

CCE3014 - Advanced Digital System Design

Practical Assignment - 2010

Dr. Ing. Carl James Debono

Department of Communications and Computer Engineering, University of Malta



Design and Implementation of a Lossless Compression Tool for Digital Audio

Objective:

To design, implement and test a lossless compression/decompression tool for digital audio. This technique is used to reduce the digital data that has to be transmitted over the communications channel. The implementation can be done on the FPGA or DSP boards.

Theory:

The paper attached proposes a solution to code digital audio signal in such a way that data is not lost during compression. The algorithm uses prediction techniques to aid compression.

Deliverables:

A report which includes

- i) A description of the implementation.
- ii) Commented programme listings.
- iii) The testing procedures adopted (with results) to ensure the proper operation.
- iv) Results summarising the performance of the algorithm.
- v) Comments regarding the efficiency of your implementation;
- vi) Comments on the techniques used and the results obtained. Possible improvements to your system should also be discussed.

A demonstration in the lab together with a short discussion is expected.

Only one report is required per group (of two students), with the names of both students included. It is expected that the work is shared equally between both members of the group.

Hard Deadline: 21st May 2010 at 12:00 – No Assignments will be accepted after the deadline. Plagiarism is unacceptable and can lead to disciplinary action.

Groups can only be of not more than two students.

Lossless Compression of Digital Audio Using Cascaded RLS-LMS Prediction

Rongshan Yu, *Member, IEEE*, and Chi Chung Ko, *Senior Member, IEEE*

Abstract—This paper proposes a cascaded RLS-LMS predictor for lossless audio coding. In this proposed predictor, a high-order LMS predictor is employed to model the ample tonal and harmonic components of the audio signal for optimal prediction gain performance. To solve the slow convergence problem of the LMS algorithm with colored inputs, a low-order RLS predictor is cascaded prior to the LMS predictor to remove the spectral tilt of the audio signal. This cascaded RLS-LMS structure effectively mitigates the slow convergence problem of the LMS algorithm and provides superior prediction gain performance compared with the conventional LMS predictor, resulting in a better overall compression performance.

Index Terms—Adaptive filter, linear prediction, LMS algorithm, lossless audio coding, RLS algorithm.

I. INTRODUCTION

DIGITAL audio waveforms possess considerable redundancy in their original PCM form and are generally compressible. Specifically, audio samples are statistical correlated and nonuniformly distributed. Most lossless audio compression algorithms [1] use a decoupled approach to exploit this redundancy, where a decorrelation pre-processor is first employed to remove the correlation between audio samples and its output is then entropy coded as a memoryless sequence. Although this is a compromised approach for building a system with tractable complexity (considering the large alphabet set of digital audio), results from [1] shows implicitly that the decoupled method may suffer only modest performance penalty when compared with some more sophisticated approaches, for example, the context model [2] based methods. A less popular method to build a lossless audio compression system is based on using linear transform [3], where the input audio sequence is transformed into the frequency domain and grouped into subsets with distinguished statistical property. Redundancy is then removed by coding each subset with a “best-fitted” entropy coder.

Generally speaking, as an effort to achieve the optimal compression performance the preprocessor in the decoupled method should be designed in such a way that the expected rate of the sequential entropy coder is minimized for audio inputs. Due to

the large alphabet set of the digital audio signal, a letter probabilities based entropy coder is generally avoided by many lossless audio coding systems to reduce the implementation complexity as well as the model cost [4] for better compression performance. Instead, a more viable approach is widely adopted where digital audio signals are described by some sorts of parametrical models, for example, Laplacian distribution [5] that leads not only to very simple statistical model description but also to a very simple structural prefix coding/decoding procedure, namely, the Golomb code [6]. For such a model, the optimal pre-processor is well approximated by one that generates output with the smallest variance. This problem is technically more tractable and has low complexity solutions that can be obtained readily from linear prediction [7]. In fact, this approach dominates the area of lossless compression of digital audio, with examples easily found from both the literature [1], [8], [9] and software in the Internet [10], [11].

