AN 8 kb/s LD-CELP WITH IMPROVED EXCITATION MODELLING

R. Soheili, A.M. Kondoz, B.G. Evans

Centre for Satellite Engineering Research
University of Surrey, U.K.

Abstract

Backward prediction of the short term redundancies in speech has resulted in very low delay algorithms, with toll quality at 16 kb/s. At medium bit rates around 8 kb/s the modelling of the excitation signal by conventional CELP techniques can result in high complexity or poor output processed speech for services such as PSTN. In this paper we propose a low delay algorithm based on a vector quantised multi-tap adaptive codebook in producing a high quality speech signal operating at 8 kb/s. A report on the comparisons with other existing standards as well as simplification techniques in realising the algorithm are presented.

1 INTRODUCTION

Interests in low bit rate speech coding for Public Switch Telephone Networks (PSTN) has taken a sharp rise. This is mainly due to the increase in the number of users and limited bandwidth available. It is hoped that with the introduction of the new standards, namely ADPCM [1], and LD-CELP [2], running at 32 Kb/s and 16 Kb/s respectively, the congestion will be reduced considerably. However demand for the use of the PSTN is rising at a higher rate than the available channel capacity, which has forced CCITT to standardize an 8 Kb/s low delay speech coder as soon as possible. Therefore interest is now being centered around medium and low delay coders running at 8 Kb/s and below. Unlike forward adaptation schemes the LD-CELP is based on a buffering of 5 samples at 8 kHz sampling rate. Frequent update of the vector quantisation of the excitation signal and a high order backward LPC synthesis filter are the basis of the LD-CELP algorithm. In order to reduce the bit rate to 8 kb/s, predictions of long term as well as short term correlations present in the speech signals, plus vector quantisation of the speech parameters are necessary to reach the system requirements. It is widely believed that [3], the future CCITT standard will be a conventional CELP algorithm [4], with the exception of backward prediction of the short term correlation rather than conventional forward techniques. However such a modelling of the excitation signal will result in large number of parameters to be encoded, thus resulting in a high transmission of side parameters. Since certain parameters are not correlated then individual coding of these parameters would result in a high bit rate system or poor output processed quality due to inaccurate modelling with fewer bits. In this paper we propose a modelling which is much simpler but updated more frequently, hence resulting in a high quality processed speech. The algorithm is based on a multi tap long term predictor "LTP", with the coefficients vector quantised, resulting in an 8 kb/s coder. The delay of the encoder is as low as 13 samples at 8 kHz sampling rate. But due to complexity restrictions the expected encoder delay is about two to three times that of proposed for real time implementation aspects, on a single Digital Signal Processor chip. In the following we briefly outline the modelling of a conventional CELP system and propose a modified simple excitation technique that is capable of producing high quality speech. This is followed by the vector quantisation aspects of the LTP tap coefficients, regarding the training and simplified search techniques.

2 CODE EXCITED LINEAR PREDICTION

In a CELP system the excitation signal is modelled by the linear combination of the outputs of a multi-tap "β_k" adaptive codebook $r(n)$, which replaces the long term redundancies of the excitation signal and a scaled "G" fixed codebook $c(n)$ filled by Gaussian sequences to overcome the inaccuracies caused by the former modellings. This excitation is then convolved with the filter response $h(n)$ of a short term predictor to produce an output speech $\hat{S}(n)$ which resembles the original signal $S(n)$ as best as possible. This can be formulated as follow ;

$$\hat{S}(n) = h(n) * [\sum_{i=-k}^{i=k} \beta_i r(n - M + k) + Gc_j(n)] \tag{1}$$

The LTP order is usually one for simplicity and the best lag position M and vector sequence c_j for the adaptive and fixed codebooks respectively can be searched in a closed loop and/or open loop methods depending on complexity and quality issues. In LD-CELP at 16 kb/s the adaptive codebook is removed and instead the order of the short term predictor filter is increased to 50. Hence the algorithm is reduced to only a fixed codebook search. The frequent update of the excitation modelling is the strong feature of this algorithm. At bit rates around 8 kb/s, such high update rates are not possible, hence for better modelling techniques one is again forced to consider conventional techniques such as equation 1.

