HYBRID LPC AND DISCRETE WAVELET TRANSFORM AUDIO CODING WITH A NOVEL BIT ALLOCATION ALGORITHM

Simon D. Boland *Mohamed Deriche*

Signal Processing Research Centre
School of Electrical and Electronic Systems Engineering
Queensland University of Technology
GPO Box 2434 Brisbane Qld 4001
email : s.boland@qut.edu.au

ABSTRACT

This paper examines a new method for coding high quality digital audio signals based on a combination of Linear Predictive Coding (LPC) and the Discrete Wavelet Transform (DWT). In this method, a linear predictor is first used to model each audio frame. Then, the prediction error is analyzed using the DWT. The LPC coefficients and DWT coefficients are quantized using a novel bit allocation scheme which minimizes the overall quantization error with respect to the masking threshold. The proposed coder is capable of delivering near-transparent audio signal quality at encoding bitrates of around 90-96 kb/s. Objective and subjective results suggest that the proposed coder operating at 90-96 kb/s has a performance comparable to that of the MPEG layer II codec operating at 128 kb/s.

1. INTRODUCTION

Uncompressed high quality audio signals are typically sampled at 44.1 kHz and encoded with 16 bits/sample PCM, resulting in the large bitrate of 705 kb/s/channel. Some applications which use high quality audio are Digital Audio Broadcasting (DAB), ISDN, Internet audio, and HDTV. For these applications, *coding* or *compression* is necessary to reduce the bitrate of 705 kb/s/channel down to a much smaller bitrate, but without compromising the audio signal quality.

A few previous audio coding schemes have proposed using LPC methods which are similar to those used in speech coding [1]-[3]. These LPC schemes are based on the well-known *Analysis-By-Synthesis (ABS)* coding technique [5]. In ABS schemes, an optimization is performed on the excitation to the filter $1/A(z)$ to minimize the weighted error between the original speech, $x(n)$, and reconstructed speech, $\hat{x}(n)$. For Multi-Pulse Excitation (MPE) and Regular Pulse Excitation (RPE) ABS schemes [5], the excitation to $1/A(z)$ is approximated by a set of pulses. In Code Excitation Linear Prediction (CELP), the excitation is found by searching a codebook of entries either randomly selected, or trained using a representative collection of speech data. Using the MPE, RPE and CELP methods, some promising results for high quality audio coding have been published in [1]-[4].

[0]This work was supported by grants from ATERB, ARC small grant scheme, and the QUT research and encouragement award scheme.

Overall though, the previous LPC-based methods have not exploited the masking properties of human hearing to the same extent that subband and transform coders do. As a result, the major limitation of the ABS schemes for audio coding is that the quantization noise cannot be as accurately controlled. Thus the coding errors do become audible at low bitrates. Some attempts to improve the performance of ABS-based techniques have been made by first performing a subband decomposition of the audio signal, and then modelling each subband signal using RPE or MPE [2]-[3]. By distributing proportionally less bits to the higher frequency subbands, a slightly lower bitrate is possible compared to an ABS coding on the full-spectrum signal. In [6], however, slightly inferior results for CELP speech coding were obtained when a subband decomposition was included in the codec design. Based on this result and our own experiments in [3], we found that subband-ABS schemes have only minor advantages over the standard ABS approach.

These limitations have meant that LPC-based methods for audio coding have, so far, been much less successful compared to those based on the traditional subband and transform coding techniques. In this paper, however, we show that LPC can be successfully applied to the problem of low bitrate audio coding. The proposed method overcomes the limitations of the ABS and subband ABS methods. By analyzing the LPC residual using the Discrete Wavelet Transform (DWT), the proposed method is able to finely shape the quantization noise spectrum. This is performed using a novel bit allocation algorithm which minimizes the difference between the quantization noise and the masking threshold. Herein, we refer to the proposed codec as the Linear Predictive Coding-Discrete Wavelet Transform (LPC-DWT) codec.

The outline of the paper is as follows. In section 2, a description of the proposed LPC-DWT audio codec is given. In section 3, the derivation of the error PSD and bit allocation scheme are explained. In section 4, the quantization of the coder parameters is detailed. Section 5 provides both objective and subjective audio coding results, and whereupon the conclusions are drawn in section 6.

2. PROPOSED LPC-DWT CODEC

2.1. LPC Analysis

The advantage of performing an LPC analysis is that audio signals with "peaky" spectra are well modelled. The disadvantage of the LPC analysis, however, is that noise-like

signals may not be as well modelled. The proposed LPC-DWT codec exploits the advantage of LPC, and at the same time, it overcomes the disadvantages. This is done by performing a DWT analysis on the LPC residual. Since the DWT contains a nonuniform time-frequency resolution, it is well suited to representing the noise-like LPC residual. It is also useful because the frequency resolution mimics the filter bank operation in the human auditory system.

