HIGH QUALITY AUDIO CODING USING MULTIPULSE LPC AND WAVELET DECOMPOSITION

S. Boland *M. Deriche*

Signal Processing Research Centre
Queensland University of Technology
GPO Box 2434 Brisbane Q 4001, AUSTRALIA

ABSTRACT

Most current work in the area of high quality audio coding falls under one of two categories: transform or sub-band coding. LPC coders since based on modelling human voice production systems are found to be inappropriate in modelling music and other non-speech sounds. A more improved model for such signals is shown to be the Multipulse LPC model. In this paper we propose to improve the quality of the Multipulse model by first passing the signal of interest through a filter bank and then extracting the Multipulse parameters from each of the bandpass filter outputs. The idea of the wavelet decomposition is utilised for the design of the filter bank. Both the Multipulse model and the wavelet decomposition are well known. But a combination of both has not been exploited yet. This combination is expected to lead to a new way in high quality low bit rate audio coding.

1. INTRODUCTION

Without compression, the bit rate needed for transmission or storage of compact disk (CD) quality audio is 705 kb/s (16 bit PCM and 44.1 kHz sampling frequency). Emerging applications such as digital audio broadcasting, multimedia/hypermedia and satellite TV require high quality audio at low bit rates. Several attempts have been investigated for "transparent" coding of CD quality digital audio at low bit rates. Most previous methods use either subband coding [1] or transform coding [2]. It is known that the MPEG layer III audio coder is capable of transparent or near transparent quality of monophonic CD audio signals at 64 kb/s. For the above applications, however, further reductions in the total bit rate is desirable. The aim of this research is to achieve lower bit rates without loss of quality in the reconstructed audio signals.

Recently, the wavelet transform has been investigated for high quality audio coding [3]. Unlike the Fourier transform, the wavelet transform is characterised by a constant relative bandwidth or "constant-Q". This is attractive for coding audio signals since audiological research has shown that the human ear behaves in a similar way. The ear integrates sounds over ranges of frequencies called "critical bands". Below 500 Hz, the critical bands are 100 Hz, while above 500 Hz, the critical bands are approximately a third octave. Bit rates of 48-66 kb/s are claimed using the wavelet transform based coder proposed in [3]. While this coder is capable of achieving lower bit rates than previous methods, the drawback of this design is the optimisation procedure that is computationally very expensive and not guaranteed to converge to the optimum solution. Another drawback is the large search of the dynamic dictionary.

Linear Predictive Coding (LPC) based coders are generally considered unsuitable for coding music and other non-speech sounds since they rely on human vocal tract models. For speech signals, LPC coders use a time-varying linear predictor to approximate the original signal. For voiced speech, the linear predictor is driven with a train of impulses separated by the pitch period, while for unvoiced speech, the excitation is white noise. *Multipulse* LPC coders, however, use an analysis-by-synthesis technique where the pulse amplitudes and pulse positions of the excitation signal are computed for each frame. Originally developed for speech coding, work performed in [4] has shown that Multipulse LPC is equally applicable for coding CD-quality audio signals down to 128 kb/s. Also a hybrid scheme consisting of a modified Multipulse LPC algorithm and a subband decomposition was proposed in [5].

In this paper we propose a hybrid audio coder using a combination of Multipulse LPC and wavelet like decomposition of the audio signal. This has not been exploited previously and is expected to achieve low bit rates while preserving high quality.

2. WAVELET TRANSFORM

In [8], Daubechies showed that in the space of square integrable functions, a signal $f(t)$ can be represented by translates and dilations of a single wavelet $W(t)$ as

$$\sum_{j=J}^{\infty}\sum_{k=-\infty}^{\infty}\sqrt{2^j}b(j,k)W(2^jt-k)+\sum_{k=-\infty}^{\infty}\sqrt{2^J}a(J,k)g(2^Jt-k) \tag{1}$$

where $b(j,k) = \int f(t)\sqrt{2^j}W(2^jt-k)dt$ and $a(j,k) = \int f(t)\sqrt{2^j}g(2^jt-k)dt$. This expansion provides a multiresolution decomposition of the signal $f(t)$. The coefficients $b(j,k)$ represent details of the original signal at different levels of resolution j and coefficients $a(J,k)$ represent an approximation of the original signal $f(t)$ at resolution J. The wavelet $W(t)$ is obtained from a scaling function $g(t)$

0-7803-2431-5/95 $4.00 © 1995 IEEE

as

$$W(t) = \sum_{k=0}^{K-1} (-1)^k c_{1-k} g(2t-k) \quad (2)$$

where the c_k are the coefficients that define the scaling function, $g(t)$, which obeys the dilation equation given by:

$$g(t) = \sum_k c_k g(2t-k) \quad (3)$$

To construct the wavelet $W(t)$, the coefficients c_k must satisfy certain conditions. In most cases, attention is restricted to wavelets with compact support i.e. c_k is nonzero only for $0 \leq k \leq K-1$.

