NEW INNOVATIONS IN MULTI-PULSE SPEECH CODING FOR BIT RATES BELOW 8kb/s

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

Department of Electronic and Electrical Engineering
University of Surrey, Guildford, Surrey, U.K.

ABSTRACT

Recently, MP-LPC [1] and other similar analysis-by-synthesis (A-by-S) coding schemes have proved that it is possible to achieve very high quality coded speech down to around 8 kb/s. In this paper we report on simulation results of a multi-pulse coder with pitch prediction at bit rates of 8 kb/s and below. The effects of various long term prediction configurations have been evaluated both subjectively and objectively. In order to reduce the bit rate and maintain the output quality, efficient quantisation of the MP-LPC parameters are required. Since in a multi-pulse coding scheme most of the bits are used to code the pulses, we have investigated various methods of adaptive pulse amplitude quantisation. Finally we report on a new joint quantisation scheme of the pulse amplitudes, which resulted in reductions of up to 1000 bits/sec with no impairment in the speech quality.

1 Introduction

In recent years the topic of low bit rate speech coding has received a great deal of interest due to the demands from the applications of low bit rate speech coding, such as mobile satellite communication systems, voice store and forward systems and digital communication networks.

For mobile satellite communication systems the resource is very limited in terms of the very small transceiver terminals requiring larger satellite power, and the very restricted bandwidth currently available. This is particularly true of the land-mobile satellite services which currently has 4 MHz allocated for primary service transmission. For such services to be economic they must employ very narrow bandwidth per channel. The competition is with analogue systems that employ single side band (ACSSB) and achieve reasonable performance at C/N’s of around 50 dB-Hz in 5 kHz transmitted bandwidth. Now in order to be competitive and to use modulation schemes that will not cause excessive distortion over the difficult land-mobile propagation channel, digital speech coding at medium bit rates is required. The performance must be better than the analogue contender in worst case degraded channels and the speech quality must be acceptable enough to be connected to PSTN transmission.

With land-mobile satellite systems proposing to operate voice services in early 1990’s, the race is on to produce digital speech coders that can meet all these requirements. Until now schemes such as MP-LPC and CELP [2] have only achieved toll qualities down to 8 kb/s and still remain fairly complex to implement.

In this paper we report on a MP-LPC coder with an improved open loop long term prediction analysis algorithm, with efficient pulse amplitude quantisation technique to produce high quality speech at 8 kb/s and below.

2 Multi-Pulse Excited Speech Coder (MPE)

In a multi-pulse scheme a sequence of pulses at positions \( m_1, m_2, \ldots, m_k \) and amplitudes \( g_1, g_2, \ldots, g_k \) are generated to replace the original excitation to an LPC synthesis filter. The location and the amplitudes of the pulses are determined sequentially, one pulse at a time, so as to minimise the mean squared weighted error between the original speech and the synthesised speech. The error signal is weighted so as to take into account the human auditory perception system. A block diagram of a MP-LPC coder in the form of an A-by-S procedure is shown in figure 1.

3 Efficient Open Loop Pitch Filter Implementation

A basic MPE produces satisfactory speech quality at medium bit rates. However as the bit rate is reduced, degradations in the speech quality become more noticeable, especially in the high pitched voiced regions. With the introduction of a simple 1-tap long term predictor in the A-by-S procedure loop, such effects can be removed.

An estimation of the long term predictor parameters, (ie delay \( \beta \) and gain \( \beta \) ) can be performed by an open loop configuration [4], or optimum values of \( \beta \) and \( \beta \) can be calculated in a closed-loop configuration [5]. The latter gives better performance at the expense of very much higher complexity.

In the open loop case, calculating the parameters from the LPC residual is only an approximation. A further point is that the error between the input signal to the pitch inverse filter and the output signal from the pitch synthesis filter is made up of two components [3]:

i) The quantisation error \( e_p \)

ii) The prediction error \( e_p \)

(The prediction error \( e_p \) is caused by the effect of the quantisation error \( e_q \) in the filter memory at the synthesis stage).

By limiting the minimum pitch filter delay to be equal to the sub-block size, we have used the same predicted values at both inverse and synthesis filter. With this restriction, the pitch filter can not be recursive in any one sub-block, and therefore we have used the pitch filter memory to obtain near optimum pitch filter parameters.
Figure 2 shows the variation of number of pulses in a sub-block versus the Segmental Signal to Noise Ratio (SNRseg), for three different pitch filter configurations. Curve (a) is for that of a MPE with an open loop pitch prediction analysis. Curve (b) is for that of the improved open loop pitch filter and finally curve (c) shows the performance of the closed loop configuration.

The improved open loop performance falls between that of the open loop and the closed loop curves and this solution provides a SNRseg gain of 0.3 to 0.5 dB. Although the gain in SNRseg is small, subjectively it is very noticeable and tests at lower bit rates (below 8 kb/s) showed significant improvements.

