[Mansoura Engineering Journal](https://mej.researchcommons.org/home)

Volume 16 | Issue 2 Article 2

8-8-2021

Sequentially Adaptive Differential Pulse Code Modulation Using Adaptive LSP Filters.

Fayez Zaki
Electrical Power Engineering Department., Faculty of Engineering., Mu'tah University., Jordan.,
fwzaki2017@gmail.com

Mansoura Engineering Journal (MEJ) Vol. 16, No. 2, Dec.1991. E.1

SEQUENTIALLY ADAPTIVE DIFFERENTIAL PULSE CODE MODULATION USING ADAPTIVE LSP FILTERS

Fayez W. Zaki

Faculty of Engineering

Mu'tah University, Jordan.

نظام تشفير تفاضلي ذو تهايؤ متتابع لمرشح أزواج الطيف

خلاصة :-

لتشفير الاشارات الصوتية ميزات عديدة منها الكفاءة العالية لاعادة توليد الاشارات المشفرة ، مقاومة الشوشرة وامكانية أرسال شفرة الاشارة مع الشفرة الخاصة بعمليات الفصل والتوصيل ولكن لسوء الحظ فان هذه الميزات لايمكن الحصول عليها إلا بزيادة حيز الاتساع لقنوات الارسال . ولهذا السبب فقد طورت طرق أخرى للتشفير يعتمد عملها أساسا على التخلص من الحشو المتواجد بالاشارات الصوتية . وقد أمكن لهذه الطرق تقليل حيز الاتساع لقنوات الارسال ولكن على حساب زيادة تعقيد الدوائر الالكترونية لهذه الانظمة .

هذا البحث يقدم طريقة متهايئة مبسطة لتقليل حيز اتساع قنوات الارسال وذلك بارسال اشارة خطأ التوقع . وفي هذه الطريقة يستخدم مرشح التوقع المتهايئ ذو أزواج خطوط الطيف وذلك للتخلص من الحشو المتواجد بالاشارات الصوتية وتتم عملية تهايؤ المرشح المذكور باستخدام إشارة خطأ التوقع والتي تكون متواجدة في المرسل والمستقبل على حد سواء .
النتائج المستخلصة من تجارب المحاكاة على الحاسب الآلي اثبتت انه يمكن الحصول على أصوات ذات صفات عالية عند مدى اتساع ما بين ٢٤٠٠٠ ـ ٣٢٠٠٠ نبضة في الثانيه ـ هذا وقد بين البحث ان النظام المقترح يعمل بصورة جيدة حتى عند تواجد خطأ في قنوات الارسال قد يصل الى ٥٪ .

ABSTRACT

The advantages of coding speech signal digitally are well known and are widely discussed in the literature [1]. Briefly, digital representation offers efficient signal regeneration, noise immunity, easy encryption, the possibility of combining transmission and switching functions, and the advantage of a uniform format for different types of signals. Unfortunately, these benefits are gained at the expense of increased transmission bandwidth. The redundancy removal systems (e.g., differential coding, linear prediction vocoders,...etc.) were developed to overcome this difficulty, although, at the expense of system complexity and speech quality.

This paper introduces a simple adaptive differential pulse code modulation (ADPCM) system for speech coding at low bit rates. In this system line spectral pair (LSP) adaptive backward predictor is used to remove the redundancy present in the speech signal. Backward adaptation of the predictor coefficients is preferred due to the fact that it does not require a portion of the transmitted data rate to be allocated to the predictor coefficients, thus allowing the use of all bits available for coding the prediction residual (error). Furthermore, backward adaptation simplifies transmitter implementation.

Computer simulation experiments using Arabic speech bandlimited to 3.5 KHz and sampled at 8 KHz, resulted in a high quality speech reproduction at bit rates between 24 - 32 Kbit/sec. Moreover, it is shown that the developed system performs well at bit error rate as high as 5%.

