

Computationally simple ADPCM based on exponential power estimator

G. Mathew
S.V. Narasimhan
A.P. Shivaprasad

Indexing terms: Modulation, Codes, Speech synthesis

Abstract: Application of the exponential power estimation (EPE) technique to develop computationally simple algorithms for adapting the quantiser and predictor of an adaptive differential pulse code modulation (ADPCM) system has been investigated. Performance of this EPE-based ADPCM system (E-ADPCM) has been studied for first order Gauss–Markov and real speech inputs (three sentences) under different channel conditions (noise-free and noisy channels), using a Gaussian non-uniform quantiser at the bit-rate of 16 kilobits per second (kbit/s). Performance of the proposed ADPCM has been compared with that of two conventional ADPCM systems. The E-ADPCM has a superior performance to the conventional systems considered, with a computational advantage of $(3N + B + 1)$ multiplications (N being the predictor order and B , the number of bits per sample used for quantising the prediction error) and one division per iteration compared with the first system and $(B + 1)$ multiplications over the second system, at the cost of one exponentiation.

1 Introduction

Use of differential pulse code modulation (DPCM) schemes for speech coding provides reduction in the bit-rate (by means of redundancy removal) compared with pulse code modulation schemes, for a specified reconstruction quality. However, it is necessary to adapt the parameters of the quantiser and predictor of the DPCM system to the input signal characteristics in order to achieve sufficient dynamic range performance and good subjective speech quality. Several algorithms have been developed for adapting the quantiser and predictor, most of which suffer from high computational complexity and are thus not easily implementable. Attempts have been made to reduce this computational complexity, and the algorithms reported in References 3, 10 and 12 are examples in this direction.

The basic principle behind adaptive quantisation (AQ) is to update the quantizer step-size in accordance with the changes in the input signal variance. This is achieved

using a short term estimate of the input variance [2, 9]. In general, an adaptive quantiser is realised by scaling the input and output levels of a fixed quantiser, that is matched to a unit variance input, by the standard deviation of the input signal. This establishes the need for variance estimation in adaptive quantisation.

Among the various algorithms available for predictor adaptation [11], the least mean square (LMS) gradient adaptive algorithm is the simplest from the computational point of view. In the LMS algorithm, at each iteration, the predictor coefficients are updated in a direction opposite to that of the instantaneous gradient of the squared prediction error surface [14]. To ensure convergence (mean convergence), the adaptation step-size is limited by the maximum eigenvalue, λ_{max} , of the autocorrelation matrix of the input signal. However, for computational simplicity, the product of the input signal power and the predictor order, which is equal to the trace of the correlation matrix, is used instead of λ_{max} . Estimation of the input signal power is thus important in predictor adaptation and its estimation complexity contributes to the complexity of the ADPCM system.

The application of the exponential power estimator (EPE) [1] is investigated for estimating the powers of the input speech and prediction error signals for use in predictor and quantiser adaptation algorithms. EPE has also been used to eliminate the multiplications which are involved in the gradient computations of the LMS algorithm. The EPE algorithm, quantiser adaptation, predictor adaptation and a discussion on the results are presented.

2 Exponential power estimation

The desired features of a power estimator are that

- (a) the estimate should follow the significant changes in the input signal
- (b) the estimation should not introduce any additional computations
- (c) use of this estimate should not lead to additional complexity.

Based on these features, the conventional block and recursive power estimators [1] make the adaptive algorithm quite complex. However, the modified EPE estimator (modified to improve some of the shortcomings of the original EPE [1, 4]), is attractive because of the following features:

- (i) the process of estimation does not involve any multiplications or divisions
- (ii) use of this estimate results in further reduction in complexity since the estimate is expressed as a power of 2 ($2^{i(n)}$, $i(n)$ being an integer)

(iii) sufficiently fast response rate owing to the exponential nature, unlike the additive nature of the conventional block and recursive estimators, and

(iv) very easily controllable response rate.