Mallat [9] has shown that the discrete wavelet transform can be implemented by a recursive algorithm. This is done by using the coefficients c_k as the low pass filter coefficients of a pair of quadrature mirror filters. The output of the low pass filter, G, is the approximation of the input signal for that level. While the output of the high pass filter, H, is the detail for that level. The impulse response of the low pass filter $\{c_k\}$ and that of the high pass filter $\{d_k\}$ are related through the equation,

$$d_k = (-1)^k c_{1-k} \quad (4)$$

In contrast to the short-time Fourier transform (STFT) where the bandwidth of the bandpass filters used for the decomposition is constant, the bandwidth of the WT bandpass filters is proportional to the central frequency or equivalently, the filter's quality factor is independent of the frequency (see Figure 1).

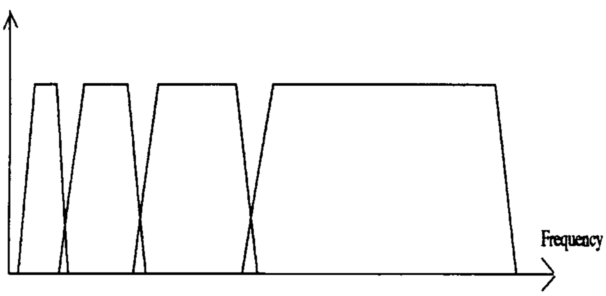

Figure 1. Constant Q Behaviour of the Wavelet Transform

3. MULTIPULSE LPC CODERS

The Multipulse coders assume that the given input of interest can be modelled by an all-pole filter driven by a train of pulses of different amplitudes and not necessarily equidistant from each other. The excitations (amplitudes and positions) are computed using an analysis-by-synthesis procedure such as the one described in [7] (see Figure 2).

The error signal between the original signal and the output of the all-pole filter excited by the train of pulses is sent back to the excitation generator. The excitation positions and amplitudes are evaluated through the minimisation of the energy of the error signal. Note that since the Multipulse LPC procedure is preceded by a frequency decomposition, the perceptual weighting of the error signal is not crucial.

Table 1. Chosen Subbands

Band (i)	*Frequency Range (Hz)*
1	0-5513
2	5513-11025
3	11025-22050

4. HYBRID AUDIO CODER

In subband coding, the frequency ranges of the decomposed subbands cannot be made the same as the critical bands. However, the discrete wavelet transform (DWT) and in particular, the wavelet packet transform (WPT), can be used to obtain a subband decomposition very close to the critical band divisions [3]. The advantage here is that the perceptibility of the quantisation noise can be controlled more accurately. The constant Q property of the wavelet transform is exploited in the proposed coder by firstly decomposing the audio signal into non-equal subbands. The full wavelet decomposition is not performed. Instead, the audio signal, which is sampled at 44.1 kHz with 16 bits/sample PCM, is decomposed into the three subbands listed in table 1. This is similar to the speech compression technique proposed in [6].

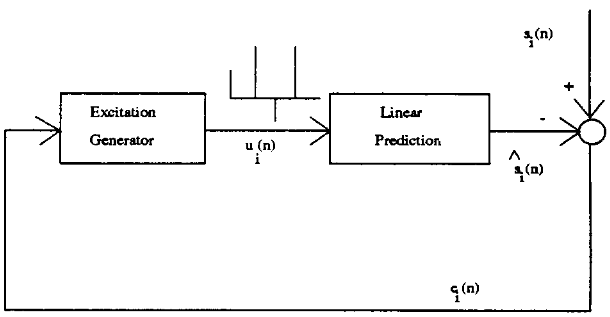

Figure 2. Block Diagram of Multipulse LPC Coder for subband i

After filtering the audio signal into subbands, each of the subbands signals, i, is modelled using the modified Multipulse LPC coder illustrated in figure 2. In the Multipulse coder, the signal is modelled by a linear filter driven by an excitation consisting of a sequence of pulses. The pulse amplitudes and locations for each excitation signal, u_i, are computed using a modified version of the algorithm proposed in [7]. The algorithm is based on the minimisation of the mean square error between the original subband signal, s_i, and the synthetic signal, $\hat{s_i}$. For each subband, the energy, the LPC coefficients, the pulse amplitudes and the pulse locations are determined, quantised, then sent to the receiver.

The advantage of decomposing the audio signal into subbands is that a different number of pulses and parameter set can be used for modelling each of the subband signals. This is opposed to the coder in [4] where the same number of pulses and LPC order are used for the entire audio spectrum. Also since the filter bank resembles the discrete wavelet transform, the advantages of the wavelet decomposition for processing audio signals are obtained.

