DIGITAL SIGNAL PROCESSING 7, 37–46 (1997)
ARTICLE NO. SP970279

A Range of Low and High Delay CELP Speech Codecs between 8 and 4 kbits/s

J. P. Woodard[1] and L. Hanzo[2]

Department of Electronics and Computer Science, University of Southampton, SO17 1BJ, United Kingdom

Woodward, J. P., and Hanzo, L., A Range of Low and High Delay CELP Speech Codecs between 8 and 4 kbits/s, *Digital Signal Processing* 7 (1997), 37–46.

In this paper we study the performance and the error sensitivities of six CELP [1] based codecs operating between 8 and 4 kbits/s. Codecs using both forward and backward adaption of the linear prediction coefficients and the long term predictor (LTP) are described. Initially we describe four low delay codecs which all use backward adaption of the LPC coefficients but which differ in their use of LTP. These codecs all have frame-lengths of 3 ms or less, and their performance at various bit rates between 8 and 4 kbits/s is examined. Next the error sensitivity of these codecs, and means of improving it, are described. Then an algebraic CELP (ACELP) [2] codec operating at 6.2 kbits/s with a frame-length of 5 ms is described. Our final codec also uses ACELP and operates between 4.7 and 7.1 kbits/s, but it is forward adaptive and so it has a much longer frame-length of up to 30 ms. After describing this codec we compare the performance of our codecs in both error-free conditions and in the presence of channel errors. Surprisingly the error sensitivity of the low delay backward adaptive codec with no LTP is similar to that of the forward adaptive, high delay, ACELP codec. © 1997 Academic Press

1. INTRODUCTION

During the past 10 years many speech codecs offering communications to toll quality reconstructed speech at bit rates of 16 kbits/s and below have been developed, and several of these are now commonly used. For example, in 1986 a 13 kbits/s regular pulse excitation (RPE) [3,4] codec was selected for use in the Pan-European GSM mobile phone network, and more recently vector sum excited linear prediction (VSELP) [5,6] codecs at 8 and 6.7 kbits/s were chosen for use in the North American IS54 and the Japanese PDC digital mobile communications networks. Also in 1991 a lower speech quality CELP codec [7] operating at 4.8 kbits/s was standardized as the U.S. Department of Defence Federal Standard 1016. Many of these, and other codecs, have been documented in books by O'Shaugnessy [8], Furui [9], Salami *et al.* [10], Anderson and Shesadri [11], Kondoz [12], and others. Also Gersho provides an excellent overview of recent work in his 1994 paper [13].

Much work has been done to produce lower bit rate speech codecs with good quality speech, and further significant advances were incorporated into half-rate speech codecs for both GSM and the Japanese PDC system. The 5.6 kbits/s half-rate GSM speech codec [14] uses VSELP with switching between four different operational modes, depending on the grade of voicing detected in the speech to be encoded, whereas the 3.45 kbits/s half-rate PDC speech codec [15] uses pitch synchronous innovation (PSI) CELP. Work is continuing on other schemes, for example prototype waveform interpolation (PWI) [16], multi-band excitation (MBE) [17], and interpolated zinc function prototype excitation (IZFPE) [18].

It is clear that a wide range of speech codecs offering different quality reconstructed speech at various bit rates are available. However, until recently most of these codecs used forward adaption (FA) to determine the short term linear prediction coefficients which are used in the encoding and decoding processes. Such codecs typically buffer about 20 or 30 ms of the input speech and use this buffered speech to determine the linear prediction coefficients (LPC). Chen *et al.*, however, argued that speech

[1] E-mail: jpw@ecs.soton.ac.uk.

[2] lh@ecs.soton.ac.uk.

1051-2004/97 $25.00

transmission systems tend to have an end to end delay of about three times the frame-length of the encoder [19], and so some of these low rate speech codecs have undesirably high one-way delays approaching 100 ms.

Therefore in recent years much work has been devoted to produce medium/low rate speech codecs with more modest delays. In 1992 a low delay 16 kbits/s CELP codec was developed by the AT&T speech compression team and standardised by the CCITT as G.728 [19,20]. This codec uses backward adaption (BA) to determine the linear prediction coefficients which are used in the encoding and decoding of the speech. Hence the LPC coefficients are derived from the past reconstructed speech rather than the future input speech, and therefore it is not necessary to buffer a long frame of the input speech for the encoding. This backward adaption allows the G.728 codec to produce toll quality reconstructed speech at 16 kbits/s with a frame-length of only 0.625 ms. More recently an 8 kbits/s codec with a frame-length of 10 ms has been developed in cooperation by the Sherbrooke speech coding team [21,22], AT&T, France Telecom/CNET, and NTT [23], which was standardised as G.729. This codec uses forward adaption of the linear prediction coefficients but manages to maintain a reasonably low delay by using a frame-length of only 10 ms, along with vector quantization of the LPC coefficients.

In this paper we seek to compare the performance and error robustness of six backward- and forward-adaptive CELP based speech codecs, Codecs A–F in Table 1, operating between 4 and 8 kbits/s. We will show that when dispensing with LTP, it is feasible to contrive low-rate backward adaptive codecs that have adequate error resilience. Initially in Section 2 we describe four backward adaptive codecs, based on the philosophy of the G.728 codec, which operate at rates between 8 and 4 kbits/s with frame-lengths between 1.5 and 3 ms. Then we investigate the error sensitivity of these codecs and describe two methods which were used to improve this error sensitivity. In Section 4 we describe a low delay algebraic CELP (ACELP) codec similar to G.729 but which uses backward adaption of its synthesis filter, allowing it to operate at a bit rate of 6.2 kbits/s and with a frame-length of 5 ms. Then in Section 5 we describe a conventional forward adaptive ACELP codec operating at 4.7 and 6.5 kbits/s with a frame-length of 30 ms and at 7.1 kbits/s with a frame-length of 20 ms. Finally in Section 6 the relative performance and error resilience of these codecs is examined.

