Lossless Audio Coding Using Code-Excited Linear Prediction with Embedded Entropy Coding

Yongmin Li and Cheung-Fat Chan

Department of Electronic Engineering, City University of Hong Kong, Hong Kong
Email: jocelyn8090@yahoo.com.cn, itcfchan@cityu.edu.hk

Abstract

A hybrid lossy/lossless audio coding is presented in this paper. The lossy coding part is based on code excitation approach where the excitation gain is adapted in a sample-by-sample fashion. Two short term predictors are used to cope with rapid time-varying nature of audio signals. The error signal between the input and the synthetic signal is sent to the decoder after arithmetic coding to achieve lossless compression. The excitation codebook is searched by an M-L tree search strategy with joint minimum error energy and minimum code length as search criteria. The advantage of the proposed coder is its low decoding complexity due to simple code excitation structure, and the compression performance achieved is comparable to other advanced lossless coders.

I. Introduction

Recently, with the rapid proliferation of large capacity storage devices and high-speed internet connections, more lossless applications are demanded for storage, transmission and processing of high-fidelity audio without distortion. For examples, music archival systems, further processing of professional audio and high-end consumer applications such as home theater systems all demand lossless audio reproduction.

State-of-the-art technologies towards lossless coding are generally categorized into two approaches, one is a decorrelation-plus-entropy coding approach and the other is a lossy-plus-entropy coding approach. In decorrelation-plus-entropy coding approach, the audio signal is firstly gone through a decorrelation process, after which the error energy is much smaller than input energy. The decorrelated signal with smaller entropy can then be efficiently encoded by an entropy coder such as arithmetic coder [1] [2]. In lossy-plus-entropy coding approach, a lossy representation of the signal is formed in the encoder, and then the difference between the reconstructed lossy signal and the original signal is sent to the decoder after entropy coding to achieve lossless coding [3]. If the channel bandwidth is not wide enough to transmit both the lossy coding parameters and the residual information to the decoder, the decoder can recover a reasonable good quality of audio signal from the received lossy coding parameters only. But in decorrelation-plus-entropy coding approach, the decorrelator parameters alone cannot be used to recover a lower quality version of original audio.

In this paper, a scalable lossless coding algorithm employs lossy-plus-entropy coding approach is proposed. In addition, a nonscalable coding scheme is introduced in section Ⅲ which is gradually developed from the scalable coding scheme.

II. Scalable Coding Scheme

The proposed lossy/lossless hybrid coding algorithm employs a joint optimization scheme, which is different from separated optimization scheme in traditional lossless coding. In the lossy coding part, a code-excited linear predictive strategy with embedded entropy coding algorithm to code the audio waveform as close to the original as possible is utilized, and then the residual signal is further encoded by an arithmetic coder. The code length after arithmetic coding will be used to assist the optimization process for the lossy coding part so as to achieve better optimization for both coding parts.

The lossy coding part is based on code excitation approach similar to code-excited linear predictive (CELP) coder for coding speech signals [4]. However, they are fundamentally different in many aspects.

1. The dimension of the stochastic codebook in CELP coding of speech is relatively high, but that in the proposed scheme for audio coding is no more than 4.

2. And in CELP, the excitation gain is unchanged within the sub-frame period; however it is adapted on a sample-by-sample (SbS) basis in the proposed scheme.

3. A synthesis filter is used in CELP to reconstruct the synthetic signal. In the proposed system, a SbS adaptive short term predictor which has sufficiently high order to capture short-time correlation could achieve the best prediction performance, such as least mean squares (LMS) algorithm used in MPEG-4 (RLS-LMS) scheme [5]. However, the encoding complexity will be extremely high because the coefficient adaptation process has to be done within the codebook search loop. Consequently, in this work, a hybrid approach using two predictors is applied. A high-order block adaptive synthesis filter is applied to follow the general waveform. And a SbS adaptive low-order predictor is used to capture the steady correlation at a short term in more detail, in which the adaptation is performed within the codebook search loop.

