AN EXPERIMENTAL STUDY OF RESIDUAL EXCITED LINEAR PREDICTIVE CODER FOR TELECOMMUNICATION PURPOSES

N. Duong
Telecom Australia

ABSTRACT: This paper describes a computer simulation study of the quality problems associated with RELP. Some observations are given on the appropriate conditions for LPC analysis.

INTRODUCTION

Applications such as digital mobile radio and digital satellite communication require a speech CODEC (coder/decoder) with a low bit rate preferably in the range from 4.8kbits/sec to 9.6 kbits/sec. To be acceptable for use in the public network such a coder must satify a wide range of performance criteria. These requirements include:

- suitable speech quality output. This requirement is sometimes loosely specified as the quality provided by 5-7 bit log PCM (Logarithmic quantized Pulse Code Modulation). In any case it exceeds that obtainable from a pitch excited Linear Predictive Coder (LPC).
- uniform quality across a large range of speakers and speaking conditions (i.e. environmental noises, microphones, transmission circuits etc...).
- low encoding delay. This requirement is due to the detrimental effect of delayed echo on the subjective quality of a speech communication circuit.
- low complexity to allow realtime implementation.

A relatively simple coder that may be able to meet these requirements is the Residual Excited Linear Predictive Coder (RELP). It is an example of a class of hybrid coders described below.

RELP PRINCIPLE

RELP is an example of a class of coders which is a cross between waveform coders and source coders. The first type (waveform coders) attempts to recreate the waveform of the input while the second type (source coders) extracts the parameters of some speech production model and employs these parameters in the reconstruction process. The slowly varying speech model parameters permit a correspondingly low sampling rate (of the order of 50 times per second) and consequently a low bit rate (in the region of 5kbits/sec). Typically, waveform coders produce high quality output speech at a high bit rate and source coders produce much lower speech quality at a lower bit rate. The third type of coders (hybrid coders) combines the techniques of the first two types in some suitable manner resulting in coders with both the speech quality and bit rate in the mid range of the source and waveform coders types (Figure 1).

In the case of RELP the LPC analysis procedure is used to generate the speech model parameters and also the excitation in the form of an LPC residual. This excitation, instead of being described simply as either unvoiced or voiced with a particular fundamental frequency as is the case of pitch excited LPC, is more faithfully represented by having its spectrum encoded by some waveform coder. This process for a basic RELP is illustrated in Figure 2.

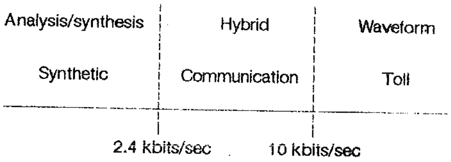

Figure 1. Classification of coders.

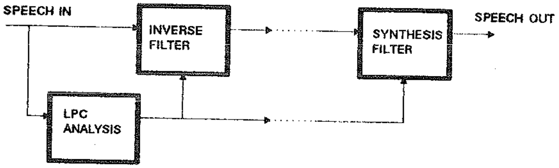

Figure 2. Basic RELP

In the absence of any quantization error on the residual signal and the reflection coeficients, the resynthesized speech is numerically identical to the original speech provided the prediction filter and the reconstruction filter are of the same type (and thus always maintain the same internal states with respect to each other irrespective of coefficient updates). In practice, however, not only is quantization noise included but the residual signal actually transmitted represents only the lower portion of the residual spectrum with the receiver relying on some sort of non-linear operation to generate the full band excitation. This procedure presumes a flat residual spectrum and thus the entire spectral structure is recoverable from the low-passed version of the spectrum.