It is well known that music signals, which are of most interest in lossless audio coding, contain abundant tonal and harmonic components whose energy can be effectively reduced by using a large order FIR predictor. Moreover, as audio signals are non-stationary, the predictor should be adaptive and possess good tracking capabilities so as to capture the local statistics of the signal. Numerous methods are available for designing and implementing an adaptive FIR predictor. Among these, the least mean square (LMS) [7] algorithm is very attractive as it provides an efficient, robust and, above all, low-complexity solution. However, the slow convergence behavior of the LMS algorithm for inputs with large eigenvalue spreads may lead to poor tracking performance and, consequently, poor prediction gain performance for audio signals that are generally highly-colored. The recursive least square (RLS) algorithm [7], on the other hand, is much less sensitive to the eigenvalue spread of the input. However, the computational complexity of RLS makes it impractical to be used as a large order predictor.

Historically, several methods have been proposed to improve the convergence performance of the LMS algorithm. Most of these are based on a two-step mechanism, where the input is first decorrelated by using a suitable transformation before applying the LMS adaptation. Examples of this approach can be found in frequency domain based FFT-LMS or DCT-LMS adaptive filters [7], which may improve convergence at the cost of greater misadjustment and complexity. Their time-domain counterparts include the use of an LMS adaptive pre-whitener [13], [14], and a cascade of independently adapting low-order LMS stages for which superior prediction gain performance for speech signal has been reported [15]. The problem of having many prediction stages, however, is that the overall combined predictor does not

Manuscript received July 27, 2002; revised February 10, 2003. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Gerald Schuller.

R. Yu is with the Institute for Infocomm Research, A*STAR, Singapore 110613 (e-mail: rsyu@i2r.a-star.edu.sg).

C. C. Ko is with the Department of Electrical and Computer Engineering, National University of Singapore, Singapore 117576 (e-mail: elekoc@nus.edu.sg).

Digital Object Identifier 10.1109/TSA.2003.818111

yield the optimal predictor. Cascaded prediction was also discussed in [16] for lossless coding of audio signals that are filtered by a psychoacoustic pre-filter [17].

In this paper, we propose a cascade RLS-LMS based adaptive linear prediction algorithm for lossless audio coding. The algorithm mitigates the slow convergence problem of the LMS predictor by cascading a small-order RLS predictor prior to it to remove the spectral tilt of the input audio signal. The spectral details of the RLS predictor output, with a reduced eigenvalue spread, is then modeled by the subsequent large-order LMS predictor for the best prediction gain performance. Since the size of the RLS predictor is small and much less than that of the subsequent LMS predictor, its computational burden is negligible and the complexity of the entire system is only slightly higher than that for a system with a conventional single stage LMS predictor. However, as will be shown, the cascaded predictor is able to achieve a superior prediction gain performance and, as a result, give a good lossless audio compression ratio performance when compared with the conventional LMS predictor.

II. RLS-LMS PREDICTOR

The structure of the proposed cascaded RLS-LMS predictor is shown in Fig. 1. In the first stage of the cascade, with $x(n)$ denoting the input to the predictor, the output $u(n)$ of the RLS predictor is given by

$$u(n) = x(n) - \mathbf{p}^T(n-1)\mathbf{x}(n) \quad (1)$$

where T denotes transposition and $\mathbf{x}(n) = [x(n-1), x(n-2), \dots, x(n-l_{RLS})]^T$. The filter tap weights $\mathbf{p}(n) = [p_1(n), p_2(n), \dots, p_{l_{RLS}}(n)]^T$ is updated using the RLS algorithm as follows:

$$\mathbf{k}(n) = \frac{\lambda^{-1}\mathbf{Q}(n-1)\mathbf{x}(n)}{1 + \lambda^{-1}\mathbf{x}^T(n)\mathbf{Q}(n-1)\mathbf{x}(n)} \quad (2)$$

$$\mathbf{p}(n) = \mathbf{p}(n-1) + \mathbf{k}(n)u(n) \quad (3)$$

and

$$\mathbf{Q}(n) = \lambda^{-1}\mathbf{Q}(n-1) - \lambda^{-1}\mathbf{k}(n)\mathbf{x}^T(n)\mathbf{Q}(n-1) \quad (4)$$

where λ is a positive value that is slightly smaller than 1. The algorithm is initialized by setting $\mathbf{Q}(-1) = \delta^{-1}\mathbf{I}$ where δ is a small positive constant and $\mathbf{p}(-1) = \mathbf{0}$.

