SAMPLE-BY-SAMPLE GAIN ADAPTIVE CELP CODING OF WIDEBAND AUDIO

Man-Tak Chu and Cheung-Fat Chan
Department of Electronic Engineering
City University of Hong Kong
83, Tat Chee Avenue, Hong Kong
Email : eecfchan@cityu.edu.hk

ABSTRACT

This paper presents a high quality wideband audio coder based on a low delay code excited linear predictive (LD-CELP) model where the excitation gain is adapted in a sample-by-sample manner. The proposed coder employs a backward adaptive predictor which introduces no extra delay to the system. A simple gain adaptive control is utilized to perform a sample-by-sample gain adaptive excitation model. In other words, the proposed coder exploits the advantages of the LD-CELP and ADPCM coding. This coder can provide transparent quality audio signals at a bitrate of 1.5 bits/sample.

INTRODUCTION

ADPCM [1] and LD-CELP[2] coders are widely used in high quality speech or wideband audio coding. However, they cannot produce transparent quality of CD audio when operate at low bit rates such as 2 bit/sample or even lower. In this paper, the coding system is designed to produce transparent quality of CD audio at such low bit rate. Figure 1 shows the encoder structure of the system.

The system composes of three main parts, a high order predictor, a psychoacoustic weighting filter and a sample-by-sample gain adaptive excitation model. Due to newly developed excitation model, the traditional Gaussian excitation codebook cannot work efficiently. To optimize the coder performance, a new codebook training algorithm is also developed. Finally, the coding system will be simulated to investigate its performance.

HIGH ORDER BACKWARD ADAPTIVE PREDICTOR

In the proposed coder, a high order backward adaptive predictor [3] is employed to remove the redundancy of the input signals. This backward adaptive predictor has been widely used in LD-CELP coder [2].

The predictor is adapted in a block-by-block manner. To adapt the predictor, LP analysis is performed on the previous reconstructed signals in every 0.72 ms (32 samples at 44.1 kHz sampling). The autocorrelation coefficients are calculated by recursive windowing [3] with window size 5.8 ms (N=256 samples). This Recursive Window has been proved to be superior than the Hamming Window when applied on backward LPC analysis. Besides, it can be implemented using a simple filter structure [3].