INTRODUCTION AND REVIEW

The use of pulse code modulation (PCM) at the standard rate of 64 Kbps demands high channel bandwidth for its transmission. In certain applications, however, channel bandwidth is at a premium, in which case there is a definite need for speech coding at " low bit rates", while maintaining acceptable fidelity or quality of reproduction. A major motivation for bit rate reduction is for secure transmission over radio channels that are inherently of low capacity. The fundamental limits on bit rate suggested by speech perception and information theory show that high quality speech coding is possible at rates considerably less than 64 Kbps (the rate may actually be as low as 2 Kbps). The price that has to be paid for attaining this advantage is increased processing complexity (and therefore increased cost of implementation). Also in many coding schemes, Increased complexity translates into increased processing delay time (delay is of no concern in applications that involve voice storage).

For coding speech at low bit rates, a waveform coder of prescribed configuration is optimised by exploiting both statistical characterisation of speech waveforms and properties of hearing. For the work reported here in particular, the design philosophy has two aims in mind:

i-To remove redundancies from the speech signal as far as possible,

ii-To assign the available bits to code the nonredundant parts of the speech signal in a perceptually efficient manner.

To reduce the bit rate from 64 Kbps (used in standard PCM) to 32, 24, 16, and 8 Kbps, the algorithms for redundancy removal and bit assignment become increasingly more sophisticated. As a rule of thumb, in the 64 to 8 Kbps range, the computational complexity (measured in terms of multiply-add operations) required to code speech increases by an order of magnitude when the bit rate is halved, for approximately equal speech quality.

Reduction in the number of bits per sample from 8 (as used in standard PCM) to 3 involves the combined use of "adaptive quantisation and adaptive prediction". In this context, the term "adaptive" means being responsive to changing level and spectrum of the input speech signal. The variation of performance with speakers and speech material, together with variations in signal level inherent in the speech communication process, make the combined use of adaptive quantisation and adaptive prediction necessary to achieve best performance over a wide range of speakers and speaking situations [2]. A digital coding scheme that uses adaptive quantisation and/or adaptive prediction is called adaptive differential pulse code modulation (ADPCM).

The term "adaptive quantisation" refers to a quantiser that operates with a time-varying step size $\Delta(n)$. At any given time identified by n, the adaptive quantiser is assumed to have a uniform transfer characteristic. The step size $\Delta(n)$ is varied so as to match the variance σ_x^2 of the input signal x(n) [3]. In particular, one can write

$$\Delta(n) = k \quad \hat{\sigma}_x(n) \qquad (1)$$

where k is a constant, and $\hat{\sigma}_x(n)$ is an estimate of the standard deviation $\sigma_x(n)$. The problem of adaptive quantisation is one of estimating $\hat{\sigma}_x(n)$ continuously in one of two ways:

1-Unquantised samples of input signal are used to derive forward estimates of $\hat{\sigma}_x(n)$,

2-Samples of the quantiser output are used to derive backward estimates of $\hat{\sigma}_x(n)$.

The respective quantisation schemes are referred to as adaptive quantisation with forward estimation (AQF) and adaptive quantisation with backward estimation (AQB) [3]. The use of AQF requires the explicit transmission of step size information (typically about 5 to 6 bits per step size sample) to a remote decoder. Also, a processing delay (on the order of 16 m.sec. for speech) in the encoding operation results from the use of AQF, which is unacceptable in some applications. The problem of side information transmission, buffering and delay intrinsic to AQF are all avoided in the AQB scheme by using the recent history of the quantiser output to extract information for the computation of the step size $\Delta(n)$. Accordingly, AQB is usually preferred over AQF in practice.

The use of adaptive prediction in ADPCM is justified because speech signals are inherently nonstationary, a phenomenon that manifests itself in the fact that the autocorrelation function and power spectral density of speech signals are time-varying functions of their respective variables. This implies that the design of predictors for such inputs should likewise be time-varying, that is, adaptive. As with adaptive quantisation, there are two schemes for performing adaptive prediction:

1-Adaptive prediction with forward estimation (APF) [4,5]; in which unquantised samples of the input signal are used to derive estimates of the predictor coefficients.

2-Adaptive prediction with backward estimation (APB) [6]; in which samples of the quantiser output and the prediction error (residual) are used to derive estimates of the predictor coefficients.

