AN EMBEDDED SCHEME FOR REGULAR PULSE EXCITED (RPE) LINEAR PREDICTIVE CODING

Shude Zhang and Gordon Lockhart

Department of Electronic and Electrical Engineering,
The University of Leeds, Leeds LS2 9JT, U.K.

ABSTRACT

The feasibility and performance of an embedded RPE (ERPE) scheme based on multistage coding is investigated. The coding efficiency of second and subsequent stages depends on the spectral envelope difference between the original speech and the error signal at each stage whereas re-use of LPC parameters derived from the original speech depends on the corresponding LPC spectral difference. Suitable measures of spectral difference are defined and simulation shows that both decrease with the perceptual weighting factor. The ERPE system requires little extra coding complexity and can be simplified further by using a partial phase adaptation procedure with marginal loss of SNR performance. The simulated ERPE system shows graceful reduction of reconstructed speech quality for bit rates from 14.8 to 6.4 kb/s in 4.2 kb/s steps.

1. INTRODUCTION

Embedded speech coding allows the output bit rate of an encoder to be progressively reduced with graceful degradation in the quality of the decoded signal. Such coders offer an effective means of controlling network or channel congestion by discarding bits, allowing communication to continue at lower bit rates with acceptable losses in quality.

Conventional PCM provides the simplest example of an embedded code, while more efficient differential schemes such as DPCM and delta modulation (DM) are not inherently embedded. Two approaches to the design of embedded DPCM are available: coarse feedback in the DPCM predictor loop [1] and the explicit noise coding technique[2], [3] which has also been used to obtain embedded DM (EDM) [4]. Bit rates for these embedded differential coders are typically between 16 and 40 kb/s. Vector quantization can also accommodate an embedded output stream and, in one proposal [5], a four-stage vector quantizer provides variable rate transmission from 32 to 8 kb/s in 8 kb/s steps reproducing toll quality speech at 32 kb/s and intelligible speech at 8 kb/s. Embedded CELP based on multistage coding has also been proposed permitting operation at bit rates of 6.4, 8 and 9.6 kb/s [6].

A novel embedded coding system based on regular pulse excited (RPE) predictive coding will be described. The RPE coder [7] is a special case of multipulse excitation (MPE) linear prediction coders where the excitation consists of a sequence of equally placed pulses. Unlike MPE, the excitation vectors in RPE are determined by solving several sets of linear equations leading to lower coding complexity than MPE but a similar performance. Because of its excellent quality and relatively low complexity, RPE has been chosen as a coding standard for the GSM European mobile radio system. Embedded coding schemes based on RPE are attractive for applications in the range of 16-7 kb/s.

2. EMBEDDED RPE CODING

The dashed box in Fig.1 illustrates a conventional RPE coder structure [7]. RPE uses a sequence of equally placed nonzero pulses for residual modelling. The LPC analysis filter $A(z)$ and the shaping filter $1/A(z/\gamma)$ have the same definition as in all analysis-by-synthesis predictive coders.

Assume that the excitation generator output of a conventional RPE coder is the sum of several, say, 3, sequences of excitation pulses. A functionally equivalent system can be formed by progressive subtraction of each sequence as illustrated in Fig. 1(a). The input to the second or the third adder (e_1 or e_2) can be viewed as the output of an error weighting filter, $W(z)$, with the corresponding non-weighted error as input. Thus, the second and third stages (outside the dashed box in Fig. 1(a)) effectively repeat the same RPE process as the first stage but with the non-weighted error signal as input. The configuration of Fig.1 can therefore be interpreted as multistage RPE coding where each stage uses the same LPC parameters obtained from the original input speech and the second and subsequent stages encode the reconstruction error from the preceding stage. The coded and transmitted excitations from each stage may be deleted in order of significance to achieve progressive bit rate reductions. At the decoder, reconstructed speech is obtained by adding together the inverse filtered decoded excitations, as illustrated in Fig.1(b). Full speech quality should be delivered when all excitations are received. The minimum transmission rate is determined by the excitation with most significance and decoded speech quality should remain acceptable when only this excitation is received.

