A RECONFIGURABLE SPEECH TRANSCEIVER

J.P. Woodard, L. Hanzo

Dept. of Electr. and Comp. Sc., Univ. of Southampton, SO17 1BJ, UK.
Tel.: +44 1703 593 125, Fax: +44 1703 593 045
Email: lh@ecs.soton.ac.uk, jpw91r@ecs.soton.ac.uk

ABSTRACT

A dual-rate 4.7/6.5 kbits/s algebraic code excited linear predictive (ACELP) codec is proposed for adaptive speech communicators, which can drop their source rate and speech quality under network control in order to invoke a more error resilient modem amongst less favourable channel conditions. Source-matched Bose-Chaudhuri-Hocquenghem (BCH) codecs combined with un-equal protection diversity- and pilot-assisted 16- and 64-level quadrature amplitude modulation (16-QAM, 64-QAM) are employed in order to accommodate both the 4.7 and the 6.5 kbits/s coded speech bits at a signaling rate of 3.1 kBd. In a bandwidth of 200 kHz 32 time slots can be created, which allows us to support in excess of 50 users, when employing packet reservation multiple access (PRMA). Good communications quality speech is delivered in an equivalent bandwidth of 4 kHz, if the channel signal-to-noise ratio (SNR) and signal-to-interference ratio (SIR) are in excess of about 15 and 25 dB for the lower and higher speech quality 16-QAM and 64-QAM systems, respectively.

1. INTRODUCTION

The near future is likely to witness the emergence of adaptive mobile speech communicators that can adopt their parameters to rapidly changing propagation environments and maintain a near-optimum combination of transceiver parameters in various scenarios. Therefore it is beneficial, if the transceiver can drop its speech source rate for example from 6.5 kbits/s to 4.7 kbits/s and invoke 16-level quadrature amplitude modulation (16-QAM) instead of 64-QAM, while maintaining the same bandwidth. In this treatise the underlying trade-offs of using such a dual-rate algebraic code excited linear predictive (ACELP) speech codec in conjunction with a diversity- and pilot-assisted coherent, re-configurable, un-equal error protection 16-QAM/64-QAM modem are investigated. Section 2 briefly describes the re-configurable transceiver scheme, Section 3 details the proposed dual-rate ACELP codec, followed by a short discussion on bit sensitivity issues in Section 4 and source-matched embedded error protection in Section 5. Finally, before concluding, the system performance is characterised in Section 6.

MOMUC-2, 2ND INTERNATIONAL WORKSHOP ON MOBILE MULTIMEDIA COMMUNICATIONS, 11-13 APR., 1995, BRISTOL, UK ©IEEE

2. THE TRANSCEIVER SCHEME

The schematic diagram of the proposed re-configurable transceiver is portrayed in Figure 1. A Voice Activity Detector (VAD) similar to that of the Pan-European GSM system [1] enables or disables the ACELP encoder [2] and queues the active speech frames in the Packet Reservation Multiple Access (PRMA) [3] contention buffer for transmission to the base station (BS). The 4.7 or 6.5 kbits/s (kbps) ACELP coded active speech frames are mapped according to their error sensitivities to n protection classes by the Bit Mapper, as shown in the Figure, and source sensitivity-matched binary Bose-Chaudhuri-Hocquenghem (BCH) encoded [4]. The 'Map & PRMA Slot Allocator' block converts the binary bit stream to 4- or 6-bit symbols, injects pilot symbols and ramp symbols and allows the packets to contend for a PRMA slot reservation. After BCH encoding the 4.7 and 6.5 kbps speech bits they are converted to 4- or 6-bit symbols, which modulate a re-configurable 16- or 64-level Quadrature Amplitude Modulation (QAM) scheme [5] will have the same signaling rate and bandwidth requirement. Therefore this transmission scheme can provide higher speech quality, if high channel signal-to-noise ratios (SNR) and signal-to-interference ratios (SIR) prevail, while it can be reconfigured under network control to deliver lower, but unimpaired speech quality amongst lower SNR and SIR conditions.

The structure of our dual rate speech encoder is shown in Figure 2. The transfer function $A(z)$ represents an all zero filter of order ten, which is used to model the short term correlation in the speech signal $s(n)$. The filter coefficients are determined for each 30 ms speech frame by minimising the variance of the prediction residual $r(n)$, using the autocorrelation method [6]. The filter $1/A(z)$, referred to as the inverse or synthesis filter is used in the decoder to produce the reconstructed speech signal from the excitation signal $u(n)$. The filter coefficients, as delivered by the autocorrelation method, are not suitable for quantization due to stability problems of the all pole synthesis filter $1/A(z)$. Therefore they are converted to Line Spectrum Frequencies (LSFs) [6], which are then scalar quantized with a total of 34 bits per frame.