2. FOUR LOW DELAY CODECS OPERATING BETWEEN 8 AND 4 kbits/s

In this section we describe four low delay CELP codecs based loosely on the philosophy of the G.728 16 kbits/s codec. The G.728 codec [19,20] uses a frame-length of five samples or 0.625 ms, with 10 bits being used to encode each five-sample frame, giving a bit rate of 16 kbits/s. Backward adaption is used to derive the short-term filter coefficients at both the encoder and the decoder, and hence, no bits need to be transmitted to specify the filter coefficients used. Therefore all 10 bits per five-sample frame are used to encode the filter excitation, which is vector quantized with a 7-bit shape codebook and a 3-bit gain codebook. For each frame the best excitation is chosen using an analysis-by-synthesis (AbS) search.

We previously showed [24] how the G.728 codec could be modified to give a variable rate codec

TABLE 1
Summary of Different Codecs Used

	Synthesis filter	Long term predictor	Excitation quantization	Frame length	Bit rate
Codec A	Backward adapted $p = 50$	None	8-bit shape plus 4-bit scalar gain	1.5–3 ms	8–4 kbits/s
Codec B	Backward adapted $p = 20$	3-tap backward adapted	8-bit shape plus 4-bit scalar gain	1.5–3 ms	8–4 kbits/s
Codec C	Backward adapted $p = 20$	Partially forward adapted	8-bit shape plus 4-bit vector gain	1.5–3 ms	8–4 kbits/s
Codec D	Backward adapted $p = 20$	Switched forward adapted	8-bit shape plus 4-bit vector gain	1.5–3 ms	8–4 kbits/s
Codec E	Backward adapted $p = 20$	Entirely forward adapted	17-bit ACELP shape plus 7-bit vector gain	5 ms	6.2 kbits/s
Codec F	Forward adapted $p = 10$	Entirely forward adapted	12-bit ACELP shape plus (5 + 3)-bit scalar gain	20–30 ms	7.1–4.7 kbits/s

between 8 and 16 kbits/s with a graceful degradation in the speech quality of the codec as the bit rate is reduced. Here we extend this work to produce four low delay codecs operating between 8 and 4 kbits/s. All four codecs use backward adaption for their synthesis filters and transmit 12 bits per frame to represent the excitation to this filter. The codecs vary their bit rates by increasing the number of speech samples coded per frame from 12 to 24 samples, giving bit rates between 8 and 4 kbits/s. Initially we describe two entirely backward adaptive codecs, the first of which follows the philosophy of G.728 and does not use LTP, whereas the second codec uses backward adaptive LTP. Then we consider forward adaption of the long term predictor, and finally, we examine the effects of using switched voiced/unvoiced gain and shape excitation codebooks.

2.1. Entirely Backward Adaptive Codecs

Our first two low delay codecs, referred to as Codec A and Codec B, both operate in an entirely backward adaptive manner and transmit no information regarding regarding either the short- or long-term filters to be used at their decoders. The difference between the two codecs lies in how they treat the long term periodicities in the speech to be encoded. Codec A follows the philosophy of G.728 and does not employ an explicit long-term predictor, but instead it uses a very high order synthesis filter. In both G.728 and our Codec A the LPC filter order of $p = 50$ is used. Codec B, on the other hand, uses a synthesis filter of order $p = 20$ but also employs a 3-tap backward adaptive LTP. The LTP delay, which takes integer values between 20 and 147, and the gain are determined at both the encoder and the decoder, based on the correlations in the previous values of the synthesis filter's excitation. The schematic of the encoder in Codec B is shown in Fig. 1, where the synthetic speech is generated by filtering the excitation through the backward adaptive synthesis filter. Observe in the figure that the vectors of the shape codebook are scaled by the gain codebook as well as by the backward adaptive gain and then filtered through the 3-tap LTP in order to generate the short-term synthesis filter's excitation. Codec A is identical except the long term filter shown in Fig. 1 is absent. In both codecs an 8-bit vector shape codebook and a 4-bit scalar gain codebook are used to represent the excitation to the synthesis filter and the LTP if it is present.

For both Codec A and Codec B, as well as the other two low delay codecs of Table 1 to be described later, the quantization of the excitation gain is assisted by backward gain adaption [25], which for each vector produces a predicted gain using an adaptive 10th-order linear prediction filter in the logarithmic domain. It is then the ratio of the predicted gain to the "optimum" gain required which is quantized, and this aids the efficiency of the gain quantization and leads to a significant improvement in the codec's performance. For all four low delay codecs the entries of the shape and gain codebooks were trained using a closed-loop training technique similar to that described in [26]. We found that this training gave a significant improvement in all the codecs' performances.

The segmental SNRs of Codec A and Codec B, as well as the two other low delay codecs of Table 1 to be described later, are shown in Fig. 2 for codecs with frame-lengths of 12, 15, 18, and 24 samples and so bit rates of 8, 6.4, 5.3, and 4 kbits/s. It can be seen by comparing Codec A to Codec B in this figure that the