It can be shown [7] that the excitation pulses, $g_2^{(k)}$ for the kth excitation pattern in the second stage, are given by

$$g_2^{(k)} = e_2^{(0)} H_k^t [\, H_k H_k^t \,]^{-1} \qquad (1)$$

where H is an upper triangular matrix containing the impulse response of the filter $1/A(z/\gamma)$. Evaluation of $H_k^t[H_k H_k^t]^{-1}$ is available from the first stage computation and $e_2^{(0)}$ is given by

0-7803-2431-5/95 $4.00

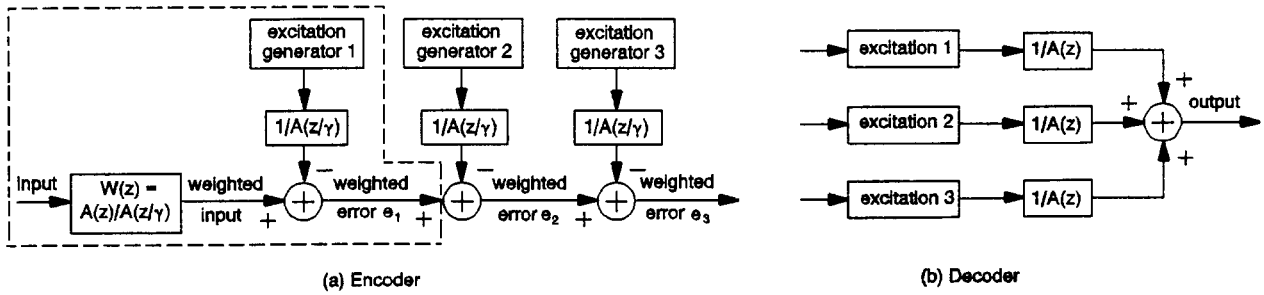


Fig. 1 Block diagram of the embedded RPE (regular pulse excited) system.

$$e_2^{(0)} = e_1 - w_2^{(0)} \tag{2}$$

where e_1 is the weighted error from the first stage and $w_2^{(0)}$ is the output of the weighting filter resulting from the initial filter state, caused by excitation only. Eqs.(1) and (2) are also used for the third stage.

3. PERFORMANCE OF MULTISTAGE RPE CODING

If the first stage of multistage RPE were efficient in removing redundancy and modelling the residual, then the reconstruction error would be noise-like and its spectrum approximately flat. Thus the capability for further SNR gains in subsequent stages would be considerably reduced. However, the spectral envelopes of the error signals are shaped by the perceptual weighting filters and therefore have some resemblance to the input signal spectrum. This increases error sample correlation and makes subsequent coding stages more effective in providing SNR gain as required for embedded coding operation. The increase in correlation determines the efficiency of subsequent coding stages and is controlled by the weighting factors. The inverse of the spectral flatness measure (*sfm*) [8] may be applied as a measure of waveform predictability but it is more useful here to define the measure of spectral envelope difference between the inputs to one stage and the next as

$$DS_k^l = \sqrt{\frac{1}{\pi}\int_0^{\pi}[\tilde{S}_k^l(\omega) - \tilde{S}_k^{l+1}(\omega)]^2\, d\omega} \tag{3a}$$

where $\tilde{S}_k^l(\omega)$ and $\tilde{S}_k^{l+1}(\omega)$ are the fluctuating components of the power spectra of the input signal to the lth and the $(l+1)$th stages of coding in the kth LPC frame and DS_k^l is the envelope difference between the two power spectra. $\tilde{S}_k^l(\omega)$ and $\tilde{S}_k^{l+1}(\omega)$ are given by

$$\tilde{S}_k^l(\omega) = S_k^l(\omega) - \frac{1}{\pi}\int_0^{\pi} S_k^l(\omega)\, d\omega \;;$$

$$\tilde{S}_k^{l+1}(\omega) = S_k^{l+1}(\omega) - \frac{1}{\pi}\int_0^{\pi} S_k^{l+1}(\omega)\, d\omega \tag{3b}$$

