LOSSLESS CODING OF AUDIO SIGNALS USING CASCADED PREDICTION

Gerald Schuller[a], *Bin Yu*[b1], *Dawei Huang*[c]

[a]Bell Labs, Murray Hill, NJ 07974, USA;
[b]Department of Statistics, University of California, Berkeley, CA 94720, USA;
[c]Bell Labs, Beijing, 100080, China;
http://www.multimedia.bell-labs.com, schuller@bell-labs.com

ABSTRACT

A novel predictive lossless coding scheme is proposed. The prediction is based on a new weighted cascaded least mean squared (WCLMS) method. WCLMS is especially designed for music/speech signals. It can be used either in combination with psycho-acoustically pre-filtered signals (an idea presented in [1]) to obtain *perceptually* lossless coding, or as a stand-alone lossless coder. Experiments on a database of moderate size and a variety of pre-filtered mono-signals show that the proposed lossless coder (which needs about 2 bit/sample for pre-filtered signals) outperforms competing lossless coders, WaveZip, Shorten, LTAC and LPAC, in terms of compression ratios.

1. INTRODUCTION

In [1, 2] a new scheme for perceptual lossless coding of audio signals was proposed. It is based on pre-filtering an audio signal with a frequency response inverse to the psycho-acoustic masked threshold. This pre-filtering, followed by a quantizer, removes the *irrelevance* of the signal. This stage is followed by a lossless coder to reduce the *redundancy* of the signal. This separation of the coder into two main stages has several advantages, e.g., both stages can be optimized independently of each other. This leads for instance to a better performance for speech signals compared to conventional audio coders, such as Bell Labs' PAC [3], which makes this scheme also more suitable for communications applications.

In [1, 2] the pre-filter is described. In the present paper a lossless coder for the redundancy reduction part is described. For bandwidth constraint applications we emphasize compression ratio over complexity. To make it suitable for communications applications, our goal is also a low possible encoding/decoding delay. Current lossless coders are usually based on blockwise prediction or transforms, which increases the encoding/decoding delay. Further, they have limited length prediction (often for complexity reasons), which limits the compression ratio. For these reasons we are looking at backward adaptive prediction (instead of blockwise prediction).

We present our lossless coder mainly in the context of pre-filtered audio signals, but we found that it is also effective as a stand-alone lossless coder for audio signals.

2. LOSSLESS CODING BASED ON WCLMS

The new prediction method Weighted Cascaded LMS (WCLMS) can be described in three steps as follows.

Normalized LMS Prediction. LMS is a well known fast stochastic gradient algorithm to minimize adaptively the least squared prediction error or residual. Its complexity is linear in the order of the predictor. Its applications have been wide, including on-line automatic control, signal processing, and acoustic echo cancellation (cf. [4]).

Let $x(n)$ be the signal at time n, and $\mathbf{x}^T(n)$ is defined as $\mathbf{x}^T(n) := [x(n-L+1), ..., x(n)]$ where L is the order of the prediction. An L'th-order predictor is of the form

$$P(\mathbf{x}(n-1)) = \mathbf{x}^T(n-1) \cdot \mathbf{h}(n), \qquad (1)$$

where $\mathbf{h}(n)$ is the L-dimensional vector of predictor coefficients at time n. We update $\mathbf{h}(n)$ with the normalized LMS:

$$\mathbf{h}(n+1) = \mathbf{h}(n) + \frac{e(n)}{1+\lambda||\mathbf{x}(n-1)||^2}\mathbf{x}(n-1). \qquad (2)$$

with $e(n)$ beeing the prediction error. This is a special case of the normalized LMS [4], i.e. we use only one tuning parameter λ to trade off adaptation speed and accuracy. Our experience shows that this works well for $15 \leq \lambda \leq 25$ and across a variety of pre-filtered sound signals, which we usually observed in the range of about -20 to 20 (determined by the pre-filters psycho-acoustic model).

[1] Work was done while with Bell Labs, Murray Hill, NJ

Cascade of the predictors. Cascaded adaptive predictors have been used and described before, e.g. in [5]. Here the prediction error of one predictor is used as input for the next predictor. These cascades have advantages in terms of adaptation speed, prediction accuracy, and numerical stability. But so far, only the output of the final stage of a cascade was used as “end result” for further processing. For our application the essential advantage of the cascade is the availability of predictors of different orders as “taps”. Speech/audio signals have varied orders of correlations. Very non-stationary signals like sounds from castanets need a short predictor that is able to track the signal fast enough, whereas more stationary signals as sounds from flutes require higher prediction orders to accurately model the signal with all its spectral details. In our predictive coding application, we apply the LMS prediction three times, leading to the predictors P_1, P_2 and P_3 as follows.