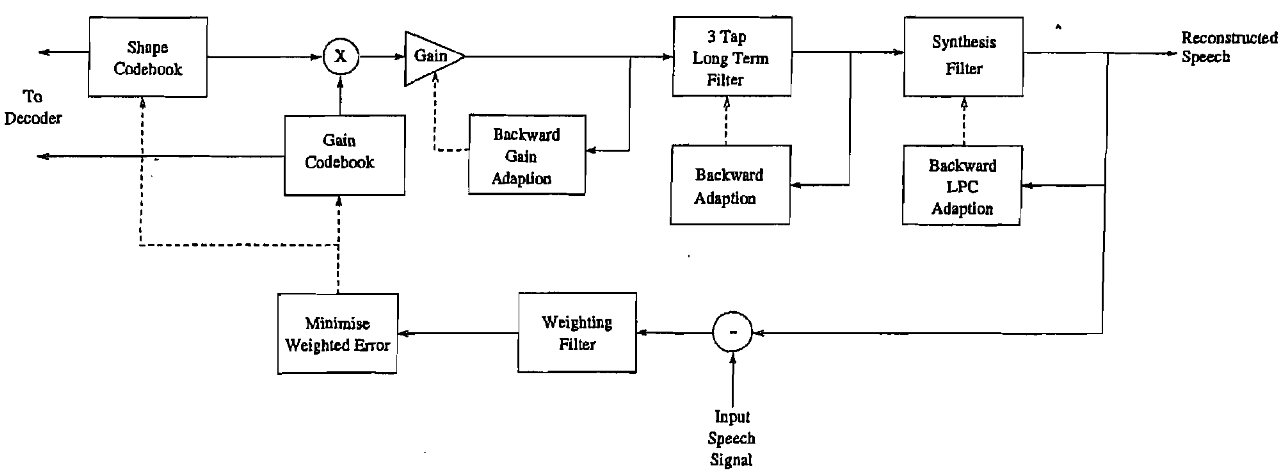


FIG. 1. "Codec B" backward adaptive CELP encoder.

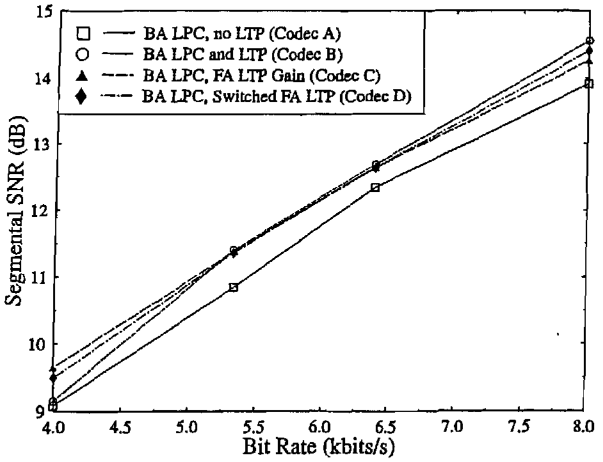


FIG. 2. Segmental SNR versus bitrate performance of our low delay CELP codecs.

addition of backward adapted LTP improves the segmental SNR of the codec by about 0.5 dB at 8 kbits/s, but as the bit rate is reduced the effectiveness of backward adapted LTP decreases. As expected, at 4 kbits/s Codec A and Codec B give almost identical segmental SNRs and the error sensitivity of the LTP is typically very high.

2.2. *The Effects of Forward Adaption of the LTP*

We also examined the effect of using forward adaption of the LTP in our low delay codecs. As seven bits would be needed to represent the LTP delay if this were forward adapted, and a total of only 12 bits are available to represent the filter excitation, we considered it impractical to use forward adaption for the LTP delay of our low bit rate, low delay, codecs. However, it is possible to use forward adaption of the LTP gain, and we implemented this in our third codec, referred to as Codec C. This codec, like Codec B, uses a short-term filter of order $p = 20$ and backward adaption of the LTP delay. However, the LTP gain is jointly determined with the fixed excitation gain in the AbS search of the excitation codebooks. The two gains are vector quantized using 4 bits, and again 8 bits are used to vector quantize the excitation shape. The structure of this codec is shown in Fig. 3.

The segmental SNR of Codec C is also shown in Fig. 2 at bit rates between 8 and 4 kbits/s. It is clear that at 8 kbits/s the entirely backward adapted LTP used in Codec B outperforms the forward adaption of the LTP gain used in Codec C. However as the bit rate is reduced the codec using entirely backward

