EXTENDED LINEAR PREDICTION TOOLS FOR LOSSLESS AUDIO CODING

*Takehiro Moriya *, Dai Tracy Yang ** and Tilman Liebchen ****

* NTT Cyber Space Labs., Tokyo, Japan
** University of Southern California, Los Angeles, USA
*** Technical University of Berlin, Berlin, Germany

ABSTRACT

Two extension tools for enhancing the compression performance of prediction-based lossless audio coding are proposed. One is progressive-order prediction of the starting samples at the random access points, where the information of previous samples is not available. The first sample is coded as is, the second is predicted by first-order prediction, the third is predicted by second-order prediction, and so on. This can be efficiently carried out with PARCOR (PARtial autoCORrelation) coefficients. The second tool is inter-channel joint coding. Both predictive coefficients and prediction error signals are efficiently coded by inter-channel differential or three-tap adaptive prediction. These new prediction tools lead to a steady reduction in bit rate when random access is activated and the inter-channel correlation is strong.

1. INTRODUCTION

For the archiving and broadband transmission of music signals, lossless reconstruction is becoming a more important feature than high efficiency in compression by means of perceptual coding as defined in MPEG standards such as MP3 or AAC. Although DVD-audio and Super CD Audio [1, 2] include proprietary lossless compression schemes, there is a demand for an open and general compression scheme among content-holders and broadcasters. In response to this demand, a new lossless coding scheme has been considered as an extension to the MPEG-4 Audio standard [3, 4].

In the course of the standardization process, a time-domain compression scheme based on linear predictive coding (LPC) has been defined as a reference model. This model is proposed by the Technical University of Berlin and the decoding process is shown in Fig. 1 [5]. For every frame, the optimum LPC coefficients are calculated and the associated PARCOR coefficients [6, 7] are quantized in an arcsine-transformed domain. The prediction error signal is derived by the quantized predictive coefficients and coded with a Rice code. For stereo signals, simple inter-channel coding is applied, where either the L-channel or R-channel together with the difference between the R- and L-channels are coded.

This paper proposes two extension tools for prediction-based lossless coding. One is progressive-order prediction to improve performance in the compression of starting samples at random-access points. The other is inter-channel joint coding for both predictive coefficients and prediction error signals. Both tools are described in detail, and the results of performance evaluation are given.

2. PROGRESSIVE ORDER PREDICTION

2.1. Random access

Samples of an audio signal usually have strong correlation in the time domain. Auto-regressive linear prediction is well-known as one of the most powerful and simple tools for reducing the amplitudes of error signals, enabling reductions of bit rate [2, 8]. However, in the editing and playback of compressed signals, the ability to start from a random access point is desirable. We thus have to reconstruct perfect signals without using any of the previous signal information. Ensuring this property for auto-regressive linear prediction leads to a significant loss of compression performance, since prediction must be shut off at the accessible points. Until now, the first p samples, where p is the prediction order, are kept unchanged and required separate coding due to a large amplitude.

2.2. Progressive prediction

For starting samples in the random access frames, progressive-order prediction is useful as a way of making full use of the available samples and thus reducing prediction error as much as possible. While it is of course impossible to predict the first sample, the second sample is predictable by first-order prediction only from the previous sample. The prediction error at the $(q+1)$-th sample is derivable by q-th order prediction in general.

For this progressive-order prediction, PARCOR coefficients are convenient, since each coefficient is independent from the prediction order p, while normal auto-regressive LPC coefficients need to be calculated for every prediction order q upto p. The associated lattice filter is shown in Fig. 2, where k_q represents the q-th PARCOR coefficient. An example procedure of PARCOR-based progressive-order prediction is shown in Fig. 3. It is understood

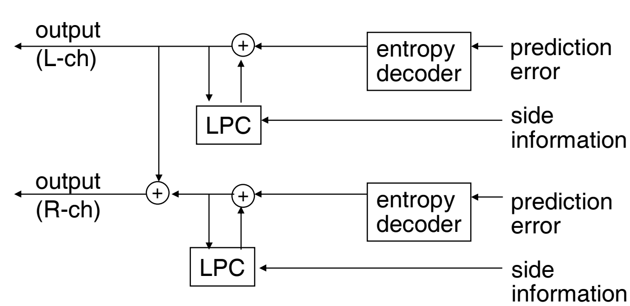

Fig. 1. Decoding process of reference predictive coding system with simple inter-channel prediction.

that the internal state is identical between the encoder and the decoder. The decoder or synthesis part is shown in Fig.4, where $y(n)$ and $x(n)$ represent the prediction error input and the reconstructed output, respectively.