5. SELECTION OF AUDIO CODER PARAMETERS

The quality of the reconstructed audio signal and the overall bit rate required are dependent upon several parameters. A 768 point frame was found to be the best choice as input to the filter bank. This was done after investigating frame sizes of 560, 768 and 960 points. Each frame is multiplied by a hamming window, decomposed into subbands and then moved 256 points each time. Decimation was used in the wavelet decomposition to give 192, 192 and 384 point subband signals in the 0-5.5 kHz, 5.5-11 kHz and 11-22 kHz subbands respectively. The excitation was computed for the middle 64 points in the 0-5.5 kHz and 5.5-11 kHz subbands, and the middle 128 points in the 11-22 kHz subband.

The quality of the reconstructed audio signal was found to largely depend upon the excitation pulse density in the 0-5.5 kHz and 5.5-11 kHz subbands. For high quality audio, 10-15 pulses were needed in both the subbands. When more than 15 pulses were used, only a very small difference in both the subjective quality and the average segmental SNR was recorded. Due to auditory masking, the excitation pulse density of the 11-22 kHz subband had a very small bearing on the quality. Consequently 10 pulses or less were typically used in the 11-22 kHz subband.

Experiments were also performed with 10th, 16th and 24th order LPC filters to investigate the effect of the predictor order. In all cases the predictor order was found to have little effect on the quality and the segmental SNR of the reconstructed audio signal. A 10th order LPC filter was selected for each subband signal, in order to obtain the lowest possible bit rate.

For quantisation of the LPC coefficients, initial experiments backed by the results in [4] suggest that differential quantisation is most appropriate. A similar approach to [4] has been adopted in quantising the pulse positions and pulse amplitudes. The actual value of the first pulse location is uniformly quantised. The remaining locations are uniformly quantised as differences from the previous location. A differential logarithmic quantiser is used to quantise a *gain* value per frame. This is the magnitude of the largest pulse in a frame. The individual pulses are then encoded with a sign bit and as a fraction of the gain value (in dB). Other possible schemes for quantising the LPC and excitation parameters are being investigated also.

6. DISCUSSION

For our experiments, the proposed audio coder was tested for several seconds of piano, bass, drums, guitar and orchestral signals. Listening tests using headphones were used to assess the quality of the reconstructed audio signals. For high quality, the bit rates of the proposed coder were found to be in the range of 80-90 kb/s. To obtain lower bit rates further areas in the coder design are currently being considered.

At this stage more work has to be performed to determine the optimal number of subbands and the frequency divisions of the subbands. In our initial experiments the three band decomposition was found to give better results than the two- and the four-band decompositions. Other decompositions are currently being investigated with respect to audio quality for a certain bit rate. However, it was found that as the number of subbands increased above five, the quality began to decrease significantly.

7. CONCLUSION

The proposed audio coder takes advantage of the "constant Q" properties of the wavelet transform by filtering the audio signal into subbands. Combined with Multipulse LPC, this is an original design for coding high quality digital audio signals. This algorithm is easy to implement and currently the proposed audio coder is being intensively tested for a variety of audio signals. Complete results will be presented at the conference.

8. ACKNOWLEDGEMENT

This work was partially supported by a grant from the Australian Telecommunications and Electronics Research Board (ATERB).

REFERENCES

[1] K. Brandenburg and G. Stoll, "The ISO/MPEG-Audio Cod ec: A generic standard for coding of high quality digital audio", Presented at 92nd AES Convention, Vienna, Austria, March 1992.

[2] J. Johnston, "Perceptual transform coding of wideband stereo signals", Proc. ICASSP, pp.1993-1995, 1989.

[3] D. Sinha and A. Tewfik, "Low bit rate transparent audio compression using adapted wavelets", IEEE Transactions on Signal Processing, pp.3464-3479, vol. 41, no. 12, 1993.

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

[5] X. Lin and R. Steele, "Subband coding with modified multipulse LPC for high quality audio", Proc. ICASSP, pp. 201-204, 1993.

[6] S. Park, "Speech compression using ARMA model and the wavelet transform", Proc. ICASSP, pp. 209-212, 1994.

[7] S. Singhal and B. Atal, "Amplitude optimization and pitch prediction in multipulse coders", IEEE Transactions on Acoustics, Speech and Signal Processing, pp.317-327, vol.37, no.3, 1989.

[8] I. Daubechies "Orthonormal bases of compactly supported wavelets", IEEE Transactions on Information Theory, Vol. 36, pp. 961-10 05, September, 1990.

[9] S. Mallat, "A theory for multiresolution signal decomposition : The wavelet representation", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 11, no. 7, 1989.