The respective schemes are shown in Figs. 1 and 2 respectively. In the APF scheme of Fig.1, N unquantised samples of the input speech are first buffered and then released after computation of M predictor coefficients that are optimised for the buffered segment of input samples. The choice of M involves a compromise between an adequate prediction gain and an acceptable amount of side information [5]. Likewise, the choice of learning period or buffer length N involves a compromise between the rate at which statistics of the input speech signal change and the rate at which information on predictor coefficients must be updated and transmitted to the receiver. For speech, a good choice of N corresponds to a 16 m.sec. buffer for a sampling rate of 8 KHz, and a choice of M=10 ensures adequate use of the short-term predictability of speech.

However, APF suffers from the same intrinsic disadvantages (side information, buffering, and delay) as AQF. These disadvantages are eliminated by using the APB scheme of Fig. 2. Since in the latter scheme, the optimum predictor coefficients are estimated on the basis of quantised and transmitted data, they can be updated as frequently as desired, e.g., from sample to sample. Moreover, APB does not require a portion of the transmitted data rate to be allocated to the predictor coefficients, thus allowing more bits to be used to code the prediction error signal and so simplifying transmitter implementation, since a homogenous bit stream is generated at the transmitter output.

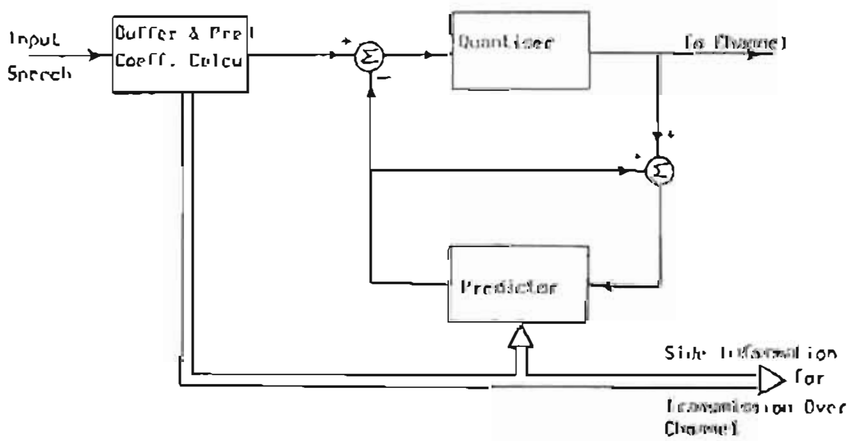
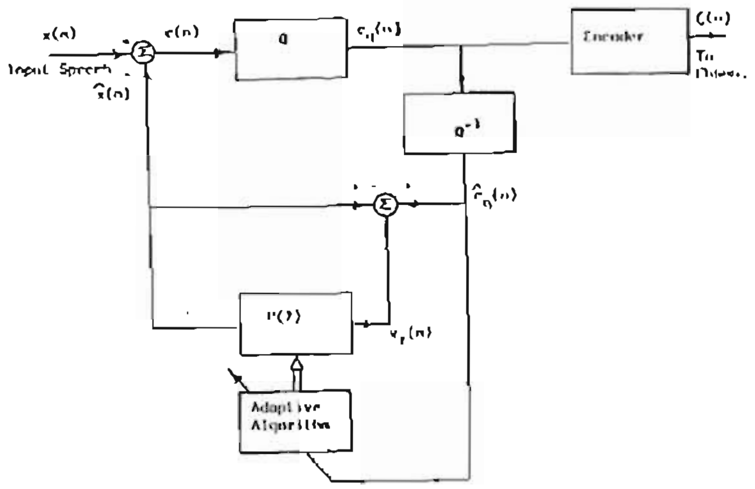
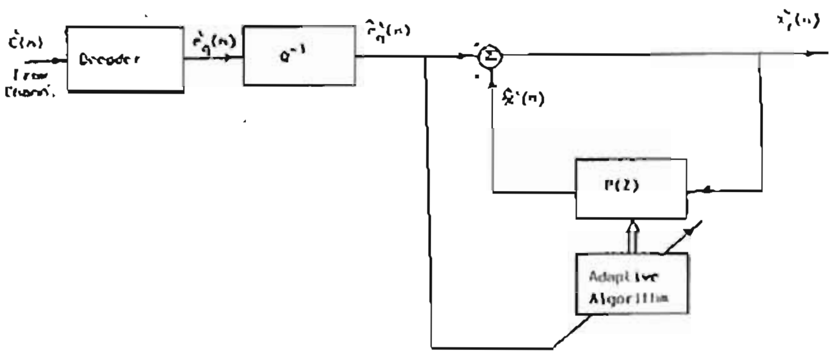