In the second stage of the cascade the residual $u(n)$ is inputted to the LMS predictor to generate the prediction error $e(n)$ according to

$$e(n) = u(n) - \mathbf{w}^T(n)\mathbf{u}(n) \quad (5)$$

where $\mathbf{u}(n) = [u(n-1), u(n-2), \dots, u(n-l_{LMS})]^T$. The tap weights $\mathbf{w}(n) = [w_1(n-1), w_2(n-1), \dots, w_{l_{LMS}}(n-1)]^T$ are updated following the normalized LMS algorithm:

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \frac{\mu}{1 + \mathbf{u}^T(n)\mathbf{u}(n)}\mathbf{u}(n)e(n) \quad (6)$$

where $\mathbf{w}(-1)$ is initialized to a zero vector and $0 < \mu > 2$ controls the adaptation step sizes of the LMS algorithm. The

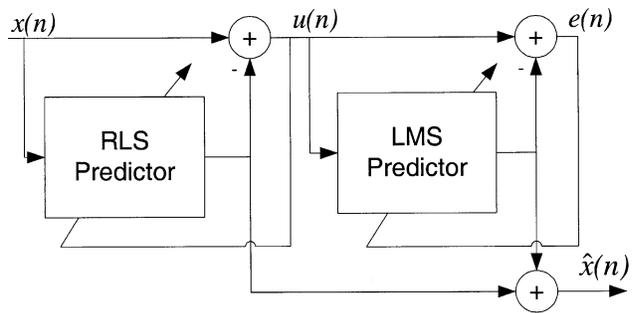


Fig. 1. Structure of the cascaded RLS-LMS predictor.

predicted value $\hat{x}(n)$ of the RLS-LMS system is given by the cascade of the RLS and LMS predictors or

$$\hat{x}(n) = \mathbf{w}^T(n)\mathbf{u}(n) + \mathbf{p}^T(n)\mathbf{x}(n). \quad (7)$$

In this paper, we will present some experimental results on the behavior and performance of the RLS-LMS predictor for the purpose of lossless audio coding. In order to examine the convergence behavior of the proposed algorithm under steady musical tones, we use the following synthesized musical tone

$$x(n) = \varepsilon(n) + \sum_{i=1}^p h(i)x(n-i) \quad (8)$$

as the input to the predictor in our experiments. The coefficients $h(1), h(2), \dots, h(p)$ define a p -order AR filter with frequency response given by

$$H(\omega) = \frac{1}{1 - \sum_{n=1}^p h(n)e^{-j\omega n}} \quad (9)$$

which controls the spectral shape, or the timbre of synthesized signal $x(n)$. The excitation signal

$$\varepsilon(n) = \sum_{i=1}^m A \cos(in\omega_0 + \varphi_i) + \gamma(n) \quad (10)$$

is obtained by summing up m harmonic sinusoids with fundamental frequency ω_0 , amplitudes A , and uniformly distributed random phases $\varphi_i \in [0, 2\pi)$, $i = 1, \dots, m$, together with a zero mean Gaussian noise $\gamma(n)$.

We define $SNR = 10 \log_{10}(mA^2/2\sigma^2)$ as the signal to noise ratio (SNR) of $\varepsilon(n)$, where $\sigma^2 = E\{\gamma^2(n)\}$ and $E\{\}$ is the expectation operator. In the experiment, the parameters are set as follows: $SNR_\varepsilon = 18$ dB, $\omega_0 = \pi/64$, $m = 32$, $\lambda = 0.9999$, $l_{RLS} = 4$ and $l_{LMS} = 60$. The effective filter length of the RLS-LMS predictor is thus $l_{RLS} + l_{LMS} = 64$ so that it provides sufficient number of poles to model the synthesized tone. We compare the performance of this predictor (RLS4-LMS60) with a normalized LMS predictor (NLMS64), an RLS predictor (RLS64) and a DCT-LMS predictor (DCT-LMS64). The orders of all the predictors are 64. The adaptation step sizes of all the LMS predictors are chosen in the same manner as in [15] to maximize the speed of convergence by increasing their values until the stability limit is reached.

