Error Masking In a Real Time Voice Codec For Mobile Satellite Communications

D Rowe*, A Perkis*, W.G.Cowley,J.A.Asenstorfer

Digital Communications Group
South Australian Institute of Technology

*Department of Electrical and Computer Engineering
The University of Wollongong

ABSTRACT - The implementation of a real time CELP based speech coder is described. The coder has been optimised for operation on a land mobile satellite channel, and full duplex operation is achieved on a single AT&T DSP32C device. The coder operates at a bit rate of around 6400bps, and produces near toll quality speech. Error masking techniques are investigated with the goal of minimising perceptual errors in the coder output in the land mobile satellite channel. The error masking is achieved with a combination of standard FEC techniques and subpacket substitution.

INTRODUCTION

A great deal of research into Code-Excited Linear Predictive coders (CELP) has been done since the introduction of the CELP algorithm (Atal, 1985), especially in the area of complexity reduction and quantization of the coder parameters. These advances coupled with the availability of powerful Digital Signal Processing (DSP) hardware have made the realisation of a real time CELP coder possible.

CELP is a low bit rate speech coding algorithm capable of delivering near toll quality speech at low bit rates (4.8 - 8 kbit/s). One potential application for CELP is in the area of land mobile satellite communications. The land mobile satellite channel is characterised by deep fades and random (Gaussian) noise. Thus there are two forms of channel errors that a speech coder must contend with, random errors and burst errors. Random errors can be dealt with by using standard Forward Error Correction (FEC) techniques based on redundant bits, and hence an increase in the overall channel bit rate.

Burst errors cannot be protected against with standard FEC techniques, as all information is lost during a burst. However, zero redundancy schemes (Perkis 1990) that repeat certain coder parameters from previous frames can be used to mask the perceptual effects of burst errors.

REAL TIME CELP CODER DESIGN

The implementation of the CELP coder is based on the SESTMP coder (Perkis 1989), and the U.S. Federal Standard 1016, 1989. The CELP algorithm is potentially very complex and implementation in real time is a non-trivial exercise. Hence several limitations were placed on the implementation because of the design goal of real time operation on a single AT&T DSP32C DSP chip.

Figure 1 indicates the basic structure of the CELP encoder. A 10th order Linear Prediction filter models the short term correlation in the speech signal. This filter effectively models the spectral envelope of the speech. The 10 LPC's are converted to 10 Line Spectrum Pair (LSP) frequencies (Itakura, 1975), which are individually quantized using non-linear Max quantizers (LSP1 - LSP10).

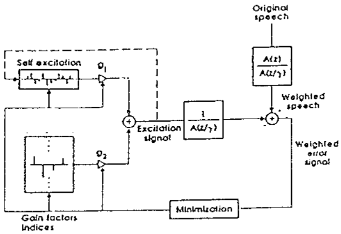

Figure 1: Basic structure of the real time BEM-CELP coder.

The excitation for the LP filter is a combination of two codebooks, the Adaptive codebook (self excitation), and the fixed or Stochastic codebook. The optimum codebook entry and gain is determined for both codebooks using the analysis by synthesis method. The Adaptive codebook is continually updated with the past excitation samples, and hence models long term periodicity in the excitation signal.

The Stochastic codebook structure in this coder is Stochastic Multipulse (STMP), consisting of an average of 4 non-zero pulses per excitation subframe (40 samples). The non-zero pulses are uniformly distributed over the codebook, and the pulse magnitudes have a unit Gaussian distribution.

The codebook indexes are integer numbers and require no further quantization, the codebook gains must be Max quantized according to their probability density functions (Perkis 1990). The excitation parameters (codebook gains and indexes) are updated every 5ms (coder subframe), while the spectral parameters (LSPs) are updated every 30ms (coder frame). There is a difference in SEGSNR of 1dB between the quantized and unquantized coder indicating satisfactory quantization.

The most computationally expensive part of the CELP coder is the Adaptive and Stochastic codebook searches, due to the analysis by synthesis search method. By using overlapping codebook entries, the codebook search complexity has been brought down to an order suitable for real time implementation. The adaptive codebook has 128 overlapping entries. The Stochastic codebook has 64 overlapping entries. The overlap between adjacent codebook entries in both cases is one sample. Codebook searches still contribute to 75% of the DSP's computational load.

The coder was implemented in C, using the AT&T DSP32C C Compiler. Computationally expensive parts of the coder, such as the codebook search and filtering operations were carefully hand crafted in DSP32C assembler, to utilise the maximum processing power of the device.

The coder runs on a IBM PC based development system. Full duplex operation (encoding and decoding) is achieved in real time, using the 8kHz PCM codec on the development card for analogue I/O. No stand alone implementation of the coder exists at this time.