4. An adaptive codebook is used to model the periodicity of voiced speech in CELP for coding speech signals. It is tried to be added in the proposed scheme, but this single-tap pitch adaptive codebook is not effective enough for coding general audio signals such as music which may have multiple tones. Since the pitch lag and pitch gain of the adaptive codebook should be sent to decoder which will increase the overhead information, and the coding complexity is increased, the adaptive codebook is not utilized in the proposed scheme.

In the second coding stage, the error signal is encoded by an arithmetic coder. The arithmetic-encoded error signal and the side information from the lossy coding part are sent to the decoder to achieve lossless coding. The proposed coding structure is shown in Fig.1 and detailed explanations for each coding block are given as follows:

A. Excitation Codebook

The excitation codebook is constructed from a collection of stochastic codewords. If the codebook has B stochastic vectors with D-dimension, the average bit rate required to encode the codebook index is $(\log_2 B)/D$ bps. The stochastic codewords are obtained by an iterative training procedure from a large collection of real audio signals. An initial codebook is constructed from uniformly distributed random codewords. The training vectors are partitioned into quantization clusters corresponding to the codewords by going through the encoding process. Then, each codeword is re-optimized from its cluster of training vectors according to the principle of minimum coding error energy. This is done by minimizing the error energy ε_m within the same cluster C_m with respect to $d(n)$.

$$\varepsilon_m = \sum_{x(n),y(n)\in C_m} \sum_{n=1}^{D} [x(n)-y(n)]^2, \tag{1}$$

where $x(n)$ is the input signal and $y(n)$ is the synthetic signal obtained as

$$y(n) = [\hat{s}(n) + d(n)\Delta(n)] * h(n), \tag{2}$$

where $d(n)$ is the excitation signal, $\Delta(n)$ is the excitation gain, $\hat{s}(n)$ is the predicted sample of $s(n)$, and $h(n)$ is the impulse response of the block adaptive filter. Assuming each element of the excitation vector can be optimized independently and by setting $\frac{\partial \varepsilon_m}{\partial d(n)} = 0$, and then $d(n)$ can be re-estimated during codebook training as

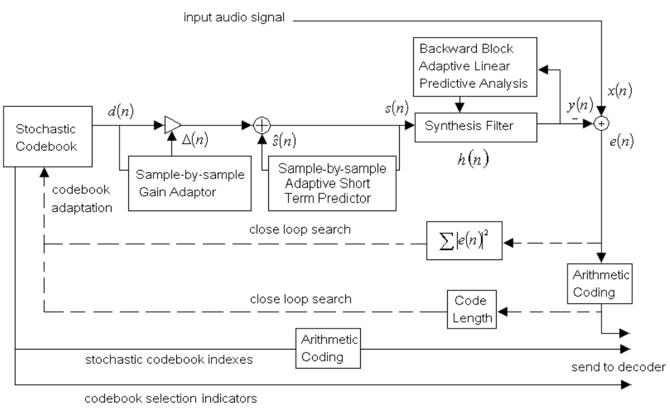

Fig.1 Proposed scalable lossless audio encoder

$$\bar{d}(n) = \frac{\sum_{x(n),\hat{s}(n),\Delta(n),h(n)\in C_m} [x(n)-\hat{s}(n)*h(n)]\,[\Delta(n)*h(n)]}{\sum_{\Delta(n),h(n)\in C_m} [\Delta(n)*h(n)]^2}, \quad n = 1,2 \ldots D. \tag{3}$$

It is found that about 1.5 dB improvement in SNR can be achieved with a trained codebook.

B. Sample-by-sample Gain Adaptation

Conventional CELP coding algorithm employs a fixed excitation gain which is quantized and sent to the decoder for each sub-frame. For speech coding using large sub-frame size the overhead is small, but it is not feasible for audio coding because the sub-frame size must be typically less than 5 samples, in order to get accurate matching of fast time-varying audio signals. In this paper, a new adaptation mechanism based on a concept similar to scale factor adaptation of ADPCM [6] coder is proposed to adapt the excitation gain in a SbS fashion directly from the codeword. In this method, the gain is adapted as:

$$\Delta(n+1) = (\beta + |d(n)|)^{\gamma}\, \Delta(n), \tag{4}$$

where β is a threshold value chosen such that if $|d(n)| > 1-\beta$, the gain is increased, otherwise it will be decreased. The rate of increasing or decreasing is an exponential function of time and γ is a modification factor to further control the adaptation rate. This adaptation formula is very simple but still capable of following rapid increases in signal magnitude during musical attacks and also allows smooth decaying during musical releases. The main reason for using such a simple adaptation formula is that this gain adaptation has to be performed for each codeword during close-loop codebook search, and the search complexity will be extremely high if a complicated rule is used.