Fig. 2 shows the MSE performance of the four algorithms. The curves are obtained from ensemble averaging over 500 runs.

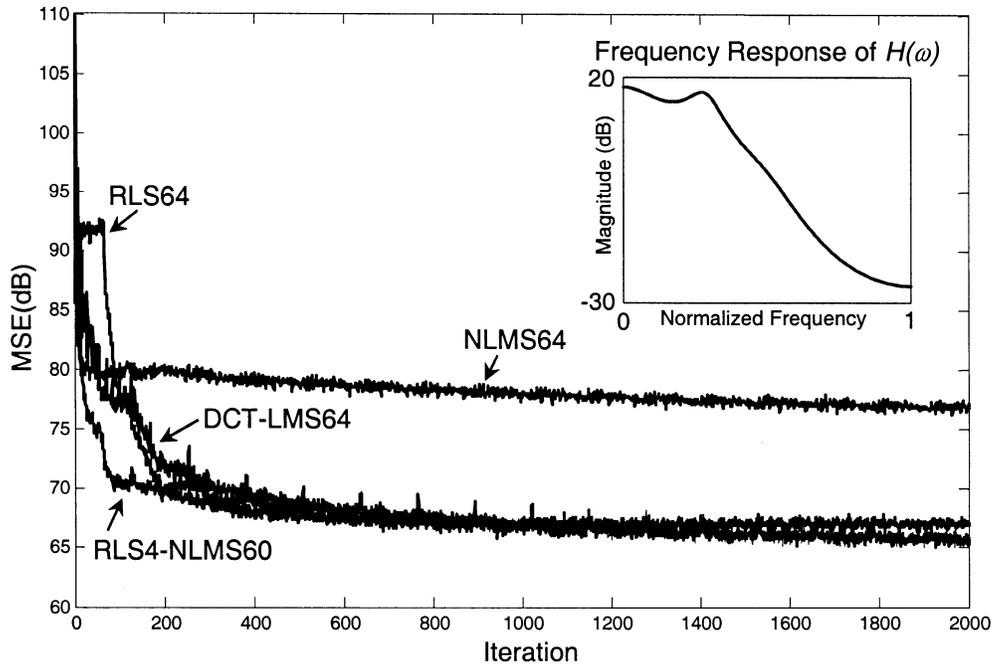


Fig. 2. MSE performance for RLS4-LMS60, NLMS64, DCT-LMS64 and RLS64 predictors. Frequency response (magnitude) of the generating filter is also shown in the subplot at the top-right corner.

Clearly, the NLMS and RLS-LMS predictors have the slowest and fastest convergence in this scenario where the input is highly colored. The DCT-LMS predictor has a slower initial convergence and a higher asymptotic MSE when compared with the RLS-LMS predictor. The full RLS algorithm has almost the same convergence speed as the RLS-LMS algorithm, but its initial convergence is slower. This is because the size of the full RLS predictor is much larger than that of the RLS predictor in the RLS-LMS algorithm, and the initial convergence is due mainly to the decoloration of the latter. Specifically, it takes about 128 iterations for the full RLS predictor to converge, compared with only eight iterations for the RLS predictor in the RLS-LMS algorithm.

To justify the effectiveness of the proposed RLS-LMS predictor, we further study the MSE performance of RLS4-LMS60 with its construction units: a 4-order RLS predictor (RLS4) and a 60-order LMS predictor (LMS60). In addition, an LMS-RLS predictor where a 60-order LMS predictor is placed before a 4-order RLS predictor (LMS60-RLS4) is also included in the comparison. The results are given in Fig. 3, which clearly shows that, separately, neither the RLS nor the LMS predictor is capable of delivering the same convergence speed as the cascaded one. Moreover, it can be observed from Fig. 3 that the LMS-RLS predictor has an inferior convergence performance, because in this configuration, the long LMS predictor will still be faced with a highly-colored signal and thus will converge very slowly.