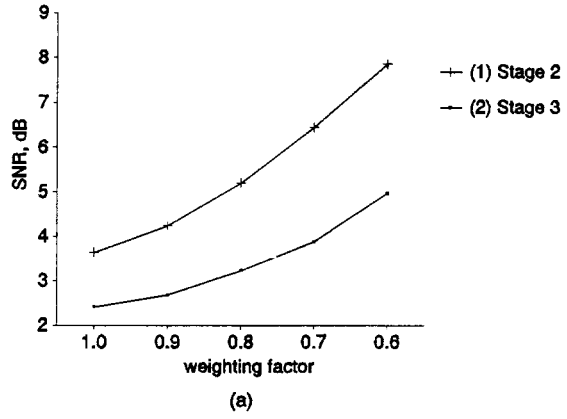
where $S_k^l(\omega)$ and $S_k^{l+1}(\omega)$ are the power spectra (in dB) of the input signal to the lth and the $(l+1)$th stages of coding in the kth LPC frame. DS, the mean value of DS_k^l over all LPC frames, provides a useful measure of the difference of the spectral envelope shape between the inputs to the different stages of coding.

The configuration of Fig.1 was simulated using 3.1 and 4.5 seconds of female and male speech respectively, bandlimited to 3.4 kHz and sampled at 8 kHz. The utterances were "The bank of England has started making a new five-pound note" and "I rode for a long distance in one of public coaches on the day preceding Christmas" by a female and male speaker respectively. The input speech is divided into LPC frames of 20 ms in length and 10th order predictor coefficients are determined using the autocorrelation method. The LPC coefficients are transformed to log area ratio (LAR) coefficients and then scalar quantized using 36 bits/LPC-frame before transmission with bit allocations: {6, 5, 4, 4, 3, 3, 3, 3, 3, 2}. Each LPC frame is further divided into 4 excitation frames of equal length. The pulse spacing $N = 8$ is used for each coding stage; i.e., there are 5 non-zero excitation pulses in each excitation frame. The excitation sequences are first normalized and the normalized values are then uniformly quantized using 3-bits each. The normalization factors are quantized using 5-bit and 3-bit logarithmic quantizers for first and subsequent stages respectively.

Let γ_1, γ_2 and γ_3 denote the perceptual weighting factors used for first, second and third stages of coding respectively. Table 1 gives the DS measure between the first and the second stages as a function of γ_1. As can be seen from Table 1, DS decreases with the values of γ_1 from 1.0 to 0.6 indicating greater correlation between contiguous samples in the error signal for smaller values of γ_1 and increasing effectiveness of

TABLE 1 DS measure between the inputs to the first and the second stages as a function of γ_1 using the female and male speech respectively.

γ_1	DS in dB	
	F	M
1.0	8.14	8.26
0.9	7.76	7.81
0.8	7.22	7.21
0.7	6.73	6.72
0.6	6.28	6.32



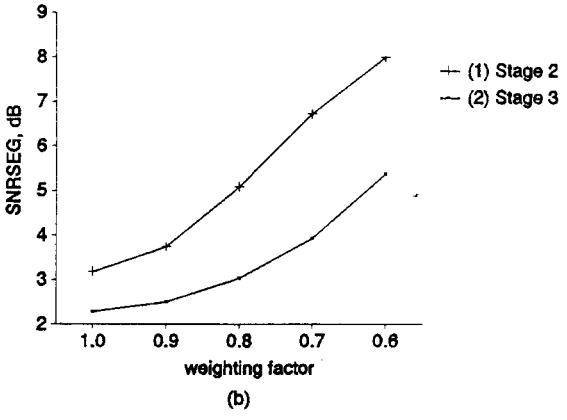


Fig. 2 Obtainable (a) SNR and (b) segmental SNR (SNRSEG) from (1) the second and (2) the third stage of multistage RPE coding (with LPC analysis in each stage) as a function of weighting factors.

the second stage of coding. Curves (1) in Figs. 2(a) and (b) illustrate SNR and segmental SNR (SNRSEG) gain as a function of γ_1 for the second stage of coding confirming that SNR gains achieved by the second stage increase with decreasing γ_1. Gains were computed using LPC parameters re-calculated from the first-stage error for the second stage and γ_2 = 1.0 was used to avoid the influence of $\gamma_2 < 1.0$ upon the obtainable SNR gain. Similar results were obtained for the third stage (Curves (2) in Figs. 2(a) and (b)).