The excitation signal $u(n)$ has two components, both of which are determined for each 5 or 7.5 ms subsegment of a 30 ms speech frame, depending on the targeted output bit rate. The first component is an entry from an adaptive codebook, which is used to model the long term periodicity of the speech signal. The other excitation component is derived from a fixed code-book which models the random, Gaussian-like prediction residual of the speech signal after both its

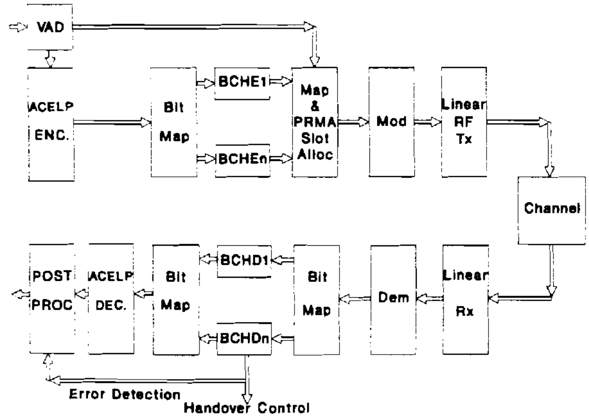

Figure 1: Transceiver Schematic

long- and short-term correlations have been removed. The excitation is described in terms of the following parameters.

- The adaptive code-book delay α that can take any integer value between 20 and 147 and hence it is represented using 7 bits.
- The adaptive code-book gain G_1 which is non-uniformly quantized with 3 bits.
- The index of the optimum fixed code-book entry $c_k(n)$, which is one of 4096 entries and therefore it is represented with 12 bits.
- The fixed code-book gain G_2, the magnitude of which is quantized with a four bit logarithmic quantizer and its sign is assigned an additional bit.

3. THE DUAL-RATE ACELP CODEC

Thus a total of 27 bits are needed to represent the excitation signal $u(n)$. In the low- and high-rate mode of the ACELP codec each 30 ms frame is split into four and six sub-frames, respectively, for each of which the excitation information is optimised and quantised using 27 bits. Thus for the low-rate mode we have a total of $(34+4 \cdot 27)=142$ bits per 30 ms frame, or a rate of about 4.7 kbits/s, while in the high rate

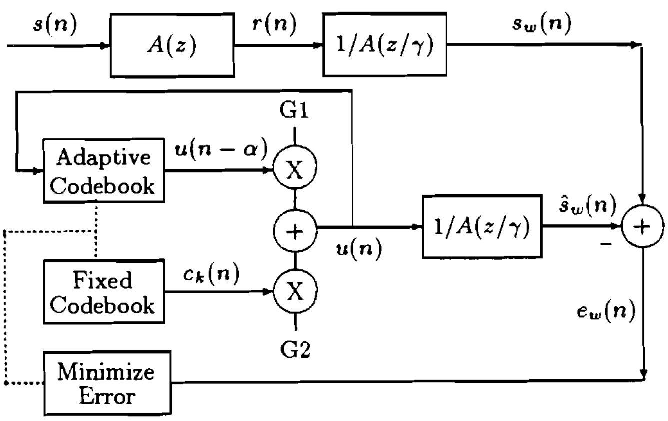

Figure 2: ACELP Encoder Structure

mode the encoder generates $(34+6 \cdot 27)=196$ bits per 30 ms frame, yielding a data rate of just over 6.5 kbits/s.

In the encoder the excitation is passed through the so-called weighted synthesis filter $1/A(z/\gamma)$. Similarly, the prediction residual $r(n)$ is passed through an identical filter to produce a weighted version $s_w(n)$ of the input speech signal $s(n)$. It is then the weighted error $e_w(n)$ between the original and reconstructed speech which the encoder attempts to minimize when choosing the excitation parameters. This weighting process deemphasizes the energy of the error signal in the formant regions, where the speech has high energy. The factor γ determines the extent of the weighting, and was set to 0.8 for the 6.5 kbits/s codec, and 0.9 for the 4.7 kbits/s codec.

The excitation parameters listed above are determined using a closed loop analysis-by-synthesis procedure. This significantly improves the codec's performance, at the expense of a major increase in complexity. First the effect of passing each possible adaptive code-book word through the weighted synthesis filter is calculated. Then the parameters G_1 and α which minimize the total squared error between the output of the filter and the original weighted speech are chosen. Finally, given the adaptive code-book parameters, the fixed code-book parameters are calculated in a similar way.