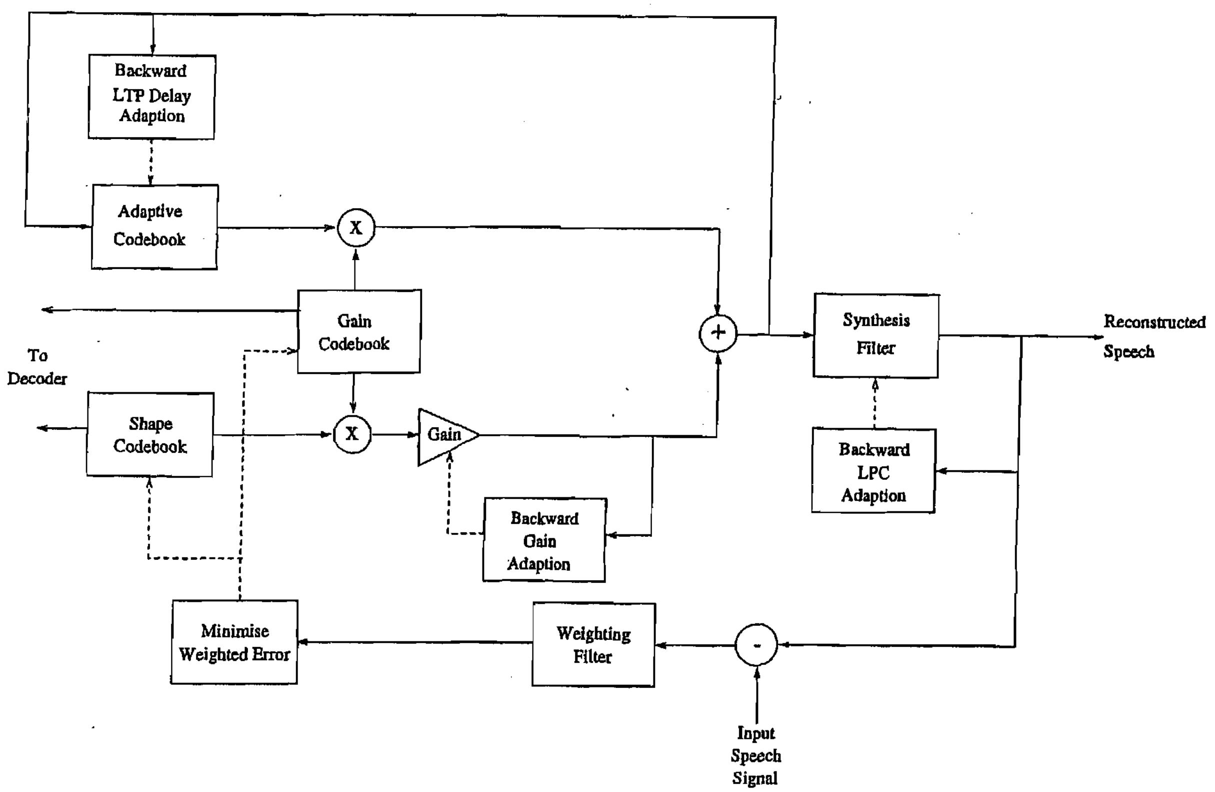
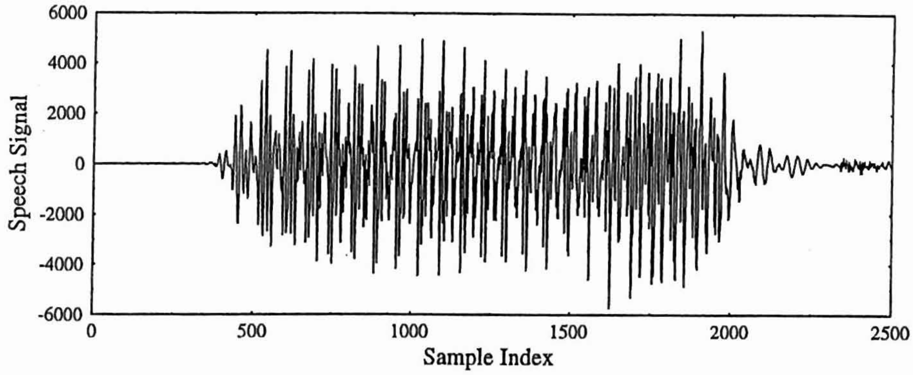


FIG. 3. "Codec C" low delay CELP encoder with forward adaptive LTP gain.



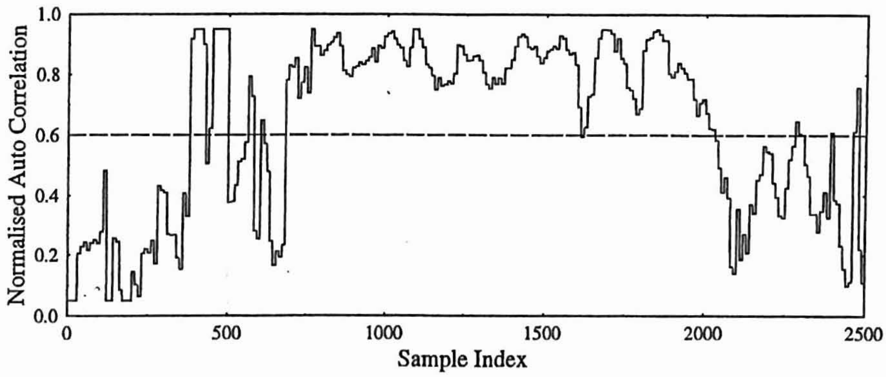


FIG. 4. Normalised autocorrelation value β_α during voiced and unvoiced speech.

adaptive LTP is more seriously affected so that at 4 kbits/s Codec C significantly outperforms Codec B.

2.3. *Switched Voiced / Unvoiced Codebooks*

An interesting feature in some recent speech codecs is the use of different codebooks to represent the excitation signal for different modes of speech [27]. For example the 5.6 kbits/s half-rate GSM speech codec [14] uses four modes depending on the voicing of the speech to be coded. Therefore, we examined the effects of the use of such switched excitation codebooks in our low delay codecs. We used separately trained shape and gain codebooks for the voiced and unvoiced segments of the speech to be coded, and made the voicing decision in a backward adaptive manner as described below. We employed the voiced/unvoiced decision with both entirely backward adaptive LTP (as used in Codec B) and in a codec using forward adaption for the LTP gain (as in Codec C). We found that switching between voiced and unvoiced shape and scalar gain codebooks in the case of entirely backward adaptive LTP gave no significant improvement in the codec's performance. However, when forward adaption was used to determine the LTP gain and a joint vector codebook was used to quantize this gain, together with the fixed excitation gain, some improvement was obtained using voiced/unvoiced switching.

Thus our final low delay codec, referred to as Codec D, uses a structure very similar to Codec C, as shown in Fig. 3. Separate shape and gain codebooks are used to code segments of speech classified as voiced and unvoiced by a backward adaptive switch. This switching is based on the voiced/unvoiced switching used in the postfilter employed in the G.728 codec [20]. In our codec the switch uses the normalized autocorrelation value of the past reconstructed speech signal $\hat{s}(n)$ at the delay α which is used by the adaptive codebook. This normalized autocorrelation value β_α is given by

$$\beta_\alpha = \frac{\sum_{n=-100}^{-1} \hat{s}(n)\hat{s}(n-\alpha)}{\sum_{n=-100}^{-1} \hat{s}^2(n-\alpha)}, \tag{1}$$

and when it is greater than a set threshold the speech is classified as voiced; otherwise the speech is classified as unvoiced. In our codec, as in the G.728 postfilter, the threshold is set to 0.6.