The block diagrams of the proposed LPC-DWT encoder and decoder are shown in figures 1(a) and 1(b) respectively. In the encoder, the original signal, $x(n)$, is processed frame-by-frame with frame lengths of 512 samples, which is approximately 12 ms for 44.1 kHz sampled signals. For the minimization of blocking artefacts, a trapezoidal window was found to give better results compared to a rectangular window. Hence a trapezoidal window is applied to each frame, and a frame overlap of 43 samples is used. At the decoder, each frame is reconstructed by performing an overlap-and-add of the first 43 samples of the current frame with the last 43 samples of the previous frame.

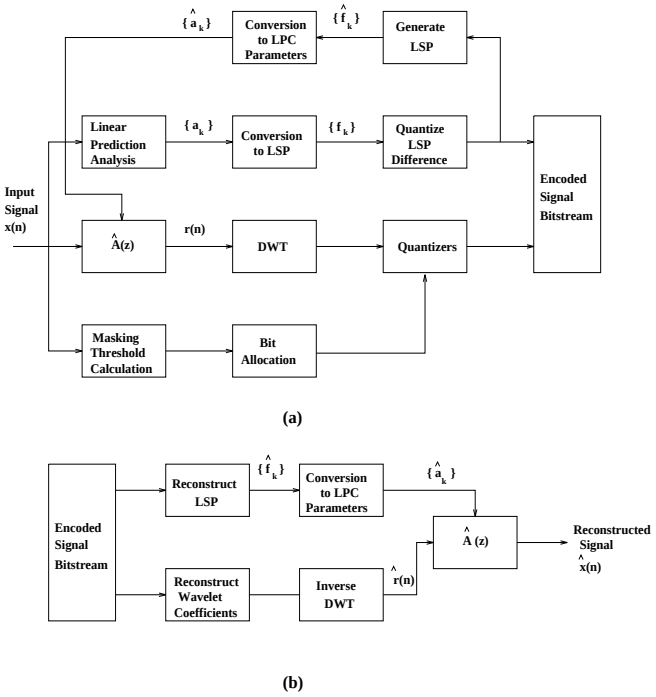

Figure 1. Block diagram of proposed LPC-DWT (a) encoder and (b) decoder

At the encoder, an LPC analysis is performed on the windowed signal, and the LPC parameters a_1, a_2,..., a_p are estimated. Quantization is performed by first converting the LPC parameters to Line Spectral Pairs (LSP) f_1, f_2,..., f_p, and afterwards, the differences between each of the parameters are quantized. The windowed signal is then passed through the quantized LPC filter, $\hat{A}(e^{j\omega})$, to obtain the LPC residual, $r[n]$. The transfer function of the quantized LPC filter is

$$\hat{A}(e^{j\omega}) = 1 - \sum_{k=1}^{p} \hat{a}_k e^{-j\omega k} \qquad (1)$$

where $\{\hat{a}_k\}$ are the quantized LPC parameters. In the Fourier domain, the expression for the LPC residual is given by

$$R(e^{j\omega}) = X(e^{j\omega})\hat{A}(e^{j\omega}) \qquad (2)$$

The LPC parameters, $\{a_k\}$, are estimated using the Autocorrelation method. A prediction order of 16 was selected after studying the performance of the *prediction gain* versus the predictor order [7].

2.2. DWT filtering of the LPC residual

The LPC residual signal, $r(n)$, is analyzed and encoded using the DWT. The wavelet coefficients are then quantized so that the error between the original signal, $x(n)$, and reconstructed signal, $\hat{x}(n)$, is shaped underneath the masking threshold. Hence, unlike the majority of the previous ABS and subband ABS methods, the quantization noise is less likely to be audible with the proposed codec. The masking calculations used in the LPC-DWT codec are based on the psychoacoustic model 2 of the MPEG codec [8]. Note that performing a DWT on the LPC residual was considered for a CELP speech coding scheme in [9]. For audio coding, however, there has been no previous work using this approach.

In the proposed coder, the LPC residual is analyzed using a 3-stage cascade of 4-band uniform filter banks. The 22050 Hz bandwidth residual signal is first decomposed into 4 bands which are each 5513 Hz in width. The low pass output (0-5513 Hz) is then decomposed further into 4 bands, and the next low pass output (0-1378 Hz) is decomposed into another 4-bands. Therefore the frequency resolution of the filter bank varies from 345 Hz up to 5513 Hz. This filter bank structure was chosen since it roughly mimics the *critical bands* [4].