Traditionally the major part of a CELP codec's complexity accrues from the fixed code-book search. For a twelve bit code-book the effect of filtering 4096 codewords through the weighted synthesis filter must be determined. In our codec we opted for an algebraic code excited linear predictive (ACELP) structure [2, 4], in which each 40 or 60 samples long excitation pattern has only four non-zero excitation pulses given by 1,-1,1, -1. Each pulse has eight legitimate positions, and the optimum positions can be determined efficiently using a series of four nested loops. This dramatically reduces the complexity of the code-book search, and also removes the need for a fixed code-book to be stored both at the encoder and decoder.

4. ERROR SENSITIVITY ISSUES

In order to achieve high robustness over fading mobile channels, in this Section ways of quantifying and improving the error sensitivity of our codec are discussed. It has been noted [7, 8] that the spectral parameters in CELP codecs are particularly sensitive to errors. This impediment can be mitigated by exploiting the ordering property of the LSFs [6]. When a non-ordered set of LSFs is received, we examine each LSF bit to check whether inverting this bit would eliminate the error. If several alternative bit inversions are found, which correct the cross-over, then the one that produces LSFs as close as possible to those in the previous frame is used. With this technique we found that when the LSFs were corrupted such that a cross-over occurred, which happened about 30% of the time, the corrupted LSF was correctly identified in more than 90% of the cases, and the corrupted bit was pinpointed about 80% of the time. When this happens, the effect of the bit error is completely removed and so 25% of all LSF errors can be entirely removed by the decoder. When un-equal error protection is used, the situation can be further improved because the algorithm will know which bits are more likely to be corrupted, and so can attempt correcting the LSF crossover by inverting these bits

Bit Index	Parameter
1 to 34	LSFs
35 to 41	Adaptive Code-book Delay
42 to 44	Adaptive Code-book Gain
45 to 56	Fixed Code-book Index
57	Fixed Code-book Gain Sign
58 to 61	Fixed Code-book Gain

Table 1: Bit Allocation Scheme

first.

The algebraic code-book structure used in our codec is inherently quite robust to channel errors. This is because if one of the code-book index bits is corrupted, the code-book entry selected at the decoder will differ from that used in the encoder only in the position of one of the four non-zero pulses. Thus the corrupted code-book entry will be similar to the original. This is in contrast to traditional CELP codecs which use a non-structured, randomly filled code-book. In such codecs when a bit of the code-book index is corrupted, a different code-book address is decoded and the code-book entry used is entirely different from the original, inflicting typically a segmental signal-to-noise ratio (SEGSNR) degradation of about 8 dB. In contrast, our codec exhibits degradation of only about 4 dB, when a code-book index bit is corrupted in every frame.

The magnitude of the fixed code-book gain tends to vary quite smoothly from one sub-frame to the next. Therefore errors in the code-book gain can be spotted using a smoother, indicating on the basis of neighbouring gains, what range of values the present gain is expected to lie within. If a code-book gain is found, which is not in this range, then it is assumed to be corrupted, and replaced with an estimated gain. After careful investigation we implemented a smoother which was able to spot almost 90% of the errors in the most significant bit of the gain level, and reduced the error sensitivity of this bit by a factor of five. The fixed code-book gain sign shows erratic behaviour and hence does not lend itself to smoothing. This bit has a high error sensitivity and has to be be well protected by the channel codec.

Seven bits per sub-frame are used to encode the adaptive code-book delay, and most of these are extremely sensitive to channel errors. An error in one of these bits produces a large SNR degradation not only in the frame in which the error occurred, but also in subsequent frames, and generally it takes more than ten frames before the effect of the error dies out. Therefore these bits should wherever possible be strongly protected. The pitch gain is much less smooth than the fixed code-book gain, and hence it is not amenable to smoothing. However its error sensitivity can be reduced by Gray-coding the quantizer levels.

Clearly, some bits are much more sensitive to channel errors than others, and so should be more heavily protected by the channel codec. However it is not obvious how the sensitivity of different bits should be quantified. A commonly used approach is, for a given bit, to invert this bit in every frame and evaluate the SEGSNR degradation. The error sensitivity for our 4.7 kbits/s ACELP codec measured in this way is shown in Figure 3 for the 34 LSF bits and the 27 first sub-frame bits. The bit allocation scheme used is shown in Table 1.