Intuitively, the threshold value β is considered to be at the midpoint of the signal magnitude range, i.e. 0.5, however, after an optimization procedure similar to codebook training is applied, an optimum value of $\beta = 0.727$ is obtained. Furthermore, from experiment $\gamma = 1$ is found to be a good compromise for its lowest adaptation complexity.

C. Sample-by-sample Adaptive Short Term Predictor

In order to remove more inter-sample correlation of the error signal, an order-2 SbS adaptive predictor is applied within the coding path. The output from the order-2 predictor is

$$\hat{s}(n) = a_1(n)s(n-1) + a_2(n)s(n-2), \tag{5}$$

where $a_1(n)$ and $a_2(n)$ are the predictor coefficients. The autocorrelations are computed recursively by using an exponential window.

D. Block Adaptive Linear Predictive Filter

To further remove short term correlation over a longer period, a high-order block adaptive predictor is necessary to predict the general shape of input waveform. By accompanying the SbS adaptive predictor, the order of the block predictor does not need to be very high. Here, an order-10 lattice synthesis filter is used. The block size is 2048 samples and block shift is 64 samples. The reflection coefficients are computed from the past synthesis samples by using an asymmetric window. They are fixed during the codebook searching process so the complexity involved is relatively low as compare to the SbS adaptation.

E. Excitation Codebook Search criterion

The search process comprises two stages: codewords search and adaptive codebook selection.

E.1. Codewords Search

During encoding, an M-L tree search technique is performed to select the "best" codeword from the codebook. In order to reduce the search complexity, the codebook size and the codevector dimension must be kept small. A number of codebooks with different combinations of dimensions and sizes have been designed. In this work, the M-L tree search range is 64 samples which are considered as a sub-block. Each sub-block is equally divided into 4 frames. In each frame, an M-L tree search technique is applied across several sub-frames. The idea is to keep the best M codewords in each search sub-frame, and with the search depth of L sub-frames, the best two paths are kept for search in the next sub-frame while other paths are released. As a compromise of search complexity and performance, M=2 and L=4 are used. Then the M-L tree search is extended into the second frame. So and so, at the end of one sub-block, 16 best paths are figure out and one of them is chosen as the best path to encode the whole sub-block.

Since the proposed coding system is a cascade of a lossy part and a lossless part, two search criteria are possible. For the lossy part, minimum error energy can be used as search criterion. For lossless compression, the ultimate performance measure is only the encoding rate because there is no audio quality issue; therefore, code length after entropy coding can be used as a search criterion. However, since the code length is dependent on the entropy of the source and the entropy is related to the statistical distribution which is a very long term measure, entropy alone can not be used as search criterion because the search is done locally with short interval; therefore, we propose a combined minimum error energy and minimum entropy search criterion. For each frame, the search criterion is minimum error energy. After dealing with a sub-block, the residual signal as a result of each coding path is arithmetic-encoded, the codewords in a path that result in the smallest code length is fetched as the best codewords. Since codebook search using minimum error energy measure already guarantees residual signal with small energy, the combination of energy/entropy search can achieve better compression performance. The M-L tree search strategy is shown in Fig. 2.

E.2. Adaptive Codebook Selection

Since the encoding bit rate is the sum of the average bit rate after arithmetic coding of the residual and the bit rate for coding the codebook indexes, an adaptive choice of codebooks is more beneficial so as to cater for various signal statistics. A total of eleven possible combinations of codebooks have been tried in turn for every sub-block separately which have encoding rates ranging from 0.75 bps to 4.5 bps. And the adopted codebook is the one that has the smallest encoding bit rate. From the observed statistic, four of the eleven codebooks are more frequently selected. In order to save the overhead, only four codebooks are set as the candidate codebooks.

