A Study of Using Least Squares Method in MPEG-4 Audio Lossless Coding

Shingchern D. You and Chau-Jia Wang
Department of Computer Science and Information Engineering,
National Taipei University of Technology, Taipei, Taiwan

Abstract—**This paper studies using least squares in the MPEG-4 audio lossless coding. The results show that it has a higher compression ratio, but it works better for large block size and moderate prediction orders.**

I. INTRODUCTION

Lossless audio coding has received wide attention recently. Unlike perceptual coding, such as MP-3 [1], lossless coding does not introduce any distortion, regardless whether perceptually noticeable or not. For preserving precious master soundtracks, it is more advantageous to use lossless coding than lossy (perceptual) coding. For this reason, MPEG-4 audio also includes the audio lossless coding (ALS) [2] tool for such applications.

The ALS tool in MPEG-4 audio supports a maximum of 65536 channels and 32 bits of resolution for sample rate up to 192 ks/s. It also provides random-access capability. Due to its lossless nature, in addition to audio signals, this tool may also be used for compressing medical signals or seismic signals.

For the lossless coding, the compression ratio is a critical issue. In this paper, we study the possibility of using the method of least squares (LS) to improve the compression performance. To be complete, we also address the relative probability that LS may fails for different conditions.

II. ALS ENCODING FLOW

The encoding flow of ALS is depicted in Fig. 1. The input samples are partitioned into frames, where the frame length is not standardized. In the experiments, we use a length of 20480 samples. Within each frame, six levels of blocks are available, from level 0 to level 5. For level 0, one block occupies one frame. For level 5, one block has a length of only 1/32 frame. It is not allowed to arbitrarily mix blocks with different levels in a frame. Some restrictions apply [2].

The second step of encoding is (short term) prediction. It is a linear-prediction-based operation to reduce the encoding bits. After linear prediction, the coefficients and prediction residual signals are encoded. In this step, the predictor coefficients are typically obtained by solving Yule-Worker equation with Levinson-Durbin (LD) algorithm [3]. With the LD algorithm, the obtained coefficients are in the form of PARCOR (partial correlation) coefficients. It has been proved that the absolute values of PARCOR coefficients are always less than one. Therefore, the ALS standard does not accept any value greater than one during entropy coding of PARCOR coefficients.

The next encoding step is a long-term prediction to extract any periodic signals out of the residual. If more than one channel is present, the next step is a joint-channel coding module for reducing inter-cannel redundancy. Then, the residual signals are encoded either by Rice code or BGMC (Block Gilbert-Moore Code). In the experiments, we use BGMC to encode residual signals for its higher compression ratio.

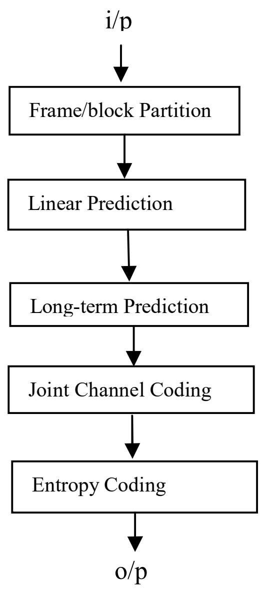

Fig. 1. The encoding flow of MPEG-4 audio lossless coding.

III. LEAST SQUARES METHOD

The Yule-Walker equation used in ALS is optimal if the signal to be predicted is from a wide-sense stationary random process. However, music signal is deterministic. Therefore, we have to estimate the autocorrelation function for the music. Considering these factors, we believe that using Yule-Walker equation does not guarantee to produce lowest residual energy.

For deterministic signals, there is another method, known as least squares (LS) [3], to find the coefficients of the linear predictor based on minimizing the squares of residual signals. The method is outlined below. Let $x[n]$ be the input signal, p be the order of the predictor, a_k be the coefficients of the predictor, and N be the length of the signal, then the error signal $v[n]$ is obtained as

$$v[n] = x[n] - \hat{x}[n] = \sum_{k=0}^{p} a_k x[n-k],\ a_0 = 1, p \le n < N \qquad (1)$$

By summing all error samples, we get the energy of total error

$$V = \frac{1}{N-p}\sum_{n=p}^{N-1} v^2[n] \quad (2)$$

By differentiating V with respect to all a_k, we can get p simultaneous equations. The coefficients of a_k can then be obtained. Since ALS only encodes coefficients of linear predictor in PARCOR form, a conversion is needed for a_k. Such a conversion can be found in [4].