The basic principle of the EPE algorithm is to track the envelope of the magnitude of the input sequence $s(n)$. The corresponding algorithm is

$$\text{compute } N(n) = |s(n)| 2^{\lfloor C - i(n-1)/F \rfloor} \quad (1)$$

$$T(n) = \begin{cases} T(n-1)(1 + 2^{IT}) & \text{if } N(n) > 1 \\ T(n-1)(1 - 2^{DT}) & \text{if } N(n) < 1 \\ T(n-1) & \text{otherwise} \end{cases} \quad (2)$$

$$T(n) = T^\beta(n) \quad (3)$$

$$\text{confine } T(n) \text{ to } T_{min} \leq T(n) \leq T_{max} \quad (4)$$

$$i(n) = \lfloor F \cdot T(n) \rfloor \quad (5)$$

$$P_{EPE}(n) = 2^{i(n)} \quad (6)$$

where $s(n)$ is the input signal whose power is to be estimated; $T(n)$ is an internal variable which is updated to track the envelope of the magnitude of the input sequence, i.e. $|s(n)|$; $i(n)$ is an integer variable; IT and DT are negative integer constants which control the amounts of increment $(T(n-1)2^{IT})$ and decrement $(T(n-1)2^{DT})$ in $T(n)$, i.e. the speed of updating; T_{min} and T_{max} are limits of $T(n)$; $P_{EPE}(n)$ is a variance estimate of $s(n)$ if $F=2$ and standard deviation estimate of $s(n)$ if $F=1$; C is a positive integer constant; β is a leakage factor ($0 < \beta < 1$) that controls the memory of the estimator, and $\lfloor \cdot \rfloor$ is the integer part of (\cdot) .

T_{max} is chosen depending on the maximum input signal level ($|s_{max}| = 2^{T_{max}}$) and then T_{min} is decided by the dynamic range (DR) requirement as $DR(dB) = 6(T_{max} - T_{min})$ [4]. The values of IT and DT are chosen depending on the required response rate of the algorithm and the ratio of increment to decrement (I/D) in $T(n)$. While operating under noisy channel conditions it is necessary to keep the memory of the estimator small, to achieve sufficient robustness to channel error effects. A leakage mechanism is thus incorporated through the parameter β (eqn. 5). This increases the computational requirement of the EPE technique to one exponentiation. However, leakage is used only in the case of quantiser adaptation and not for prediction.

3 Quantiser adaptation

There are three different classes of adaptive quantisation schemes: instantaneous, syllabic and hybrid approaches. The instantaneous schemes update the step-size at every sampling instant according to the changes in the input signal. The syllabic schemes update the step-size only at a fairly slow rate. In these two schemes, it is assumed that the range of step-size values required to obtain good quantiser performance is the same for all segments of speech. Since speech is a quasi-stationary signal; the

signal power may vary among different speech sounds and speakers. Hence, the optimum range of step-size values may be different for different segments of speech [13]. The hybrid approach overcomes this problem by varying the adaptation range of the instantaneous scheme according to the input signal envelope level, by making use of the information from the syllabic scheme. The corresponding algorithm is [8]

$$\Delta l_{min}(n) = \Delta_{min} \sigma_{ns}(n-1) \quad (9)$$

and

$$\Delta l_{max}(n) = \Delta l_{min}(n) R_l \quad (10)$$

where $\sigma_{ns}(n-1)$ is the normalised syllabic standard deviation, i.e. the envelope information, and this is proportional to $P_{EPE}(n)$ with $F=1$. $\Delta l_{min}(n)$ and $\Delta l_{max}(n)$ are the optimum lower and upper limits of the local step-size range for any speech segment, and

$$R_l = \frac{\Delta l_{max}(n)}{\Delta l_{min}(n)} \quad (11)$$

is the local dynamic range.

The step-size calculated using the instantaneous scheme is confined to the local limits to obtain the final step-size. The quantities $\Delta l_{max}(n)$, $\Delta l_{min}(n)$ and R_l are such that

$$\Delta l_{max}(n) \leq \Delta_{max}$$

$$\Delta l_{min}(n) \geq \Delta_{min}$$

$$R_l < R_t$$

where Δ_{min} and Δ_{max} are the absolute limits of the step-sizes, and

$$R_t = \frac{\Delta_{max}}{\Delta_{min}} \quad (12)$$

is the total dynamic range of the quantiser.

