CASCADED RLS-LMS PREDICTION IN MPEG-4 LOSSLESS AUDIO CODING

H. Huang, S. Rahardja, X. Lin, R. Yu

Institute for Infocomm Research
21 Heng Mui Keng Terrace
Singapore 119613

P. Franti

University of Joensuu
Department of Computer Science
Finland FI-80101

ABSTRACT

A new MPEG-4 standard for lossless audio coding is going to be published in 2006. This coming international standard consists of two parts: the transform-domain Scalable to Lossless coding (SLS), and the time-domain Audio Lossless coding (ALS). In ALS, linear prediction is used to compress the dynamic ranges of the input audio signal. The prediction residual is coded by an entropy coder with either Rice code or arithmetic code. There are two prediction modes in ALS: linear predictive coding (LPC) and cascaded RLS-LMS. As the developer of the RLS-LMS prediction, we present this technology in this paper. In RLS-LMS prediction, the input audio samples go through the cascaded DPCM, RLS, and LMS predictors, whose output predictions are linearly combined to generate a prediction for the current input sample. Through MPEG testings, it has been found that ALS with RLS-LMS prediction provides the best lossless compression ratio compared with SLS, ALS with LPC, and several non-MPEG codecs.

1. INTRODUCTION

Lossless audio compression, as the name suggests, converts digital audio files in raw PCM format into compressed files with smaller size. The original audio file can be perfectly recovered from the compressed file. With the ever-increasing hard disk capacity and widespread broadband access, lossless audio compression has become more and more popular nowadays. It can be used to archive music CDs into hard disk, share your favorite songs with friends over internet, or simply play an album in portable music players.

So far various lossless audio compression codecs have been developed by individuals, interested groups, and commercial companies. For example, APE (by Monkey's Audio) [1] and FLAC (Free Lossless Audio Coding) [2] are popular for file sharing on internet. OptimFrog [3] and LA (lossless Audio) [4] provide leading compression ratios. Companies like Apple, Microsoft, and Real Networks also have their own lossless coding formats. All these codecs are proprietary and lack wide industrial adoption. Responding to industrial requirements for a standardized lossless audio coding scheme, MPEG has issued a call for proposal in 2002 [5]. Various parties responded and eventually two schemes, SLS and ALS, were selected by MPEG as the up-coming international standard for lossless audio coding.

In SLS [6], the input audio frames are transformed losslessly by using the integer modified discrete cosine transform (IntMDCT) to generate the transform coefficients. In the core Advance Audio Coding (AAC) encoder, these coefficients are scaled, quantized and coded with perceptual weighting so that the noise introduced is best masked by the masking threshold of the human auditory system. The resultant core layer bitstream thus constitutes the minimum rate/quality unit of the final lossless bitstream. For optimal coding efficiency, an error-mapping procedure is employed to remove the information that has been coded in the core layer from the IntMDCT coefficients, which are subsequently noiselessly coded by bitplane Golomb code (BPGC) to form lossless bitstream. SLS provides fine-granular quality/rate scalability. It is especially useful for audio streaming where the bitstream can be dynamically truncated according to the available bandwidth.

In ALS [7], linear prediction is performed on the input audio samples to generate a prediction residual which has smaller dynamic range than the input signal. The prediction residual is coded by an entropy coder with either Rice code or block Gilbert-Moore code (BGMC). For linear prediction, ALS has two modes: LPC and RLS-LMS. In the former mode, input audio samples are divided into blocks and analyzed by LPC. The LPC coefficients are coded together with the prediction residual into the lossless bitstream. To achieve optimum coding efficiency, the order of LPC as well as the coding block length have to be properly tuned to the input signal.

In the RLS-LMS mode, the predictor uses past samples to predict the value of the current sample. Only the residual is coded and sent to the decoder. There is no need to code the predictor coefficients, as the decoder uses a 'mirror' predictor which maintains exactly the same states as that in encoder. The predictor consists of cascaded DPCM, RLS, and LMS prediction stages. A linear combiner generates the prediction of the current sample by summing up the output predictions of the prediction stages (after proper weighting). The RLS-LMS predictor uses standard RLS and normalized LMS (NLMS) algorithms to adjust the predictor coefficients in the cascade stages. The coefficients of the linear combiner are updated by the sign-sign LMS algorithm.

The rest of the paper is organized as follows: Section 2 explains the concept of ALS with RLS-LMS prediction, and what is the cascaded predictor structure. Section 3 lists the adaptive algorithms for updating the predictor coefficients. The following Section 4 introduces joint-stereo prediction which explores the cross-channel correlations between stereo audio channels. After that, Section 5 provides results of lossless compression. Finally, a conclusion is given in Section 6.