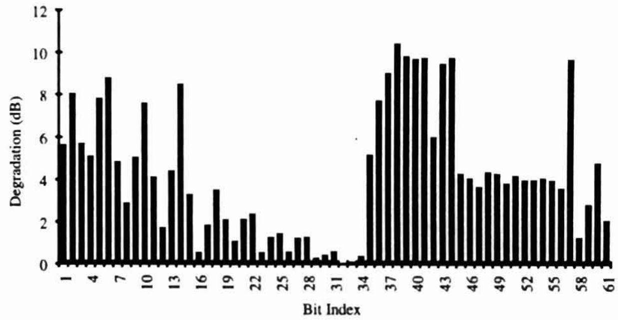

Figure 3: SEGSNR Degradations of the 4.7 kbps codec due to 100% Bit Error Rate

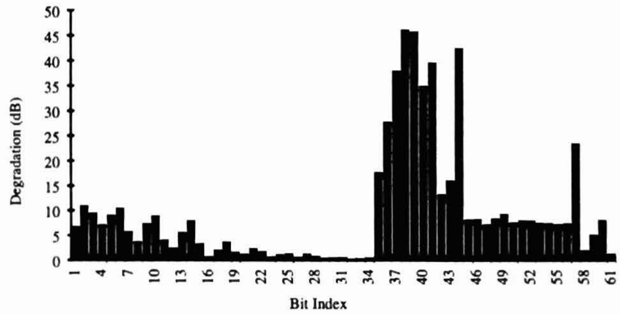

Figure 4: SNR Degradation of the 4.7 kbps codec due to Single Errors in Various Bits

The problem with this approach of quantifying error sensitivity is that it does not take adequate account of the different error propagation properties of different bits. This means that if, instead of corrupting a bit in every frame it is corrupted randomly with some error probability, then the relative sensitivity of different bits will change. Hence we propose a new measure of error sensitivity. For each bit we find the average SNR degradation due a single bit error both in the frame in which the error occurs and in consecutive frames. The total SNR degradation is then found by adding or integrating the degradations caused in the erroneous frame and in all the following frames which are affected by the error. Figure 4 demonstrates the total SNR degradations evaluated in this way. The graph is significantly different from that in Figure 3. In particular, the importance of the adaptive code-book delay bits due to their memory propagation becomes more explicit.

Our final error sensitivity measure is based on the total SEGSNR degradations described above and on a similar measure for the total Cepstral Distance degradation. The two sets of degradation figures are given equal weight and combined to produce an overall sensitivity measure, which was invoked to assign the speech bits to various bit protection classes.

5. EMBEDDED ERROR PROTECTION

5.1. Low-quality Mode

In this Section we highlight our code design approach using the 4.7 kbits/s codec and note that similar principles were followed in case of the 6.5 kbits/s codec. The sensitivity of the 4.7 kbps ACELP source bits was evaluated in the

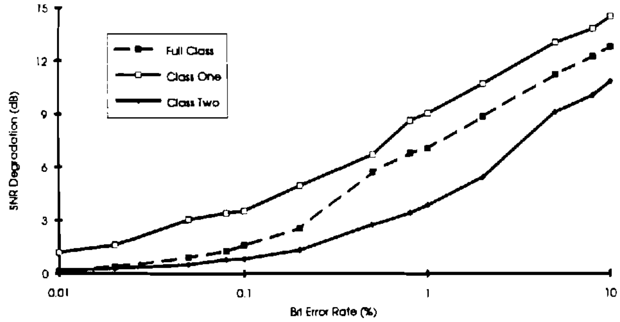

Figure 5: SEGSNR Degradation versus Bit Error Rate for the 4.7 kbps ACELP codec

previous Section in Figures 3 and 4, but the number of bit protection classes n still remains to be resolved. Intuitively, one would expect that the more closely the FEC protection power is matched to the source sensitivity, the higher the robustness. In order to limit the system's complexity and the variety of candidate schemes, in case of the 4.7 kbits/s ACELP codec we have experimented with a full-class BCH codec, a twin-class and a quad-class scheme, while maintaining the same coding rate.

Non-coherent QAM modems are less complex to implement, but typically require higher SNR and SIR values than their coherent counterparts. Hence in our proposed scheme second-order switched-diversity assisted coherent Pilot Symbol Assisted Modulation (PSAM) using the maximum-minimum-distance square QAM constellation is preferred [5]. Our propagation conditions are characterised by a pedestrian speed of 3 mph, propagation frequency of 1.8 GHz, pilot symbol spacing of P=10 and a signaling rate of 100 kBd, which will fit in a bandwidth of 200 kHz, when using a unity roll-off factor. For a channel SNR of about 20 dB this modem provides two independent QAM subchannels exhibiting different bit error rates (BERs). The BER is about a factor three to four times lower for the higher integrity path referred to as the Class 1 (C1) subchannel than for the C2 subchannel over Rayleigh-fading channels [9]. We capitalise on this feature to provide un-equal source sensitivity-matched error protection combined with different BCH codecs for our ACELP codecs.