The coder is also useful as a non-real time simulation tool. For this purpose the coder is separated into an encoder and a decoder program, and the DSP32C development system is given access to the peripherals of the host PC via interface software. The encoder converts disk files of speech samples into fully quantized files of bits, which are then run through the decoder, which has facilities for introducing random and burst errors. The decoder can also perform objective measures of the speech quality (SEGSNR and Cepstral Distance (CD)), and generate output speech files for subjective listening tests.

This system enables extremely fast simulations. The decoder actually runs faster than real time! Bit allocations for the coder are presented in Tables 1 and 2. A total of 193 bits per 30ms frame are used for the coder, including the parity bits that are used for Forward Error Correction (FEC). The overall bit rate (channel bit rate) is 6433 bits/s. The coder bit rate with no FEC is 5333 bits/s.

Table 1: LSP and Ex. bit allocation

LSPs	Excitation
LSP1 - 3 bits	L - 7 bits
LSP2 - 4 bits	β - 3 bits
LSP3 - 4 bits	i - 6 bits
LSP4 - 4 bits	α - 5 bits
LSP5 - 4 bits	
LSP6 - 3 bits	
LSP7 - 3 bits	
LSP8 - 3 bits	
LSP9 - 3 bits	
LSP10 - 3 bits	

Table 2: Bit allocation for 30ms frame

Bits	Allocation
0 - 33	LSPs
34 - 44	Parity 1
45 - 65	Excitation 1
66 - 86	Excitation 2
87 - 107	Excitation 3
108 - 118	Parity 2
119 - 139	Excitation 4
140 - 160	Excitation 5
161 - 181	Excitation 6
182 - 192	Parity 3

APPLICATION OF FEC TO CODER IN RANDOM ERROR CHANNELS

Adding FEC to the coder equates to an increase in bit rate. We wish to minimise the redundancy in the coded signal but maximise the protection given to the coder in the random error channel. A unequal error protection scheme has been adopted, based on the perceptual importance of each bit in the random error environment. The FEC code chosen is the (23,12) Golay code, with hard decision decoding. This code requires 11 parity bits to protect 12 message bits. It is capable of correcting a maximum of 3 bit errors in a 23 bit message/parity codeword.

To evaluate the coder performance a test database of 4 speakers was used, 2 male adults and 2 female adults, speaking 2 phonetically balanced Harvard sentences each. The total length of the test database is 24 seconds. The database was used for all of the results presented in this paper. The SEGSNR of each test is presented, and objective results are confirmed with informal subjective listening tests.

The coder parameters have been placed into 3 groups for the application of FEC. Group 1 is the LSP's, which carry the spectral envelope information of the coded speech signal. They are protected by the Parity 1 bits. Group 2 and Group 3 are the excitation information for the frame, composed of 6 excitation subframes. Excitation subframes 1 - 3 are protected by the Parity 2 bits, and Excitation subframes 4 - 6 are protected by the Parity 3 bits

LSP SORTING ALOGORITHMS

The Line Spectral Pairs (LSPs) have a built in error detection property. When they are transmitted by the encoder, they will always be monotonic in frequency, $\omega_i < \omega_{i+1}$. The decoder must ensure monotonicity in the received LSPs by re-arranging them if necessary, to avoid instability in the IIR synthesis filter at the decoder. These methods have zero redundancy as no FEC bits are required. Some method of maintaining synthesis filter stability is essential, as the subjective effects of instability are extremely unpleasant.

Two methods of dealing with non-monotonicity have been investigated. Sort A separates non-monotonic LSPs by the minimum amount possible. This amount is set by the LSP quantizers for the two non-monotonic LSPs. Effectively, this is separating the two non-monotonic LSP's by the bandwidth expansion of the LSP quantizers. The Sort B method swaps the frequencies of the non-monotonic LSPs to ensure stability. Figure 2 demonstrates the performance of the Sort A and Sort B algorithms, against the no sorting case. Note that only the LSP bits (bits 0-33) are subjected to bit errors in this test. The Sort A algorithm is clearly superior, and was adopted as the LSP sorting algorithm for the coder.

