Perceptual Audio Coding using Adaptive Pre- and Post-Filters and Lossless Compression

Gerald Schuller[1], Bin Yu[2], Dawei Huang[3], and Bernd Edler[4]
Bell Labs, Lucent Technologies
600 Mountain Ave., Murray Hill, NJ
schuller@agere.com

Abstract

This paper proposes a versatile perceptual audio coding method that achieves high compression ratios and is capable of low encoding/decoding delay. It accommodates a variety of source signals (including both music and speech) with different sampling rates. It is based on separating irrelevance and redundancy reductions into independent functional units. This contrasts traditional audio coding where both are integrated within the same subband decomposition. The separation allows for the independent optimization of the irrelevance and redundancy reduction units. For both reductions, we rely on adaptive filtering and predictive coding as much as possible to minimize the delay. A psycho-acoustically controlled adaptive linear filter is used for the irrelevance reduction, and the redundancy reduction is carried out by a predictive lossless coding scheme, which is termed weighted cascaded least mean squared (WCLMS) method. Experiments are carried out on a database of moderate size which contains mono-signals of different sampling rates and varying nature (music, speech, or mixed). They show that the proposed WCLMS lossless coder outperforms other competing lossless coders in terms of compression ratios and delay, as applied to the pre-filtered signal. Moreover, a subjective listening test of the combined pre-filter/lossless coder and a state-of-the-art perceptual audio coder (PAC) shows that the new method achieves a comparable compression ratio and audio quality with a lower delay.

Keywords

Lossless coding, perceptual audio coding, prediction, LMS algorithm

I. Introduction

Perceptual audio coding removes both "irrelevance" and "redundancy" from a signal. The former is defined as signal components undetectable by the receiver (the ear). Psycho-acoustics defines the masked threshold as the threshold below which distortions cannot be heard. This threshold is time- and frequency-dependent, as well as signal dependent. Perceptual audio coding keeps only audible signal components by hiding quantization distortions below the threshold, which is estimated based on a psycho-acoustic model in the encoder. "Redundancy", on the other hand, refers to the predictability or statistical dependencies in the signal, and can be removed via lossless compression.

The typical sampling rate of a high quality audio signal is 32 - 48 kHz with an accuracy of 16 - 24 bits per sample. In particular, a CD with a 44.1 kHz sampling rate and a 16 bits/sample per

[1]Now with Agere Systems, Murray Hill, NJ

[2]Now with the Department of Statistics, University of California, Berkeley, CA 94720, USA;

[3]Now with Bell Labs, Beijing, 100080, China

[4]Now with the University of Hannover, Hannover, Germany

stereo channel leads to a bit rate of 2×705.1 kb/s, or 1.41 Mb/s. However, a much lower bit rate is often desired, ranging from 16 kb/s for internet streaming with modems to a couple of hundred kb/s for higher speed connections. Perceptual audio coding provides the most effective tool to achieve sufficiently high compression ratios while maintaining a good audio quality for many applications. One example is digital broadcasting, where the audio signals contain both music and speech, demanding a coder that performs well for both types of signals. Additionally, a low encoding/decoding delay is desirable in communications applications such as video conferencing.

The goal of this paper is to present an audio coding method that provides a very low encoding/decoding delay without compromising the compression performance. This makes it suitable for real time communications applications such as high quality audio for next generation wireless networks, high quality video conferencing, and musicians playing together over long distances. In particular we make two new contributions. The first is a psycho-acoustically controlled pre-filter based on an adaptive linear filter for the irrelevance reduction. The second is a low delay lossless audio coder based on cascaded prediction and backward adaptation for the redundancy reduction.

The paper is organized as follows. Section II describes our new framework where we start by a brief review of traditional audio coding methods. Section III gives a detailed description of the psycho-acoustically controlled pre- and post-filter. Section IV contains the lossless coder based on Weighted Cascade Least Mean Squares (WCLMS) prediction. Experimental results appear in Section V, followed by conclusions in Section VI.

II. A New Framework: Separation of Irrelevance and Redundancy Reductions

A. Traditional Audio Coding

The popular MP3 (short for MPEG1 Layer 3 [1]) coder was developed in the late 80's and early 90's. It needs roughly twice as many bits for a comparable quality as the present state-of-the-art coders such as MPEG2/4 AAC [1], AC3 [2], ATRAC [3] and PAC [4], which typically achieve "CD quality" at roughly 64 kb/s for a mono signal. They use analysis filter banks to decompose the signal into subbands. These subband signals are then critically downsampled and quantized. The quantization step-size is controlled by a psycho-acoustic model, and removes or reduces the irrelevant portions of a signal. After quantization, entropy coding such as Huffman coding is applied to remove or reduce the redundancy in the signal. The decoder consists of an entropy decoder followed by the synthesis filter bank which reconstructs the time domain signal from the subbands. Note that the same subband resolution is used for both irrelevance and redundancy reductions.

The goal of a high compression ratio in perceptual coding has historically led to the use of large transforms or filter banks with many bands. They are suitable to obtain high coding gains for the mostly stationary parts in music signals. The large number of subbands, however, leads to audible "pre-echo" artifacts for very non-stationary signals as the attacks of castanets. Hence there are usually two modes for the filter bank: one mode with a small number of bands (typically 128 bands) for very non-stationary parts of the signals, and another with a large number of bands (typically 1024 bands) for the more stationary parts of the signal. The large number of bands also contributes to a high encoding/decoding delay, which is undesirable for communications applications. The delay of a coder depends on the filter bank size, the size of the look-ahead block for mode switch decisions, and buffering for constant bit-rate channels. For coders like MPEG2/4 AAC or PAC, the delays caused by the first two factors are 2047 and 576 samples respectively. The delay caused by buffering could be a few thousand samples due to the high bit-rate peaks usually associated with the 128 band mode. This is too long for communications applications. A remedy for the filter bank delay is to use switchable low-delay filter banks [5], [6] instead of the traditionally used MDCT filter bank. This low-delay filter bank also has the two modes of 128 and 1024 bands. However, they can only reduce the delay down to the downsampling rate, which is equal to the number of subbands (down to 1023 from 2047 samples delay). A good example of an audio coder intended for communications applications is the MPEG-4 low delay coder [7]. Its delay is only about 960 samples (about 30 ms at 32 kHz sampling rate), and is achieved mainly by reducing the number of subbands and avoiding switching the number of bands. This reduced number of bands in turn leads to a decreased compression ratio compared to MPEG2/4-AAC. Its delay is still not low enough for more time-critical applications such as musicians playing together over long distances [8]. Speech coders, on the other hand, handle speech signals well and have a short encoding/decoding delay, but they do not perform well on non-speech signals like music or room noise.

In this paper, we target for a delay of about 10 ms at 32 kHz sampling rate (320 samples), which would be sufficient for the musicians application [8] and on the low end of speech coders.

B. The New Approach

As discussed in the previous subsection, the pitfalls of traditional transform coding are (1) the use of the same transform for both irrelevance and redundancy reductions which leads to the necessity of having two modes for the number of bands, and (2) the relatively long delay because of the large number of subbands in the filter bank, which is especially a problem in our targeted communications applications. Our solution is first to separate irrelevance reduction from the redundancy reduction. Then for both reductions, we rely on adaptive filtering and predictive coding as much as possible to minimize the delay. It is known, that predictive coding has the same asymptotic coding gain as transform coding [9], [10], but unlike transform coding, predictive coding has no system inherent delay.

Our irrelevance reduction unit consists of a psycho-acoustically controlled time-varying pre-filter followed by a quantizer. The psycho-acoustic part is block based, but the block is made very short (128 samples) to reduce delay. This is illustrated in Fig. 1. The pre-filter has a frequency response inverse to the masked threshold. The post-filter in the decoder is the inverse of the pre-filter, and hence has a frequency response like the masked threshold. To obtain the inverse filter in the decoder, the frequency response function of the pre-filter has to be parameterized and transmitted as side information to the decoder [11], [12]. The effect of the pre-filter can be seen as a normalization of the signal to its masked threshold so that the level of the quantization distortions can be made constant in time and frequency. Since our signal is in the time domain (not in subbands) this can be accomplished with a simple uniform constant step size quantizer, as shown in Fig. 1. In our system this quantizer is a simple rounding operation to the nearest integer. This way, (ideally) all the irrelevance has been removed, and a lossless compression scheme needs to be applied to remove the remaining redundancy in the pre-filtered and quantized integer-valued signal.

