaving. This scheme thus gives

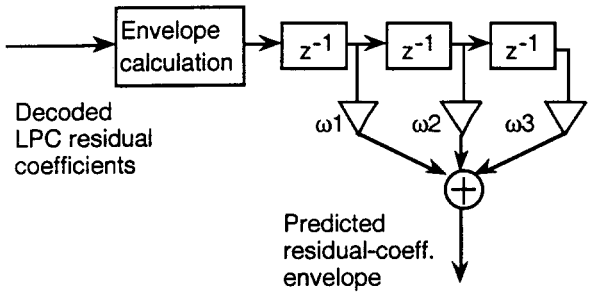

Fig. 2. Interframe backward prediction.

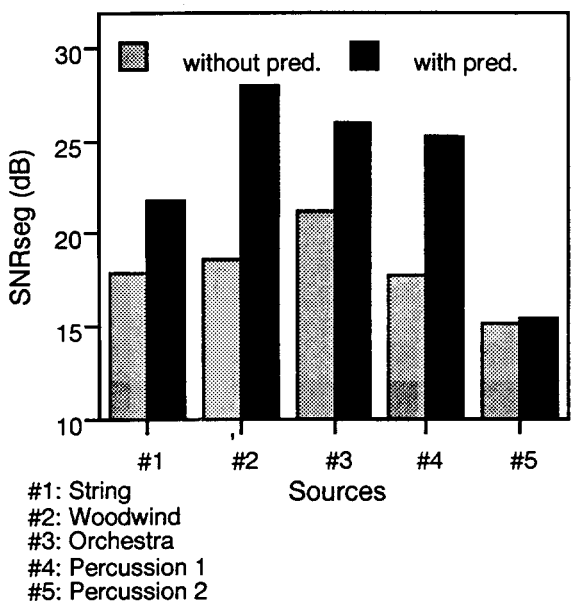

Fig. 3. Effect of interframe prediction.

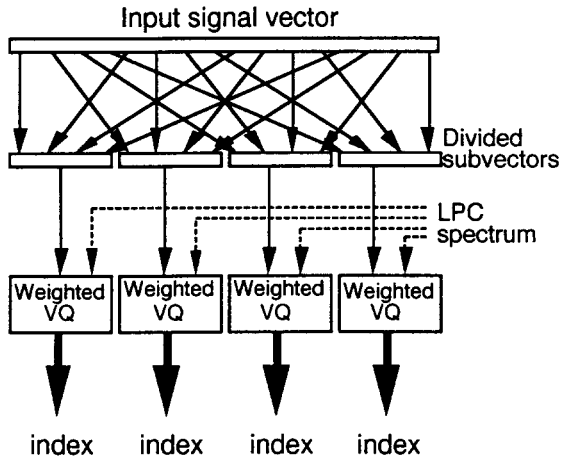

Fig. 4. Weighted interleave vector quantization.

flexibility to frequency characteristics with reasonable computational complexity.

6. SUBJECTIVE EVALUATION TESTS

Subjective evaluation tests were done using twenty-four listeners. Ten kinds of monaural sources were presented. Each source was sampled at two sampling rates: 48 kHz and 32 kHz. The 48-kHz sources were coded at 64 kbit/s by using TwinVQ. The 32-kHz sources were coded at 32, 16, and 8 kbit/s. The analysis frame length of the MDCT was 1024; the LPC order was 20; and the interframe prediction order was 4. Table 1 lists the number of bits per frame allocated to each code. The sources were also coded by using MPEG Layer II in 48-kHz / 64-kbps and 32-kHz / 32-kbps modes, and they were added to the test set for comparison.

Three signals were presented in one evaluation set. The order was one of the following patterns: either the original signal for reference / coded signal / original signal or the original signal for reference / original signal / coded signal. Listeners were not told which pattern would be presented. The listeners were told to give a score of five to the second or third signal if they thought it was the original, and to evaluate the distortion of the other signal using the following subjective grades:

5. Imperceptible
4. Perceptible, but not annoying
3. Slightly annoying
2. Annoying
1. Very annoying

After the tests, a normalized score was calculated for each evaluation set by subtracting the score of the original signal from the score of the coded signal. Then, the mean normalized score for each coding mode was calculated. Figure 5 shows the mean normalized scores. In the 48-kHz / 64-kbps mode, the score of TwinVQ slightly exceeds that of MPEG Layer II. In the 32-kHz / 32-kbps mode, the score of TwinVQ obviously exceeds that of MPEG Layer II. From these results, it can be said that TwinVQ can compress audio signals with high quality particularly at lower bitrates.