Fig. 1, Adaptive Prediction With Forward Estimation



a)Transmitter



b)Receiver

Fig. 2, Adaptive Prediction With Backward Estimation

In the APB scheme, the updating of the prediction filter is performed by some form of steepest descent algorithms [7,8]. The prediction error signal e(n) is the only function that need be quantised, coded, and transmitted. At the receiver the output speech is reconstructed by another adaptive prediction filter arranged in the feedback loop as shown in Fig. 2. Again this adaptive prediction filter updates its coefficients on a sample-by-sample basis using the received error signal.

In 1972, Moye [9,10] reported a system similar to that shown in Fig.2 for transmitting speech at 9.6 Kbps. The adaptive predictor used was a tapped-delay line self adaptive filter [7]. In his report, Moye pointed out the most difficult problem inherent in his system, that is: due to the slow convergence of the tapped delay line, the predictor at the transmitter removes the first formant almost completely, leaving mostly the second formant in the prediction error. The receiver filter then amplifies the second formant to make it larger in the output speech. This uncontrollable problem would remain unless the coefficients at both transmitter and receiver are reset from time to time.

In 1974, Gibson et al [6] reported a sequentially adaptive prediction system using adaptive Kalman filtering algorithm [11] and stochastic approximation algorithm. A bit rate of 16 Kbps was suggested using minimum mean square error quantisers [12,13]. It was concluded that the Kalman filtering algorithm performs better than the stochastic approximation algorithm. Furthermore, the Laplacian quantiser is more effictive than the Gaussian quantiser. Later Cohn and Melsa [14] studied the performance of the above system [6] using an adaptive quantiser and variable length coding. Their results showed that the system provides 5 dB gain in SNR over adaptive DPCM with fixed predictor. They claimed that a channel error rate as high as 10^{-3} does not produce noticeable degradation in speech quality and the system can still work with error rates up to 5%

In 1978, Gibson [15] reported a comparison study between his system [6] and ADPCM with fixed predictor. An adaptive quantiser with one-word memory [17] was used to quantise the prediction error signal. It was concluded that for bit rates from 16 to 18.4 Kbps, the sequentially ADPCM system using stochastic approximation algorithm was preferred to ADPCM with 2nd. order and 4th. order fixed predictors. At higher bit rates, ADPCM with 2nd. order fixed predictor performs better than a sequentially ADPCM system using stochastic approximation algorithm. A 4th. order sequential ADPCM system using Kalman algorithm provides better performance over ADPCM with any order of fixed predictor.

In 1980, Gibson et al [16] reported a study of backward adaptive predictor with Kalman algorithm and modified pitch compensating quantiser (Kalman/MPCQ). Although, the system complexity has greatly increased, they claimed that the Kalman predictor with MPCQ in ADPCM produce high quality output speech and outperforms (in terms of SNR) the fixed-tap/MPCQ and the Kalman/robust Jayant systems. Moreover, the catastrophic effect of bit error is eliminated by either setting the predictor coefficients to zero or replacing it with a fixed second order predictor (for certain period of time) depending on some criteria. This again increases the system complexity.

In this paper, a sequentially backward adaptive DPCM system for speech coding at bit rates between 24 - 32 K bit/sec. is introduced. In this system the adaptive predictor structure used is the "Line Spectral Pair (LSP)" adaptive filter developed by Zaki [18,20] This adaptive filter structure is proved to have superior convergence properties over Lattice structure, which in turn have higher convergence rate than tapped delay line structure [19].