To achieve an efficient reduction of redundancy and a low delay, we designed a novel lossless coder based on integer prediction and Huffman coding of the integer residuals. It is important to note that this lossless compression scheme can also be used as a stand-alone lossless coder, just as we can use other lossless coders for this stage. Our prediction scheme is based on the well-known LMS (Least Mean Squares) algorithm. However, two new ingredients are added in the prediction that are essential for the coder's performance. The first is the cascading of predictors to derive three predictors of different orders; the second is the weighting of the three predictors using

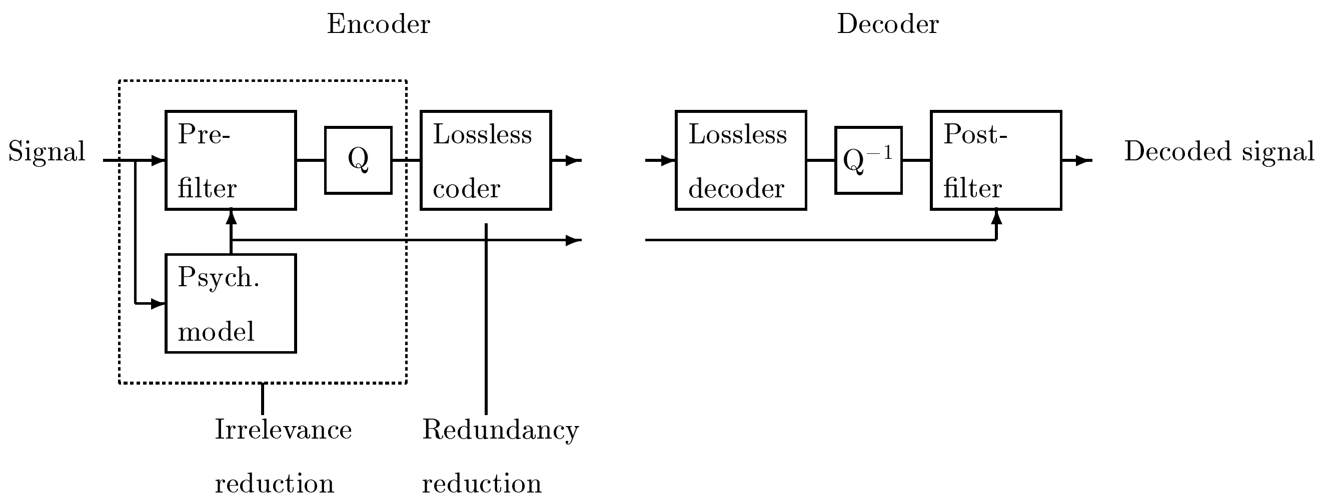

Fig. 1. The audio coding scheme with separated irrelevance and redundancy reduction, using a psycho-acoustic pre- and post-filter and lossless compression.

measures of their past performances via the Predictive Minimum Description Length (PMDL) principle to form our final predictor. The weighting allows a *soft* switching between the three predictors of different orders.

III. The Pre- and Post- filter

In this section, we design a predictive pre-filter such that its transfer function matches the inverse of the estimated masked threshold from the psycho-acoustic model.

All psycho-acoustic models are block based. To minimize the delay that they introduce, but at the same time provide a sufficient accuracy for stationary signals, the psycho-acoustic model in [13] is used for our pre-filter and is based on 128 subbands. This choice is made for two reasons. One is that subjective evaluations have shown that an update interval of the masked threshold of approximately 2 to 4 ms is appropriate for achieving a high audio quality (64 to 128 samples at 32 kHz sampling); the second is that the 128 band mode in traditional audio coders has a sufficient time resolution for non-stationary signals. Most psycho-acoustic models use a tonality estimation to obtain the masked threshold. A tonality estimation is more difficult at a lower frequency resolution of 128 bands. However, the model in [13] does not need a tonality estimation for the higher frequencies. Moreover, for the lower frequencies we use a predictability measure to improve the tonality estimation. It is worth noting that the particular psycho-acoustic model used is not important here, it can as well be a modified version of traditional audio coders.

For the psycho-acoustic model we divide the input signal $x(n)$ into blocks of size 128 and let t be the block index. Then the output of the psycho-acoustic model is the masked threshold $M(f,t)$ (dependent on frequency f). We compute this threshold for every consecutive block of 128 input samples. Now we need to find a pre-filter so that its time dependent transfer function $H(f,t)$ satisfies

$$H(f,t) = \frac{1}{|M(f,t)|} \tag{1}$$

To obtain this frequency response or a close approximation, we apply an adaptive filter structure as used in linear predictive coding (LPC). Its filter coefficients are computed with techniques from LPC analysis, using the masked threshold $|M(f,t)|^2$ as short-term power spectrum. If our filter order is K, then its output $x(n)$ is related to its input $s(n)$ through equation

$$x(n) = s(n) - \sum_{k=1}^{K} a_k^t s(n-k) \ . \tag{2}$$

We now describe how the a_k^t are obtained. The inverse DFT of $|M(f,t)|^2$ over frequency for block t gives the target auto-correlation function $r_{mm}^t(n)$. Then a_k^t are obtained by solving the linear equation system

$$\sum_{k=0}^{K-1} r_{mm}^t(|k-n|) a_k^t = r_{mm}^t(n+1) \ , \quad 0 \leq n < K \ . \tag{3}$$

However, subjective evaluation experiments show that simple switching of filter parameter sets $\{a_k^t\}$ from one block t to the next block $t+1$ leads to audible artifacts. The most obvious approach for avoiding rapid changes of filter coefficients is a direct interpolation. First let's re-index the filter coefficients in terms of the sample number n by defining $a_k(n) = a_k^t$ if n falls into the middle of the tth block of size 128. Otherwise the filter coefficient is given by the linear interpolation of these middle-of-the-block values. This simple linear interpolation in the filter coefficient domain does not work because the post-filter is a filter with an infinite impulse response (IIR) which can become instable. Experiments also show that this occurs in practice and leads to audible artifacts.

The remedy is to use a lattice structure for the filter [14]. Then the filter coefficients are re-parametrized in the lattice structure into reflection coefficients. These coefficients lead to stable filters, and the stability is guaranteed for the linear interpolation between parameter sets of stable systems. Moreover, they can be directly used in the lattice filter structure so that no complex conversions are necessary. The conversion to reflection coefficients only needs to be done at the boundaries of the blocks of 128 samples. Subjective evaluations indicate that the transition problems are eliminated [12].

The pre-filter structure is shown in Fig. 2. The psycho-acoustic model has an inherent delay of 128 samples due to the blocking for the computation of the masked threshold. Therefore, to obtain a precise correspondence between the filtered audio signal and the output from the psycho-acoustic model, a corresponding delay should be introduced before the filter. This is the block labeled "Delay" in Fig. 2. Since the pre-filter coefficients a_k^t need to be transmitted to the decoder as side information, we are interested in approximating the masked threshold $M(f,t)$ with the maximum accuracy and with the lowest number K of coefficients. The masked threshold $M(f,t)$ has more spectral detail at lower frequencies than at higher frequencies due to the properties of hearing. Thus we use another interesting technique known from prediction, the so-called frequency-warping. All delay elements of the FIR pre-filter are replaced by suitable all-pass filters [15], [16]. This "warps" the filters frequency scale such that we obtain a higher spectral resolution at low frequencies than at higher frequencies. We find that a pre-filter order of $K = 12$ is sufficient using warping. Also for the coding and transmission of the filter parameters, techniques known from speech coding can be adapted. We use line spectral frequency (LSF) parameters [12], because they reduce the effect of quantization on the resulting frequency response. To increase the efficiency of the parameter coding, we transmit a new set of parameters only if there is a sufficient change compared to the previous parameter set. This works because in stationary audio segments the masked threshold changes very little. In our implementation and depending on the signals, the bit-rate for the side information (coefficients, gain factor, and update bits) is in the range from 0.03 to 0.2 bit/sample. Fig. 3 shows an example for the magnitude response of the post-filter, the masked threshold, and the signal. It is clear that we have done quite well matching the magnitude response of the post-filter with the masked threshold.