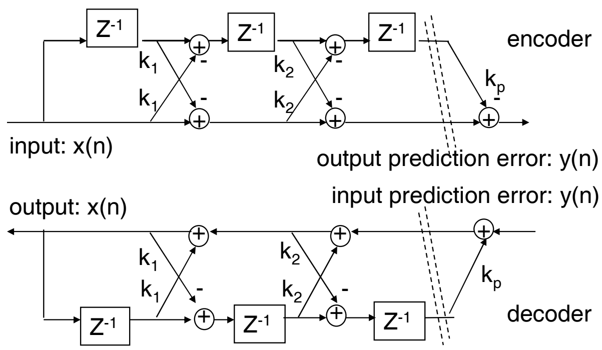

Fig. 2. General forms of an encoder and decoder based on a linear-prediction scheme with PARCOR coefficients.

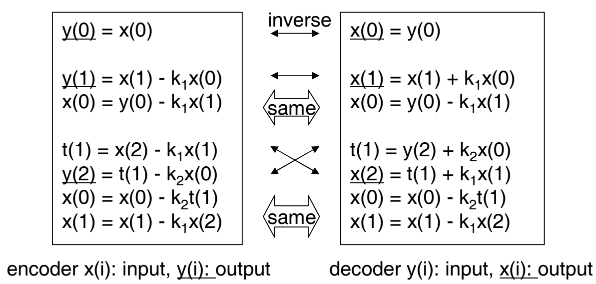

Fig. 3. Relationship between the encoder and decoder of the PARCOR-based progressive-order prediction.

Examples of the waveform around a random access point are shown in Fig.5. Waveform (a) represents the original input signal, waveform (b) represents the normal prediction error with p non-predicted samples after the random access point, and waveform (c) has the prediction errors that result from progressive-order prediction. For the first sample in (c), we need to use special entropy coding with a significantly larger amplitude than for the later samples. For the second and third samples which are the prediction errors with the first and the second order prediction, we need a Rice code with lower amplitude than the first sample. From the fourth sample and beyond, we have observed that the prediction errors become small enough that we are able to use the same Rice code for typical continuous-prediction errors.

3. INTER-CHANNEL JOINT CODING

3.1. Differential coding of PARCOR coefficients

In the reference system, the PARCOR coefficients are independently quantized for each of the channels. There is a strong similarity between the PARCOR coefficients for the two channels of a stereo signal. One way to take advantage of this is to reuse the coefficients of one channel on the other, saving bit rates for the coefficients at the cost of a greater amplitude for the prediction error signals. The other is to use differential coding of the PARCOR

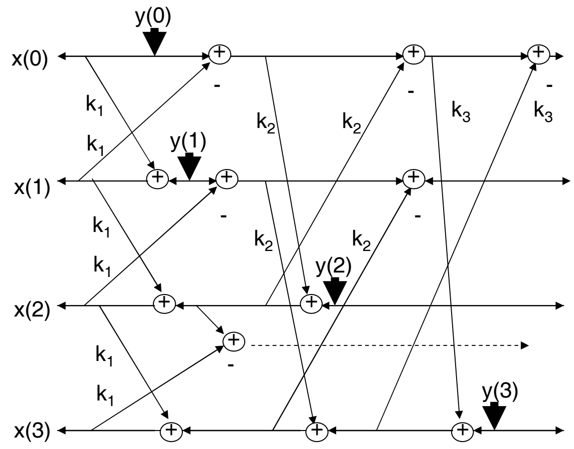

Fig. 4. Decoding process for PARCOR-based progressive-order prediction.