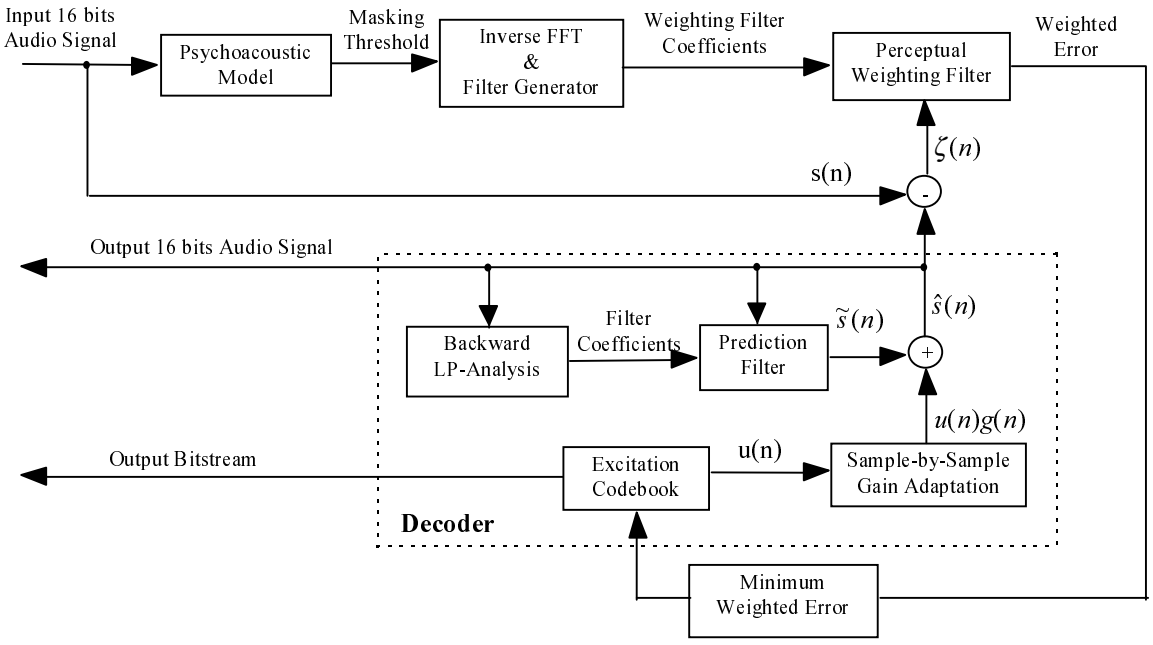


Figure 1. Encoder Structure of the proposed coding system

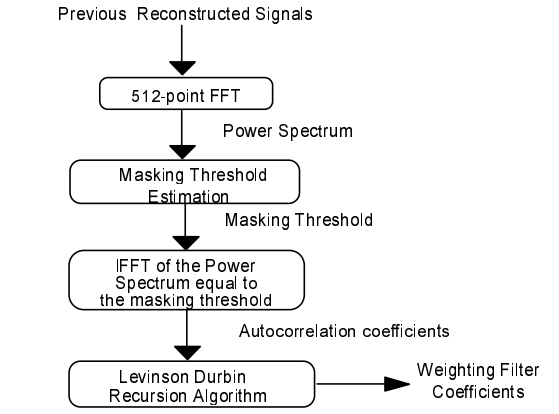
Since the LP analysis is only performed on the previous reconstructed outputs, there is no extra delay introduced to the coding system and no additional information about the predictor should be transmitted to the decoder. In order to obtain higher prediction gain, a higher order predictor should be used. However, the complexity will grow as the order increased. In the proposed coding system, an order 20 prediction filter is employed and which can provide good performance with acceptable complexity.

PSYCHOACOUSTIC WEIGHTING FILTER

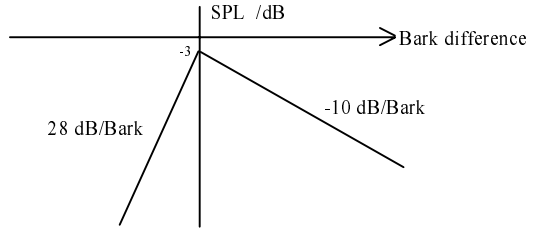
In conventional CELP coder, a perceptual weighting filter is utilized to reduce the subjective loudness of the noise. The weighting filter can be obtained by performing LP analysis on the input audio and has the general form

$$W(z) = \frac{1 - A(z/\beta)}{1 - A(z/\gamma)} \quad \text{where } 0<\gamma,\beta<1.$$

This weighting filter is generated based on the spectrum of the input signals. Obviously, it is not optimum; the optimum weighting filter should be generated based on the desired noise spectrum.



(a) Flow diagram of the psychoacoustic weighting filter estimation



(b) Spreading function used in the psychoacoustic model.

Figure 2

Therefore, the proposed coding system makes use of a psychoacoustic weighting filter. This psychoacoustic weighting filter exploits the masking effect of the human auditory system [4,5] to the coding system. It shapes the quantization noise spectrum below the Just Noticeable Loudness so that the noise becomes inaudible.

The coefficients of weighting filter are obtained from using frequency-domain analysis according to figure 2(a). In order to maintain the coding system in low delay, the weighting filter is calculated from the previous reconstructed signals. The signals are firstly windowed by the recursive window which is also used in the backward adaptive predictor and introduced no extra delay. Then a 512 point FFT is performed to produce the signal spectrum Y. The analysis frame is overlapped by 50%, i.e. the number of new samples in each analysis frame is 256. The critical band power density P [6] can be calculated by partitioning the spectrum Y into 25 critical bands,

$$P(j) = 20\log\left(\frac{1}{n(j)}\sum_{i=l(j)}^{h(j)} Y(i)^2\right)$$

where n(j) is the number of spectrum lines belong to critical band j, l(j) and h(j) are the lower and higher boundaries of the critical band j. The information of the boundaries and the critical band rate can be obtained in [5].

Due to the spreading function shown in figure 2(b), the contribution to the masking threshold from the power density of critical band i on critical band j can be expressed as

$$s(i,j) = \begin{cases} P(i) - 28(i-j) & \text{if i > j} \\ P(i) - 3 & \text{if i = j} \\ P(i) + 10(i-j) & \text{if i < j} \end{cases} \qquad (1)$$

Then, the masking thresholds T can be calculated from (1) as,

$$T(j) = 20\log\sum_{i=1}^{25} 10^{s(i,j)/20} \qquad (2)$$

The masking thresholds from (2) are finally compared to the absolute threshold [5] of hearing to ensure that the resultant masking thresholds are not below the minimum hearing level.

Converting these masking thresholds from bark domain to frequency domain can yield a power spectrum. Taking inverse FFT of this power spectrum produces the required autocorrelation coefficients. Finally, the filter coefficients can be obtained by the Levinson-Durbin Recursion Algorithm [7]. The order of the weighting filter employing by the proposed coding system is 20.