Our coding unit of a pre-filter, a quantizer, and a lossless coder produces a signal-dependent bit-stream. Often it is desirable to control the resulting bit-rate. We achieve this simply by adding an attenuation factor between the pre-filter output and the quantizer as shown in Fig. 4. With a factor of 1, the quantization noise is (ideally) right at the masked threshold. If the factor is smaller than 1, this factor increases the effective step-size of the quantizer. This means the resulting quantization noise is uniformly above the masked threshold leading to audible

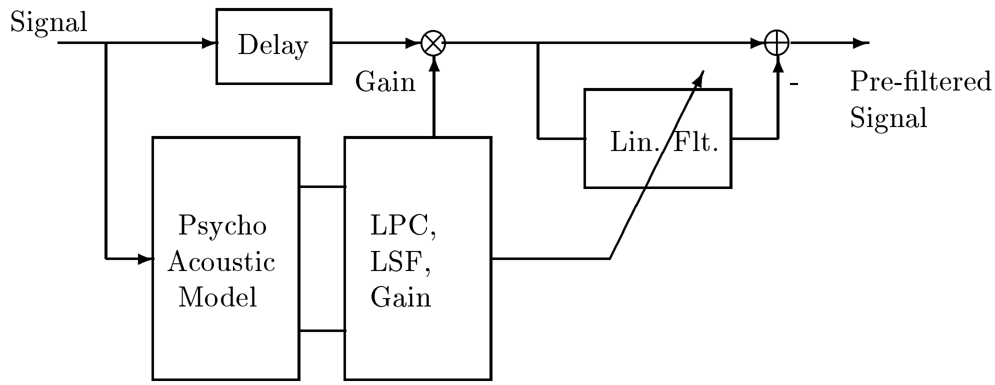

Fig. 2. Structure of the implementation of our pre-filter. It results in a low delay.

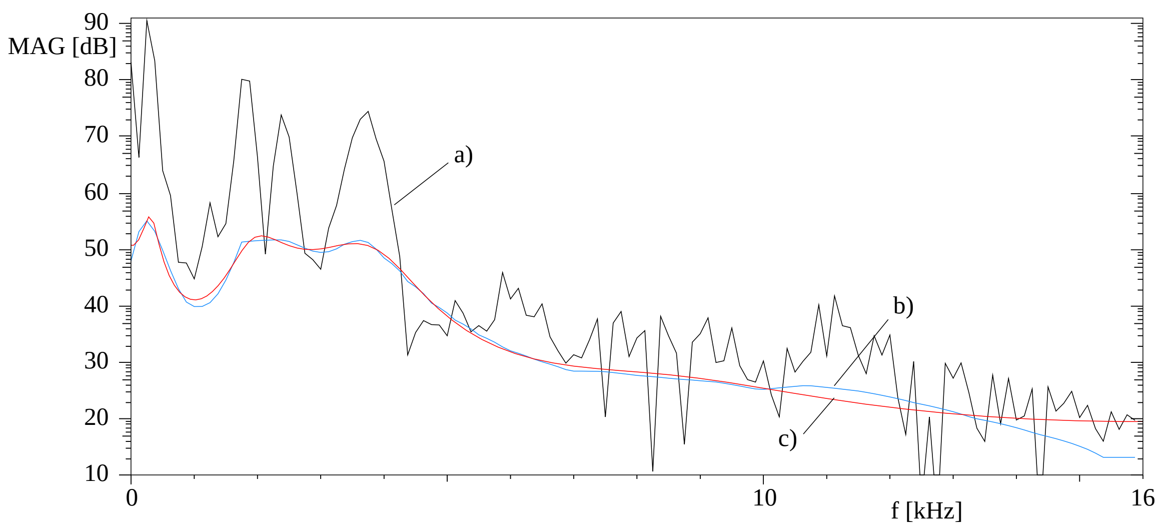

Fig. 3. 128 point spectrum of a signal (a), masked threshold as computed by the psycho-acoustic model (b) and the magnitude response the Post-filter (c).

distortions, but resulting in a reduced bit-rate.

IV. Lossless coding based on Weighted Cascade LMS (WCLMS) prediction

After pre-filtering, there is still considerable correlation or dependencies left in the signal. These dependencies are to be removed or reduced as much as possible in the redundancy reduction unit using lossless compression.

Current lossless audio coders include Shorten [17], LPAC [18], LTAC[19], and WaveZip [20]. Shorten and LPAC are based on block-wise forward prediction. In particular, Shorten uses a linear or polynomial prediction within blocks of typically 256 samples [17]. The prediction coefficients are transmitted as overhead, and the residuals are Huffman coded and transmitted.

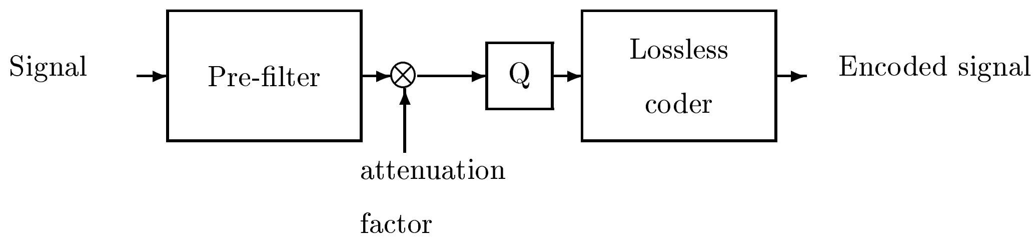

Fig. 4. Structure with a control for the bit-rate.

LPAC uses a block size of typically 1024, and it uses an adaptive prediction order up to 30. LTAC uses transforms on blocks of typically 4096 samples for compression and hence is close to traditional audio coding. They all introduce a delay of at least the size of the block. WaveZip is a very popular lossless compression program. It is claimed to have a low computational complexity, but no exact documentation is available in the literature. Meridian Lossless Packing (MLP) is another lossless coder based on forward prediction, and has been adopted for DVD audio [21].

These coders are typically intended for file compression, where delay is of no concern, and where the computational complexity is of some importance because it determines the compression time. We believe that for future communication applications the compression ratio and encoding and decoding delays will become increasingly critical and that more complexity will be tolerable. These considerations motivate the proposal of a backward adaptive prediction scheme as opposed to block-wise or forward prediction.

A. Weighted Cascaded LMS Predictors (WCLMS)

Our new causal prediction method has three ingredients: (i) normalized LMS (Least Mean Square); (ii) cascading of the normalized LMS predictors; and (iii) predictive Minimum Description Length (PMDL) weighting of the cascaded predictors.

Normalized LMS Prediction

LMS is an old but efficient stochastic gradient algorithm that minimizes adaptively the least squared error. Its complexity is linear in the order of the predictor, and its applications have been wide and varying, including on-line automatic control, signal processing, and acoustic echo cancellation (cf. [22], [23]). Let $x(n)$ be the signal at time n, $\mathbf{x}(n) = (x(1), ..., x(n))$, and $\mathbf{x}_L(n) = (x(n-L+1), ..., x(n))$. Then an L'th-order single stage LMS predictor is of the form

$$\hat{x}(n) = \mathbf{x}_L(n-1) \cdot \mathbf{h}^T(n), \tag{4}$$

where $\mathbf{h}(n)$ is the L-dimensional row vector of predictor coefficients at time n.

We initialize with $\mathbf{x}_L(0) = (0, 0, ..., 0)$, $\mathbf{h}(0) = (1/L, ..., 1/L)$ and update $\mathbf{h}(n)$ as follows:

$$\mathbf{h}(n+1) \quad = \quad \mathbf{h}(n) + \frac{x(n) - \hat{x}(n)}{1 + \lambda||\mathbf{x}_L(n-1)||^2}\mathbf{x}_L(n-1). \tag{5}$$

Equation (5) is a special case of the normalized LMS presented in [22] (pp. 432-447), with one tuning parameter instead of two. Our experience shows that, for audio signals, this prediction scheme works well for $15 \leq \lambda \leq 25$.