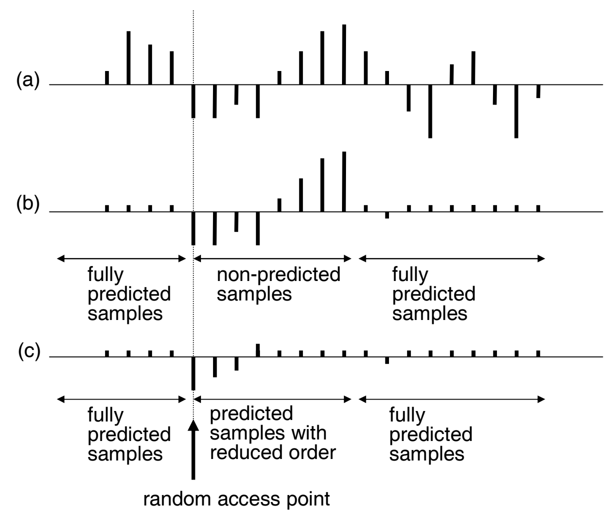

Fig. 5. Examples of waveforms in the vicinity of the random access point: (a) input signal, (b) error signal for non-progressive prediction, and (c) the error signal for progressive prediction.

coefficients or coding coefficients for the R- and L-channels in sequence. In this case, we need entropy coding to exploit the lower amplitude of the difference data. Both configrations are shown in Fig. 6.

3.2. Inter-channel prediction of prediction error

For some stereo signals, i.e. those where the inter-channel correlation is particularly rich, simple differential coding in the time domain is useful. However, the adjacent samples of most signals are correlated with each other, so it is generally more effective to use time-domain linear prediction. The prediction error signals then still reflect the correlation between channels, and prediction based on the cross-correlation function can further reduce the amplitude of the error. We use $\mathbf{X}(x(0), \ldots, x(N-1))$ to denote the first or L-channel and $\mathbf{Y}(y(0), \ldots, y(N-1))$ to denote the second or R-channel. N is the number of samples and γ represents

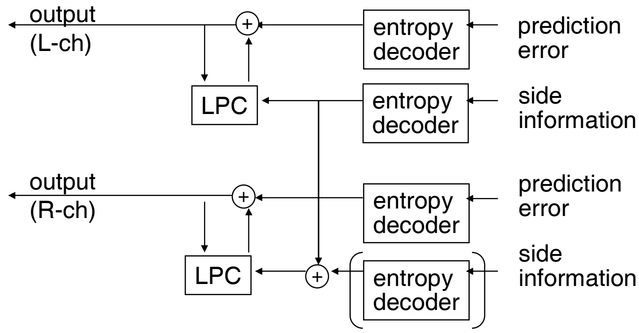

Fig. 6. Decoding process for the inter-channel prediction of PARCOR coefficients: R-channel PARCOR coefficients may be differentially coded or skipped.

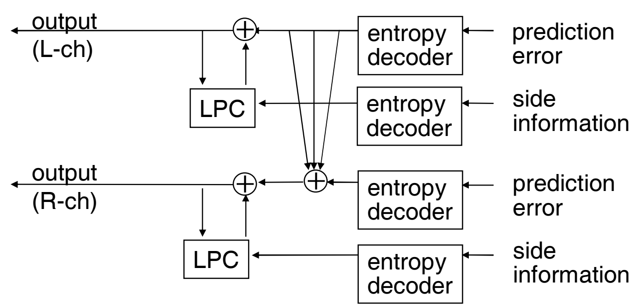

Fig. 7. Decoding process for inter-channel prediction of the prediction error signal when three-tap-based weighting is used.

the optimum coefficient to minimize the distortion d in eq. (1).

$$d = ||\mathbf{Y} - \gamma \mathbf{X}||^{\mathbf{2}} \tag{1}$$

$$\gamma = \frac{\mathbf{X}_0^T \mathbf{Y_0}}{\mathbf{X_0^T X_0}}, \text{ where } \mathbf{X}_0^{\mathbf{T}} \mathbf{Y}_0 = \sum_{i=0}^{N-1} x(i)y(i). \tag{2}$$

Inter-channel prediction can be extended to multi-tap cases, such as the three-tap case as shown in Fig. 7. Multi-tap prediction may compensate for the small phase difference between channels. The optimum coefficients γ are found by solving the equation eq. (6) which minimizes the distortion d between the L-channel prediction error and the weighted sum of the R-channel prediction error. γ_1 and γ_{-1} usually have a smaller amplitude than γ_0. γ_1 and γ_{-1} can be quantized with two bits each and γ_0 can be quantized with four bits representing values in the range from 0 to 0.8.