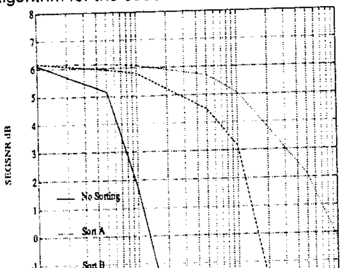

Figure 2: LSP monotonicty algorithms

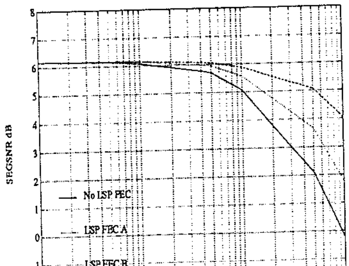

Figure 3: Application of FEC to LSPs

APPLICATION OF FEC TO THE LSPS

Using the Parity 1 bits, we can only protect 12 bits out of the 34 used to transmit the quantized LSPs. It is therefore necessary to determine which bits are the most important to the perceived quality of the decoded speech. This was achieved by subjecting each LSP bit to a bit error rate of 0.5, and measuring the SEGSNR and the Cepstral Distance (CD) of the decoder with the test database described above. All other bits in the coder were left error free, except for the particular LSP bit being tested. From this test, the sensitivity of the LSP bits can be graded, from most sensitive, to the least sensitive.

Two grading scales were obtained, one in order of decreasing CD (LSP FEC scheme A), and the other in order of increasing SEGSNR, (LSP FEC scheme B), which corresponds to the most sensitive bit to the least sensitive bit. Note that CD decreases with increasing speech quality, while SEGSNR increases. The 12 most sensitive bits were protected using the Golay code, and the performance determined for each LSP FEC scheme. Only the LSP bits were subjected to bit errors. The results are presented in Figure 3.

Figure 3 indicates that on a SEGSNR scale, LSP scheme B is superior, which is expected because the bits were protected according to SEGSNR sensitivity for scheme B. However, listening tests proved that in fact LSP scheme A is superior, despite the SEGSNR results. The LSPs are spectral parameters, they describe the short term spectrum of the speech signal. Their sensitivity to bit errors is thus best measured by a spectral measure, such as CD.

APPLICATION OF FEC TO THE EXCITATION PARAMETERS

Figure 4 indicates the sensitivity of the 4 excitation parameters to random errors. From this graph we can deduce that the most important parameters are L, the Adaptive codebook index, and α, the Stochastic codebook gain. To produce this plot each excitation parameter was subjected to random errors. When one parameter was being tested, all bits in that parameter were subjected to the same BER, and all other coder parameters were error free. Note that each of these parameters occurs 6 times in a frame, at the subframe rate of 5ms.

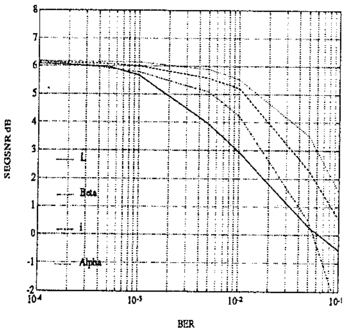

Figure 4: Sensitivity of excitation parameters to random errors

Bit error sensitivity tests were applied to each bit in the excitation parameters. In this case, only the SEGSNR measure in meaningful, because no errors at all were introduced into the LSPs, from which the CD measures are obtained. The tests were performed in a similar way to the LSPs, when a bit was tested, all other bits were kept error free. Note that when an excitation bit was subjected to bit errors, the same bit in the other 5 excitation subframes was also subjected. Thus only 21 separate bit error sensitivity measures were taken for the excitation parameters. The BER applied to the bits was 0.1 for the excitation bit error sensitivity tests.

Of the 126 bits used to transmit the excitation information, we can protect only 24, using the Parity 2 and Parity 3 bits, or 4 bits out of every 21 bit subframe. Another FEC scheme for the excitation was investigated, that used the Parity 1 bits to protect another 2 bits per subframe, or 36 bits out of the 126 excitation bits. Note that this scheme would remove all FEC protection from the LSPs. The results of the 24 bit FEC scheme (excitation FEC scheme A), and the 36 bit FEC scheme (excitation FEC scheme B) are presented in Figure 5. Bit errors are applied only to the excitation bits. There is a significant improvement in SEGSNR for high BER, and as expected the FEC B scheme performs better than the FEC A scheme.

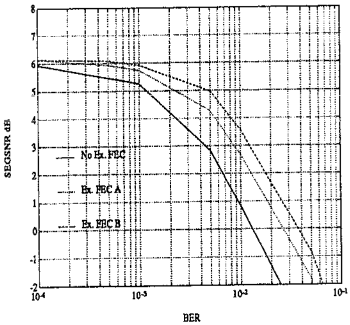

Figure 5: Application of FEC to excitation parameters

Subjectively, the results are less impressive, with only a small difference between excitation FEC A and excitation FEC B. Generally, errors in the excitation are perceptually less important than errors in the spectral parameters (LSPs). Excitation errors cause the coder to sound "rougher" or more "hoarse", and there is a corresponding loss of intelligibility.