Cascading the LMS Predictors

When the prediction error from one predictor is used as the input to the next predictor, the predictors are said to be cascaded. Cascaded adaptive predictors have been studied in [24], where it is shown in a special case that cascades are advantageous in terms of adaptation speed, prediction accuracy, and numerical stability. Cascading once the same predictor has also been used in statistical analysis by Tukey under the name "twicing" ([25]). All the existing cascading schemes use only the output of the final stage as the "end result" for further processing. However, cascading can be used in a different way. We can take advantage of the availability of predictors of different orders as additional outputs with an economical computational cost (because the computation for the next stage of cascading is built upon that from the earlier stages). The different orders of the predictors from different cascading stages enable us to adapt to the varied windows of stationarity in speech and music signals. For our predictive coding purpose, we apply the normalized LMS predictor three times, leading to the predictors P_1, P_2 and P_3 as described below (see also Fig. 5). In the following we use the term "predictor" for the three outputs of the cascade which predict $x(n)$, as opposed to "LMS predictor" which denotes the individual LMS stages within the cascade. We find that cascading three predictors is sufficient, and adding more predictors does not do much to improve the prediction. The use of cascading LMS predictors in the encoder is depicted in Fig. 6, and in the decoder in Fig. 7.

Since the residuals or the prediction errors from each LMS predictor are not integers but real numbers, they cannot be reproduced and stored in finite precision without losing accuracy. The encoder and decoder must use the same arithmetic throughout the prediction process. One option is to use a standard arithmetic package. For the results we discuss, we limit the precision of the residuals by using 8 bit precision after the fractional point. Observe that this only affects the prediction, not the lossless property of the lossless coder. More generally, for any real number

x, let $[x]$ denote the closest integer to x; and for a positive integer A, define $[x]_A$ by

$$[x]_A = A^{-1}[Ax]. \tag{6}$$

Then using 8 bit precision is equivalent to choosing $A = 256$. We find that this precision is sufficient for a good prediction, and that it results in the same predicted values at the encoder and decoder. In our cascade, the first predictor P_1 of order L_1 of $x(n)$ is a finite precision version of (4):

$$P_1(\mathbf{x}(n-1)) = [\mathbf{x}(n-1) \cdot \mathbf{h}^T(n)]_A. \tag{7}$$

Since $x(n)$ is integer valued, hence of finite precision, the residual $e_1(n) = x(n) - P_1(\mathbf{x}(n-1))$ of the first predictor is also of finite precision. It serves as the input to the second LMS predictor, which is of order L_2. Let $\hat{e}_1(n)$ denote its output, the *finite precision* predicted value of $e_1(n)$. We obtain the second predictor P_2 of $x(n)$ as the sum of the first predictor and the predicted prediction error,

$$P_2(\mathbf{x}(n-1)) = P_1(\mathbf{x}(n-1)) + \hat{e}_1(n), \tag{8}$$

where the effective order of predictor P_2 is the sum of the two first stages. We denote the finite precision residual associated with the second LMS predictor by

$$e_2(n) = e_1(n) - \hat{e}_1(n). \tag{9}$$

to which we apply a normalized LMS of order L_3 to get the third predictor

$$P_3(\mathbf{x}(n-1)) = P_2(\mathbf{x}(n-1)) + \hat{e}_2(n)$$

as seen in Fig. 5.

Predictive Minimum Description Length (MDL) Weighting.

By using the cascade, three predictors are at our disposal. We now have to combine these predictors in a way that optimizes the prediction accuracy or resulting coding rate. For this purpose we look at Bayesian statistics (cf. [26]) for inspiration, which motivates the use of a weighted combination of predictors for an improved prediction performance. In particular, we combine the three predictors into a final predictor $P(\mathbf{x}(n-1))$ by weighting:

$$P(\mathbf{x}(n-1)) = \sum_{i=1}^{3} w_i(n-1) P_i(\mathbf{x}(n-1)), \tag{10}$$

$$w_i(n-1) \geq 0, \;\; \sum_{i=1}^{3} w_i(n-1) = 1.$$

Each $w_i(n-1)$ measures how well predictor P_i has predicted the signal in the past. The relative weights are updated every time a prediction is made. Our choice of $w_i(n-1)$ is based on the so-called Predictive Minimum Description Length (PMDL) principle (see, e.g., [27], [28]), which has a close connection to Bayesian statistics. To be precise, we construct a joint probability density of $x(1), x(2), ..., x(n-1), x(n)$ in a predictive way. Since the prediction residual $e_i(n) = x(n) - P_i(\mathbf{x}(n-1))$ at time n follows roughly a Laplacian distribution, we model the conditional probability density function $f_{n,i}$ of $x(n)$ given $\mathbf{x}(n-1) = (x(1), ..., x(n-1))$ as

$$\begin{aligned} f_{n,i}(x(n)|x(1), ..., x(n-1)) &\propto \exp(-c|e_i(n)|) \\ &= \exp(-c|x(n) - P_i(\mathbf{x}(n-1))|), \end{aligned} \tag{11}$$

for some positive parameter c and with $i = 1, 2, 3$. Then at time n the joint probability of $(x(1), x(2), ..., x(n))$ is the product of the conditional probabilities $f_{n,i}$. This joint probability is called the *PMDL weight*. Since our signals are non-stationary, we introduce a "forgetting parameter" μ to emphasize the performance for recent samples. The product of conditional Laplacian expressions (11) together with the forgetting parameter μ leads to our final $PMDL$ weights:

$$w_i(n-1) \propto e^{-c(1-\mu)\sum_{i=1}^{n-1} |e_i(n-i)| \cdot \mu^{i-1}}. \tag{12}$$

Note that c and μ are tuning parameters and will be fixed as $c = 2$ and $\mu = 0.9$ in our implementation of WCLMS for the results section V. The weights are normalized to sum to 1 and initialized with 1/3. We do not quantize w_i, because in our experiments they have led to the same final integer-valued predictor P for both the sender and the receiver. Since $x(n)$ is integer valued, the resulting prediction error or residual

$$e(n) = x(n) - [P(\mathbf{x}(n-1))] \tag{13}$$

is also integer valued, and it is entropy coded and transmitted to the receiver. Note that $[P(\mathbf{x}(n-1))]$, which is based on past values, is available at both the sender and the receiver. So signal $x(n)$ can be easily recovered at the decoder or receiver from $e(n)$ via

$$x(n) = e(n) + [P(\mathbf{x}(n-1))] \tag{14}$$

Possible transmission errors of the residual can propagate, because backward prediction is not block-based. However, this could be countered for instance by a periodic reset of the predictor.

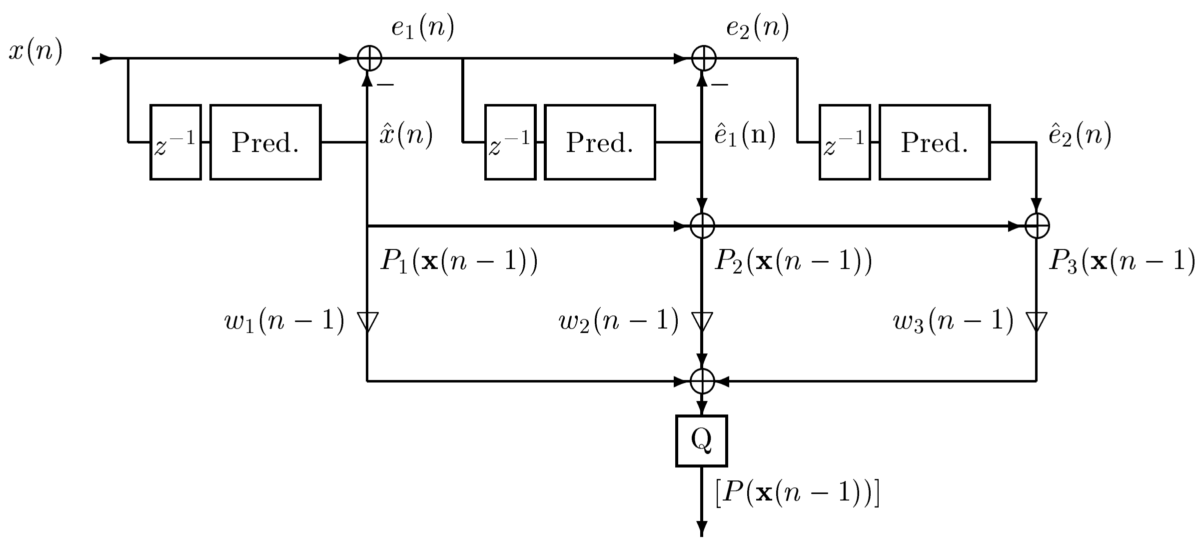

Fig. 5. The WCLMS predictor. Input $x(n)$, output $P(\mathbf{x}(n-1))$ ($\bigtriangledown$ symbolizes multiplication).