Figure 4 shows a segment of the original speech and the normalised autocorrelation value β_α calculated from the reconstructed speech of our 8 kbits/s codec. To aid the clarity of this graph the values of β_α have been limited to lie between 0.05 and 0.95. It can be seen that the condition $\beta_\alpha > 0.6$ gives a good indication of whether the speech is voiced or unvoiced.

The segmental SNR of Codec D is shown along with those of Codecs A–C in Fig. 2. It can be seen by comparing this curve to the performance of Codec C that at 8 kbits/s the backward adaptive switching between specially trained voiced and unvoiced gain and shape codebooks improves the performance of the codec. However, as the bit rate is reduced, the gain due to this codebook switching is eroded, and at 4 kbits/s Codec D gives a lower segmental SNR than Codec C. This is due to inaccuracies in the backward adaptive voicing decisions at the lower bit rates—we found that at 4 kbits/s the condition $\beta_\alpha > 0.6$ did not give a good indication of the voicing of the speech to be encoded.

In informal listening tests we found that all four low delay codecs described in this section gave near toll quality speech at 8 kbits/s, with differences between the codecs being difficult to distinguish. However, at 4 kbits/s Codec C sounded clearly better than the other codecs and gave reconstructed speech of communications quality.

3. THE ERROR SENSITIVITY OF THE LOW DELAY CODECS

In this section we consider the error sensitivity of our low delay codecs. For simplicity, only the error sensitivities of the codecs operating with a framelength of 15 samples and a bit rate of 6.4 kbits/s are detailed in this section. However, similar results also apply at the other bit rates.

It is well known that codecs using backward adaption for both the LTP delay and gain are very sensitive to bit errors, and this is why LTP was not used in G.728 [19]. Thus, as expected, we found that Codec B gave a very poor performance, when subjected to even a relatively low bit error rate (BER). Unfortunately, we also found similar results for Codec C and Codec D which, although they used backward adaption for the LTP delay, used forward adaption for the LTP gain. We therefore decided that neither Codec B, Codec C, nor Codec D were suitable for use over error-prone channels, and we examined the error sensitivity of Codec A, which does not use LTP. At 6.4 kbits/s this codec transmits only 12 bits per 15 sample frame from the encoder to the decoder. Of these 12 bits eight are used to represent the index of the shape codebook, and the remaining four bits are used to represent the index of the gain codebook entry used. The error resilience of these bits can be significantly improved by careful assignment of codebook indices to the various codebook entries. Ideally, each codebook entry would be assigned an index so that corruption of any of the bits representing this index will result in another entry being selected in the decoder's codebook which is in someway "close" to the intended codebook entry. If this ideal can be achieved, then the effects of errors in the bits representing the codebook indices will be minimised.

Consider first the 8-bit shape codebook. Initially the 256 available codebook indices are effectively randomly distributed amongst the codebook entries. We seek to rearrange these codebook indices so that when the index representing a codebook entry is corrupted, the new index will represent a codebook entry that is "close" to the original entry. In our work we chose to measure this "closeness" by the squared error between the original and the corrupted codebook entries. We considered only the effects of single bit errors among the eight codebook bits because at reasonable BERs the probability of two or more errors occurring in eight bits will be small. Thus for each codebook entry the "closeness" produced by a certain arrangement of codebook entries is given by the sum of the squared errors between the original codebook entry and the eight corrupted entries that would be produced by inverting each of the eight bits representing the entry's index. The overall "cost" of a given arrangement of codebook indices is then given by the closeness for each codebook entry, weighted by the probability of that codebook entry being used. Thus the cost we seek to minimise is given by

$$\text{Cost} = \sum_{j=0}^{255} P(j) \left[\sum_{i=1}^{8} \left(\sum_{n=1}^{15} (c_j(n) - c_j^i(n))^2 \right) \right], \quad (2)$$

where $P(j)$ is the probability of the jth codebook entry being used, $c_j(n)$, $n = 1 \ldots 15$, is the jth codebook entry and $c_j^i(n)$ is the entry that will be received if the index j is transmitted but the ith bit of this index is corrupted.

The problem of choosing the best arrangement of the 256 codebook indices among the codebook entries is similar to the well-known travelling salesman problem. The minimization method of simulated annealing has been successfully applied to this problem [28] and has also been used by other researchers as a method of improving the error resilience of quantizers [29]. The optimization commences in an initial state, which in our situation is an initial assignment of the 256 codebook indices to the codebook entries. Random changes in the state of the system are generated by randomly choosing two codebook entries and swapping the indices of these two entries, and all changes which reduce the cost in Eq. (2) are accepted while some which increase the cost are also accepted. As the optimization progresses fewer changes which increase the cost are accepted, and eventually a minimum of the cost function is reached which we hope is the global minimum.

The effectiveness of the simulated annealing method in reducing the cost-function given in Eq. (2) is shown in Fig. 5. This graph shows the cost of the present arrangement of codebook indices against the number of arrangements of codebook indices which have been attempted by the minimization process. As seen in the figure, the initial randomly assigned arrangement of indices to codebook entries gives a "cost" of 1915, while the cost of the final arrangement of codebook indices is 1077, which corresponds to a reduction of about 44%.

The effectiveness of this rearrangement of codebook indices in increasing the resilience of the codec to errors in the bit stream between its encoder and decoder can be seen in Fig. 6. This graph shows the variation in the segmental SNR of our 6.4 kbits/s low delay Codec A versus the BER between its encoder and decoder using randomly distributed errors. The solid line shows the performance of the codec with the original codebook index assignment, and the lower dashed line shows the performance when the shape codebook indices are rearranged as described above. It can be seen that at BERs of between 0.1% and 1% the codec with the rearranged codebook indices has a segmental SNR of about 0.5 to 1 dB higher than the original codec.