Since the residuals $e_1(n)$ of the first predictor are not integers but floating point numbers, they cannot be reproduced and stored in finite precision without losing accuracy. This was not a problem for a single LMS since its input $x(n)$ are integers (PCM signals). However, in the second and third stages in cascading LMS, the non-integer residuals are chosen as the inputs to improve the accuracy of their prediction. But when the encoding and decoding sides have different rounding precisions, we are not able to synchronize the two sides and the encoder and decoder will produce different outputs. We solve this problem by limiting the precision of the residuals in a defined manner, e.g. using 8 bit precision after the fractional point.

Predictive Minimum Description Length weighting. By using the cascade of predictors one of the main issues is how to select or combine these predictors. One powerful technique based on Bayesian statistics uses weighted combinations for a superior prediction performance (cf. [6]). Using this approach the model-based predictors P_i can be combined into

$$\sum_i w_i P_i, \;\; w_i \geq 0, \; \sum_i w_i = 1 \tag{3}$$

where w_i is the posterior (i.e. based on the observed data) probability that P_i is “correct” given data to date, which can be viewed as a measure of the goodness-of-fit of the model or predictor P_i. This can be seen as a ”soft” model switching. Model based prediction uses a joint distribution postulation of the signals, unlike our non-model based LMS prediction. To obtain or estimate the probability w_i for non-model based predictor P_i we will use the weights based on the predictive Minimum Description Length (MDL) principle or the PMDL weights. When the predictors are model-based, the precise asymptotic equivalence of PMDL and the Bayesian approach can be found e.g. in [7], but PMDL also covers the non-model based predictors as follows. We found that the probability density function (pdf) of the prediction error $e_i(n) = x(n) - P_i(\mathbf{x}(n-1))$ can be approximated well with a Laplacian distribution for our data:

$$\text{pdf}(e_i(n)) \propto e^{-c|e_i(n)|},$$

with some (real) parameter $c > 0$. This exponentiated prediction error based on Laplace distribution can now in turn be used as the PMDL weight w_i of predictor P_i. To adapt to the nonstationarity of the signal, we consider a vector of the past prediction error values (with a “forgetting factor” μ) and its joint pdf. This leads to our actual estimate of the weight w_i:

$$w_i(n) \propto e^{-c(1-\mu)\sum_{i=1}^{\infty} |e_i(n-i)|\cdot\mu^{(i-1)}}. \tag{4}$$

Heuristically, these weights reward predictors with good past prediction performance. For our experiments we chose $c = 2$ and $\mu = 0.9$.

Since the input signal is integer valued, we round the weighted predicted value to obtain an integer-valued final WCLMS predicted value. The entire process is shown in the diagram of the WCLMS predictor, Fig. 1, with the rounding in block Q. The encoder is depicted in Fig. 2 and the decoder is shown in Fig. 3. A sensitivity to transmission errors can be countered e.g. by periodic reset of the predictor.

The integer valued residuals after prediction can be encoded using entropy coding. We used a block based Huffman coder, but for instance an adaptive Huffman coder can be used as well. As can be seen in the figures, the predicted value is available to both ends, so that the transmission of the residuals is enough to recover the signal $x(n)$.

3. APPLICATION TO PRE-FILTERED SIGNALS

This section assesses the performance of the WCLMS coder when applied to mono-signals processed by the psycho-acoustic pre- and post-filter described in [1, 2]. We compare our method with the benchmark lossless schemes LTAC [8], which is a Transform based lossless coder, LPAC [9], which is based on (block) prediction, Shorten [10], which is based on polynomial (block) prediction, and Wavezip [11]. LTAC has a coding part closest to traditional audio coders, because it uses a transform for compression. Meridian Lossless Packing is a lossless coder (also based on prediction) which was recently adopted for use on DVD audio [12]. But since it is more intended for higher sampling rates and we had no evaluation copy available, we did not include it in our comparison. A database of moderate size is chosen to assess WCLMS and the benchmark coders performance in terms of bit rates. The database contains about 50 pieces of music, speech and mixed music/speech with sampling rates 8, 16, and 32 kHz.