2. ALS WITH CASCADED RLS-LMS PREDICTION

Figure 1 shows the structure of ALS encoder and decoder with the RLS-LMS predictor. In the encoder, a prediction of the current sample is generated by the RLS-LMS predictor by using the past samples. The prediction is subtracted from the current sample to generate the residual, which is subsequently coded by an entropy coder with either Rice code or BGMC to form the ALS bitstream.

In the decoder, a reverse process is performed. The ALS bitstream is decoded by the entropy decoder into the prediction residual, which is added to the prediction from the RLS-LMS predictor to re-generate the original audio signal. The predictor in the decoder is a mirror of that in the encoder. They run synchronously so that at all times identical internal states, coefficients, and outputs are maintained. The synchronization is ensured by the control bits embedded in the ALS bitstream.

The ALS encoder and decoder are illustrated by the following equations:

$$\hat{x}(n) = \sum_{m=1}^{M} a_m x(n-m) \tag{1}$$

$$e(n) = x(n) - \hat{x}(n) \tag{2}$$

$$x(n) = \hat{x}(n) + e(n) \tag{3}$$

In the predictor (1), the prediction $\hat{x}(n)$ is generated for the current input sample $x(n)$ by linearly combining the past M samples of the audio sequence x. a_m are the coefficients of the Mth-order predictor. In encoder (2), the prediction residual $e(n)$ is obtained by subtracting the prediction from the current sample. The decoding process is shown by (3) where the residual is added to the prediction to recover the original input sample.

Figure 2 shows the structure of the cascaded RLS-LMS predictor. The input samples pass through a series of simple adaptive prediction stages. The prediction residual of one stage serves as the input signal to the next stage. The output predictions of the stages are weighted and added together by a

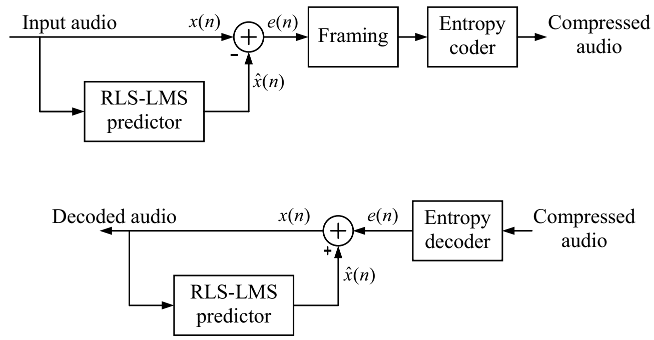

Fig. 1. Structure of ALS encoder (top) and decoder (bottom) with RLS-LMS predictor

linear combiner to generate the final prediction of the current input sample according to

$$\hat{x}(n) = \sum_{k=1}^{K} c_k y_k(n) \tag{4}$$

where c_k are linear combiner coefficients, $K = 5$ is the total number of prediction stages, and $y_k(n)$ are output predictions of the stages. For the k-th stage, we have

$$y_k(n) = \sum_{m=1}^{M_k} a_{k,m} e_{k-1}(n-m) \tag{5}$$

where M_k is the predictor order, $a_{k,m}$ are the coefficients, and $e_{k-1}(n-m)$ are past samples of the input sequence, which is also the prediction residual from the previous stage. The prediction residual of the k-th stage is given by

$$e_k(n) = e_{k-1}(n) - y_k(n) \tag{6}$$

3. ADAPTIVE COEFFICIENTS UPDATING

The RLS-LMS predictor consists of five adaptive prediction stages and an adaptive linear combiner. Except the first DPCM stage which uses a single fixed coefficient, the coefficients of the remaining stages and the combiner are adjusted adaptively. In the second stage the standard RLS algorithm is applied. The NLMS algorithm is used in the remaining three stages. The linear combiner adopts the sign-sign LMS to for coefficients updating.