ADAPTIVE DIFFERENTIAL PULSE CODE MODULATION SYSTEM

A block diagram of the adaptive differential pulse code modulation (ADPCM) system is shown in Fig. 2. In the figure, Q denotes the quantiser, P(Z) denotes the predictor, Q^{-1} represents an inverse quantisation operation. The encoder at the transmitter transforms the quantiser levels into a binary data stream and the decoder at the receiver transforms the binary data back to quantiser levels. The adaptive algorithm (at both transmitter and receiver) is a process that updates the predictor coefficients on the basis of quantised prediction error. It is important to note for noiseless channel that $\hat{e}'_q(n)=\hat{e}_q(n)$, $x'_r(n)=x_r(n)$, $\hat{x}'(n)=\hat{x}(n)$, and $C'(n)=C(n)$. In the transmitter, Fig. 2(a), the predictor forms an estimate $\hat{x}(n)$ of the incoming speech sample $x(n)$ based on a set of past samples $\{x_r(n-1), x_r(n-2),.....\}$. The difference between the input speech sample and its predicted value defined as prediction error

$$e(n) = x(n) - \hat{x}(n) \qquad (2)$$

is computed and quantised to obtain $e_q(n)$. An inversely quantised version $\hat{e}_q(n)$ is given by

$$\hat{e}_q(n) = e(n) + n_q(n) \qquad (3)$$

where $n_q(n)$ represents the quantisation noise. The signal $x_r(n)$ is then obtained as

$$x_r(n) = \hat{x}(n) + \hat{e}_q(n) \qquad (4)$$

At the receiver, the decoded and inversely quantised error signal $\hat{e}_q'(n)$ is added to the predicted value $\hat{x}'(n)$ to obtain $x_r'(n)$. Note that the predictors in the transmitter and receiver are identical, and that both predictors estimate the speech signal from the same sample sequence $x_r(n)$, since $x_r'(n)=x_r(n)$ for noiseless channel. Therefore,

$$x_r(n) = \hat{x}(n) + \hat{e}_q(n)$$

$$= \hat{x}(n) + e(n) + n_q(n) \qquad (5)$$

Applying Eq.(2) into Eq.(5), then the received signal is given by

$$x_r(n) = x(n) + n_q(n) \qquad (6)$$

Equation (6) is true for all predictors and all quantisers, and says that the reconstructed speech signal at the receiver is equal to the transmitted signal plus quantisation noise of the quantiser. Furthermore, if the quantisation noise can be reduced, a better reproduction of the transmitted signal will be obtained at the receiver output.

The signal-to-quantising noise ratio of the system of Fig. 2 is given by

$$\mathrm{SNR} = \frac{E\,[x^2(n)]}{E\,[n_q^{\,2}(n)]} = \frac{\sigma_x^2}{\sigma_n^2} \qquad (7)$$

where E[.] denotes expectation operation and σ_x^2 and σ_n^2 are the variances of input signal and quantisation noise respectively. Dividing and multiplying Eq.(7) by the variance of the prediction error σ_e^2 yields

$$\mathrm{SNR} = \frac{\sigma_x^2}{\sigma_e^2} \cdot \frac{\sigma_e^2}{\sigma_n^2}$$

$$= G_p \cdot (\mathrm{SNR})_q \qquad (8)$$

where

$$(\mathrm{SNR})_q = \frac{\sigma_e^2}{\sigma_n^2} \qquad (9)$$

is the signal-to-quantising noise ratio of the quantiser, and the quantity

$$G_p = \frac{\sigma_x^2}{\sigma_e^2} \qquad (10)$$

is defined as the gain due to the differential configuration.

The quantity $(\mathrm{SNR})_q$ is dependent upon the particular quantiser that is used, and, given knowledge of the properties of e(n), $(\mathrm{SNR})_q$ can be maximised by using nonliear or adaptive quantisers. The quantity G_p, if greater than unity, represents the gain in SNR that is due to the differential scheme. Clearly, our objective should be to maximise G_p by appropriate choice of the predictor P(Z). For a given signal x(n), σ_x^2 is a fixed quantity so that G_p can only be maximised by minimising the denominator of Eq.(10), i.e., by minimising the variance of the prediction error.