3 VECTOR QUANTISED LTP EXCITATION MODELLING

In order to reduce the bit rate to 8 kb/s and maintain a high quality, the introduction of the LTP is absolutely necessary. At the same time high update rates of all the parameters is vital in achieving high quality standards. Employing equation 2 in a backward short term prediction system would atleast need the side transmission of four parameters, $\beta_k, M, G,$ and c_j. At bit rates around 8 kb/s, the available number of bits would limit the modelling of the excitation signal over shorter periods, thereby resulting in less frequent update rates. Another problem is the individual coding of certain parameters which inhibits the use of vector quantisation techniques. Thus resulting in an inefficient use of the channel capacity. With all the above in mind, a simple experiment was set up to see if two code books of different nature were really needed. A simple backward short term synthesis filter is excited with an excitation signal according to equation 1. The long term prediction search was limited to 128 integer lag positions utilising a simple 1-tap model in a closed loop manner. The fixed code book contained 256 entries with a single scaling factor. No quantisation of the gains was considered. The aim of the experiment was to see which of the codebooks would contribute more to the modelling of the excitation signal at different update rates. The results are given in table 1.

Update Rate	STP Contribution	LTP Contribution	EXC Contribution
10 samples	6 %	76 %	18 %
20 samples	6 %	75 %	19 %
30 samples	7 %	72 %	22 %

Table 1: Overall signal to noise ratio, percentage contribution of short term, long term and excitation modellings in CELP coders at different update rates

As expected it is clear that the contribution of the LTP is by far greater, at all update rates. This suggests more emphasis should be placed on the LTP contribution rather than the fixed codebook output. For this reason the fixed codebook contribution is removed from equation 1, resulting in a purely adaptive LTP excitation modelling. The side information is now reduced to β_k and M. The bits gained from the loss of the fixed codebook parameters can now be utilised in a higher order modelling of the LTP filter or more frequent update rates of the parameters. For the best choices on the number of taps and lag search intervals, objective measurements in the form of segmental signal to noise ratio was considered. Table 2 gives the results obtained for different values of M and β_k for 3 male and 3 female talkers over one minute of conversation. For a comparison at a given bit rate the LTP coefficients are assumed to take 3 bits for the quantisation process. Hence the results are for different buffering delays. From the results it is clear that extreme lower and higher LTP orders results in poor quality. Low LTP orders are not efficient in modelling the excitation signal, where as higher order filters are suffering from the less frequent update rate of the LPC and excitation parameters due to their buffering delays. This suggests the use of 3 to 5 tap coefficients for efficient modelling. In subjective tests, an LTP filter of order 3 and lag search interval of 64 was considered as the best compromise. Figure 1 shows a schematic block diagram of the encoder.

Lag Interval	LTP Filter Order			
	1	3	5	7
64	12.90	13.63	12.98	12.55
128	12.81	13.94	13.41	12.63
256	13.17	14.15	13.50	13.14

Table 2: Segmental SNR performance of the LTP excitation modelling in dB

So a low delay adaptive CELP was constructed utilising a 3 tap LTP filter with a lag search interval of 64 samples. The total encoding delay is 13 samples at 8 kHz sampling frequency. The excitation is then fed to a 10^{th} order synthesis filter whose coefficients are adaptive in a backward manner.

4 LTP COEFFICIENTS QUANTISATION

For efficient use of the channel capacity and the high correlations that exists among the LTP coefficients, vector quantisation of the parameters are very common and necessary. An efficient vector quantisation would require a properly set codebook that can represent the parameters

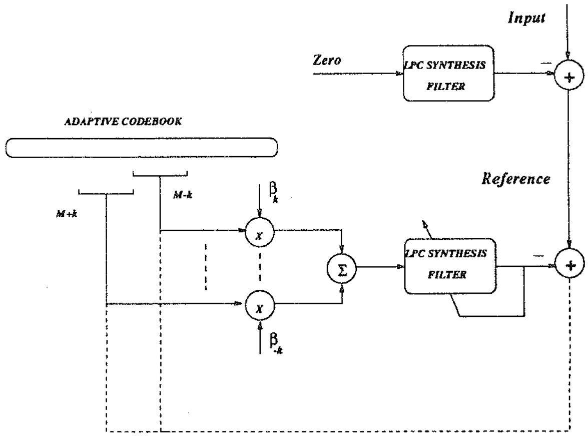

Figure 1: Block diagram of the proposed VQ LTP excitation model

with as little distortions as possible. This would of course very much depend on the parameter correlations and the codebook size. In training our codebook we have incorporated the closed loop technique reported in [2], for constructing an efficient codebook to represent the LTP coefficients. Assuming that S is the input signal, Y the excitation vectors at lag values $M-k$ to $M+k$, and H is a lower triangular matrix with the sub-diagonals equal to the samples of the STP impulse response, then the total accumulated distortion due to the j^{th} cluster corresponding to β_j is given by ;

$$D_j = \sum_{n \in N_j} ||S - HY\beta_j||^2 \tag{2}$$

To minimise the distortion due to the j^{th} cluster, we differentiate the above equation with respect to β_j and set it to zero. This results in a new j^{th} centroid given by;