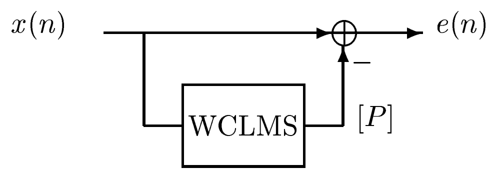

Fig. 6. The WCLMS lossless encoder (input $x(n)$, output $e(n)$).

We find that resetting the predictors at every 4096th sample does not degrade compression performance much.

B. Entropy Coding of Prediction Errors

The integer valued residuals or prediction errors after WCLMS

$$e(n) = x(n) - [w_1(n-1) \cdot P_1(\mathbf{x}(n-1)) + w_2(n-1) \cdot P_2(\mathbf{x}(n-1)) + w_3(n-1) \cdot P_3(\mathbf{x}(n-1))]$$

are entropy coded and transmitted. For simplicity we first used a block-based Huffman coder, for which the experimental results are shown in the next section. For this scheme, we divide the integer residual stream into blocks of length 4096. Then we pair all two consecutive symbols in which the first one is zero. The empirical probabilities of these symbols are calculated over the block. Based on these probabilities, a standard Huffman code is constructed. We transmit this Huffman table as an overhead and the residuals coded in this Huffman code. The coding of

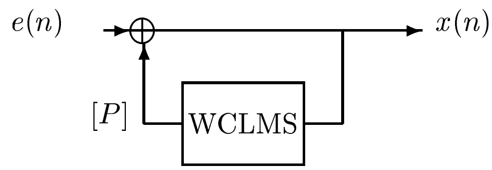

Fig. 7. The WCLMS lossless decoder (input $e(n)$, output $x(n)$).

zero-started pairs reduces the bit rate. Usually the count of zero residuals is more than half, so standard Huffman coding without pairing could be inefficient since it assigns at least one bit to the zero residual.

Observe that the Huffman processing in blocks of 4096 has the disadvantage of an according delay. Since the WCLMS prediction introduces no delay, the delay of the lossless coding unit is determined by the entropy coding part. To obtain much lower delays than with this block-based Huffman approach, several alternative entropy coding schemes were investigated as described in [29]. Surprisingly, we find that an adaptive Huffman coding scheme, with a delay of only 17 samples, achieves comparable bit-rates. An adaptive arithmetic coding scheme, with a delay of about 100 samples, even improves the bit-rate by about 2% over the block-based Huffman coder.

Since the WCLMS prediction combined with the adaptive Huffman coding introduces only a very short delay, the overall coding delay is mainly determined by the irrelevance reduction unit with its pre-filter and psycho-acoustic model, which in our setup introduced a delay of 128 samples. The combination of the pre- and post-filter (PPF) with the WCLMS lossless unit then leads to, depending on the adaptive Huffman coding or arithmetic coding, a delay of $128 + 17$ or $128 + 100$, which are both in the order of 200 samples. Since the decoder does not introduce additional delay, this is about 6 ms encoding/decoding delay at 32 kHz sampling rate, if no bit-rate buffering is used. Hence it is even below our targeted delay of 10 ms at 32 kHz.

V. Experimental Results

This section makes comparisons of our proposed methods PPF and WCLMS at three different levels on test signals from a database of about 140 pieces of music, speech and mixed music/speech. These pieces vary in length from 10 to 16 s and with sampling rates of 8, 16, 32, and 44 kHz. Firstly, using the outputs of these pieces from the pre-filter and quantizer, we compare the bit rates of WCLMS of various cascade predictor orders with the (normalized) LMS of various orders in terms of bit-rate, when the same (Huffman) entropy coding is applied to the residuals of both prediction methods. Secondly, we compare our best WCLMS lossless coder with other lossless coders. These comparisons are only about the lossless unit, which affects the bit-rate but not the sound quality. Thirdly, we compare the complete PPF-WCLMS coder with PAC [4], a traditional state-of-the-art audio coder.

For all the WCLMS results to follow, the tuning parameters are set to be fixed at $\lambda = 20$, $c = 2$, and $\mu = 0.9$. We found that these values lead to good compression results, but that their exact value is not critical. The side information for the post-filter is not included in the result tables since it is the same for all lossless coders and depends on the actual parameterization of the transfer function (as mentioned, about 0.03 to 0.2 bit/sample).

Now let us look at the bit-rate comparisons results to the (normalized) LMS. To show the variation of obtained bit-rates for signals of different characteristics, Table I contains the results of comparisons for four individual signals from the database. In the table, signal "chart" and "chart44" are pop music; "jazz" and "jazz44" are classical jazz; "mixed" is speech with background music; "spot" and "spot44" are commercials containing speech; and mspeech is male speech. This table contains results for a fixed length LMS prediction, compared to WCLMS implemented with predictors of unequal orders. Observe that the best compression is obtained with the highest order prediction in the first stage of the cascade (200,80,40), which means order $L_1 = 200$ in the first stage, $L_2 = 80$ in the second stage, and order $L_3 = 40$ in the final third stage.

Based on the same individual pieces, Table II shows a bit-rate comparison of our best lossless coder WCLMS (200,80,40) to widely used general purpose lossless audio coders, applied to the output of the psycho-acoustic pre-filter. These lossless coders are the earlier mentioned LTAC, LPAC, Shorten and WaveZip. Meridian Lossless Packing (MLP) is not included in our comparison since no evaluation copy is available. Moreover, it is also intended for higher sampling rates than we treat in this paper. Clearly WCLMS (200,80,40) gives the best performance in terms of bit rate: based on the averages for the shown test signals about 13% improvement over LPAC, 24% over LTAC, 23% over SHORTEN, and 38% over WaveZip. 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. Delays are 1023 samples for LPAC, 4095 samples for LTAC, and 255 samples for Shorten. The delay for WaveZip is unknown due to the lack of public documentation. For WCLMS (200,80,40), as mentioned, a delay of only 17 to 100 samples can be achieved using adaptive entropy coding methods.

WCLMS (200,80,40) gives the best performances also for the entire database of signals relative to LMS and other WCLMS based coders, as shown in Table III. The n's in the brackets are the sizes of the sub-categories. For example, in the second row there are 19 pieces in the music category at 32kHz sampling rate. Note that the WCLMS (200,80,40) with a combined order of 320 achieves a better compression ratio than the LMS of order 400. Moreover, we also compare with the best other lossless coder LPAC on the data base in the last column. The average improvement of WCLMS(200,80,40) over LPAC is 15%.

We obtained the combination (200,80,40) by first looking at the bit-rates for fixed length predictors. In Table I and Table III it can be seen that speech signals have a minimum bit-rate around order 40 to 80. Observe that there is a less than linear dependence on the sampling rate, because the higher frequency signal components don't need to be encoded anymore by going to a lower sampling rate. Music signals have a minimum around order 200 to 400. For that reason we combined predictors of orders of that magnitude into WCLMS, which can be seen in the next Table IV. It compares different WCLMS' of different order combinations, with the WCLMS columns of the previous table repeated as the first two columns here. The (200,80,40) order stands out as the best WCLMS order which beats all other order combinations and single LMS's, in terms of average coding rates. The scatter diagrams in figures 8 and 9 show that this improved performance in terms of the average coding rate is the result of an improved performance for almost all the 140 individual signal pieces in the database (including different types and different sampling rates). This is because the data points in the figures fall below the diagonal line, implying that the x-coordinates for LMS are larger than the y-coordinates for WCLMS. The two plots in Figure 9 show that for all signal pieces, with the same total order, the WCLMS with decreasing order is better than the one with increasing order: (200,80,40) as opposed to (40,80,200) for the first plot, and (100,60,40) for the second plot.