SAMPLE-BY-SAMPLE GAIN ADAPTIVE EXCITATION MODEL

In the conventional CELP coder, the excitations are obtained by multiplying a vector codeword with a gain value. The vector codeword is in general selected from a random Gaussian codebook. The gain value is the quantized version of the root mean square of the original residue signal vector. The index of the

codeword which minimized the weighted mean square error and the gain value are transmitted to the decoder. Some later version of CELP or LD-CELP coders are achieved to adapt the excitation gain by using a gain adapter which can be a fixed-coefficient and first-order or an adaptive-coefficient and higher order predictor in the logarithm domain. However, until now, all the CELP coders update the excitation gain in a vector-by-vector fashion. For every excitation vector, one gain value is generated. This approach has been proved to be efficient for high-quality speech coder but not for transparent CD-quality audio coder.

In this paper, the proposed coder utilizes a different approach called sample-by-sample gain adaptive excitation model. The difference between this excitation model and the conventional one is that the excitation gain is adapted in a sample-by-sample approach, rather than a vector-by-vector approach. In other words, for every excitation vector, a gain vector is generated. The proposed gain adaptation has an instantaneous response to any rapid changing of the input variance. The idea of this adaptation approach is similar to the step-size control logic used by the ADPCM coding [1].

Gain Adaptation

Suppose $\mathbf{u_n}$ and $\mathbf{g_n}$ be the N-dimension optimal codeword and gain vector for the excitation vector $\mathbf{e_n}$, we have

$$e_n(i) = u_n(i)g_n(i) \qquad 1 \le i \le N$$

where

$$\begin{aligned} g_n(i) &= L[d_n(i)] \\ d_n(i) &= d_n(i-1) + f(u_n(i-1)) \end{aligned} \qquad (3)$$

The gain value $g_n(i)$ is obtained from a look-up table L and is located by an addressing index $d_n(i)$. The values stored in L are in ascending order and they are selected to span the dynamic range of the prediction error.

In (3), when i=1, $d_n(0)$ and $u_n(0)$ can be obtained as the $d_{n-1}(N)$ and $u_{n-1}(N)$ from the previous $\mathbf{d_{n-1}}$ and $\mathbf{u_{n-1}}$. The adaptation rate is driven by the index driving function $f(\mu)$ which is defined as

$$f(u) = \begin{cases} -1 & \text{if } u \le D \\ (u-D)S & \text{if } u > \text{D} \end{cases} \qquad (4)$$

According to (3), the magnitude of the input codeword element is examined by a pre-defined decision threshold D. If it is larger than the threshold, it implies that the desired excitation signals are high energy relative to the current gain value. Thus, the gain value should be increased, otherwise it should be reduced. The asymmetric structure of (4) explores the characteristics that the gain increasing rate is faster than the decreasing rate. The driving function f(u) returns a value which is then added as an offset to the current addressing index $d_n(i)$ to produce the next $d_n(i+1)$. The appropriate parameters of the proposed coding system are chosen as D=0.5 and S=2.

Since all the parameters required to update the excitation gain can be obtained from the previous outputs, no additional information should be transmitted to the decoder. Besides, using the table look-up procedure makes the gain adaptation more robustness.

Delay Search of the Codebook

Delay-search coding [7] has been proved to be effective to increase the performance of an analysis-by-synthesis coder. The benefit of this approach is to increase the coding gain without increasing the bit rate. Besides, only the encoder will be modified, the structure of the decoder remains unchanged.

Since an exhaustive search algorithm consumes high computation, an non-exhaustive search algorithm is used. Although the final output sequences are not optimum, the performance loss can be ignored if the non-exhaustive search algorithm is chosen appropriately.

In the proposed coder, the vector codewords are searched by the (M,L)-algorithm [7] to optimize the excitation model. To obtain significant improvement in coding gain with an acceptable increase in complexity, both M and L are set to 4.

EXCITATION CODEBOOK TRAINING

In the proposed coder, the conventional Gaussian codebook is no longer effective because the sample-by-sample gain adaptive model is used. The training procedure of the codebook should take the effect of the gain adaptation into account. Hence, a new codebook training algorithm is designed to fit the proposed coding system.

For a given signal vector $\mathbf{s_n}$, according to the minimum MSE rule, the reconstructed output of signal vector $\hat{\mathbf{s}}_\mathbf{n}$ can be obtained by the synthesis equation,

$$\hat{s}_n(j) = \tilde{s}_n(j) + u_i(j)g_n(j) \qquad 1 \le j \le N$$

where $\tilde{s}_n(j)$, $u_i(j)$ and $g_n(j)$ are the corresponding element of the predictor output, selected codeword and gain vector.

Suppose ζ_i be the overall distortion produced by those input signals mapping to the quantization cluster C_i of the given codebook.

$$\zeta_i = \sum_{n \in C_i} |\mathbf{s_n} - \hat{\mathbf{s}}_\mathbf{n}|^2 \qquad (5)$$

Obviously, the optimal codebook is the one that produces minimum distortion for any input signals.