The fast convergence behavior of the RLS-LMS predictor for the synthetic musical signal above suggests its potential in tracking real nonstationary audio signals and providing superior prediction gain performance. This is confirmed by extensive experiments where we used some real audio signals obtained from typical music CDs as the inputs to the predictor.

As an example, Fig. 4 shows the variation of the prediction gain for a short audio segment for the four predictors. Clearly, the NLMS predictor has the worst performance due to its slow convergence, while the performance of both the RLS-LMS and RLS predictors are much better and almost the same. In comparison, although the DCT-LMS predictor has good performance in the steady part of the audio segment, it does not perform as well at the beginning or during the transient part of the audio segment (sample 2000 to 4000). These results are consistent with the convergence behavior of the predictors as indicated in Fig. 2.

III. APPLICATION TO LOSSLESS AUDIO CODING

We build a simple experimental lossless audio codec to evaluate the practical performance of the proposed RLS-LMS predictor. The structure of this is given in Fig. 5. In this experimental codec, the complexity of the RLS-LMS predictor is further reduced by using the sign algorithm (SA) [18] to replace the LMS weight update procedure given by (6) as follows:

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \alpha \operatorname{sgn}[\mathbf{u}(n)] \operatorname{sgn}[e(n)] \quad (11)$$

where the “sgn” function is defined as

$$\operatorname{sgn}[r] = \begin{cases} 1; & r > 0, \\ 0; & r = 0, \\ -1; & r < 0, \end{cases} \quad (12)$$

$\alpha = 2^{-1}$ and $l \gg 1$ is an integer so that the sign algorithm can be implemented by using only add and bit shift operations [19]. In the experiments, l is set to 11 for best compression performance. It is found in our experiments that this modification introduces only a small performance penalty while it reduces

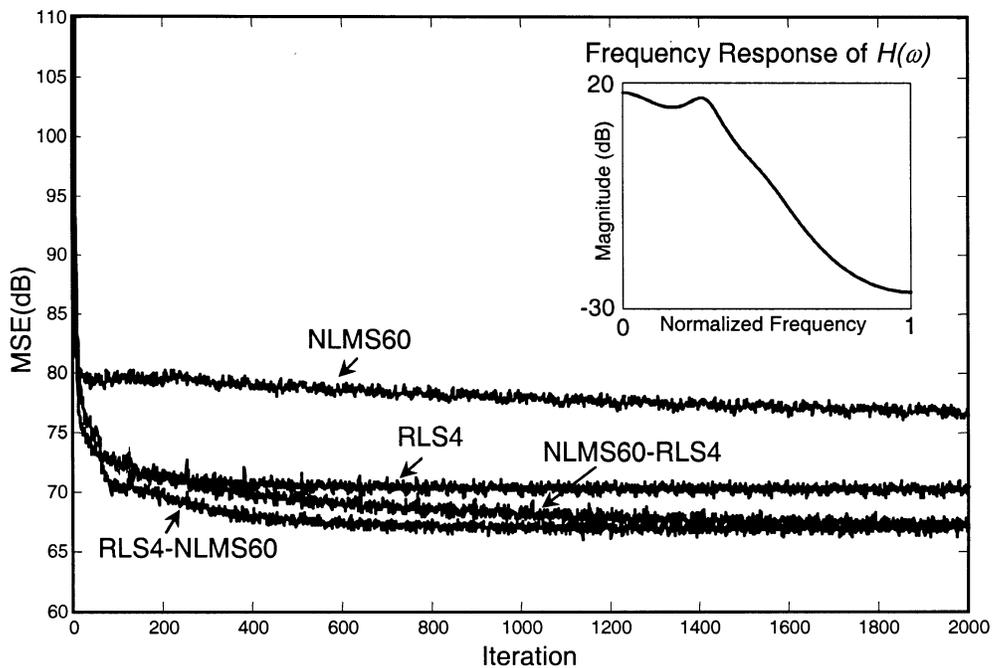


Fig. 3. MSE performance for RLS4-LMS60, RLS4, LMS60 and LMS60-RLS4 predictors. Frequency response (magnitude) of the generating filter is also shown in the subplot at the top-right corner.