To proceed, we need to specify the nature of the predictor P(Z). If the predictor is a simple delay, $P(Z)=Z^{-1}$, a differential pulse code modulation (DPCM) results. In order to improve the prediction gain in Eq.(10), hence SNR in Eq.(8), a linear predictor of length four was used in the feedback loop around the quantiser. The output of this predictor $\hat{x}(n)$, is a linear combination of past quantised values, that is

$$\hat{x}(n) = \sum_{i=1}^{4} a_i x_r(n-i) \qquad (11)$$

where a_i, i=1,2,3,4 are the predictor coefficients. The predicted value is thus the output of a finite impulse response (FIR) filter whose system function is

$$P(Z) = \sum_{i=1}^{4} a_i Z^{-i} \qquad (12)$$

and whose input is the reconstructed quantised signal $x_r(n)$. Moreover, the reconstructed signal $x_r(n)$ is the output of a system whose system function is

$$H(Z) = \frac{1}{1 - \sum_{i=1}^{4} a_i Z^{-i}} \qquad (13)$$

and whose input is the quantised difference (prediction error) signal $\hat{e}_q(n)$.

The predictor coefficients a_i's may be calculated using block methods (e.g. autocorrelation, covariance, and PARCOR [3]) or sequential adaptive methods (e.g. least mean square, Kalman, stochastic,...etc.). In sequential adaptive prediction methods, the FIR filter may be implemented as Ladder, Lattice, or Line Spectral Pair (LSP) structure [see references 5,7,18].

In this work, the prediction filter chosen is the LSP structure with least mean square (LMS) updating algorithm [18]. This adaptive filter structure has been shown to provide higher convergence rate and less misadjustment than both Ladder and Lattice structures. These are the features that we depend upon to rectify the uncontrollable divergence problem noticed in other systems (e.g. Moye [9,10]). Moreover, the LMS algorithm requires less computation complexity than both Kalman and stochastic approximation algorithms used elsewhere [15].

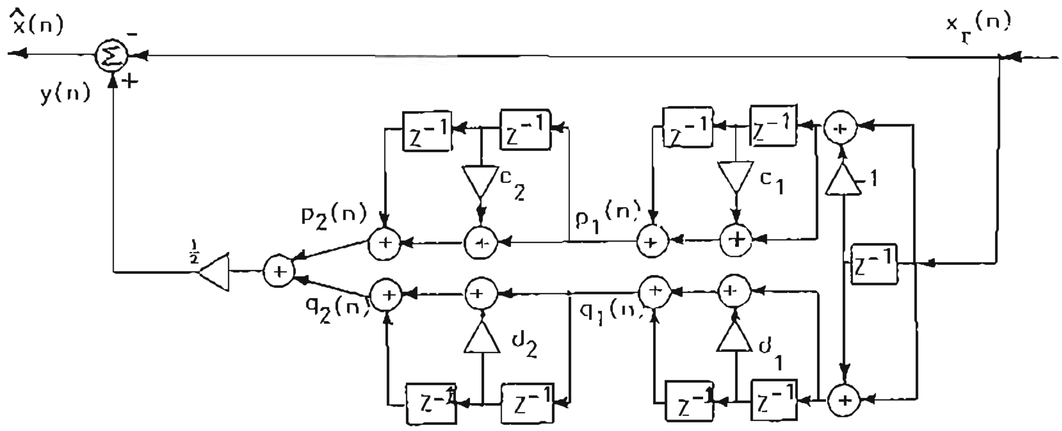


Fig. 3, Line Spectral Pair Predictor Structure

Fig.3 shows the LSP predictor used. In this figure, the output $y(n)$ is expressed as