The time series plots in Fig. 10 give an idea about how the weights w_1, w_2, and w_3 change for one sample in the case of WCLMS (200,80,40). It is generated from a piece of the jazz signal with 32 kHz sampling rate in table I, and with 320,000 samples. Only samples 50000 to 51000 are shown here. A similar pattern holds for the entire 320,000 samples. Hence throughout the whole jazz piece, all three predictors have similar weights – none of them is dominant. This indicates that hard-switching among the three predictors would not achieve the same good prediction as the weighting in WCLMS.

So far we have evaluated the lossless compression unit. To give an impression of the performance of the combined system PPF-WCLMS in terms of bit-rate and audio quality, we compare it with a state-of-the-art audio coder, PAC, in mono-mode. We use a subjective listening test on a set of 10 test signals. The 10 test signals are chosen from a set of 73 signals by several experienced listeners to be particularly critical (coding artifacts are more pronounced) for both coders or either one. They consist of speech signals (mspeech, spot), single instruments (tink, castanet, triangle, oboe), music with several instruments (chart, jazz), and mixed speech and music (mixed). Both coders are used without an output bit-rate buffer, as they could be used for transmission channels with variable bit-rate (e.g. packet networks). Not using a bit-rate buffer is also helping our goal of a low encoding/decoding delay. We set both coders such that they use the same average bit-rate over the length of each individual signal. This is done by adjusting the attenuation factor in Fig. 4 for the PPF-WCLMS, or adjusting the target bit-rate for PAC. The adjustment is done such that the bit-rate is not too far from the starting point given by their psycho-acoustic model (for most signals this starting point is quite similar between them), and such that it is between 1.5 and 2.4 bits/sample. For our PPF-WCLMS coder this bit-rate includes the side information for the post-filter. Table V shows the used test signals and their corresponding bit-rates for both coders.

We use a test method called RAB test, as described by the ITU [30]. It is a comparison of a coded/decoded signal with the original signal in a triplet. The first signal of this triplet is the known original or reference. The following two signals in the triplet are called A and B, and one is the original or hidden reference, and the other the encoded/decoded signal, in a random order. The test subject is then asked to evaluate each, A and B, in comparison with the known original. For the evaluation the ITU five-grade impairment scale is used, with 5.0 meaning an imperceptible difference, 4.0 perceptible but not annoying, 3.0 slightly annoying, 2.0 annoying, 1.0 very annoying. For each of the 10 test pieces, we have two coded signals, and each coded signal is used in both orders of hidden reference and coded/decoded in the RAB triplet, leading to a set of 40 triplets for each subject to listen to. For each subject a different random order was used.

The differences between the original and our encoded/decoded signal are often very subtle.

Since expert listeners are more sensitive in detecting and more reliable in evaluating distortions than naive listeners, we use 5 expert listeners in our test. The listening test is conducted in a sound proof booth, and with STAX Lambda Pro headphones. The results are displayed in Fig. 11, where the difference grading is the difference between the grade a subject gives for the original and for the encoded/decoded signal (it is possible that the hidden reference is graded worse than the encoded/decoded signal). The circles show the average grading for the PPF-WCLMS coder, the squares the averages for the PAC coder. On 6 of the signals the PPF-WCLMS has higher averages, on 4 signals PAC has higher averages. It can be seen that for most signals there is no statistically significant difference in the evaluation of the two coders, since their 95%-confidence intervals overlap. Only for signal H (Oboe) there is a bigger difference, where PAC mono performs better. Since this is a mostly stationary signal, the difference might have been caused by the precision of the tonality estimation of our psycho-acoustic model. However, overall there is no statistically significant difference between the two coders. Recall that PPF-WCLMS has a delay of about 200 samples compared with the 2047+576=2623 sample delay of PAC. Hence we conclude that we can indeed significantly reduce our encoding/decoding delay without sacrificing quality or compression performance compared to traditional audio coders.

As a side note, we observe the bit-rate variation over time. We find that the variation of the bit-rate using PPF-WCLMS in general is more limited than for the PAC coder (before buffering). This is an effect of using the pre-filter. In traditional audio coding the switching to the 128 band mode causes the highest peaks in the bit-rate. This switching is not present with the pre-filter. The limited variation is an advantage for applications with fixed bit-rates, since it will only require smaller buffers and hence have smaller additional buffering delay.

VI. Conclusions

We presented a new approach to perceptual audio coding. It is based on the separation of the irrelevance reduction and redundancy reduction into separate functional units. They are connected by a full band audio signal, making the application of lossless audio coders possible. This separation enables us to optimize each unit independently, and to obtain a much lower encoding/decoding delay than traditional audio coders. For the irrelevance reduction unit, an adaptive pre- and post-filter, controlled by a psycho-acoustic model was used. This part has a delay of 128 samples, needed for the psycho-acoustic model. For the redundancy reduction unit, instead of the conventional transform coding, we applied WCLMS lossless predictive coding with advantages on both compression ratio and coding delay. At the expense of a moderate increase in computational complexity over other lossless coders, cascading LMS predictors and the weighting scheme from the predictive Minimum Description Length principle held the key to the good performance for the prediction of the WCLMS lossless unit and hence results in higher compression ratios. Despite of a higher sensitivity to transmission errors, backward adaptation (in contrast to block or forward adaptation) and adaptive entropy coding are responsible for the possibility of a low encoding/decoding delay. For a database of music, speech, and mixed mono-signals of moderate size and different sampling rates, the WCLMS(200,80,40) lossless coder shows improved performance relative to its lossless competitors in terms of bit rates and coding delays.

To date our implementation of the combined PPF-WCLMS perceptually lossless coder is for mono signals. At the same bit-rate it yields an audio quality comparable to the state-of-the-art coder PAC, as a subjective listening test showed. This shows that we pay no penalty in audio quality or compression performance by obtaining a much lower encoding/decoding delay than traditional audio coders. The increased design flexibility of our scheme can be used to obtain a low delay of around 200 samples or 6 ms at 32 kHz sampling rate for the combined PPF-WCLMS coder.

Possible future work includes the extension to multi-channel signals, increased robustness against transmission errors, especially at the bit-level, and lower complexity versions.

Acknowledgements: We would like to thank Serap Savari, Sean Dorward, Peter Kroon, Christof Faller, and Frank Baumgarte for their help. Bin Yu would like to acknowledge the partial support of grants FD98-02314, FD01-12731, and DMS-9803063 from the National Science Foundation and grants DAAG55-98-1-0341 and DAAD19-01-1-0643 from the Army Research Office.

References

[1] P. Noll, "MPEG digital audio coding standards," in *The Digital Signal Processing Handbook* (V. Madisetti and D. B. Williams, eds.), ch. 40, Boca Raton, Florida: CRC Press, IEEE Press, 1998.

[2] G. A. Davidson, "Digital audio coding: Dolby AC-3," in *The Digital Signal Processing Handbook* (V. Madisetti and D. B. Williams, eds.), ch. 41, Boca Raton, Florida: CRC Press, IEEE Press, 1998.

[3] K. Akagiri *et al.*, "Sony systems," in *The Digital Signal Processing Handbook* (V. Madisetti and D. B. Williams, eds.), ch. 42, Boca Raton, Florida: CRC Press, IEEE Press, 1998.