EXPERIMENTAL WORK

In order to examine the quality problems associated with RELP coders we are conducting an experimental study based on computer simulation. In this study we hope to be able to isolate the degradations of the various impediments introduced in a RELP coder, the basic design parameters to be adopted for RELP, and the quality obtainable at the various bit rates. In short we would like to ascertain the suitablity of RELP as a coding technique for telecommunication purposes. Toward this end a suite of programs has been devised to simulate the coder structure shown below (Figure 3). The programs implement the following functions:

1. Short term spectral prediction: the autocorrelation (Markel & Gray, 1976) method is used to compute the LPC reflection coefficients. The quantized (or un-quantized for some subjective tests) reflection coefficients are then used to generate the residual error. The shifting interval as well as the analysis frame size are variable. The analysis data is multiplied by a Hamming window prior to analysis. The order of the inverse filter can be any value up to 16. The reflection coefficients are log-area quantized. While the number of bits assigned to each reflection coefficient is variable, the ranges over which the coefficients are quantized are fixed.

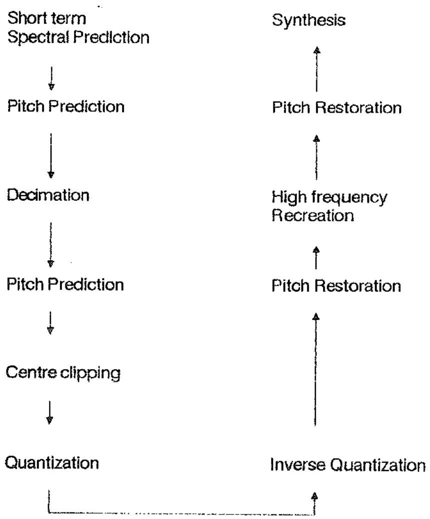

Figure 3. The structure of the simulated coder.

2. First long term pitch prediction: this functional blocks attempts to remove the periodic structure of the residual signal. This process supposedly eliminates the so called tonal noises created when the spectral folding process is used to recreate the full band excition signal (Sluyter, Bosscha & Schmitz, 1984). Pitch prediction is done by simple peak picking of the correlation function of the down sampled residual (1:4 decimation). Interpolation is then done on three correlation values in order to estimate the true pitch. Tests carried out showed that at the original sampling rate of 8kHz this process is adequate. In the regions of strong voicing the pitches computed by the two methods (no decimation and 1:4 decimation) differ by at most one sample. Except for a few frames (out of a total of 100 frames) the prediction gain does not seem to suffer because of this (this does not suggest that prediction gain is a meaningful parameter for this predictor).

3. Decimation: this block decimates the low pass filtered full band pitch predicted residual. Sixth order elliptic low pass filters with cut off frequencies corresponding to decimation rates of 1:2, 1:3, 1:4 are used.

4. Second long term pitch prediction: this block minimizes the variance of the decimated residual to aid in the quantization process.

5. Centre clipper: it has been noted elsewhere (Atal, 1982) that severe centre-clipping can be done on the full band residual signal without loss of subjective quality in the reconstructed speech signal. This block centre clips the decimated residual using a variable threshold based on the residual energy.

6. Quantizer: a number of quantization procedures using both the feed-forward and feeback (Jayant & Noll, 1984) configurations are implemented. The number of levels are 2,3,4 and 8 corresponding to 1,1.6,2 and 3bits/samnple. The probability density function assumed for the Max-Lloyd quantizer is either Gaussian or Laplacian.

7. The decoder simply inverts the procedures used in the coder.

DISCUSSION OF SOME PRELIMINARY RESULTS

At the moment only very informal listening tests have been carried out. Once the number of variables involved have been reduced to a more manageable size we plan to do more rigorous subjective testing in order to determine the output quality of the resulting coders. Even with the limited testing that we have done a number of observations can be made:

- the normal procedure for spectral prediction as discussed in literature seems to recommend pre-emphasis (or pre-whitening). This has not been our experience in the case of RELP. In a very simple test with just the spectral prediction (filter order = 16, no coefficient quantization), the spectral decimation (and subsequent folding) and the reconstruction blocks switched on, the quality of the reconstructed speech was better when no pre-emphasis was used. It seems that for the few frames where the LPC model breaks down, the de-emphasis that is carried out will only exaggerate the spectral distortion that is already present in a not very flat residual signal. For the case where 16 bit integer arithmetic is employed (as is true in most real time situations) this conclusion may not apply. We have yet to simulate this condition. A simple remedy is some kind of test on a frame by frame basis to determine if pre-emphasis is to be used or not (e.g. one bit to indicate 90% or 0% emphasis). However, for the case of realtime implementation where buffer memory is limited it may not be a practical solution.