$$y(n) = \{p_2(n) + q_2(n)\}/2 \tag{14}$$

where

$$p_i(n) = p_{i-1}(n) + c_i p_{i-1}(n-1) + p_{i-1}(n-2) \tag{15}$$

$$q_i(n) = q_{i-1}(n) + d_i q_{i-1}(n-1) + q_{i-1}(n-2) \tag{16}$$

$$p_0(n) = x_r(n) - x_r(n-1) \tag{17}$$

and

$$q_0(n) = x_r(n) + x_r(n-1) \tag{18}$$

The prediction $\hat{x}(n)$ for $x(n)$ is given by

$$\hat{x}(n) = y(n) - x_r(n) \tag{19}$$

Applying Eqs. (15), (16), (17), and (18) into Eq.(14) and applying Eq.(14) into Eq.(19) with some algebraic manipulations, then

$$\hat{x}(n) = \sum_{i=1}^{4} a_i x_r(n-i) \tag{20}$$

where

$$a_1 = (c_1 + c_2 + d_1 + d_2)/2 \tag{21}$$

$$a_2 = (4 - c_1 - c_2 + c_1 c_2 + d_1 + d_2 + d_1 d_2)/2 \tag{22}$$

$$a_3 = (c_1 - c_2 - c_1 c_2 + d_1 + d_2 + d_1 d_2)/2 \tag{23}$$

$$a_4 = (2 - c_1 - c_2 + d_1 + d_2)/2 \tag{24}$$

Note that Eq.(20) is the same as Eq.(11) given previously. Applying Eq.(20) into Eq.(2), the prediction error may be expressed as

$$e(n) = x(n) - \sum_{i=1}^{4} a_i x_r(n-i) \tag{25}$$

For Ladder or tapped-delay-line adaptive predictor, the coefficients $\{a_i, i=1,2,3,4\}$ are updated so that the mean square value of $e(n)$ is minimised. However, for LSP structure shown in Fig.3, the coefficients $\{c_i, d_i, i=1,2\}$ are updated instead so that the mean square value of prediction error is minimised. To make the algorithm reported in [18] suitable for our application, $e(n)$ is replaced by its quantised version $\hat{e}_q(n)$, since this quantity is available at both transmitter and receiver. With this change, the LMS updating algorithm for the LSP predictor shown in Fig.3, will be

$$c_i(n+1) = c_i(n) - 2\mu\, \hat{e}_q(n) p_{i-1}(n-1) \tag{26}$$

and

$$d_i(n+1) = d_i(n) - 2\mu\, \hat{e}_q(n) q_{i-1}(n-1) \tag{27}$$

where $i=1,2$, and μ is a quantity that controls stability and rate of convergence of the algorithm. To maintain minimum phase condition for the FIR filter, it must be ensured that the condition

$$-2 < d_1 < c_1 < d_2 < c_2 < 2 \tag{28}$$

is satisfied at all times.

SYSTEM SIMULATION AND RESULTS

The ultimate measure of performance of a speech digitisation scheme is the level of user satisfaction when the system is actually operative. Prior to that time, performance can at best be predicted by computer simulation experiments. Although, subjective listening tests are, of course, preferable, the most common parameter of performance prediction is the signal-to-quantisation error-ratio (SNR) as defined by Eq.(8). Moreover, some comments based on:

i-flatness of the short-time spectral of the prediction error signal,
ii-short-time spectral of reconstructed speech as compared to that of the original speech, and
iii-informal listening tests

are included in the analysis of the experimental results.

The results presented here are based on the four Arabic speech words شمال، صعود، سريع، يسين with bit rates lie in the range of 24 to 32 K bits/sec. The data library for these words was prepared as follows. Two different male speakers spoke into a high quality dynamic microphone in a normal laboratory environment. The amplified microphone signal was lowpass filtered at 3.5 KHz, sampled and converted into digital form by a 12 bits/sample linear A/D converter operating at 8 KHz sampling frequency, and finally written onto floppy disks.