For all the WCLMS results, the tuning parameters are set to be the fixed $\lambda = 20$ and $c = 2$. Extensive experiments on the data base found that a good combination of orders is 200 for the first stage, 80 for the next, and 40 for the final stage. This combination works well for different signals and at different sampling rates. A heuristic explanation for this combination is as follows. For most if not all the pieces in the database, there is a dominant pattern which is close to stationary and hence requires a high order to capture. After this dominant pattern is removed with the first stage LMS predictor of order 200, other features get to show. So the first residual predictor of order 80 at stage two tailors itself to this feature, which is more detailed, followed by the final predictor of order 40. We found it interesting to observe that the reverse sequence of orders does not perform as well.

Table 1 gives a comparison of our WCLMS scheme with fixed LMS predictors for a subset of our database for different sampling rates (32, 16, and 8 kHz). Here it can be seen that the best fixed predictor order varies depending on the signal. But in all cases the WCLMS leads to lower bit-rates, even though the order of the highest fixed order predictor is higher than the total order of the WCLMS (400 vs. 320), and the lowest order predictor is lower than the lowest order section in WCLMS (10 vs. 40).

Table 2 shows a comparison of our lossless compression scheme WCLMS (200,80,40) to the other widely used general purpose lossless coders, applied to the output of the psycho-acoustic pre-filter. In this table chart is pop music; 16cj is classical jazz; mixed is speech with background music; spot2 is a commercial containing speech.

Clearly, our WCLMS coder gives the best coding rate for every signal in the table: roughly, a 10 % improvement over the second best LPAC, a 20% improvement over LTAC, a 25% improvement over Shorten, and a 35% improvement over WaveZip which is a widely used PC sound compression software. Similar results hold for other samples in our database. We observed that the peak rates do not rise much above the average values because of the “compressing” nature of the pre-filter. Moreover, it is interesting to observe here that LPAC, which is similar to LTAC but based on prediction, performs better for most signals than the transform based LTAC.

4. CONCLUSIONS

We presented a lossless compression scheme which is based on weighted cascaded backward prediction. The weighting is derived from Bayesian statistics. Although this scheme has a higher complexity than other lossless coders, and depending on the implementation can be sensitive to bit-errors, we found that it has a clearly higher compression ratio, which is important e.g. for bandwidth constraint channels. Further it allows for a low encoding/decoding delay, which is important for communications applications.

Order	LMS 10	LMS 80	LMS 400	WCLMS (200,80,40)
32kHz				
chart	2.18	2.11	2.02	1.94
16cj	2.38	2.21	2.10	1.99
mixed	2.29	2.25	2.25	2.16
spot2	2.06	2.06	2.08	1.96
16kHz				
chart	2.38	2.22	2.19	2.01
16cj	2.66	2.30	2.29	2.07
mixed	2.40	2.36	2.37	2.28
spot2	2.32	2.32	2.35	2.21
8kHz				
chart	2.50	2.22	2.26	2.03
16cj	2.83	2.25	2.35	2.06
mixed	2.42	2.37	2.41	2.31
spot2	2.39	2.37	2.43	2.31

Table 1: The resulting bit-rate in bit/sample for different fixed length LMS predictors and for weighted cascaded LMS with sections of length 200,80,40, on pre-filtered signals

REFERENCES

[1] B. Edler and G. Schuller, ”Audio Coding Using a Psychoacoustic Pre- and Post-Filter”, ICASSP 2000, Istanbul, Turkey, pp. II-881:884.

[2] B. Edler, C. Faller, G. Schuller, “Perceptual Audio Coding Using a Time-Varying Linear Pre- and Post-filter”, AES Symposium, Los Angeles, CA, Sep. 2000

[3] V. Madisetti, D. B. Williams, eds., “The Digital Signal Processing Handbook”, Chapter 42, D. Sinha et al., “The Perceptual Audio Coder (PAC)”, CRC Press, Boca Raton, Fl., 1998.

[4] S. S. Haykin (1999). *Adaptive Filter Theory.* Englewood Cliffs, N.J. : Prentice Hall.

[5] P. Prandoni and M. Vetterli, “An FIR cascade structure for adaptive linear prediction”, IEEE Trans. Sig. Proc., Sep. 1998, **46**, 2566-2671.

[6] A. Gelman, H. Stein, and D. Rubin (1995). *Bayesian data analysis*, New York : Chapman & Hall.