Order	LMS 10	40	80	200	400	WCLMS 40,80,200	WCLMS 200,80,40
44kHz							
chart44	2.05	1.98	1.98	1.93	1.92	1.87	1.81
jazz44	2.08	2.05	1.98	1.89	1.89	1.86	1.75
mspeech	2.11	2.06	2.07	2.09	2.10	1.99	1.98
spot44	2.01	1.99	2.00	2.01	2.01	1.90	1.88
32kHz							
chart	2.18	2.13	2.11	2.04	2.02	2.01	1.94
jazz	2.39	2.34	2.22	2.11	2.10	2.10	1.99
mixed	2.29	2.26	2.26	2.25	2.25	2.18	2.16
spot	2.07	2.05	2.06	2.07	2.09	1.99	1.96
16kHz							
chart	2.39	2.33	2.22	2.17	2.19	2.10	2.02
jazz	2.7	2.43	2.30	2.27	2.29	2.15	2.08
mixed	2.40	2.38	2.37	2.36	2.37	2.31	2.28
spot	2.32	2.33	2.33	2.32	2.35	2.25	2.21
8kHz							
chart	2.51	2.28	2.22	2.21	2.27	2.09	2.03
jazz	2.83	2.31	2.25	2.27	2.35	2.10	2.06
mixed	2.42	2.39	2.37	2.39	2.41	2.33	2.31
spot	2.40	2.39	2.37	2.41	2.44	2.33	2.32

TABLE I

THE RESULTING BIT-RATE IN BIT/SAMPLE FOR DIFFERENT FIXED LENGTH LMS PREDICTORS AND FOR TWO WEIGHTED CASCADED LMS OF DIFFERENT LENGTHS

	WCLMS 200,80,40	LPAC	LTAC	Sho.	WZ.
44kHz					
chart44	1.81	2.1	2.33	2.47	3.14
jazz44	1.75	2.09	2.30	2.45	3.14
mspeech	1.98	2.17	2.48	2.49	3.06
spot44	1.88	2.04	2.36	2.42	3.01
32kHz					
chart	1.94	2.23	2.36	2.52	3.22
jazz	1.99	2.48	2.42	2.67	3.35
mixed	2.16	2.35	2.59	2.58	3.19
spot	1.96	2.12	2.42	2.48	3.09
16kHz					
chart	2.02	2.50	2.56	2.68	3.42
jazz	2.08	2.64	2.57	2.85	3.48
mixed	2.28	2.50	2.80	2.67	3.23
spot	2.21	2.38	2.76	2.63	3.27
8kHz					
chart	2.03	2.58	3.10	2.89	3.67
jazz	2.06	2.34	3.04	3.11	3.77
mixed	2.31	2.56	3.37	2.78	3.46
spot	2.32	2.54	3.38	2.77	3.46
average	2.05	2.35	2.68	2.65	3.31

TABLE II

Comparison of the best weighted cascaded prediction and coding with other widely used lossless compression schemes, in bit/sample. Sho.: Shorten, WZ.: WaveZip.

Order	40	80	200	400	40,80,200	200,80,40	LPAC
44kHz							
mixed (n=4)	2.02	2.01	1.98	1.98	1.90	1.86	2.10
32kHz							
music (n=19)	2.13	2.09	2.05	2.04	1.99	1.93	2.23
voice (n=9)	2.07	2.08	2.09	2.10	1.99	1.97	2.15
16kHz							
music (n=45)	2.32	2.26	2.21	2.20	2.15	2.09	2.49
voice (n=9)	2.32	2.31	2.32	2.34	2.24	2.20	2.39
8kHz							
music (n=45)	2.34	2.26	2.22	2.24	2.17	2.10	2.59
voice (n=9)	2.34	2.34	2.37	2.40	2.28	2.27	2.48
Overall (n=140)	2.28	2.23	2.19	2.20	2.13	2.08	2.45

TABLE III

THE AVERAGE RESULTING BIT-RATE IN BIT/SAMPLE FOR DIFFERENT FIXED LENGTH LMS PREDICTORS, FOR WEIGHTED CASCADED LMS OF DIFFERENT LENGTHS, AND FOR LPAC AS A COMPARISON. n INDICATES THE NUMBER OF ITEMS IN A CATEGORY.

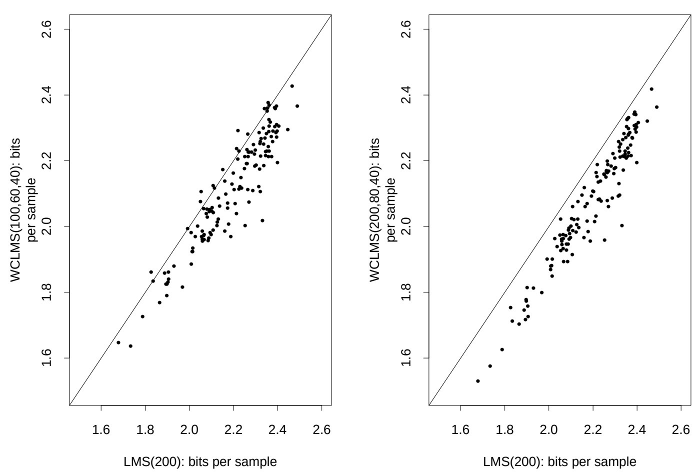

Fig. 8. Comparison of the bit-rates for LMS based prediction to WCLMS prediction for the signals in our test (each dot represents one signal).

Order	40,80,200	200,80,40	40,60,100	100,60,40	120,60,20	150,30,20
44kHz						
mixed (n=4)	1.90	1.86	1.92	1.90	1.88	1.87
32kHz						
music (n=19)	1.99	1.93	2.01	1.97	1.96	1.96
voice (n=9)	1.99	1.97	1.99	1.98	1.97	1.97
16kHz						
music (n=45)	2.15	2.09	2.18	2.14	2.13	2.13
voice (n=9)	2.24	2.20	2.23	2.20	2.19	2.17
8kHz						
music (n=45)	2.17	2.10	2.20	2.15	2.14	2.14
voice (n=9)	2.28	2.27	2.27	2.25	2.25	2.24
Overall (n=140)	2.13	2.08	2.15	2.11	2.10	2.10

TABLE IV

The average resulting bit-rate in bit/sample for weighted cascaded section of different lengths

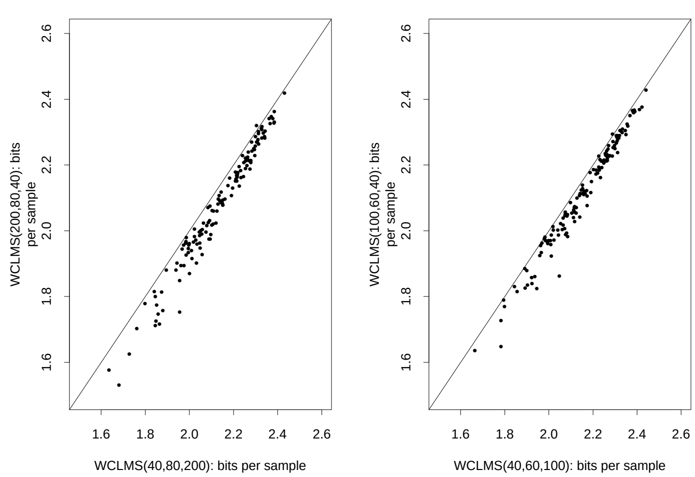

Fig. 9. Comparison of the effect of different ordering of the unequal size prediction segments (each dot represents one signal).