Define ζ as the total distortion produced by quantizing a training set of audio signals $\{\mathbf{s_n}\}$, i.e.

$$\zeta = \sum_i \zeta_i = \sum_i \sum_{n \in C_i} |\mathbf{s_n} - \hat{\mathbf{s}}_\mathbf{n}|^2 = \sum_i \sum_{n \in C_i} \sum_{j=1}^{N} [s_n(j) - (\tilde{s}_n(j) + u_n(j) g_n(j))]^2 \quad (6)$$

To minimize the overall distortion with respect to the codeword element $\mu_i(j)$, set

$$\frac{\partial \zeta}{\partial u_i(j)} = \frac{\partial \zeta_i}{\partial u_i(j)} = 0$$

which yields

$$u_i(j)_{optimal} = \frac{\sum_{n \in C_i} [s_n(j) - \tilde{s}_n(j)] g_n(j)}{\sum_{n \in C_i} [g_n(j)]^2} \quad (7)$$

Based on the equation (6) & (7), the training process can be implemented by the following steps:

1. Given an initial codebook $\Psi = \{u_i(j)\}$;
2. Quantize the training signals by the coder. In every quantization cluster C_i, a set of signal sample $s_n(j)$, predicted sample $\tilde{s}_n(j)$ and gain value $g_n(j)$ can be obtained according to (6);
3. Using equation (7), the new optimal codeword of each quantization cluster can be calculated;
4. Using the new codebook generated from step 3 as Ψ and goto step 1.

The codebook design begins with an initial codebook which can be a random Gaussian codebook or generated by distributing the codewords uniformly in the codeword space. If at any iteration, codeword i has not been used, i.e. an empty cluster situation appears, the corresponding codeword should be replaced or slightly changed. The iteration continues until the distortion D is no further reduction or below a specified value.

SIMULATION

The proposed coding system is simulated by computer programming. The simulated coding system is designed to operate at 1.5 bits/sample. Since all information except the excitation codeword can be obtained from the quantized output, only the index of the codeword should be transmitted. The excitation codebook is 4-dimensional with 64 codewords. The codewords are searched by (M,L)-Algorithm where M and L are both set to 4, this causes a delay of 16 samples or 0.3628 ms at 44.1 kHz sampling rate. The predictor order is 20 and updates in every 32 samples. The order of the psychoacoustic weighting filter is also 20 and updates in every 256 samples.

To investigate the performance of the coding system, different audio segments are coded. All segments are digitized from CDs at 44.1 kHz sampling frequency with 16 bits ADC. The signals are selected to contain different material, including classical and popular audio segments, songs with male and female singers. Table 2 shows the objective signal-to-noise results of the coding system. Formal subjective listening tests indicate that the coding system can provide transparent qualities for all coded audio segments.

Table 2. SNR results of the proposed coder.

Music Signal	SNR /dB
Male singer	21.90
Female singer	20.35
Pop music	25.06
Orchestra I (Mozart)	26.10
Orchestra II (Beethoven)	25.79
Average	23.84

CONCLUSION

A new coding system has been presented in this paper. This coding system bases on a LD-CELP model employs a sample-by-sample gain adaptive excitation model and a psychoacoustic weighting filter. Subjective listening tests reveal that the proposed coding system can produce transparent quality of CD audio at a bitrate of 1.5 bit/sample.

REFERENCES

[1] J.R. Boddie, J.D. Johnson, C.A. McGonegal, J.W. Upton, D.A. Berkley, R.E. Crochiere and J.L. Flanagan, "Adaptive Differential Pulse-Coded-Modulation Coding", AT&T The Bell System Technical Journal, VOL. 60, No. 7, Sept. 1981.

[2] Juin-Hwey Chen, "A Robust Low-Delay CELP Speech Coder at 16 kb/s", Advances in Speech Coding, pp.25-35, 1990.

[3] Thomas P. Barnwell, "Recursive Windowing for Generating Autocorrelation Coefficients for LPC Analysis", IEEE Trans. on ASSP, VOL. ASSP-29, NO. 5, pp. 1062-1066, October 1981.

[4] M. R. Schroeder, B. S. Atal, J. L. Hall, "Optimizing digital speech coders by exploiting masking properties of the human ear", Journal of Acoustical Society of America, Vol 66, No. 6, Dec. 1979.

[5] Jerry V. Tabias, "Foundations of Modern Auditory Theory I & II", Academic Press, 1972.

[6] T. Sporer , U. Gbur, J. Herre, R. Kapust, "Evaluating a Measurement System.", Journals of Audio Eng. Soc., VOL 43, No.5, pp. 353-363, 1995 May.

[7] N.S. Jayant, Peter Noll, "Digital Coding of Waveforms : Principles and Applications to Speech and Video", Prentice-Hall, Englewood Cliffs, New Jersey, 1984.