Apart from the eight shape codebook bits which Codec A transmits from its encoder to the decoder,

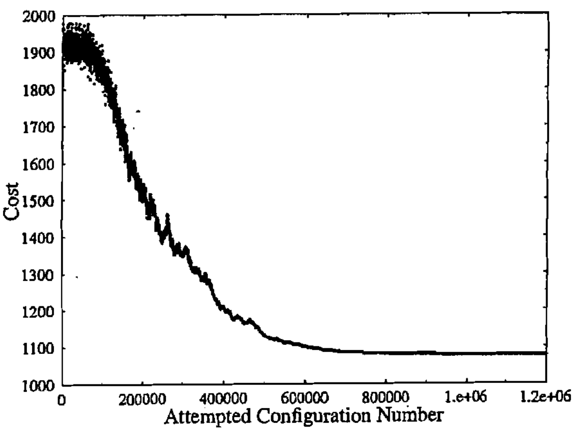


FIG. 5. Reduction in cost using simulated annealing.

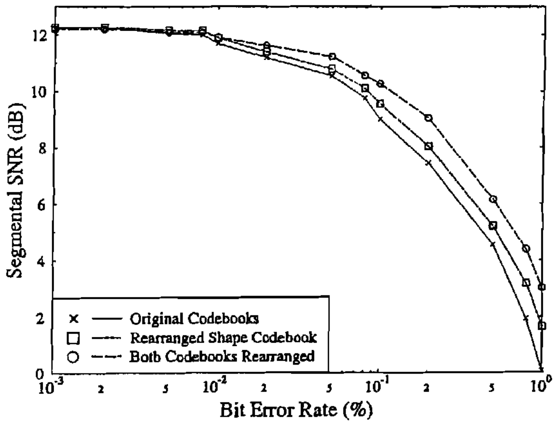


FIG. 6. The error sensitivity of our low delay 6.4 kbits/s Codec A.

the only other information that is explicitly transmitted are the four bits representing the gain codebook entry selected. Initially indices were assigned to the 16 gain codebook entries using the simple natural binary code (NBC). However, since the gain codebook levels do not have a uniform relative frequency, this simple assignment can be improved upon in a similar fashion to that described for the shape codebook above. Again, we defined a cost function that was to be minimised. This cost function was similar to that given in Eq. (2). However, since the gain codebook is scalar, whereas the shape codebook has a vector dimension of 15, no summation over n is needed in the cost function for the gain codebook index arrangement. We used simulated annealing again to reduce the cost function over that given using a NBC and found that we were able to reduce the cost by over 60%. The effect of this rearrangement of the gain codebook indices is shown by the top curve in Fig. 6, which gives the performance of Codec A with both the gain and shape codebooks rearranged. It can be seen that the rearrangement of the gain codebook indices gives a further improvement in the error resilience of the codec and that the codec with both the shape and gain codebooks rearranged has a segmental SNR more than 1 dB higher than the original codec at BERs around 0.1%.

4. A 6.2 kbits/s ACELP CODEC WITH A 5 ms FRAME LENGTH

In this section we discuss the development of a 6.2 kbits/s codec which is loosely based on the philosophy of the G.729 codec [21–23]. The main difference between the G.729 codec and our 6.2 kbits/s codec is that the G.729 codec uses forward adaption to determine the LPC synthesis filter coefficients, whereas our codec uses backward adaption.

The G.729 codec uses a 10-ms frame to determine the LPC coefficients and vector quantizes these coefficients using 18 bits. Each 10-ms frame is split into two 5-ms subframes, and for each of these subframes 17 bits are used to transmit a codebook index from an algebraic codebook [30], an average of 7 bits are used to represent a forward adapted long term predictor (LTP) delay, and 7 bits are used to give an index from a vector quantizer, which quantizes both the LTP and the ACELP gains. Thus a total of 80 bits are used for each 10-ms subframe, giving a codec with a bit rate of 8 kbits/s and a buffering delay of 10 ms.

Our 6.2 kbits/s scheme is similar to the G.729 codec, except it uses backward adaption to determine the LPC coefficients. This implies that it does not transmit the 18 bits per 10 ms that G.729 uses to represent the LPC parameters, and hence, it operates at a bit rate 1.8 kbits/s lower. Furthermore, its buffering delay is halved to only 5 ms. We found that this codec, which we refer to as Codec E in Table 1, gave reconstructed speech with a segmental SNR of 12.1 dB. This is compared to the segmental SNR from Codec A above, and Codec F to be described in the next section, in Fig. 7. It can be seen that irrespective of the different nature of the ACELP and vector-quantized excitations, the segmental SNR of the 6.2 kbps Codec E is in line with the performance of the similar-rate Codec A from above, but subjectively lower than that of the 8-kbps G.729 codec.

We rearranged the 7-bit vector gain quantizer of Codec E to improve its resilience to channel errors using simulated annealing as described in Section 3 and found that again this gave a significant improvement in the codecs error resilience. The segmental SNR of Codec E against the BER between its encoder and decoder is detailed in Section 6. In the next section we briefly describe a forward adaptive ACELP codec operating between 4.7 and 7.1 kbits/s.

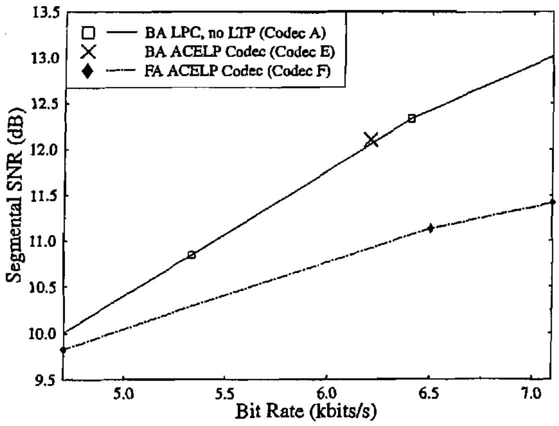


FIG. 7. Segmental SNR versus bitrate performance of various low and high delay CELP codecs.