3.1. DPCM

Fixed coefficient of 1 is used in the first-order DPCM predictor

$$y_1(n) = x(n-1) \tag{7}$$

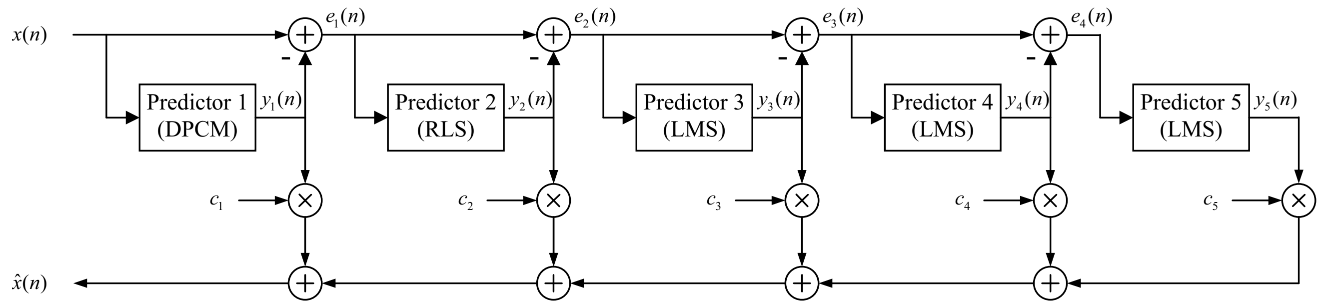

Fig. 2. Structure of the cascaded RLS-LMS predictor

3.2. RLS

The coefficients of the predictor are adjusted by the RLS algorithm as follows

$$\mathbf{v}(n) = \mathbf{P}(n-1)\mathbf{e}_1(n) \quad (8)$$

$$\mathbf{k}(n) = \frac{\lambda^{-1}\mathbf{v}(n)}{1+\lambda^{-1}\mathbf{e}_1^T(n)\mathbf{v}(n)} \quad (9)$$

$$e_2(n) = e_1(n) - \mathbf{a}_2^T(n-1)\mathbf{e}_1(n) \quad (10)$$

$$\mathbf{a}_2(n) = \mathbf{a}_2(n-1) + \mathbf{k}(n)e_2(n) \quad (11)$$

$$\mathbf{P}(n) = \lambda^{-1}\mathbf{P}(n-1) - \lambda^{-1}\mathbf{k}(n)\mathbf{v}^T(n) \quad (12)$$

where

$$\mathbf{e}_1(n) = [e_1(n-1), e_1(n-2), \ldots, e_1(n-M_2)]^T$$

$$\mathbf{a}_2(n) = [a_{2,1}(n), e_{2,2}(n), \ldots, e_{2,M_2}(n)]^T$$

λ is the forgetting factor which determines the convergence rate of the RLS algorithm. $\mathbf{P}(n)$ is the inverse auto-correlation matrix which is initialized to

$$\mathbf{P}(0) = \delta\mathbf{I}$$

where δ is a small number and $\mathbf{I}$ is the identity matrix.

3.3. NLMS

The coefficients of the LMS predictor in the k-th stage are updated according to

$$\mathbf{a}_k(n) = \mathbf{a}_k(n-1) + \frac{e_k\mathbf{e}_{k-1}(n)}{1+\mu_k\mathbf{e}_{k-1}^T(n)\mathbf{e}_{k-1}(n)} \quad (13)$$

where

$$\mathbf{e}_{k-1}(n) = [e_{k-1}(n-1), e_{k-1}(n-2), \ldots, e_{k-1}(n-M_k)]^T$$

$$\mathbf{a}_k(n) = [a_{k,1}(n), e_{k,2}(n), \ldots, e_{k,M_k}(n)]^T$$

The convergence rate of the LMS filter is controlled by the stepsize μ_k.

3.4. Sign-Sign LMS

The linear combiner uses the sign-sign LMS algorithm for the coefficients updating

$$\mathbf{c}(n) = \mathbf{c}(n-1) + \alpha\,\mathrm{sgn}[\mathbf{y}(n)]\,\mathrm{sgn}[e(n)] \quad (14)$$

where

$$\mathbf{c}(n) = [c_1(n), c_2(n), \ldots, c_K(n)]^T$$

$$\mathbf{y}(n) = [y_1(n), y_2(n), \ldots, y_K(n)]^T$$

and the sgn function is defined as

$$\mathrm{sgn}[r] = \begin{cases} 1 & r > 0 \\ 0 & r = 0 \\ -1 & r < 0 \end{cases} \quad (15)$$

and α is the stepsize.

4. JOINT-STEREO PREDICTION

For stereo audio signals, the samples in one channel usually have some correlations to those in the other channel. The RLS-LMS predictor exploits this cross-channel correlation by using joint-stereo prediction. In joint-stereo prediction, samples from both channels are used in the prediction. For example, to generate the prediction $y_L(n)$ for the current left channel input sample $x_L(n)$, past samples from both channels are used. This is shown by the following equation:

$$y_L(n) = \sum_{m=1}^{M_a} a_{L,m}x_L(n-m) + \sum_{m=1}^{M_b} b_{L,m}x_R(n-m) \quad (16)$$

where $a_{L,m}$ and $b_{L,m}$ are coefficients of the intra- and inter-channel predictors for the left channel, respectively. M_a and M_b are orders of the predictors.