- frame position, frame size and frame update rate: the same test as above was employed in order to determine these three parameters. For both male and female voices speaking a Harvard test phrase "The birch canoes slid on the smooth planks" serious difficulty was encountered in regions of non-stationarity, especially after the affricate 'tʃ' in "birch" and the plosive 'd' in "slid" when very low overlapping or non-centralized data segments are used. For male voice a data segment of 200 samples (25 msec) centralized in an analysis window of 300 samples (37.5 msec) was found to be adequate. For the female voice a larger analysis window size was required. The analysis window size had to be increased to 400 samples (50msec), for the same data segment size, in order to obtain a reasonable result. Another solution for the female voice is to decrease the data segment size to just 100 samples (12.5 msec) while keeping the same analysis window size of 300 samples. The quality obtained was much superior to the previous case. The frame rate is now much higher (80 frames/sec). The difficulty experienced with both the male and female voice was due to the breaking down of the LPC model in regions of non-stationarity where the data at the frame edges can not be predicted adequately, resulting in very large residual error with a lot of energy concentrated in the low frequency region (probably in area near the first formant). The decimation and spectral folding process of RELP essentially duplicates this energy resulting in very annoying impulsive noises. The extra difficulty encountered with the female voice is due to the already well known problem of inverse filtering high pitched voices where the estimated formants tend to shift toward a harmonic. A higher frame repetition rate and a large analysis window size (as compared to the data size) is very effective in discarding the large residual errors at the frame edges. For RELP with a bit rate around 9.6 kbits/sec a large fraction of the available bits are used for the residual quantization. As these bits are unaffected by the frame rate, a higher repetition rate is possible. For lower bit rate RELP it is only possible to extend the analysis window size. There is a slight increase in encoding delay for this case.

- centre clipping: both sample by sample and block estimation of residual energy were used in setting up the clipping level. There was not much difference in either the speech quality or the variation in the number of samples discarded per frame. The maximum clipping level that could be used with no perceptual degradation in the speech quality was found to be around 0.7 of the rms of the frame energy. At this level roughly 60% of the residual samples were discarded. Even though we do not plan to implement variable rate coders, this data is important in determining the optimum quantizer decision levels.

CONCLUSION

As is apparent from the foregoing discussion we are less interested at this stage in the performance of a particular RELP coder than in the quality problems that affect all RELP coders. In our opinion there is not much to be gained (compared to the effort expended) in constructing a realtime version that is no more than just another variation on the theme of RELP. On the other hand, the final goal of any coding algorithm development for telecommunication purposes is a hardware version that operates in realtime. A compromise has been reached by ensuring that none of the procedures used in the simulation study will present a problem regarding realtime implementation. This study is continuing and there are more questions to be answered before we can be reasonably assured of the suitablility or otherwise of RELP for the low to medium bit rate speech coders for use in the public telecommunication networks.

ACKNOWLEDGEMENT

The permission of the Director of Research, Telecom Australia to present this paper is hereby acknowledged.

REFERENCES

ATAL, BISHNU S. (1982) "Predictive coding of speech at low bit rate", IEEE Transactions on commumications, Vol. COM-30, No. 4, April 1982, pp. 600-614.

JAYANT, N.S., NOLL, P. (1984) "Digital coding of waveforms, Principles and applications to speech and video", (Prentice-Hall)

MARKEL, J.D., GRAY, A.H. (1976) "Linear prediction of speech", (Springer Verlag: Berlin)

SLUYTER, R.J., BOSSCHA, G.J, SCHMITZ, H.M.P.T. (1984) "A 9.6 kbit/s speech coder for mobile radio applications", Proceedings of the international conference on communications, ICC-84.