5. A FORWARD ADAPTIVE ACELP CODEC

The final codec in our comparison was a standard forward adaptive ACELP codec, which we refer to as Codec F. This codec, which is described in detail in [31], operates at 6.5 and 4.7 kbits/s with a 30-ms frame-length and at 7.1 kbits/s with a 20-ms frame-length. In each frame 34 bits are used to quantize the forward adapted LPC coefficients. Each frame is split into either 4 or 6 subframes, depending on the bit rate, and for each subframe 12 bits are used to represent the algebraic codebook entry selected, 7 bits are used to represent the forward adapted LTP delay, and a total of 8 bits are used to scalar quantize the LTP and ACELP gains. This gives a total of either 196 or 142 bits per 20- or 30-ms frame, giving bit rates of 4.7, 6.5, and 7.1 kbits/s.

The segmental SNR of this codec is shown in Fig. 7, along with the segmental SNRs from Codec E and Codec A. It can be seen that Codec F gives a significantly lower segmental SNR than the low delay codecs, but that this difference decreases as the bit rate is reduced. This is due to the fact that at very low rates the backward adaptive scheme fails to adequately reproduce the speech spectrum. At very low rates forward adaption of the synthesis filter parameters gives an improvement in the spectral match that can be achieved between the original and the reconstructed speech in comparison to backward adaptive schemes. This improvement in the spectral match is not adequately reflected in the segmental SNR. Thus although, as indicated by Fig. 7, the backward adaptive codecs do give better reconstructed speech quality than Codec F at high rates; at lower rates the forward adaptive Codec F provides the best speech quality.

The error sensitivity of Codec F was improved using the techniques described in [31,32]. These techniques include a way to correct 25% of all bit errors that occur in the 34 bits representing the LPC coefficients of the codec and smoothing of the ACELP gains to correct errors that occur in these bits. The error sensitivity of the resulting codec is detailed in the next section, where it is compared to that of two of the backward adaptive codecs.

6. RELATIVE PERFORMANCES AND ERROR SENSITIVITIES OF THE FORWARD AND BACKWARD ADAPTIVE CODECS

The main features of the six codecs described in this paper are summarized in Table 1. As noted above, in error-free conditions at high bit rates the backward adaptive codecs give a superior performance to the forward adaptive Codec F. However, as the bit rate is reduced toward 4 kbits/s, the backward adaptive codecs are most seriously affected, and so at low bit rates, although the backward adaptive codecs give higher segmental SNRs than Codec F, it is the forward adaptive Codec F that provides the highest subjective speech quality.

As noted in Section 3, Codec B, Codec C, and Codec D, which all use long-term prediction and employ backward adaption of the LTP delay, are extremely sensitive to channel errors. However, the 6.4 kbit/s Codec A is much more robust, and its segmental SNR at various BERs is compared in Fig. 8 to that of the backward adaptive 6.2 kbits/s ACELP Codec E and the 6.5 kbits/s forward adaptive Codec F. As noted above, at 0% BER the two backward adaptive Codecs A and E give similar segmental SNRs, with the forward adaptive Codec F giving a segmental SNR of about 1 dB lower. As the BER is increased, the backward adaptive ACELP Codec E is the worst affected, but surprisingly, the other backward adaptive Codec A is almost as robust to channel errors as the forward adaptive Codec F. Both Codec A and Codec F give a graceful degradation in their reconstructed speech quality at BERs up to about 0.1%, but they provide impaired reconstructed speech for BERs much above this.

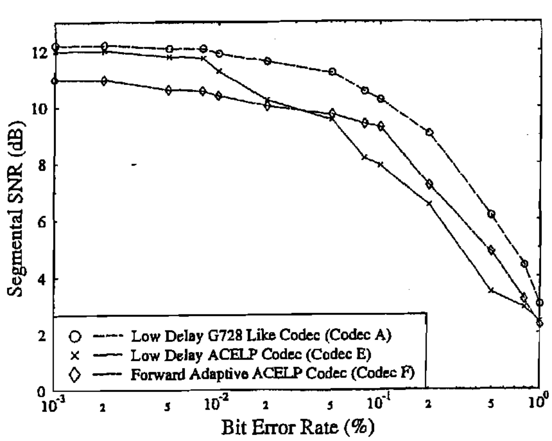


FIG. 8. A comparison of the bit error sensitivities Codecs A, E, and F.

7. CONCLUSIONS

In this paper we have detailed four backward adaptive, low delay, CELP codecs operating between 8 and 4 kbits/s. Furthermore, details were given of a 6.2 kbits/s low delay ACELP codec, similar to G.729 but using backward adaption of the synthesis filter and, thus, operating with a lower bit rate and with a lower delay. Both the reconstructed speech quality and the error sensitivity of these codecs were compared to that of a much higher delay forward adaptive ACELP codec. We found that at high bit rates the backward adaptive codecs provide superior speech quality, but as the bit rate is reduced toward 4 kbits/s the forward adaptive ACELP codec provides better speech quality. Last, we investigated and compared the error sensitivity of the codecs and found that error resilience similar to that of forward adaptive codecs can be achieved with backward adaptive codecs.

ACKNOWLEDGMENT

The financial support of the SERC, UK (GR/K74043), and that of the Department of Education, Northern Ireland are gratefully acknowledged.