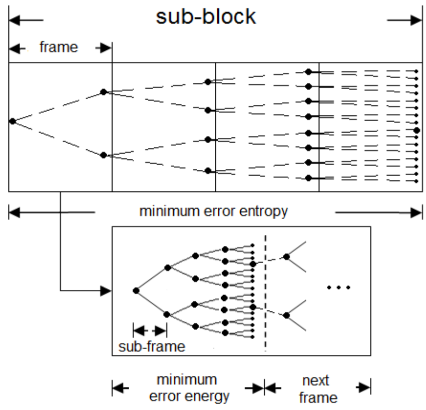

Fig.2 M-L tree search strategy

F. Arithmetic Coding

An adaptive arithmetic coder is applied to code the error signal losslessly. The symbol probabilities of the error signal are computed using a sliding window approach and updated for each input symbol. The probability table represents the statistics of the past 500 samples; this provides the best performance for accurately matching to the local statistics of the input signal. The arithmetic coder is implemented using an incremental shift out algorithm with low complexity.

III. NONSCALABLE CODING SCHEME

Because the prediction coefficients of the block adaptive predictor are calculated from the past synthesis samples, they will be affected by quantization error. And the output of the block adaptive predictor is predicted by the previous synthesis samples, the prediction error will be fed back to the predictor. Therefore we propose to use the past input samples to replace the past synthesis samples for block adaptive prediction in order to achieve better compression performance. However since the past input signal is used for encoding, the lossy coding mode is not applicable in this case because a lossless signal is required in the decoder to do the same prediction process as in the encoder. Therefore this is a nonscalable lossless coding scheme. Because the SbS adaptive predictor is also affected by the prediction error, a clean input of it is more beneficial. The block diagram of the nonscalable coding scheme is shown in Fig. 3.

In the nonscalable coding scheme, if no codebook is used, the error signal is the residual signal $e_1(n)$ shown in Fig. 3. Observing from experiments, the compression performance of the scheme without codebook is better than the scheme with codebook in the nonscalable coding mode.

IV. EVALUATION RESULTS

The performance of proposed coding system is evaluated by encoding various audio pieces composed of Pop music, jazz music, classical music and audio files come from MPEG test

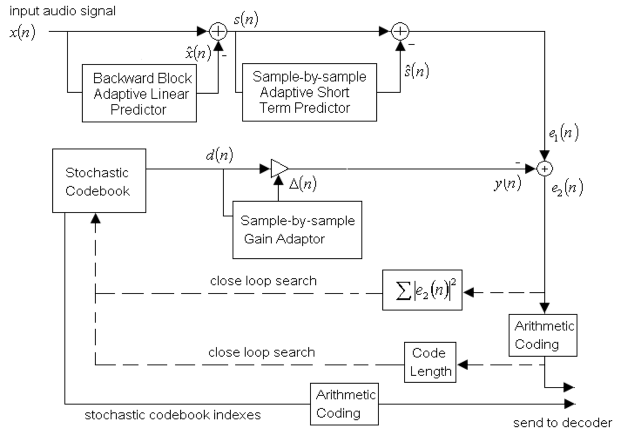

Fig.3 Proposed nonscalable lossless audio encoder

set. All audio signals are sampled at 44.1 kHz with 16 bit A/D conversion. For lossy coding part, three performance indicators are used, SNR and entropy of the residual signal, and bit rate for lossy coding. Table I lists the performance indicators obtained from encoding four categories of audio pieces by the proposed scalable coding algorithm.

Table I. Performance measure for lossy coding part

	MPEG test set	jazz music	Pop music	classical music
Entropy(bit)	5.074	6.037	7.644	8.092
SNR(dB)	35.553	47.545	28.929	33.574
Bit rate for lossy coding (bps)	3.184	3.373	3.673	3.458

For lossless coding part, two popular lossless audio coders FLAC [7] and MPEG-4 ALS [5], are used for comparison. Table Ⅱ shows the average compression ratios achieved by FLAC, MPEG-4 ALS and our proposed coding algorithm. In Table Ⅱ, MPEG-4 ALS (LPC) stands for ALS forward prediction scheme, and MPEG-4 ALS (RLS-LMS) stands for ALS cascaded recursive least square-least mean square prediction scheme.