For the subband filters, 32-tap PR and Linear Phase subband filters were designed using the method in [10]. This method designs the subband filters by iteratively solving a set of linear equations which are constructed from the time-domain constraints on the filters. We used this method because filters with the desired characteristics could be designed for easily. The desirable characteristics of the subband filters include Linear-Phase, high sidelobe attenuation, and the Perfect Reconstruction (PR) property. From extensive simulations, the overall codec performance did not improve significantly when the number of taps was increased above 32.

3. BIT ALLOCATION ALGORITHM

After computing the masking threshold for each frame, the next step is to allocate the available bits to the subbands. Denoting the error between the original and reconstructed signals as $e(n)$, the Fourier Transform of the error is $E(e^{j\omega}) = \hat{X}(e^{j\omega}) - X(e^{j\omega})$. In figure 1(a), the original signal $x(n)$ is filtered through the quantized LPC filter $\hat{A}(e^{j\omega})$. This LPC residual $r(n)$ is then analyzed using the DWT. At the decoder, an approximation to LPC residual is obtained and this is denoted in figure 1(b) as $\hat{r}(n)$. Hence the Fourier Transform of the error can be written as :

$$E(e^{j\omega}) = \frac{1}{\hat{A}(e^{j\omega})}\left[\hat{R}(e^{j\omega}) - R(e^{j\omega})\right] \quad (3)$$

The term in brackets is the filter bank error term. Denoting this by $E_{FB}(e^{j\omega})$ then $E(e^{j\omega}) = E_{FB}(e^{j\omega})/\hat{A}(e^{j\omega})$.

For transparent coding, the PSD of the overall error between the original and reconstructed signal, denoted here as $S_E(e^{j\omega})$, must be below the masking threshold, which we define here as $T(e^{j\omega})$. The PSD of the overall error is given as :

$$S_E(e^{j\omega}) = E(e^{j\omega})[E(e^{j\omega})]^* \quad (4)$$

Hence the PSD of the overall error becomes

$$S_E(e^{j\omega}) = S_{FB}(e^{j\omega})|H(e^{j\omega})|^2 \quad (5)$$

where $H(e^{j\omega}) = 1/\hat{A}(e^{j\omega})$. Since filters obeying the PR requirements are used here, the only filter bank error is from the quantization of the wavelet coefficients, the PSD of the overall error becomes :

$$\hat{S}_{FB}(e^{j\omega}) = \sum_{i=0}^{M-1} q_i^2 \mid F_i(e^{j\omega}) \mid^2 \quad (6)$$

where q_i^2 is the variance of the quantization noise for subband i, M is the number of subbands and $F_i(e^{j\omega})$ is the Fourier Transform of the synthesis filter i. For uniform quantization of each subband, the quantization noise variance is $q_i^2 = \Delta_i^2/12$, where Δ_i is the step-size for a B-bit quantizer with peak-to-peak quantizer range $2X_{max}$ and $\Delta_i = 2X_{max}/2^B$. Hence to obtain transparency, it is necessary to have :

$$\sum_{i=0}^{M-1} \frac{\Delta_i^2}{12} \mid F_i(e^{j\omega}) \mid^2 |H(e^{j\omega})|^2 < T(e^{j\omega}) \quad (7)$$

Using (7) , bits are allocated to the DWT subbands using an iterative approach. For each iteration, one bit is allocated to the subband with the smallest *Mask-to-Noise Ratio (MNR)*. The MNR of each subband is defined as the difference between the minimum masking threshold, and the quantization noise power. Thus, the bit allocation procedure aims to minimize the audibility of the overall error between the input and output signals.

4. QUANTIZATION

4.1. Quantization of LPC parameters

In speech coding, the most popular transformation used for the quantizating the LPC parameters are the *Line Spectrum Pairs (LSP)* [7]. For speech coding, the number of bits necessary for encoding the LSPs can be reduced significantly if the differences between the adjacent LSP frequencies, i.e. $df_k = f_{k+1} - f_k$, are encoded. This is since LSP frequency differences have a lower variance than the absolute frequencies. The same approach was followed for the quantization of the LSP parameters here.