Table 1. Lossless Compression Ratios

MPEG-4 Test Set	SLS RM8	ALS RM16 LPC	ALS RM16 RLS-LMS	OptimFrog v4.509	Monkey's Audio v3.99
48kHz/16bit	2.201	2.238	2.278	**2.283**	2.250
48kHz/24bit	1.584	1.595	**1.609**	**1.609**	1.573
96kHz/24bit	2.134	2.163	**2.190**	2.187	2.090
192kHz/24bit	2.626	2.658	**2.681**	2.677	2.570
Total	2.123	2.148	**2.172**	2.171	2.097

Similarly, for the right channel, we have

$$y_R(n) = \sum_{m=1}^{M_a} a_{R,m} x_R(n-m) + \sum_{m=0}^{M_b-1} b_{R,m} x_L(n-m) \quad (17)$$

The decoder processes samples in the order of $LRLRLR\cdots$. The current left channel sample $x_L(n)$ is always decoded before its right channel counterpart $x_R(n)$. Therefore, $x_L(n)$ can be used in the prediction of $x_R(n)$. This explains why the second summation in (17) starts from 0. In joint-stereo prediction, the RLS algorithm is used to adapt the coefficients of the intra- and inter-channel predictors.

5. LOSSLESS COMPRESSION RESULTS

The most important performance index for lossless compression is the compression ratio, defined according to

$$\text{compression ratio} = \frac{\text{original filesize}}{\text{compressed filesize}} \quad (18)$$

To evaluate lossless audio compression schemes, MPEG use a testing set which contains stereo audio sequences of 15 types of music. The testing sequences have sampling rates of 48, 96, and 192 kHz and resolutions of 16 and 24 bits/sample.

Table 1 lists the compression ratios for five lossless audio compression codecs, which are:

- SLS RM8 (RM8 = Reference Model 8)
- ALS RM16 (with LPC)
- ALS RM16 (with RLSLMS)
- OptimFrog v.4.509 ('-bestnew')
- Monkey's Audio v.3.99 ('-insane')

The RLS-LMS predictor uses the following parameter sets: RLS ($M_a = M_b = 8$, $\lambda = 0.99$), LMS ($M = 512, 128, 16$, $\mu = 10, 20, 30$), Combiner ($K = 5$, $\alpha = 0.001$).

In Table 1, the best compression ratio in each row is shown in boldface. Clearly, ALS with RLS-LMS prediction provides the best overall compression ratio. OptimFrog does slightly better for the 48kHz/16bit subset. In the test, ALS RM16 with RLS-LMS prediction runs at about $0.2\times$ realtime on a Pentium-4 2.4GHz PC, which is the slowest among all the codecs. This is due to the high computational complexity of the RLS filter and the long LMS filters. One way to speed up the RLS-LMS predictor is optimizing encoder parameters, especially the adaptation stepsizes, so that orders of the predictors can be reduced. Optimizing the fixed-point implementation could also help.

6. CONCLUSION

This paper presented the cascaded RLS-LMS prediction in MPEG-4 lossless audio coding. Cascaded prediction has become the state-of-the-art technology in lossless audio compression. MPEG-4 ALS with RLS-LMS prediction has demonstrated top-notch compression ratios. Since the standard specifies only the decoder, there remain rooms for further encoder optimization, especially for improving the encoding speed.

7. REFERENCES

[1] Monkey's Audio, [Online]. Available: http://www.monkeysaudio.com/

[2] Free Lossless Audio Codec, [Online]. Available: http://flac.sourceforge.net/

[3] OptimFROG, [Online]. Available: http://www.losslessaudio.org/

[4] Lossless Audio, [Online]. Available: http://www.lossless-audio.com/

[5] "Call for Proposals on MPEG-4 Lossless Audio Coding," ISO/IEC JTC1/SC29/WG11 (MPEG) N5040, Awaji, Japan, Dec. 2002.

[6] R. Geiger, R. Yu,"MPEG-4 Scalable to Lossless Audio Coding," *117th AES Convention*, PrePrint 6183, San Francisco, Oct. 2004.

[7] T. Liebchen,"MPEG-4 Lossless Audio Coding," *116th AES Convention*, PrePrint 6047, Berlin, May 2004.