4 Quantisation of MPE parameters.

In a MPE scheme at low bit rates, the overall transmission capacity does not have the capability to accommodate sufficient numbers of pulses for high quality speech. At bit rates below 8 kb/s, about 40% or more of the transmission rate is occupied by voice model parameters and the rest is used to code the excitation pulses. It is therefore obvious that with the restrictions imposed on the bit rate by the channel capacity, efficient coding of these parameters are necessary.

Efficient coding of the short term predictor parameters are crucial for good coder performance. Direct quantisation of the LPC parameters using uniform or non-uniform quantisers lead to high bit rates. Transformation of the LPC parameters into another set can be employed for stability and efficient quantisation reasons. The LPC quantisation was carried out on the Log Area Ratio’s (LAR) using a 40 bit linear quantiser with non-uniform bit allocation for 10 LPC parameters, with update rate of 40 Hz.

For the long term predictor values, the delay "8" was quantised using 6 bits to cover the pitch delay values in the range of 50 - 114. The pitch gain "B" was quantised using a 3 bit non-uniform quantiser. Altogether 9 bits were considered as the best compromise for an update rate of 160 Hz.

The coding of the pulse positions can be performed by considering the number of different possibilities of placing "L" pulses in a frame of "N" samples [6].

4.1 Pulse Amplitude Coding

For the coding of the pulse amplitudes different methods of normalisation schemes were considered [7]. Efficient normalisation is necessary because the pulse amplitudes have a large dynamic range and direct quantisation would require large numbers of bits. Efficient normalisation can be carried out by the root mean square (rms) of the amplitudes. In such methods the rms value must also be included in the transmission. This inevitably leads to higher bit rates. In most MP-LPC designs, quantisation of the first pulse is accomplished by incorporating a high number of non-uniform quantiser levels (usually 5 bits or more) and the rest of the pulses in the sub-block are normalised and coded using fewer quantisation levels (typically 3 bits) [8].

In the following we look at four new amplitude quantisation schemes which require no side information. The four adaptation schemes are as follows:

a) First Pulse Magnitude Adaptive
b) Previous Pulse Magnitude Adaptive
c) Previous Sub-Block Energy Adaptive
d) Pitch Filter Memory Adaptive

Assuming that A_i is the magnitude of the i_th pulse then in scheme (a), the magnitude A_i for i > 1 are normalised with respect to the magnitude of the first pulse A_1 in the block. The first pulse itself is coded using either methods (c) or (d). (see below). With the second procedure (b), the magnitudes A_i (for i > 1) are normalised with respect to the previous optimum pulse magnitude in the current block. Again schemes (c) and (d) are used to code A_1. Method (c) uses the block energy from the previous sub-block to normalise the magnitudes and finally in (d) the energy in the pitch filter memory is used for the normalisation of all the pulses, including A_1. A similar procedure has been used to normalise vector gains in [9].

Figures 3a and 3b, show the Probability Density Function plots (PDF) for the pulse magnitudes using the above adaptation technique for 15 seconds of speech data containing female and male speakers.

The first two schemes considered for i > 1, are very similar to one another and are the currently used techniques. From the PDF plots (figure 3.b) and subjective tests, 3 bits were found to be adequate for satisfactory speech quality.

From the PDF plots, method (c) looks very promising. This technique gives a better normalisation range (ie a small dynamic range), if the sub-block size is smaller and more pulses are included. For larger sub-block sizes and fewer number of pulses, the variation in the energy will be greater. Hence this technique will not perform satisfactorily unless certain limits are imposed.

From both subjective and objective tests carried out, method (d) gave the best overall adaptation technique, which led to a finer quantisation of the amplitudes. During our investigations we noticed that during active speech regions, the pitch filter energy was very similar to the input signal energy. In forming the reference signal, usually more than half the energy from the input signal was removed. Hence the remaining energy would be reflected in the pulse magnitudes. Therefore normalising the pulse magnitudes with respect to the pitch filter energy limits the normalised magnitudes to a range between 0 and around 1.5 . (see figures 3a and 3b). This technique was finally chosen for the quantisation of the pulse amplitudes with 4 bits allocated for the first pulse amplitude and 3 bits for each of the remaining pulse amplitudes.

One major problem encountered with such a scheme is that the memories of the long term predictors in the encoder and decoder are different, probably due to the loss of clock synchronisation at the decoder. Such a condition may lead to the instability of the decoder. In order to overcome this problem the memories of the long term predictors, both in the encoder and decoder, should be reset periodically. This is done by initialising the memories of the filters at known intervals. In a normal real time implementation, it would be likely to find Voice Activity Detectors (VAD) within the coder. Since there is a high probability of finding silence regions of duration less than one second, then initialisation can take place during these regions. If no VADs are included then the resetting process should take place regularly at known intervals, (ie after certain number of frames).
4.2 Joint Quantisation of Pulse Amplitudes.