Use of EPE in quantiser adaptation limits the performance of the quantiser since EPE allows only a limited number of distinct quantiser step-sizes. It is thus necessary to go for the hybrid approach so as to make the best use of the available finite number of step-sizes. The EPE technique is used to estimate the instantaneous and syllabic standard deviations of the quantiser input. The EPE-quantiser (EPEQ) is further modified by introducing additional (intermediate) step-sizes while retaining the simplicity [4].

4 Predictor adaptation

The LMS algorithm used for updating the predictor coefficients of the ADPCM system (Fig. 1) is given by

$$a_j(n+1) = a_j(n) + \frac{\mu}{N P_x(n)} e_q(n) \tilde{x}(n-j) \quad 1 \leq j \leq N \quad (13)$$

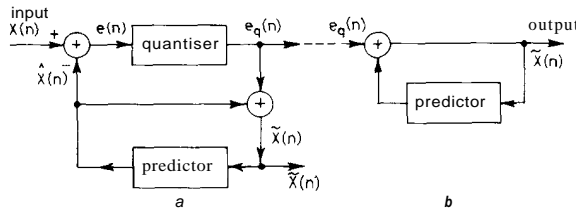


Fig. 1 Differential pulse code modulation system
a Encoder b Decoder

where $\tilde{x}(n)$ is the input signal to the predictor and also the reconstructed signal from the ADPCM system; $a_j(n)$ is the j th predictor coefficient at the n th instant; $e_q(n)$ is the quantised prediction error; $P_{\tilde{x}}(n)$ is the power estimate of the input signal; N is the predictor order and $0 < \alpha < 1$.

The power estimate $P_{\tilde{x}}(n)$ is obtained using the EPE algorithm as $P_{\tilde{x}}(n) = 2^{i(n)}$. The integer $i(n)$ is computed in such a way that the quantity $2^{i(n)}$ is the estimate for $\tilde{x}^2(n)$. Hence, $2^{i(n)/2}$ will be the corresponding estimate for $|\tilde{x}(n)|$. Substituting for $P_{\tilde{x}}(n)$ and $\% (n - j)$ in eqn. 13

$$a_j(n+1) = a_j(n) + \frac{\alpha}{N} 2^{1-i(n)+(i(n-j)/2)} \times e_q(n) \operatorname{sgn}(\tilde{x}(n-j)) \quad 1 \leq j \leq N \quad (14)$$

where $\operatorname{sgn}(\cdot)$ is the sign of the quantity (\cdot) . It must be noted that the implementation of the adaptation algorithm involves no multiplications or divisions. The performance of the EPE based ADPCM (E-ADPCM) has been compared with two conventional ADPCM systems described below.

4.1 Block power estimator based ADPCM (B-ADPCM)

The adaptive quantiser used in B-ADPCM is the robust adaptive quantiser (RAQ) [5]. This quantiser is nothing but the Jayant's one-word memory adaptive quantiser [6] incorporating the exponential leakage mechanism to minimise the effects of channel errors when working under noisy channel conditions. Since the performance of this instantaneous quantiser and its hybrid version are reported to be identical [8], only the instantaneous RAQ is used for comparison purposes. However, for the comparison to be fair, the leakages provided in the EPEQ and RAQ should be equivalent to each other. An approximately equivalent leakage factor has thus been derived for the RAQ with respect to the leakage factor of the EPEQ [4]. The RAQ algorithm can be given as

$$\Delta_R(n) = \Delta_R^{2^n}(n-1)M(|H(n-1)|) \quad (15)$$

where, $\Delta_R(n)$ is the RAQ step size at instant n ; β_R is the leakage factor ($0 < \beta_R < 1$); $M(|H(n-1)|)$ is a time invariant multiplication factor, which depends upon the quantiser level corresponding to the previous sample; $H(n-1)$ is an integer in the range $(-2^{B-1}, 2^{B-1})$ representing the quantiser level occupied by the prediction error sample at the $(n-1)$ th instant, where B is the number of bits per sample used for quantising the prediction error.