As described previously, the Yule-Walker equation may not be optimal in the ALS encoding. Thus, we expect that LS method yields better prediction. However, the PARCOR coefficients converted from LS coefficients do not guarantee to have absolute values of less than one. Therefore, if the converted PARCOR coefficient is greater than one, this value (and LS method) should not be used, or extremely large residual signal may appear.

IV. EXPERIMENTS AND RESULTS

A. Comparison of residual bits

The first experiment compares the used bits for encoding residual signals by LD and LS methods. Since we use the same orders for these two methods, they use the same number of bits to encode PARCOR coefficients. The results for block level zero are shown in Fig. 2. Other block levels have similar results. In producing Fig. 2, if LS produces any PARCOR coefficient whose value is greater than one for a particular prediction order, the number of residual bits associated with this order is not used in the figure. A smooth line is used to link the missing points.

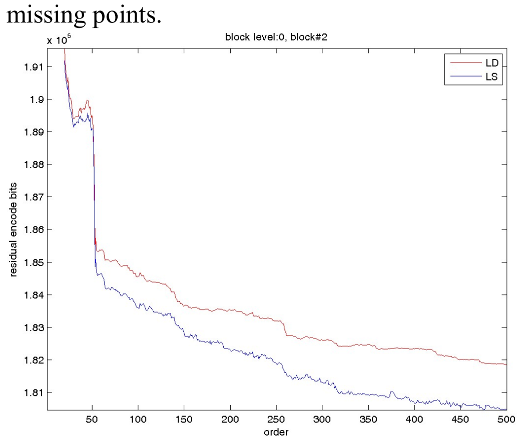

Fig. 2. The bits to encode residual signal for both LD and LS methods. The upper line is LD and the lower line is LS.

The results show that LS actually uses fewer bits in encoding residual signal, especially in higher orders. Therefore, LS is better than LD in ALS.

B. Possibility of PARCOR coefficients greater than one in different block levels

The possibility that PARCOR coefficients are greater than one is always a problem of LS method. To quantitatively assess the chance, we encode four different pieces of music with different block levels and calculate the number of blocks that has this problem. The results are given in Table I.

TABLE I
PERCENTAGE OF PARCOR COEFFICIENTS GREATER THAN ONE

Block level	Music 1	Music 2	Music 3	Music 4
0	2.4 %	0.0 %	0.0 %	0.0 %
1	6.4 %	5.3 %	0.7 %	19.2 %
2	15.7 %	15.2 %	9.6 %	20.8 %
3	35.8 %	37.8 %	27.7 %	30.1 %
4	66.7 %	64.8 %	61.3 %	53.6 %
5	79.7 %	76.8 %	78.2 %	72.6 %

It can be seen that as the block level increases, the possibility that PARCOR coefficients are greater than one also increases. Since block level is inversely proportional to the block length, the results implies that LS is more suitable for longer block.

C. Possibility of PARCOR greater than one for different orders

The order of the predictor also has an influence on the possibility that PARCOR coefficients are greater than one. The results, as shown in Fig. 3, show that higher-order predictors have a much higher tendency to have this problem.

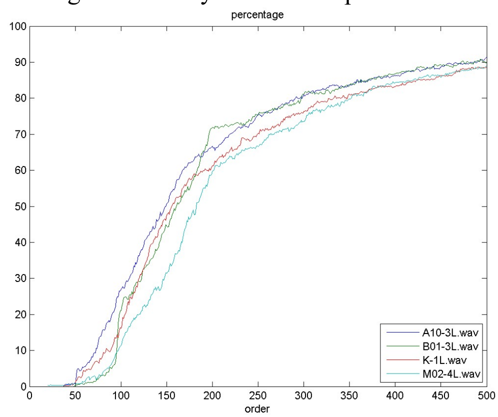

Fig. 3. Percentage of PARCOR coefficients greater than one versus predictor order. The unit in the vertical axis is percent.

V. CONCLUSIONS

This paper studies the possibility of using LS method in ALS coding. The results show that LS method can provide higher compression ratio. However, this approach is suitable for long block size and moderate predictor orders. For short blocks and/or very high prediction order, LS should not be used.

REFERENCE

[1] ISO/IEC, Information technology - Coding of moving pictures and associated audio for digital storage media at up to about 1.5 Mbit/s - Part 3: Audio, IS 11172-3, 1993.

[2] ISO/IEC, "ISO/IEC 14496-3:2005/Amd 2:2006, Audio Lossless Coding (ALS), new audio profiles and BSAC extensions, 2005.

[3] B. Porat, A Course in Digital Signal Processing, John Wiley & Sons, New York, NY, 1997.

[4] S. Haykin, Adaptive Filter Theory, pp. 205, 2nd Ed., Prentice Hall, Englewood Cliffs, NJ 1996.