7. CONCLUSION

In this paper, we propose a new audio-coding method called TwinVQ. In this method, the input signal is transformed using MDCT, and the MDCT coefficients are flattened at two stages. For flattening using LPC in the first stage, less side information has to be transferred, and spectrum analysis of LPC coefficients enables this method

to be applied to MDCT, which has time domain aliasing. Interframe backward prediction, at the second flattening stage, reduces the fine structure that LPC-spectrum flattening cannot reduce; it improves segmental SNRs of reconstructed signals by 1-10 dB. The flattened coefficients are applied to the weighted interleave vector quantization. This scheme gives flexibility to frequency characteristics of input signal with reasonable computational complexity. According to the subjective evaluation tests, TwinVQ achieves higher quality reproduction than MPEG Layer II particularly at lower bitrates.

8. ACKNOWLEDGEMENTS

The authors would like to thank Dr. Nobuhiro Kitawaki, Director of the Speech and Acoustics Laboratory, and Takao Kaneko, Leader of the Speech Information Processing Group, for their helpful guidance in this research.

REFERENCES

[1] "Coding of Moving Pictures and Associated Audio for Digital Storage Media up to about 1.5 Mbit/s," ISO/IEC 11172, 1993.

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

[3] J. M. Tribolet and R. E. Crochiere, "Frequency Domain Coding of Speech Signals," IEEE Trans. ASSP, vol. ASSP-27, pp. 512-530, 1979.

[4] B. Mazor and W. Pearlman, "An Optimal Transform Trellis Code with Application of Speech," IEEE Trans. Commun., vol. COM-33, pp.1109-1116, 1985

[5] K. Brandenburg, J. Herre, and J. D. Johnston, "ASPEC: Adaptive Spectral Perceptual entropy coding of High Quality Music Signals," 90th AES Convention, Preprint, 1991.

[6] I. M. Trancoso and B. S. Atal, "Efficient Procedures for Finding the Optimum Innovation in Stochastic Coders," Proc. ICASSP '86, pp. 2375-2378, 1986.

[7] T. Moriya and M. Honda, "Transform Coding of Speech Using a Weighted Vector Quantizer," IEEE Trans. JSAC, vol. JSAC-6, pp. 425-631, 1988.

[8] J. Princen, A. Johnson, and A. Bradley, "Adaptive transform coding incorporating time domain aliasing cancellation," Speech Commun., vol. 6, pp. 299-308, 1987.

[9] T. Moriya, "Two-Channel Conjugate Vector Quantizer for Noisy Channel Speech Coding," IEEE JSAC, vol. JSAC 10, pp. 866-874, 1992.

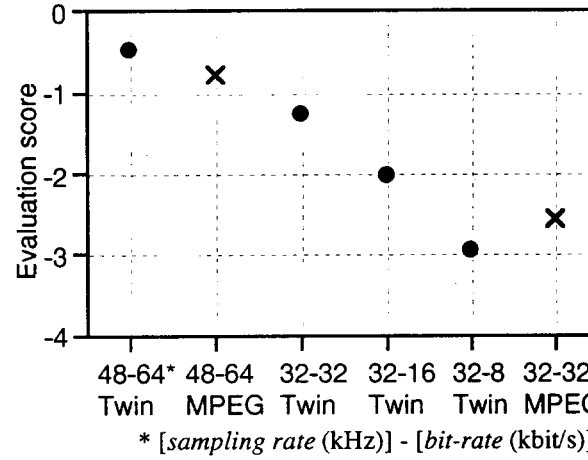

Fig. 5. Results of subjective evaluation tests.

Table 1. Bit allocations of codes per frame (1024 samples).

	48-kHz / 64-kbps	32-kHz / 32-kbps	32-kHz / 16-kbps	32-kHz / 8-kbps
LSP	28 bits	20 bits	20 bits	20 bits
Weighted VQ (for each subvector *)	8 + 8 bits	8 + 8 bits or 8 + 7 bits	8 + 8 bits or 8 + 7 bits	8 + 8 bits or 8 + 7 bits
Weighted VQ (total)	1328 bits {8+8 bits x 83}	995 bits {(8+8) bits x 50 + (8+7) bits x 13}	483 bits {(8+8) bits x 18 + (8+7) bits x 13}	227 bits {(8+8) bits x 2 + (8+7) bits x 13}
Total	1364 bits	1023 bits	511 bits	255 bits

* Using two-channel conjugate vector quantization [9].