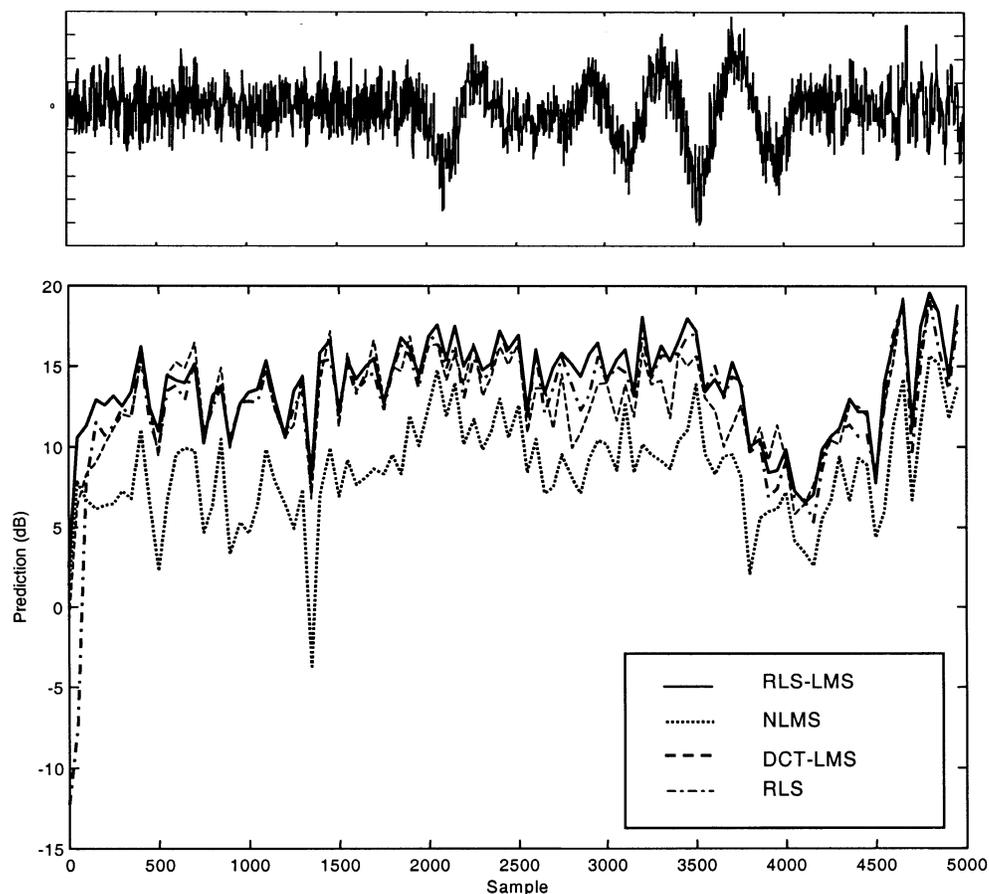


Fig. 4. Prediction gain performance of RLS-LMS (solid line), NLMS (dotted), DCT-LMS (dashed) and RLS (dash-dotted) predictors for a piece of audio signal.

the complexity significantly. The orders of the RLS and the SA predictor are 4 and 128, respectively. The output of the cascade predictor is segmented into frames with 1024 samples and en-

ropy coded using the Golomb code as in [1]. In the decoder, the signal flow chart is reversed to obtain a lossless copy of the original audio.

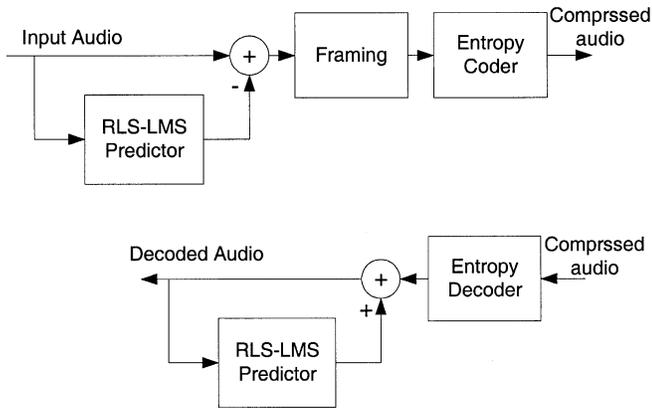


Fig. 5. Structure of the lossless audio encoder (top) and the lossless audio decoder (bottom).