If the ratio of the BERs of these QAM subchannels does not match the sensitivity constraints of the ACELP codec, it can be 'fine-tuned' by the help of different BCH codecs, while maintaining the same number of BCH-coded bits in both subchannels. However, the increased number of redundancy bits of stronger BCH codecs requires that a higher number of sensitive bits are directed to the lower integrity C2 subchannel, whose coding power must be concurrently reduced in order to accommodate more source bits. This non-linear optimisation problem can only be solved experimentally, assuming a certain sub-division of the source bits, which would match a given pair of BCH codecs.

For the full-class system we have decided to use the approximately half-rate BCH(127,71,9) codec in both subchannels, which can correct 9 errors in each 127-bits block, while encoding 71 primary information bits. The coding rate is R=71/127 $\approx$ 0.56 and the error correction capability is about 7 %. When splitting the ACELP source bits into two classes each hosting 71 bits, the more sensitive bits require almost an order of magnitude lower BER than the more robust bits, in order to inflict a similar SEGSNR penalty. Hence both classes must be protected by different codes, and after some experimentation we found that the BCH(127,57,11) and BCH(127,85,6) codes employed in the C1 and C2 16-QAM subchannels provide the required integrity. The SEGSNR degradation caused by a certain fixed BER assuming randomly distributed errors is portrayed in Figure 5 for both the full-class and the above twin-class system, where the number of ACELP bits in these protection classes is 57 and 85, respectively. Note that the coding rate of this system is the same as that of the full-class scheme and each 142-bit ACELP frame is encoded by two BCH codewords, yielding $2 \cdot 127$=254 encoded bits and curtailing error propagation at the block boundaries. The FEC-coded bit rate became $\approx$8.5 kbps.

With the incentive of perfectly matching the FEC coding power and the number of bits in the distinct protection classes to the ACELP source sensitivity requirements we also designed a quad-class system, while maintaining the same coding rate. We used the BCH(63,24,7), BCH(63,30,6), BCH(63,36,5) and BCH(63,51,2) codes and transmitted the most sensitive bits over the C1 16-QAM subchannel using the two strongest codes and relegated the rest of them to the C2 subchannel, protected by the two weaker codes.

The PRMA control header was allocated a BCH(63,24,7) code in case of all schemes and hence the total PRMA framelength became 317 bits, representing 30 ms speech and yielding a bit rate of $\approx$ 10.57 kbps. The 317 bits constitute 80 16-QAM symbols, and 9 pilot symbols as well as 2+2=4 ramp symbols must be added, resulting in a PRMA framelength of 93 symbols per 30 ms slot. Hence the signaling rate becomes 3.1 kBd. Using a PRMA bandwidth of 200 kHz, similarly to the Pan-European GSM system, and a filtering excess bandwidth of 100 % allowed us to accommodate 100 kBd/3.1 kBd $\approx$ 32 PRMA slots.

5.2. High-quality Mode

Following the approach proposed in the previous Subsection we designed a triple-class source-matched protection scheme for the 6.5 kbps ACELP codec. The reason for using three protection classes this time is that the 6.5 kbps ACELP codec's higher bit rate must be accommodated by a 64-level QAM constellation, which inherently provides three different subchannels. When using second-order switched-diversity and pilot-symbol assisted coherent square-constellation 64-QAM [5] amongst our previously stipulated propagation conditions with a pilot-spacing of P=5 and channel SNR of about 25 dB, the C1, C2 and C3 subchannels have BERs of about 10^{-3}, 10^{-2} and $2 \cdot 10^{-2}$, respectively [8].

The source sensitivity-matched codes for the C1, C2 and C3 subchannels are BCH(126,49,13), BCH(126,63,10) and BCH(126,84,6), while the packet header was allocated again a BCH(63,24,7) code. The total number of BCH-coded bits becomes $3 \cdot 126 + 63$=441/30 ms, yielding a bit rate of 14.7 kbps. The resulting 74 64-QAM symbols are amalgamated with 15 pilot and 4 ramp symbols, giving 93 symbols/30 ms, which is equivalent to a signaling rate of 3.1 kBd, as in case of the low-quality mode of operation. Again, 32 PRMA slots can be created, as for the low-quality system, accommodating more than 50 speech users in a bandwidth of 200 kHz and yielding a speech user bandwidth of about