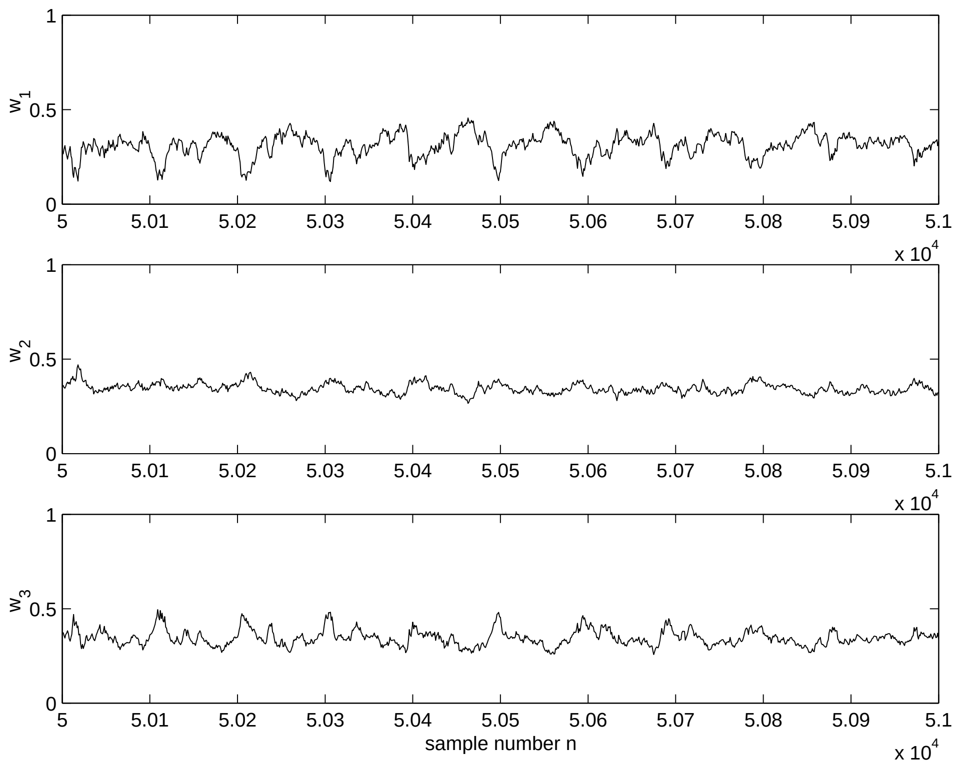

Fig. 10. Timeseries plots of w_1, w_2, w_3 values for the jazz signal.

Signal	Bit-rate
A tink	1.625
B chart	2.0625
C jazz	2.0625
D castanet	2.0625
E harps	1.8437
F mixed	2.375
G mspeech	2.3125
H oboe	1.5937
I spot	2.25
K triangle	1.5937

TABLE V

THE SIGNALS FOR THE SUBJECTIVE COMPARISON TEST AND THEIR BIT-RATE (IN BIT PER SAMPLE) FOR BOTH CODERS, AT 32 KHZ SAMPLING RATE.

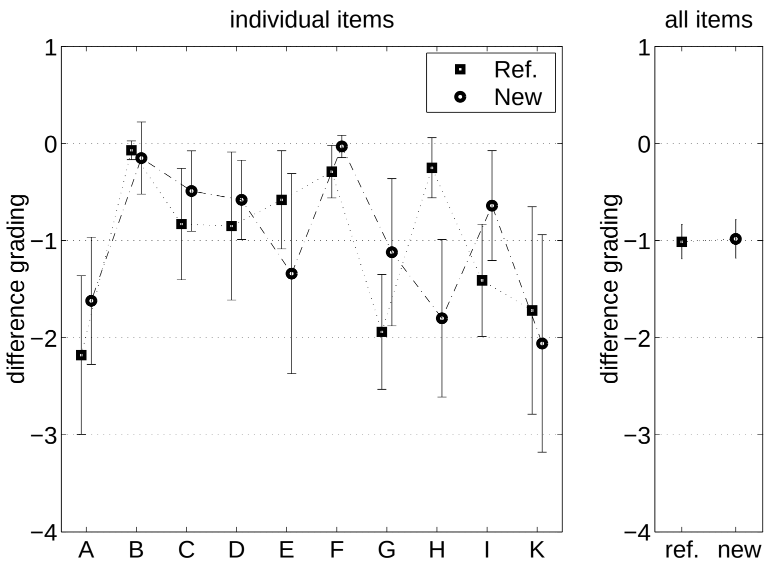

Fig. 11. The result of our listening test. "Ref." is the PAC mono coder, "New" is our PPF-WCLMS coder. The vertical bars around each value show the 95 % confidence interval.

[4] D. Sinha, J. D. Johnston, S. Dorward, and S. Quackenbush, "The perceptual audio coder (PAC)," in *The Digital Signal Processing Handbook* (V. Madisetti and D. B. Williams, eds.), ch. 42, Boca Raton, Florida: CRC Press, IEEE Press, 1998.

[5] G. Schuller, "Time-varying filter banks with variable system delay," in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, (Munich, Germany), pp. III–2469:2472, Apr. 1997.

[6] G. Schuller and T. Karp, "Modulated filter banks with arbitrary system delay: Efficient implementations and the time-varying case," *IEEE Transactions on Signal Processing*, vol. 48, pp. 737–748, March 2000.

[7] E. Allamanche, R. Geiger, J. Herre, and T. Sporer, "MPEG-4 low delay audio coding based on the AAC codec," in *106th AES Convention*, (Munich, Germany), May 1999.

[8] A. Harma and U. Laine, "Warped low-delay CELP for wideband audio coding," in *AES 17th Int. Conf.*, (Florence, Italy), pp. 207–215, Sep. 1999.

[9] K. Nitadori, "Linear transform coding and predictive coding," *IECE Japan*, Feb. 1970.

[10] N. Jayant and P. Noll, *Digital Coding of Waveforms.* Englewood Cliffs, New Jersey: Prentice Hall, 1984.

[11] B. Edler and G. Schuller, "Audio coding using a psychoacoustic pre- and post-filter," in *IEEE Int. Conf. Acoustics, Speech, Signal Processing*, (Istanbul, Turkey), pp. II–881:884, 2000.

[12] B. Edler, C. Faller, and G. Schuller, "Perceptual audio coding using a time-varying linear pre- and post-filter," in *AES Symposium*, (Los Angeles, CA), Sep. 2000.

[13] H. F. F. Baumgarte, C. Ferekidis, "A nonlinear psychoacoustic model applied to the iso mpeg layer 3 coder," in *99th AES Symposium*, (New York, NY), Oct. 1995.

[14] F. Itakura and S. Saito, "Digital filtering techniques for speech analysis and synthesis," in *7th Int. Congr. Acoustics*, 1971.

[15] U. K. Laine, M. Karjalainen, and T. Altosaar, "Warped linear prediction (WLP) in speech and audio

processing," in *IEEE Int. Conf. Acoustics, Speech, Signal Processing*, pp. III–349 – III–352, 1994.

[16] H. W. Strube, "Linear prediction on a warped frequency scale," *J. of the Acoust. Soc. Am.*, vol. 68, pp. 1071–1076, 1980.

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

[18] T. Liebchen, Technical University of Berlin, Germany, LPAC, version 0.99h, http://www-ft.ee.tu-berlin.de/~liebchen/lpac.html.

[19] T. Liebchen, Technical University of Berlin, Germany, LTAC, version 1.71, http://www-ft.ee.tu-berlin.de/~liebchen/ltac.html.

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

[21] M. Gerzon *et al.*, "The MLP lossless compression system," in *AES 17th Int. Conf.*, (Florence, Italy), pp. 61–75, Sep. 1999.

[22] S. Haykin, *Adaptive Filter Theory.* Englewood Cliffs, NJ 07632: Prentice Hall, 1999.

[23] B. Widrow and M. H. Jr., "Adaptive switching circuits," in *IRE WESCON*, pp. 96–104, 1960.

[24] P. Prandoni and M. Vetterli, "An FIR cascade structure for adaptive linear prediction," *IEEE Trans. Signal Processing*, vol. 46, No. 9, pp. 2566–2671, Sep. 1998.

[25] J. W. Tukey, *Exploratory data analysis.* Addison-Wesley, Reading, MA, 1977.

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

[27] A. Barron, J. Rissanen, and B. Yu, "The minimum description length principle in coding and modeling," *IEEE. Trans. on Information Theory (Special Commemorative Issue: Information Theory 1948-1998)*, vol. 44, pp. 2743–2760, 1998.

[28] M. Hansen and B. Yu, "Model selection and the principle of minimum description length," *J. Amer. Statist. Assoc.*, vol. 96, pp. 746–774, 2001.

[29] S. Dorward, D. Huang, S. Savari, G. Schuller, and B. Yu, "Low delay perceptual lossless coding of audio signals," in *submitted to IEEE Data Compression Conference*, (Snowbird, Utah), March 2001.

[30] ITU-R, *Methods for the subjective assessment of small impairments in audio systems including multichannel sound systems.* Geneva: Rec. ITU-R BS.1116-1, 1997.