TABLE I

AVERAGE BITRATES IN BITS/SAMPLE FOR THE EXPERIMENTAL CODEC WITH DIFFERENT PREDICTORS. ALL THE TESTING ITEMS ARE SAMPLED AT 44.1 KHz UNLESS STATED OTHERWISE

Item	LPC	LMS	RLS+LMS
Solo cello	6.88	6.72	6.24
Indian sitar	8.16	8.16	7.68
Mandolin	8.64	8.48	7.68
Pop	11.52	11.52	11.04
Rock	9.60	9.76	9.28
Software synthesized music	7.04	6.56	6.08
Ballet	9.44	9.60	8.96
Symphony	8.48	8.48	8.16
Average:	8.72	8.66	8.14
Clarinet (96kHz)	6.65	6.08	4.77
Flute (96kHz)	5.52	4.80	4.32
Violin (96kHz)	6.56	6.40	5.06
Average: (96kHz)	6.24	5.76	4.72

In our experiments, we first replace the RLS-LMS predictor in the experimental codec with an NLMS (order 132) predictor and a block-adaptive LPC (order 16) predictor and compared their compression performance. A total of 11 music files selected from a variety of music types including vocal, pop, rock, musical instrument and computer synthesized music are used in our tests. Among them, eight music files are sampled at 44.1 kHz and the rests are sampled at 96 kHz. All the music files are quantized at 16 bits/sample. For simplicity, we use only mono files in our experiments, although it is straightforward to extend the results to stereo or multi-channel cases. The average bitrate (in bit/sample) of the compressed files are compared in Table I. As can be seen, the NLMS codec has a similar compression performance as LPC although it has a much larger prediction order. On the other hand, the RLS-LMS codec performs much better and successfully reduces the average bitrate by 0.3–1.0 bit/sample in all the test files.

We further compare the compression performance of the RLS-LMS codec with three widely used lossless audio codecs, namely, Shorten [8], LPAC [11] and WA [12]. Note that, as report in [1], these three codecs give a good reflection of the state of the art compression performance for lossless audio coding. All the benchmark codecs run under their highest compression ratio setting in our tests. The results are given in Table II, which clearly show that the best compression performance is achieved

TABLE II

THE AVERAGE BITRATES IN BITS/SAMPLE FOR DIFFERENT LOSSLESS CODECS. ALL THE TESTING ITEMS ARE SAMPLED AT 44.1 KHz UNLESS STATED OTHERWISE

Items	SHORTEN	LPAC	WA	RLS+LMS
Solo cello	6.72	6.46	6.40	6.24
Indian sitar	9.12	7.87	8.16	7.68
Mandolin	8.64	8.10	8.16	7.68
Pop	12.16	11.26	11.52	11.04
Rock	10.56	9.29	9.44	9.28
Software synthesized music	7.20	6.61	6.72	6.08
Ballet	9.76	9.15	9.28	8.96
Symphony	8.96	8.37	8.32	8.16
Average:	9.14	8.39	8.50	8.14
Clarinet (96kHz)	5.92	4.92	4.96	4.77
Flute (96kHz)	4.96	4.52	4.48	4.32
Violin (96kHz)	6.08	5.22	5.12	5.06
Average: (96kHz)	5.65	4.89	4.85	4.71

by the RLS-LMS codec with an average bitrate reduction ranging from 0.2–0.6 bit/sample compared with the next best algorithm, LPAC, for most testing items.

IV. CONCLUSION

This paper introduces a cascade RLS-LMS predictor, which provides rapid convergence speed and superior prediction gain performance for input with large eigenvalue spread. Its application to lossless audio coding is discussed where a high-order RLS-LMS predictor is employed to model the abundant tonal and harmonic structures of audio signals for best coding performance.