REFERENCES

1. Schroeder, M. R., and Atal, B. S. Code excited linear prediction (CELP): High-quality speech at very low bit rates. In *Proc. ICASSP, 1985,* pp. 937–940.
2. Adoul, J. P., Mabilleau, P., Delprat, M., and Morissette, S. Fast CELP coding based on algebraic codes. In *Proc. ICASSP, April 1987,* pp. 1957–1960.
3. Kroon, P., and Deprettere, E. F. Regular pulse excitation—A novel approach to effective efficient multipulse coding of speech. *IEEE Trans. Acoust. Speech Signal Process.* (1986), 1054–1063.
4. Vary, P., Hellwig, K., Hofmann, R., Sluyter, R., Galland, C., and Rosso, M. Speech codec for the European mobile radio system. In *Proc. ICASSP, April 1988,* pp. 227–230.
5. Gerson, I. A., and Jasuik, M. A. Vector sum excited linear prediction (VSELP) speech coding at 8 kbps. *IEEE J. Selected Areas Commun.* (1990), 461–464.
6. Gerson, I. A., and Jasiuk, M. A. Vector sum excited linear prediction (VSELP). In *Advances in Speech Coding* (B. S. Atal, V. Cuperman, and A. Gersho, Eds.), pp. 69–80. Kluwer Academic, Amsterdam, 1991.
7. Campbell, J. Welch, V., and Tremain, T. An expandable error-protected 4800 bps CELP coder (U.S. Federal Standard 4800 bps Voice Coder). In *Proc. ICASSP, 1989,* pp. 735–738.
8. O'Shaughnessy, D. *Speech Communication, Human and Machine.* Addison–Wesley, Reading, MA, 1987.
9. Furui, S. *Digital Speech Processing, Synthesis and Recognition.* Dekker, New York, 1989.

10. Salami, R. A., *et al.*, Speech Coding. In *Mobile Radio Communications* (R. Steele, Ed.). IEEE Press–Pentech, New York, 1992.
11. Anderson, J. B., and Mohan, S. *Source and Channel Coding—An Algorithmic Approach.* Kluwer Academic, Dordrecht 1991.
12. Kondoz, A. M. *Digital Speech: Coding for Low Bit Rate Communications Systems.* Wiley, New York, 1994.
13. Gersho, A. Advances in speech and audio compression. *Proc. IEEE* (1994), 900–918.
14. Gerson, I. A., Jasiuk, M. A., Muller, J.-M., Nowack, J. M., and Winter, E. H. Speech and channel coding for the half-rate GSM channel. In *Proceedings ITG-Fachbericht 130, VDE-Verlag, Berlin,* Nov. 1994, pp. 225–233.
15. Ohya, T. Suda, H., and Miki, T. 5.6 kbits/s PSI-CELP of the half-rate PDC speech coding standard. In *Proceedings of the IEEE Conference on Vehicular Technology, June* 1994, pp. 1680–1684.
16. Bastiaan Kleijn, W. Encoding speech using prototype waveforms. *IEEE Trans. Acoust. Speech Signal Process.* (1993), 386–399.
17. Griffin, D. W., and Lim, J. S. Multiband excitation vocoder. *IEEE Trans. Acoust. Speech Signal Process.* (1988), 1223–1235.
18. Hiotakakos, D. J., and Xydeas, C. S. Low bit rate coding using an interpolated zinc excitation model. In *Proceedings of the IEEE Singapore International Conference on Communications Systems.* (1994), 865–869.
19. Chen, J.-H., Cox, R. V., Lin, Y.-C., Jayant, N., and Melchner, M. J. A low-delay CELP coder for the CCITT 16 kb/s speech coding standard. *IEEE J. Selected Areas Commun.* (1991), 830–849.
20. Coding of speech at 16 kbit/s using low-delay code excited linear prediction. CCITT Recommendation G.728, 1992.
21. Salami, R., Laflamme, C., Adoul, J.-P., and Massaloux, D. A toll quality 8 Kb/s speech codec for the personal communications system (PCS), *IEEE Trans. Vehicular Technology* (1994), 808–816.
22. "Coding of speech at 8 kbit/s using conjugate-structure algebraic code-excited linear prediction (CS-ACELP). ITU Draft Recommendation G.729, February 1996.
23. Kataoka, A., Adoul, J.-P., Combescure, P., and Kroon, P. ITU-T 8-kbits/s standard speech codec for personal communication services. In *Proceedings of International Conference on Universal Personal Communications, Tokyo, Japan, Nov. 1995,* pp. 818–822.
24. Woodard, J. P., Torrance, J. M., and Hanzo, L. A low delay multimode speech terminal. To be published in the *Proceedings of the IEEE Vehicular Technology Conference, April 28–May* 1, 1996.
25. Chen, J.-H., and Gersho, A. Gain-adaptive vector quantization with application to speech coding, *IEEE Trans. Commun.* (1987), 918–930.
26. Chen, J.-H., High-quality 16 Kb/s speech coding with a one-way delay less than 2 ms. In *Proceedings of the IEEE Conference on Acoustics, Speech and Signal Processing, 1990,* pp. 453–456.
27. Wang, S., and Gersho, A. Phonetic segmentation for low rate speech coding. In *Advances in Speech Coding* (B. S. Atal, V. Cuperman, and A. Gersho, Eds.), pp. 257–266. Kluwer Academic, Dordrecht, 1991.
28. Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P. *Numerical Recipes in C.* Cambridge Univ. Press, Cambridge, 1992.
29. Dall'Agnol, S. L. Q., De Marca, J. R. B., and Alcaim, A. On the use of simulated annealing for error protection of CELP coders employing LSF vector quantizers. In *Proceedings of the IEEE Conference on Vehicular Technology, June 1994,* Vol. 3, pp. 1699–1703.
30. Laflamme, C., Adoul, J.-P., Su, H. Y., and Morissette, S. On reducing the complexity of codebook search in CELP through the use of algebraic codes, In *Proc. ICASSP, 1990,* pp. 177–180.
31. Woodard, J. P. *Digital Coding of Speech Using Code Excited Linear Prediction.* Ph.D. thesis, University of Southampton, 1995.
32. Hanzo, L., and Woodard, J. P. An intelligent multimode voice communications system for indoor communications, *IEEE Trans. Vehicular Technology* (1995), 735–748.