4. LPC PARAMETER RE-USE

To achieve the greatest coding efficiency, LPC analysis should be performed independently for each coding stage using the corresponding input. (Note that the SNR and SNRSEG gain resulting from the second and the third stages shown in Fig 2 were obtained in this way.) However, such an approach is not acceptable for ERPE based on the multistage coding structure of Fig. 1, where LPC parameters derived from the original speech are used in the later stages and therefore the question arises as to the penalty incurred due to LPC spectral mismatch when the same LPC parameters are re-used by subsequent stages with differing input spectra.

Because the spectral envelopes of the error signal in subsequent stages become closer to that of the original speech as DS decreases with γ, LPC spectra derived directly from the corresponding error signal should also behave in the same way. The LPC spectral difference between the lth and the $(l+1)$th stage for the kth frame, DL_k^l is defined as

$$DL_k^l = \sqrt{\frac{1}{\pi}\int_0^{\pi}[H_k^l(\omega) - H_k^{l+1}(\omega)]^2\, d\omega} \qquad (4a)$$

where $H_k^l(\omega)$ and $H_k^{l+1}(\omega)$ are the LPC power spectra in dB of the kth LPC frame in the lth and the $(l+1)$th stages of coding given by

$$H_k^l(\omega) = 1/|A_k^l(e^{j\omega})|^2;\ H_k^{l+1}(\omega) = 1/|A_k^{l+1}(e^{j\omega})|^2 \qquad (4b)$$

$A_k^l(z)$ and $A_k^{l+1}(z)$ are the LPC polynomials for the kth LPC frame in the lth and the $(l+1)$th stages of coding respectively. DL, the average, is computed over all LPC frames and used to measure the LPC spectral difference between different stages of coding. (A spectral distortion measure similar to Eq. (4) has been widely used for measuring LPC quantization performance: e.g., [9].)

Table 2 gives the DL measure between the first and the second stages as a function of γ_1. The DL measure between the first and the second stages decreases with γ_1 until the LPC spectral envelope of the error signal is very similar to that of the original signal. In this case, LPC parameters derived from the original speech can be used for the second stage of coding with negligible loss of obtainable SNR gain.

SNR and SNRSEG values from the second and third stage of coding were obtained as a function of γ_1 and γ_2 (third stage

TABLE 2 DL measure between the inputs to the first and the second stages as a function of γ_1 using the female and malespeech respectively.

γ_1	DL in dB	
	F	M
1.0	5.24	5.00
0.9	4.66	4.45
0.8	3.89	3.68
0.7	3.18	3.00
0.6	2.59	2.48

only) using the LPC parameters derived from the original speech. The SNR difference with no LPC analysis in second and third stages is high in comparison with the SNR obtained using stage-by-stage LP analysis (Fig.2) only when the values of γ_1 and γ_2 are equal or close to 1 (e.g., $\gamma_1 = \gamma_2 = 0.9$). The SNR difference becomes smaller with the decrease of γ_1 to 0.8 (γ_1 and γ_2 to 0.7 for the third stage) and there is almost no difference when $\gamma_1 = 0.7$ and 0.6 ($\gamma_1 = \gamma_2 = 0.6$ for the third stage).

Although the original purpose of the weighting filter is to shape the noise spectrum to achieve noise masking effects, it plays additional important roles in multistage ERPE. Later coding stages can produce sufficient SNR gain using the shaped noise as input, the SNR value depending on the degree of shaping. Shaping also makes possible LPC parameter re-use by the later stages. However, the decrease in the value of the weighting factor is limited by the auditory masking effects and a suitable compromise value must be selected. Taking into consideration the various factors discussed above, values of $\gamma_1 = \gamma_2 = \gamma_3 = 0.75$ were chosen for the ERPE system.