$$d = (\sum_{i=1}^{N-2} (y(i) - \sum_{j=-1}^{1} (\gamma_j x(i-j)))^2 \tag{3}$$

$$R_{-1,-1} = \mathbf{X_{-1}^T X_{-1}} \tag{4}$$

$$U_{-1} = \mathbf{X_{-1}^T Y_0} \tag{5}$$

$$\begin{bmatrix} \gamma_{-1} \\ \gamma_0 \\ \gamma_1 \end{bmatrix} = \begin{bmatrix} R_{-1,-1} & R_{-1,0} & R_{-1,1} \\ R_{-1,0} & R_{0,0} & R_{1,0} \\ R_{-1,1} & R_{1,0} & R_{1,1} \end{bmatrix}^{-1} \begin{bmatrix} U_{-1} \\ U_0 \\ U_1 \end{bmatrix} \tag{6}$$

4. PERFORMANCE EVALUATION

4.1. Progressive-order prediction

Compression experiments were carried out with the conditions summarized in Table 1. Performance improvements (relative improvement ratios as percentages) in terms of compression ratio are shown in Figs. 8 and 9.

$$\text{improvement\_ratio} = \frac{\text{reference\_filesize} - \text{tested\_filesize}}{\text{reference\_filesize}}$$

The 15 test items (**a** to **o**, each 30-sec long) were provided by Matsushita Corp. for use in standardizing MPEG-4 lossless coding. For each item, the left-hand bar shows the performance improvement in compression ratio with full continuous prediction, compared with the reference condition where the first p samples are not predicted. The right-hand bar shows the improvement in performance with the proposed form of progressive-order prediction. We see that the proposed system consistently improves compression to an extent approaching that of a continuous prediction system which does not provide random-access capability.

Table 1. Experimental conditions.

sampling rate	48 and 96 [kHz]
amplitude resolution	16 and 24 [bit]
prediction order	30
random-access interval	100 [ms]
number of input files	15 (30 [s] each)

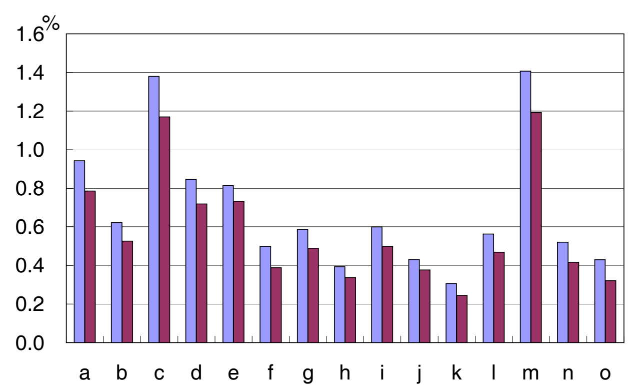

Fig. 8. Improvement ratio in comparison with the reference system. Input audio files are sampled at 48 kHz with 16-bit amplitude resolution and labeled as **a** through **o**. The left-hand bars show the improvement ratio with continuous prediction and the right-hand bars show the ratio with the proposed prediction.

4.2. Inter-channel coding

Performance improvement with the combined form of inter-channel coding are shown in Figs. 10 and 11. The combinations are listed in Table 2, in which a two-bit code indicates whether or not inter-channel coding of the coefficients (first bit) and inter-channel coding of the prediction errors (second bit) are in use. The best combination is selected and the corresponding two-bit code is included as side information. For the selection, we tried

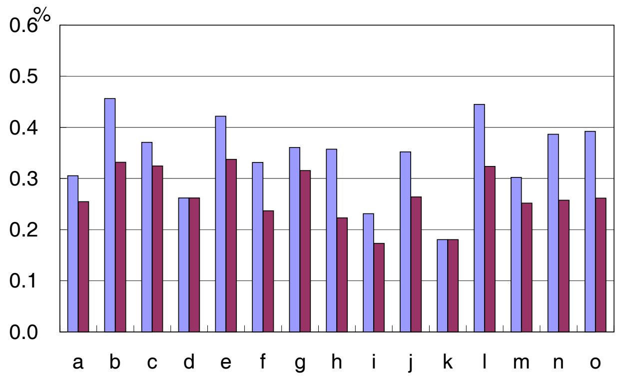

Fig. 9. Improvement ratio for 96 kHz sampling rate with 24-bit resolution. Others are identical to those of Fig. 8.