$$[\sum_{n \in N_j} Y^T H^T H Y]\beta_j^* = \sum_{n \in N_j} Y^T H^T S \tag{3}$$

In constructing an efficient codebook search algorithm the vectors are placed in a sequence according to the sum of the absolute β_k values. This is featured in Figure 2. This way the quantisation process can be placed within the analysis by synthesis loop and only a limited number of quantisation vectors are considered depending on how well the absolute sum of the calculated coefficients matched the absolute sum of the codebook vectors. This resulted in a very small drop in objective measurements and insignificant subjective performance loss. Above all the complexity of the coder is greatly decreased. A further reduction in complexity can be achieved if the LPC coefficients are updated every other frame. This way the second frame would only require to synthesise a small number of vectors depending on the input frames length. High computational savings in cross-correlations and auto- correlations for calculating β_k can be achieved. The penalty paid is increased delay at the encoder.

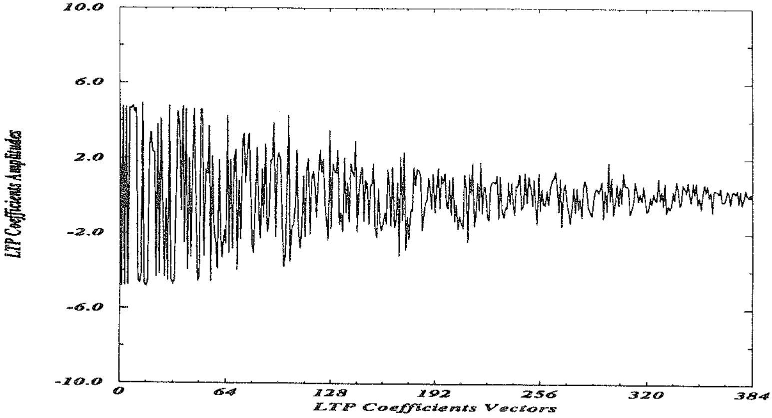

Figure 2: Ordered 3-tap LTP vector codebook, (7 bit)

5 PERFORMANCE

A full quantised version of the proposed algorithm was compared with two existing standards. The LD-CELP of CCITT and the RPE-LTP [5] of the full rate GSM at 16 kb/s and 13 kb/s respectively. In objective testings the algorithm was superior to the GSM by an average of 2 dB, for a wide variety of talkers. With comparisons to the LD-CELP, the algorithm was lacking on average 6 dB in SNR measurements. This was emphasised by the subjective testings which indicated the superiority of the LD-CELP quality, nevertheless the coder outperformed the GSM output quality. The inclusion of post filtering techniques did improve the quality, but it was significantly lower in quality than LD-CELP. In further objective tests, it was observed that higher objective values were obtained at places where the input residual frame contained the pitch pulses. This indicated the limitations of the excitation modelling for unvoiced input frames.

6 CONCLUSION

In this paper, a simplified excitation modelling based on the long term prediction of the excitation signal was proposed for low delay encoding of speech signals. The algorithm can produce high quality speech at bit rates of 8 kb/s, but it is still lacking the toll quality of LD-CELP at 16 kb/s. The performance results indicated a need for several type of excitation modellings, depending on the nature of input signal.

The problem of toll quality at 8 kb/s is still an outstanding issue. Since no forward techniques at 8 kb/s has resulted in toll quality outputs, then is it possible for backward techniques to achieve such a highly set quality target?

7 REFERENCES

[1] J.R. Boddie, J.D. Johnston, C.A. McGonegal, J.W.Upton, D.A.Berkley. R.E.Crochiere, J.L. Flanagan, "Digital Signal Processor: Adaptive Differential Pulse Code Modulation Coding", Bell System Tech.J., pp. 1547-1562, sept. 1981.

[2] J.H. Chen, "A robust low-delay CELP speech coder at 16 Kb/s", Proc. IEEE Global Commun. Conf., Dallas, Texas Nov. 1989, pp. 34.1.1-34.1.5.

[3] S.Hayashi, M.Taka, "Standardization Activities On 8-kbits/s Speech Coding in CCITT SGXV", Proc. Int. Conf. on Selected Topics in Wireless Communications, pp. 188-192, June 1992.

[4] M.R. Shroeder, B.S.Atal, "Code-Excited Linear Prediction (CELP): High Quality Speech at very Low Bit Rates," Proc. Int. Conf. Acoustics, Speech, and Signal Processing, pp. 937-940, Mar. 1985.

[5] GSM Recommendations 06.10: GSM Full Rate Speech Transcoding 1988