COMBINED FEC

Figure 6 presents the performance of the coder with 3 FEC schemes that are derived from the LSP and excitation schemes discussed above. Note that there is very little difference in the SEGSNR plot for the 3 schemes. This indicates the inadequacy of SEGSNR. Subjectively, the best scheme was LSP FEC A (LSPs protected according to CD), and excitation FEC A. Figure 7 demonstrates the effectiveness of the Cepstral Distance (CD) measure, plotting the same schemes on a CD vs. BER plot instead of SEGSNR vs. CD. The order of preference on the CD plot agrees with the results obtained from informal listening tests.

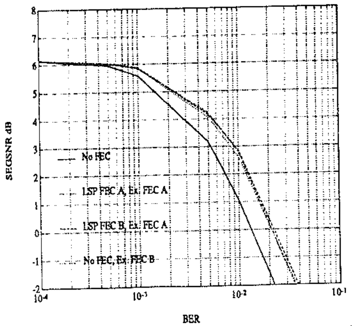

Figure 6: Combined FEC on SEGSNR scale

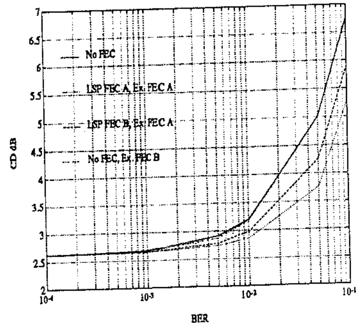

Figure 7: Combined FEC on CD scale

BURST ERROR MASKING

Burst errors involve the near total destruction of frames, rendering most FEC schemes useless. In fact, FEC codes such as the Golay (23,12) code implemented here actually create more errors than they correct, once a certain threshold BER has been reached.

A simple method for masking the effects of burst errors in CELP coders has been suggested (Perkis, 1990). In the presence of an error burst, the current frame is considered unusable due the high BER. A previous, hopefully less corrupted frame is substituted for the current one. Note that it is assumed that information on the channel state is available to the decoder, such as a signal strength indication from the demodulator.

For the simulation of the bursty channel, a 4 state Markov model was used (Bundrock, 1990). This model is considered to be a statistically accurate simulation of the burst error channel obtained in the Australian land mobile satellite environment. The worst case, Wellington 2 model is used.

The packet substitution scheme used is based on the 3 groups defined previously for the purposes of FEC. If the number of errors in a group (sub-packet) exceeds a given threshold, then the corresponding sub-packet from the previous frame is substituted.

The results are impressive. Without any Burst Error Masking (BEM), the coder sounds very poor in the burst error channel, despite the FEC. BEM improves the intelligibility significantly, removing some of the perceptually annoying effects of burst error channels. A bit error threshold of 10% for subpacket substitution was found to be optimum for this coder. Table 3 reflects the improvement using objective measurements Once again, the subjective speech quality is not well represented by SEGSNR.

Table 3: Effect of Burst Error Masking (BEM) on coder performance

Error Masking	SEGSNR dB	CD dB
None	-3.20	4.58
FEC	-2.43	3.91
BEM	-3.08	4.18
FEC & BEM	-2.30	3.55

CONCLUSION

A real time, Burst Error Masking CELP voice codec (BEM-CELP) that runs in real time on a single DSP is presented that is robust to the random and burst error environments. Both redundant, and non redundant error masking schemes have been implemented, with the objective of minimising the perceptual error of the decoded speech in the random and burst error channels. The zero redundancy schemes investigated include algorithms to ensure LSP monotonicity, and burst error masking algorithms based of the repetition of subpackets in the presence of burst errors. The implementation of an unequal FEC scheme based on the individual bit error sensitivity of the parameters in the coder has resulted in a coder that delivers good near toll quality speech at a random error BER of 2%.

REFERENCES

Atal, B.S, (1985), "*Code-Excited Linear Prediction (CELP): High Quality Speech at Very Low Bit Rates*", Proc ICASSP, pp937-940.

Perkis, A. (1990), "*Speech coding for mobile satellite communications; a novel scenario*", ISSPA 90, Gold Coast, August 1990.

Perkis, A., Ribbum, B., "*A good quality, low complexity 4.8kbit/s stochastic multipulse coder*", Proc ICASSP.

Itakura, P, (1975), "*Line Spectrum Pair representation of linear prediction coefficients of speech signals*", I. Acoust. Soc. Am.

(1989) Proposed U.S. Federal Standard 1016: "*Telecommunications: Analog to digital conversion of radio voice by 4,800 bit/second code excited linear prediction (CELP)*"", Second draft.

Bundrock, T., Wilkinson, M., (1990), "*Evaluation of voice codecs for the Australian Mobile Satellite System*", Ottawa, June 1990.