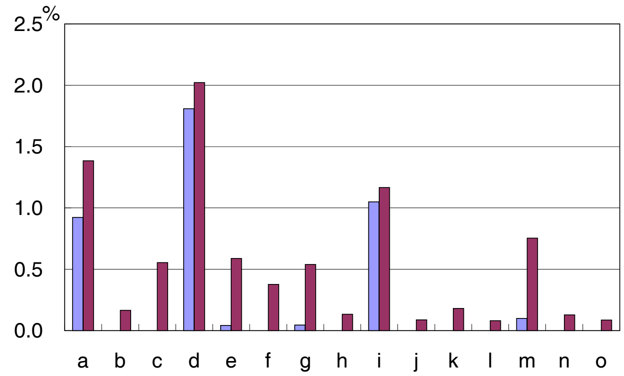

Fig. 11. Improvement ratio for 96-kHz sampling rate and 24-bit resolution. Others are identical to those for Fig. 10.

the prediction twice and the compression of prediction error four times, although clever estimation may be practical. The test items and test conditions are identical with those in the previous section. Left-hand bars show the performance improvement for the reference form of inter-channel coding in comparison with independent channel coding. The right-hand bars show the improvement over independent coding gained by the proposed coding tools. Note that for some input signals, such as **b**, **c** and **e**, simple inter-channel coding delivers no improvement in compression ratio. In contrast to this, the proposed tools are in all cases more effective than independent coding and simple inter-channel coding.

Table 2. Choice of combined inter-channel tools

tools	no prediction for prediction error	prediction for prediction error
substitution of coefficients	00	01
difference of coefficients	10	11

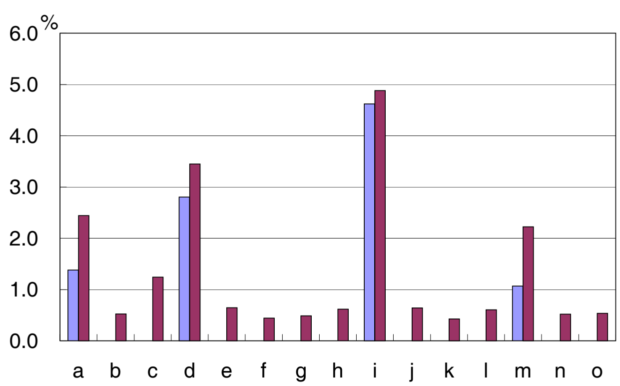

Fig. 10. Improvement ratio in comparison with the independent channel coding for 48-kHz sampled 16-bit stereo input. The left-hand bars show the improvement ratio for simple inter-channel coding in the reference system and the right-hand bars show the improvement with the proposed inter-channel coding.

5. CONCLUSION

We have proposed two new prediction tools for enhancing the compression performance of prediction-based lossless audio coding. The use of a PARCOR-based progressive-order prediction tool for the starting samples in each random-access frame has been investigated. Differential coding of PARCOR coefficients and the 3-tap prediction of prediction error signals were proposed as inter-channel coding tools. Compression testing demonstrated that these proposed tools improve the performance in compression under all input and test conditions.

Note that progressive-order coding is also useful as a measure for reducing noise due to packet loss in a predictive speech-coding scheme. The form of inter-channel coding we have described is extensible to the multi-channel case. The overall coding system is flexible and is applicable to various signals, including bio-medical signals and environmental monitoring signals from sensor networks.

6. REFERENCES

[1] K. Konstantinides, “An introduction to super audio CD and DVD-Audio,” *IEEE Signal Proc. Mag.*, vol. 20, no. 4, pp. 71–82, 2003.

[2] M. Hans and R. W. Schafer, “Lossless compression of digital audio,” *IEEE Signal Proc. Mag.*, vol. 18, no. 4, pp. 21–32, 2001.

[3] ISO/IEC JTC 1/SC29/WG11 N5040, *Call for proposals on MPEG-4 lossless audio coding*, Klagenfurt, AT, July 2002.

[4] ISO/IEC JTC 1/SC29/WG11 N5208, *Revised call for proposals on MPEG-4 lossless audio coding*, Shanghai, China, October 2002.

[5] T. Liebchen, “MPEG-4 lossless coding for high-definition audio,” in *AES 115th Convention paper 5872*, New York, USA, October 2003.

[6] F. Itakura and S. Saito, “Digital filtering techniques for speech analysis and syhthesis,” in *Proc. Int. Acoust. Cong. 25C-1*, Budapest, Hungary, 1971.

[7] J. D. Markel and Jr. R. M. Gray, *Linear Prediction of Speech*, Springer-Verlag, New York, USA, 1976.

[8] Monkeys Audio, http://www.monkeysaudio.com.