The number of bits and the quantizer bounds must be chosen carefully so as to minimize both the overload distortion and the granular distortion. Compared to speech, the LSP differences for audio exhibit a greater variance. To minimize the effects of both overload and granular distortions, an adaptive quantization scheme is used. For this method, 4 possible regions were chosen empirically after studying the distributions of the LSP differences. These were **(1)** $0 \leq df_k < 0.025$, **(2)** $0.025 \leq df_k < 0.050$, **(3)** $0.050 \leq df_k < 0.100$, and **(4)** $0.100 \leq df_k < 0.175$.

The quantizer bounds for the LSP differences, df_k, depend on which region the differences lie in. All of the differences are quantized with 5 bits. Therefore the differences falling in the first 2 regions are quantized more accurately than the other 2. The reasoning is that the $df_k \geq 0.05$ mostly occur for frames with transient-like behaviour. This approach improves on a simple uniform quantizer over the complete range of the LSP differences, but it comes at the cost of an extra 2 bits/coefficient for the region information. Hence 7 bits/coefficient are needed for the quantization of the LPC parameters.

4.2. Quantization of Wavelet Coefficients

For each subband of the LPC residual $r(n)$, a *scalefactor* is extracted using a similar method to the MPEG codec. Each subband scalefactor is computed by finding the power of two which is just greater than the maximum of the absolute value of the subband vector. The scalefactor is quantized by computing $log_2(gain)$, and we found that 5 bits were necessary to represent the dynamic range of each scalefactor. Each subband is then divided by its respective scalefactor to produce values ranging between -1 and 1. The normalized coefficients are uniformly quantized using the number of bits determined from the bit allocation algorithm. Rate reduction is then obtained by Huffman coding the quantized scalefactors, and also the normalized coefficients. This information is sent to the decoder, together with bit allocation side information of 4 bits/subband.

Compared to the subbands from $0-11$ kHz, the subbands from $11-22$ kHz of the LPC residual do not contribute as significantly to the overall quality of the reconstructed signal. However, if the wavelet coefficients of the $11-22$ kHz subbands are ignored and set to zero in the decoder, the result is a lowpassed, muffled audio quality. A significant improvement in the audio quality was possible if the two subbands from $11-22$ kHz are set to random noise. Using a normally distributed random number generator with zero mean and unit variance, good results were obtained if the noise was multiplied by 1 % of the standard deviation of the original wavelet coefficients. For the two $11-22$ kHz subbands, this information is sent to the decoder using 10 bits/subband.

5. CODER ASSESSMENT

The source signals used for the coder evaluations are taken from the European Broadcasting Union SQAM CD. Each signal was a monophonic recording sampled at 44.1 kHz with 16 bits/sample. Comparisons to original source material were made for signals encoded and decoded with the LPC-DWT. The MPEG layer II codec was selected as a benchmark. The encoding bitrate chosen for the MPEG codec was 128 kb/s. For the LPC-DWT codec, a total bitrate of 112 kb/s was firstly selected to encode each signal. Huffman coding of the quantized DWT coefficients reduced this bitrate down to an average variable bitrate of 90-96

kb/s. The quantized LPC information accounts for 10.5 kb/s, while the remainder is due to the quantization of the DWT coefficients. For each signal, the performance of the LPC-DWT and MPEG layer II codecs are assessed using both objective and subjective measures.

5.1. Objective Measurement

For each of the test signals, the Segmental Signal-to-Noise Ratio (Seg.SNR), and the *Generalized Bark Spectral Distortion (GBSD)* are given in table 1. It is well-known that an increase in the Seg. SNR does not always correlate with an increase in the coder performance. For this reason, the GBSD measure was also included for an objective measurement. The GBSD is a perceptually-motivated objective measure which compares the spectral difference between the original and coded signals using the Bark scale [4]. In table 1, a smaller GBSD value correlates with an increase in the perceived quality of the coded signal.

Similar Seg. SNR values and GBSD values were recorded for both codecs. From our experiments, the maximum Seg. SNR difference between the two codecs is 3 dB, while the maximum GBSD difference is less than an order of magnitude. A reliable conclusion cannot be made based solely on the Seg. SNR values because of its limitations for assessing audio codec performance. However, based on the similar results of both the Seg. SNR and the GBSD, we conclude that the two codecs perform nearly the same in achieving high quality audio at low bitrates. Note this is despite the lower bitrate offered by the proposed codec (about 30-40 kb/s less).

5.2. Informal Subjective Measurement

As well as objective measures, informal listening tests were performed by using the *A-B-C* double blind stimulus. In this sequence, A is the original source, while B and C could be one of either, **(1)** the identical original source or **(2)** the same signal after encoding-decoding. Separate tests were performed for the MPEG layer II codec and the proposed LPC-DWT codec. For each codec and for each test signal, listeners were asked to compare the quality of the two coded signals, B and C using a 41-point impairment ranging from 1.0 up to 5.0 [11]. A value of 5.0 is given to the signal, B or C, which the listener believes to be the original. The signal which the listener believes to be the coded signal is assigned a value from 1.0 to 4.9. The impairments ratings range from 5.0 for impairments that are imperceptible, down to 1.0 for impairments that are very annoying.