5. COMPLEXITY REDUCTION

Although the coding complexity of the ERPE system is almost the same as that of the original coder when Eqs.(1) and (2) are used for the second and third stages of coding and LPC parameters are re-used, a further reduction in complexity can be achieved by using a partial phase adaptation procedure to place constraints on the choice of optimum excitation patterns. A relatively small loss of SNR and SNRSEG performance of approximately 0.3, 0.5 and 0.7 dB at the three operating rates of 6.2, 10.2 and 14.2 kb/s respectively is incurred but coding complexity is significantly reduced and the transmitted bit rates decrease by 0.2, 0.4, and 0.6 kb/s respectively. Listening tests showed that the resulting degradation in decoded speech quality is marginal.

6. PERFORMANCE COMPARISON

ERPE based on multistage coding can be viewed as a progressive addition by subsequent stages of non-zero excitation pulses to the excitation frame of the initial coding stage. The performance of a single-stage RPE coder with approximately the same total number of nonzero pulses in its excitation frame and $\gamma = 0.75$, can therefore be compared with the ERPE system. Comparison results show that ERPE using two or three-stage coding (equivalent to a total of 10 or 15 non-zero pulses respectively in a 40 sample excitation frame) gives SNR and SNRSEG values almost equivalent to that of the conventional RPE coder with a pulse spacing of respectively 4 or 2 (respectively 10 or 20 non-zero pulses per excitation frame). Listening tests indicate almost no difference between the reproduced speech using ERPE with two stages and conventional RPE. The speech quality of ERPE using the three-stage coding scheme with a total of 15 non-zero excitation pulses is only slightly inferior to conventional RPE with 20 non-zero pulses. Listening results also show that ERPE exhibits a graceful degradation in decoded speech quality over the range from 14.8 to 6.4 kb/s in 4.2 kb/s steps.

REFERENCES

[1] D. J. Goodman, "Embedded DPCM for variable bit rate transmission," *IEEE Trans. Commun.*, vol. COM-28, July 1980, pp. 1040-1046.

[2] N. S. Jayant, "Variable rate ADPCM based on explicit noise coding," *Bell Syst. Tech. J.*, vol. 62, Mar. 1983, pp. 657-677.

[3] S. Zhang and G. B. Lockhart, "Design and performance of robust embedded ADPCM coder," *Electron. Lett.* vol. 27, Sept. 1991, pp. 1786-1788.

[4] I. J. Wassell, D. J. Goodman, and R. Steel, "Embedded delta modulation," *IEEE Trans. Acoust. Speech, Signal Processing*, vol. 36, Aug. 1988, PP. 1236-1243.

[5] A. Haoui and D. G. Messerschmitt, "Embedded coding of speech: a vector quantization approach," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, Mar. 1985, pp. 1703-1706.

[6] R. D. D. Lacovo and D. Sereno, "Embedded CELP coding for variable bit-rate between 6.4 and 9.6 kbit/s," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, 1991, pp. 681-684.

[7] P. Kroon, E. F. Deprettere, and R. J. Sluyter, "Regular-pulse excitation: A novel approach to effective and efficient multipulse coding of speech," *IEEE Trans. Acoust. Speech, Signal Processing*, vol. ASSP-34, Oct. 1986, pp. 1054-1063.

[8] N. S. Jayant and P. Noll, *Digital Coding of Waveforms, Principles and Applications to Speech and Video*, Englewood Cliffs, NJ: Prentice-Hall, 1984.

[9] W. P. LeBlanc, B. Bhattacharya, S. A. Mahmoud, and V. Cuperman, "Efficient search and design procedures for robust multi-stage VQ of LPC parameters for 4 kb/s speech coding," *IEEE Trans. Speech, Audio Processing*, vol. 1, Oct. 1993, pp. 373-385.