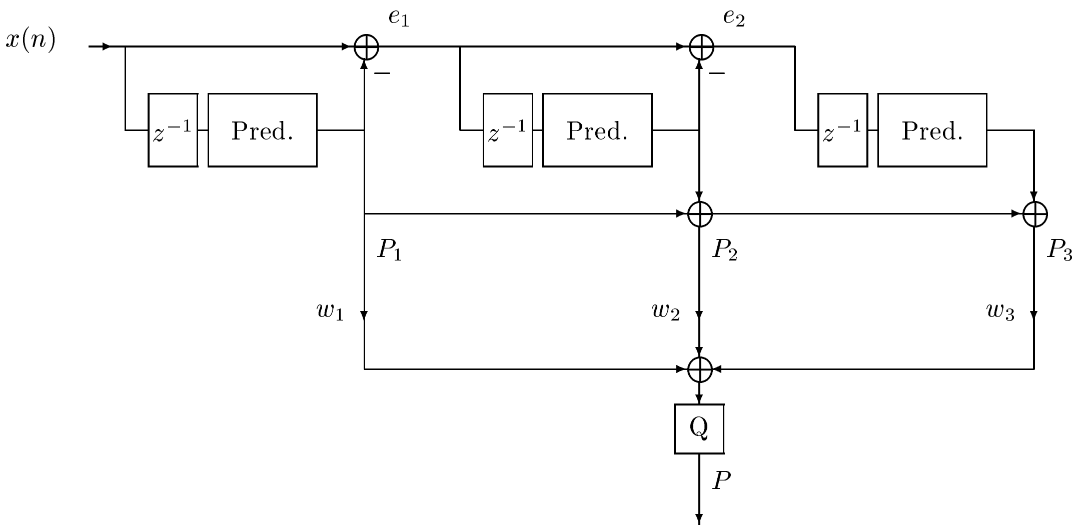

Figure 1: The WCLMS predictor. Input $x(n)$, output $P(\mathbf{x}(n-1))$.

	WCLMS	LPAC	LTAC	Sho.	WZ.
32kHz					
chart	1.94	2.23	2.36	2.51	3.22
16cj	1.99	2.47	2.42	2.67	3.35
mixed	2.16	2.34	2.59	2.58	3.19
spot2	1.96	2.12	2.42	2.47	3.09
16kHz					
chart	2.01	2.49	2.55	2.68	3.42
16cj	2.08	2.64	2.56	2.85	3.48
mixed	2.27	2.50	2.80	2.67	3.23
spot2	2.21	2.38	2.75	2.63	3.27
8kHz					
chart	2.03	2.58	3.10	2.89	3.67
16cj	2.06	2.33	3.04	3.11	3.77
mixed	2.31	2.56	3.36	2.78	3.46
spot2	2.31	2.53	3.38	2.76	3.46

Table 2: Comparison of the weighted cascaded prediction (200,80,40) and coding with other widely used lossless compression schemes, using pre-filtered signals. Bit-rate in bit/sample. Sho.: Shorten, WZ.: WaveZip.

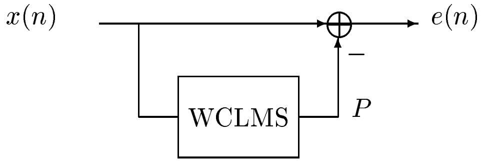

Figure 2: The WCLMS lossless encoder (input $x(n)$, output $e(n)$).

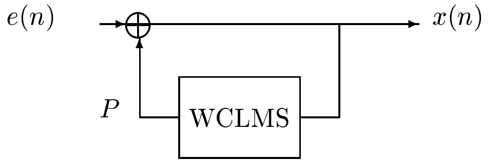

Figure 3: The WCLMS lossless decoder (input $e(n)$, output $x(n)$).

[7] A. Barron, J. Rissanen, and B. Yu (1998). The Minimum Description Length principle in coding and modeling. (Special Commemorative Issue: Information Theory: 1948-1998) *IEEE. Trans. Inform. Th.*, **44**, 2743-2760.

[8] T. Liebchen, LTAC, version 1.71, blocksize 4096. http://www-ft.ee.tu-berlin.de/~liebchen/ltac.html

[9] T. Liebchen, LPAC, version 0.99h, setting “Extra High Compression”, TU Berlin, Germany, http://www-ft.ee.tu-berlin.de/~liebchen/lpac.html

[10] Softsound, Great Britain, Shorten, version 1.03, default setting (polynomial prediction), http://www.softsound.com/Shorten.html

[11] WaveZip, version 2.00 uses MUSICompress of Soundspace, Sunnyvale, CA. http://www.gadgetlabs.com/wavezip.htm

[12] M.A. Gerzon et al.,”The MLP Lossless Compression System”, AES 17th Int. Conf., Florence, Italy, Sep. 1999, pp. 61-75.