REFERENCES

- [1] M. Hans and R. W. Schaffer, "Lossless compression of digital audio," *IEEE Signal Processing Mag.*, pp. 21–32, July 2001.
- [2] J. Rissanen, "A universal data compression system," *IEEE Trans. Inform. Theory*, vol. 29, pp. 656–664, Sept. 1983.
- [3] M. Purat *et al.*, "Lossless transform coding of audio signal," in Proc. 102nd AES Conv., Munich, Germany, 1997, preprint 4414.
- [4] J. Rissanen, "Universal coding, information, prediction and estimation," *IEEE Trans. Inform. Theory*, vol. 30, pp. 629–636, July 1984.
- [5] M. D. Paez and T. H. Glisson, "Minimum mean-squared-error quantization in speech PCM and DPCM systems," *IEEE Trans. Commun.*, vol. 20, pp. 225–230, Apr. 1972.
- [6] S. W. Golomb, "Run-length encodings," *IEEE Trans. Inform. Theory*, vol. 12, pp. 399–401, July 1966.
- [7] S. Haykin, *Adaptive Filter Theory*, 3rd ed. Englewood Cliffs, NJ: Prentice-Hall, 1996.
- [8] T. Robinson. (1994) SHORTEN: Simple Lossless and Near-Lossless Waveform Compression. Cambridge Univ. Eng. Dept, Cambridge, U.K. [Online]. Available: <http://www.softsound.com/Shorten.html>
- [9] L. Xiao, L. Gang, L. Zhengguo, C. T. King, and Y. A. Ling, "A novel prediction scheme for lossless compression of audio waveform," in Proc. *IEEE ICME2001*, Japan, Aug.
- [10] "WaveZip," [Online], Available: <http://www.gadgetlabs.com/wavezip.html>.
- [11] T. Liebchen. LPAC. [Online]. Available: <http://www-ft.ee.tu-berlin.de/~liebchen/>
- [12] D. Lee. Wave Archive (WA). [Online]. Available: <http://www.firstpr.com.au/audiocomp/>
- [13] M. Mboup *et al.*, "LMS coupled adaptive prediction and system identification: A statistical model and transient mean analysis," *IEEE Trans. Signal Processing*, vol. 42, pp. 2607–2615, Oct. 1994.
- [14] E. Hansler, "Adaptive echo compensation applied to the hands-free telephone problem," in Proc. *ISCAS 1990*, 1990, pp. 279–282.
- [15] P. Prandoni and M. Vetterli, "An FIR cascade structure for adaptive linear prediction," *IEEE Trans. Signal Processing*, vol. 46, pp. 2566–2571, Sept. 1998.
- [16] G. Schuller *et al.*, "Lossless coding of audio signals using cascade prediction," in Proc. *ICASSP 2001*, 2001, pp. 3273–3277.

- [17] B. Edler and G. Schuller, "Audio coding using a psychoacoustic pre- and post-filter," in *Proc. ICASSP 2000*, 2000, pp. 881–884.
- [18] P. M. Clarkson, *Optimal and Adaptive Signal Processing*. Boca Raton, FL: CRC, 1993.
- [19] N. Jayant, *Digital Coding of Waveforms—Principles and Applications to Speech and Video*. Englewood Cliffs, NJ: Prentice-Hall, 1984, ch. 6.



Rongshan Yu (M'01) received the B.Eng. degree from Shanghai Jiaotong University, Shanghai, China, and the M.Eng. degree from National University of Singapore, Singapore, in 1995 and 2000, respectively.

He was with the Centre for Signal Processing, School of Electrical and Electronics, Nanyang Technological University, Singapore from 1999–2001. He is presently an Associate Member, Research Staff at the Institute for Infocomm Research (I^2R), A*STAR, Singapore. His main research interests

include data compression and digital signal processing.

Chi Chung Ko (M'82–SM'93) received the B.Sc. (first class honors) and Ph.D. degrees in the electrical engineering from Loughborough University of Technology, Loughborough, U.K.

In 1982, he joined the Department of Electrical and Computer Engineering, National University of Singapore. His current research interests include digital signal processing, speech processing, adaptive arrays and computer networks. He has written over 100 technical publication in these areas.

Dr. Ko is an Associate Editor of the IEEE TRANSACTION ON ANTENNAS AND PROPAGATION.