Table Ⅱ. Compression ratios obtained by FLAC -2, MPEG-4 ALS and our proposed coding scheme

	MPEG test set	jazz music	Pop music	classical music	average
FLAC -2	2.039	1.675	1.399	1.382	1.624
MPEG-4 ALS (LPC)	2.311	1.786	1.510	1.444	1.763
MPEG-4 ALS (RLS-LMS)	2.594	1.833	1.543	1.469	1.860
Proposed scalable scheme	2.071	1.689	1.431	1.381	1.643
Proposed nonscalable scheme	2.318	1.793	1.516	1.448	1.769

Under evaluation with compression ratio, the scalable scheme performs slightly better than FLAC -2 and it performs 11.65% worse than ALS (RLS-LMS). Nonetheless the scalable scheme has advantage of low decoding complexity as well as its lossy coding functionalities. The nonscalable scheme has comparable compression performance with ALS (LPC). In conclusion, the proposed coder comprises a scalable scheme which has lossy coding functionalities and a nonscalable scheme whose compression performance is comparable with ALS (LPC).

The decoding time of the proposed scalable decoder is 51.57% less than that of ALS (RLS-LMS) decoder as shown in Table Ⅲ, while the proposed decoder has not been optimized for speed. All the tests were conducted on a 2.4GHz Intel Core (TM) 2 Quad CPU PC, with 3.25G of memory.

Table Ⅲ. Decoding time (in seconds) by MPEG-4 ALS (RLS-LMS) and our scalable coding scheme

	MPEG test set	Pop music	classical music	jazz music
Audio length (s)	142	15.67	30.07	30.47
Proposed scalable coding scheme (s)	35.12	4.74	8.95	8.27
MPEG-4 ALS (RLS-LMS) (s)	75.91	8.55	16.45	16.94
Decoding time saving (%)	53.73	44.56	45.59	51.18

V. CONCLUSIONS

A hybrid lossy/lossless audio coding algorithm is proposed. The algorithm is a cascade of a lossy coding part and a lossless coding part to achieve scalable coding. The lossy coding part is based on a modified code-excitation approach where the excitation gain is adapted in a SbS fashion. A close-loop codebook search strategy using a combined minimum error energy and minimum code length criterion is applied to guarantee good coding performance. In addition, a nonscalable scheme is proposed to achieve better compression performance. The decoding complexity is relatively low.

REFERENCES

[1] T. Liebchen. Tech.Univ. Berlin, Berlin,Germany, LPAC, version 0.99h, http://www-ft.ee.tu-erlin.de/~liebchen/lpac.html.

[2] Eric Knapen, Derk Reefman, Erwin Janssen, and Fons Bruekers, "Lossless compression of one-bit audio," *JAES*, vol. 52, pp. 2, Feb. 2004.

[3] M. Hans and R. W. Schafer, "Lossless compression of digital audio," *IEEE Signal Processing Magazine*, pp. 21-32, Jul. 2001.

[4] M. R. Schroeder and B. S. Atal, "Code-excited linear prediction (CELP): high-quality speech at very low bit rates," in *Proceedings of ICASSP*, vol. 10, pp. 937-940, 1985.

[5] "MPEG-4 Audio Lossless Coding (ALS)," Internet: http://www.nue.tu-berlin.de/menue/forschung/projekte/parameter/en/ Nov. 13, 2009 [Feb. 24, 2011].

[6] P. Cummiskey, N. S. Jayant, and J. L. Flanagan, "Adaptive quantilation in differential PCM coding of speech," *Bell Syst. Tech. J.*, vol. 52, pp. 1105-1118, Sep. 1973.

[7] "free lossless audio codec (FLAC)," Internet: http://flac.sourceforge.net/ Dec. 07, 2009 [Feb. 24, 2011].