Predictor adaptation is performed by the LMS algorithm in which the power estimate $P_{\tilde{x}}(n)$ is obtained using the conventional block power estimator given by

$$P_{\tilde{x}}(n) = \frac{1}{N_1} \sum_{i=1}^{N_1} \tilde{x}^2(n-i) + K \quad (16)$$

where, N_1 is the block length of the data used for power estimation, and K is a positive constant to keep the estimate away from zero during the silence intervals.

4.2 Sign-sign LMS based ADPCM (S-ADPCM)

The quantiser is the same as the RAQ. The predictor adaptation is performed using the sign-sign LMS algorithm [10] given by

$$a_j(n+1) = a_j(n) + \gamma \operatorname{sgn}(e_q(n)) \operatorname{sgn}(\tilde{x}(n-j)) \quad 0 < \gamma < 1 \quad 1 \leq j \leq N \quad (17)$$

5 Results and discussion

Performance of the three ADPCM systems (E-, B- and S-ADPCMs) have been evaluated for first order Gauss-Markov and real speech (three sentences) inputs under noise-free and noisy channel conditions. The correlated Gaussian signal is generated as

$$x(n) = \rho x(n-1) + \sqrt{\{1-\rho^2\}}r(n) \quad (18)$$

where $r(n)$ is a zero mean and unit variance uncorrelated Gaussian sequence, and ρ is the correlation coefficient, and is chosen to be 0.85.

The speech sentences used are:

- (a) 'The pipe began to rust while new'
- (b) 'Move the vat over the hot fire'
- (c) 'Oak is strong and also gives shade'

spoken by a male speaker, bandlimited to 3.2 kHz, sampled at 8 kHz and digitised into 12 bits, thus resulting in an input signal range of $(-2048, 2048)$. Segmented signal-to-noise ratio (SEGSNR) with non-overlapping segments of length 160 samples was used as the performance measure. Non-uniform (Gaussian) [7] quantiser characteristic at 16 kbit/s (2 bits/sample) has been used for the prediction error quantisation.

The ADPCM system parameters as used in the simulation study are [4]

(a) *Quantiser:*

$$R_i(\text{dB}) = 60; R_r(\text{dB}) = 40$$

$$\Lambda_m = 1024; \Delta_{min} = 1; \beta_R = 0.987509$$

EPEQ: Modifications were also performed to improve the dynamic range and noisy channel performance of the EPEQ [4], resulting in the following set of parameters:

(i) *Instantaneous-EPE*

$$C = 0; T_{max} = 10 + C = 10$$

$$T_{min} = T_{max} - 10 = 0; \beta = 127/128$$

$$DT = -5 \text{ and } IT = -2 \text{ for } T(n) \leq 3$$

$$DT = -6 \text{ and } IT = -3 \text{ for } 4 \leq T(n) < 8$$

$$DT = -6 \text{ and } IT = -4 \text{ for } T(n) \geq 8$$

(ii) *Syllabic-EPE*

$$C = 1; T_{max} = 10 + C = 11$$

$$T_{min} = T_{max} - 10 = 1; \beta = 1023/1024$$

$$DT = -8 \text{ and } IT = -6 \text{ for } T(n) \leq 3$$

$$DT = -9 \text{ and } IT = -8 \text{ for } T(n) \geq 4$$

(b) *Predictor:*

$$\alpha = 2^{-8}; N = N_1 = 4; \gamma = 2^{-11}; K = 100$$

$$\text{dynamic range} = 60 \text{ dB}; C = 0; T_{max} = 11 + C = 11$$

$$T_{min} = T_{max} - 10 = 1; I/D = 4; IT = -5; DT = -7$$

Figs. 2 and 3 show the ideal or noise-free channel performance of E-, B- and S-ADPCM systems for Gaussian and speech inputs (sentence n). The E-ADPCM performs superior to both B- and S-ADPCMs for about a signal range of 40 dB. It becomes slightly inferior at the lower signal side. This inferior performance at the lower signal side is caused by the fact that the response of the EPE estimator is slow for very low signal levels.