During the investigations of method (b), simulation results showed that for about 95% of the time the first pulse has the highest magnitude. This is followed by the second pulse, then the third and so on in order of computation, (ie \( A_3 > A_1 > A_2 \ldots \)). This means that a correlation exists among the pulse magnitudes which has led to a new joint magnitude quantisation technique. A simple way of representing the magnitudes \( A_i \) for \( i \geq 1 \), is to find a single coefficient which, with the knowledge of \( A_0 \) will recreate the magnitudes \( A_i \) for \( i > 1 \). This correlation among the pulses can be written as:

\[
A_i = \lambda_i A_{i-1} \quad \text{for} \quad i \geq 1
\]  

Our assumption is based on the investigations that all \( \lambda_i \) are the same for \( i \geq 1 \). Hence equation (1) can be modified to:

\[
A_i = \lambda A_{i-1} \quad \text{for} \quad i \geq 1
\]  

A simple approach to calculate a value for \( \lambda \) is to assume that \( \lambda \) is the average of the ratios between consecutive pulses. Therefore \( \lambda \) is given by:

\[
\lambda = \frac{\sum_{i=1}^{L} A_i - A_{i-1}}{L-1}
\]  

The relation in equation (3) gives equal weighting to all the pulses. Since this value of \( \lambda \) does not minimise the error between the pulse magnitudes, it is not an optimum value. A single optimum coefficient is obtained by using a simple linear predictor for the representation of the pulse magnitudes rather than the magnitudes \( A_i \) for \( i \geq 1 \), of the conventional case. In order to find an optimum value for \( \lambda \), the squared error is employed. The error from equation (2) is:

\[
E = \sum_{i=1}^{L-1} (A_i - \lambda A_{i-1})^2
\]  

The minimisation of \( E \) with respect to \( \lambda \) results in:

\[
\lambda = \frac{\sum_{i=1}^{L-1} A_i A_{i-1}}{\sum_{i=1}^{L-1} A_{i-1} A_{i-1}}
\]  

With this approach, the sign bits of each pulse, together with the magnitude of the first pulse using method (d) are first coded. Then the single coefficient obtained is coded using 3 bits, which represents the magnitudes of the pulses with respect to the previous pulse magnitude.

Figure 4 shows the unquantised pulse magnitudes along with the quantised pulse magnitudes using equation (5). Simulation results show that with this approach 28 dB, 21.5 dB, 20 dB etc signal to quantisation noise ratio can be achieved for the \( 1^\text{st} \), \( 2^\text{nd} \), \( 3^\text{rd} \) etc pulses respectively. For pulse rates of more than 3 pulses/sub-block savings around 500 to 1000 bits/sec are made. Joint quantisation of the amplitudes using equation (5) showed that the quantisation noise increased rapidly for values of \( L \geq 5 \). This is because the magnitudes decayed in an exponential manner. Hence for best results an optimum value of \( \lambda \) was calculated to represent the magnitudes \( A_1 \) to \( A_4 \). For values of \( L > 5 \), a second linear prediction is needed to reduce the quantisation noise.

SIMULATION RESULTS

Figure 5 shows the objective performance of three fully quantised MPE's with open loop analysis of the long term pitch predictor. Curve (a) is for the rms technique for amplitude quantisation and curve (b) is for the improved open loop pitch prediction analysis and quantisation of the pulse amplitudes according to [7]. Curve (c) is very similar to (b) except that the amplitudes (for \( i \geq 1 \)) are jointly quantised using equation (5). Clearly the performance of both curves (b) and (c) are better than (a) at all bit rates, and at bit rates of around 8 kb/s, SNR gain of about 1.5 - 2 dB is obtained. Curve (c) shows significant gain at very high bit rates. Hence it can be deduced that joint quantisation of the amplitudes performs better when 3 - 5 pulses are incorporated per sub-block.

CONCLUSION

In this paper we have investigated a MP-LPC coder with Pitch Prediction at 8 kb/s and below. We have discussed an efficient pitch filter implementation. Its performance has been shown to be nearly equal to the closed loop solution both subjectively and objectively. Finally efficient quantisation of the pulse amplitudes were investigated under four different algorithms. A proposed method known as Pitch Filter Memory Adaptive was shown to be the best. The algorithm is more robust to noisy channel environments than the other schemes considered, and two methods of avoiding the instability problem were reported.

Finally we looked at a new joint amplitude quantisation scheme. The new technique is very promising and reductions in the bit rate by about 1 kb/s can be achieved with no decrease in the quality or increase in the complexity. A basic MPE combined with the improvements presented in this paper can produce high quality speech at bit rates around 7 kb/s.

REFERENCES


A block diagram of an A-by-S procedure after B.S. Atal (1)

Figure 1

PDF plots of magnitude normalisation of the first pulse

Figure 3.a

PDF plots of magnitude normalisation of the pulses (excluding the first pulse)

Figure 3.b

Pulses in order of computation
Joint quantisation of amplitudes

Figure 4

Performance of MPE with Pitch Prediction at various bit rates with three different magnitude quantisation techniques.

Figure 5