Eight subjects, including an expert listener, were selected for the informal listening tests. The tests were conducted with headphones in a quiet office environment. All of the subjects were researchers working in speech or audio processing related fields. Consequently all subjects have experience to some degree in identifying degradations in audio recordings. Given in table 1 are the Mean Opinion Scores (MOS) obtained for the LPC-DWT codec and MPEG layer II codec. From the informal listening tests, we conclude the listeners found the proposed LPC-DWT codec to be similar in performance to the MPEG layer II codec. Again note that this is despite the much lower bitrate offered by the proposed method.

Audio Signal	LPC-DWT (90-96 kb/s)	MPEG (128 kb/s)
Eddie Rabbit	$22/8.2\times10^{-10}/4.93$	$22/2.5\times10^{-9}/4.93$
Castanets	$16/5.4\times10^{-9}/4.00$	$15/1.4\times10^{-9}/4.43$
Female Speech	$23/9.9\times10^{-9}/4.61$	$24/1.7\times10^{-9}/4.85$
Male Speech	$25/6.5\times10^{-10}/4.09$	$26/9.4\times10^{-10}/4.38$
Triangle	$24/1.0\times10^{-9}/4.65$	$26/7.9\times10^{-10}/4.95$
Guitar	$27/3.4\times10^{-10}/3.91$	$25/9.6\times10^{-10}/4.43$
Violoncello	$29/3.4\times10^{-10}/4.56$	$28/4.4\times10^{-10}/4.73$

Table 1. SegSNR(dB)/GBSD/Informal MOS performance of LPC-DWT and MPEG layer II.

6. CONCLUSIONS

The similar objective and subjective performance of the LPC-DWT and MPEG layer II codecs, suggest that excellent signal quality is possible with the proposed codec. This conclusion is made after comparing the small differences in the recorded Seg. SNR and GBSD measures, and the equally small differences in the recorded informal MOS. A couple of ways for obtaining even lower bitrates with the proposed codec may be possible with further work. The first possibility is the investigation of a more sophisticated scheme for the quantization of the LPC parameters. Nonuniform quantization and entropy coding of the parameters may result in even lower bitrates. Similarly, a more sophisticated approach for quantizing the upper subbands of the LPC error could lead to an improved coded signal quality.

REFERENCES

[1] S. Singhal, "High quality audio coding using multipulse LPC," *Proc. ICASSP*, pp. 1101-1104, 1990.

[2] X. Lin et al, "Subband-multipulse digital audio broadcasting for mobile receivers," *IEEE Trans. Broadcast.*, 39:4:373-382, 1993.

[3] S. Boland and M. Deriche, "High quality audio coding using multipulse LPC and the wavelet transform," *Proc. ICASSP*, pp. 3067-3069, 1995.

[4] W. Chang and C. Wang, "A masking-threshold adapted weighting filter for excitation search," *IEEE Trans. SAP*, 4(2):124-132, 1996.

[5] P. Kroon and E. F. Deprettere, "A class of analysis-by-synthesis predictive coders for high quality speech coding at rates between 4.8 and 16 kbits/s," *IEEE J. Sel. Areas in Commun.*, 6(2):353-363, February 1988.

[6] A. Benyassine and A. N. Akansu, "Subspectral Modeling in Filter Banks," *IEEE Trans. Signal Process.*, 43(12):3050-3053, 1993.

[7] R. Steele (editor), *Mobile Radio Communications*, Pentech Press, London, 1992.

[8] ISO/IEC JTC1/SC29/WG11 MPEG, International Standard IS 11172-3. Coding of moving pictures and associated audio for digital storage media at up to about 1.5 Mbit/s, Part 3: Audio.

[9] J. Ooi and V. Viswanathan, "A computationally efficient wavelet transform CELP coder," *Proc. ICASSP*, pp. 101-104, 1994.

[10] M. Ikehara and T. Q. Nguyen, "Time-Domain Design of Linear-Phase PR Filter Banks," *Proc. ICASSP*, pp. 2077-2080, 1997.

[11] T. Grusec, L. Thibault and G. Soulodre, "Subjective evaluation of high quality audio coding systems : methods and results in the two-channel case," *New York AES convention 1995*, preprint 4065.