Numerous computer simulation runs were conducted to establish the objective and subjective performance of the ADPCM system introduced in this paper. For comparison purpose, a fixed-weight 4th. order predictor was considered along with the adaptive LSP 4th. order predictor updated by the LMS algorithm. The coefficients of the fixed-weight optimum 4th. order predictor were taken from [19] and shown in table 1. These coefficients were calculated by the autocorrelation method and averaged over a wide range of speech data.

In all experiments, three types of quantisers were used to quantise the prediction error signal. The first two are linear quantisers with 4 bits/sample (16-levels) and 3 bits/sample (8-levels) respectively. The third one is a 3 bits/sample nonlinear quantiser. The optimum 8-levels for the nonlinear quantiser were obtained from [13] and shown in table 2. Note that these numbers are derived assuming Gamma distributed signal with unit variance. If the variance of the prediction error is σ_e^2, then the numbers in the table should be multiplied by the standard deviation σ_e.

Table 1, Optimum Fixed-Weight 4th. Order Predictor Coefficients [19].

a_1	a_2	a_3	a_4
1.793	-1.401	0.566	-0.147

Table 2, Optimum Quantiser Levels for Signals with Gamma Density, Mean = 0 and $\sigma^2=1$ [13].

Input	output
0.504	0.149
1.401	0.859
2.872	1.944
∞	3.799

Tables 3, 4, and 5 show the resulting SNR in dB (for each of the four Arabic words) as provided by DPCM with fixed 4th. order optimum predictor and ADPCM with 4th. order adaptive LSP predictor.

Table 3, SNR for 4-bits/sample Linear Quantiser

	Word1 شمال	Word2 يمين	Word3 صعود	Word4 سريع	Average SNR
Fixed Predictor	15.439	13.76	20.368	19.349	17.23
Adaptive Predictor	19.188	18.586	24.982	21.23	21.0

Table 4, SNR for 3-bits/sample Linear Quantiser

	Word1 شمال	Word2 يمين	Word3 صعود	Word4 سريع	Average SNR
Fixed Predictor	11.634	9.532	15.327	14.59	12.77
Adaptive Predictor	13.914	13.688	18.01	16.2	15.453

Table 5, SNR for 3-bits/sample Gamma Quantiser

	Word1 شمال	Word2 يمين	Word3 صعود	Word4 سريع	Average SNR
Fixed Predictor	13.326	11.532	18.268	16.369	14.87
Adaptive Predictor	16.972	16.247	20.644	18.544	18.102

Inspection of these tables reveals that the ADPCM system introduced here has an SNR that is about 3 to 4 dB better than that given by a fixed DPCM of the same predictor order. This difference is accounted for by the adaptive predictor. This improvement in SNR exhibited by the ADPCM system makes it appealing for use at bit rates from 24 to 32 K bits/Sec. because of the improvement in quality for a modest increase in complexity as compared to the fixed-tap DPCM system. It is important to note that the SNR is computed for active portions of the signal in all cases, i.e., silence is discarded.

A series of experiments have been carried out to study the properties of the reconstructed (received) speech signal and the prediction error signal. The results of these experiments are now considered. Fig. 4, shows the waveform of the original speech signal for the word "shamaal" (شمال). Fig. 5, shows the corresponding reconstructed signal from both DPCM system with fixed optimum predictor and ADPCM system with adaptive LSP predictor. Parts (a), (b), and (c) show the output from the DPCM system using 4-bits/sample linear, 3-bits/sample linear, and 3-bits/sample Gamma quantisers respectively, whereas, parts (d), (e), and (f) show the corresponding output from the ADPCM system. Comparing Fig. 5(a), (b), and (c) with Fig. 4, it can be seen that the output of the DPCM system for the unvoiced sound /sh/ (شَ) between samples 0 and 1000 is almost destroyed, especially in part (b) where a "shot-noise" like signal is noticed. This has been observed in all experiments and click sounds were noticed during listening tests. This problem is not present in the output of the ADPCM system in Fig. 5(d), (e), and (f). Comparing Fig. 5(e) with Fig. 5(f), it can be seen that the nonlinear distribution of the quantiser levels reduces the effect of quantisation error significantly. Note that, a further reduction in quantisation error and/ or further reduction in bit rate may be accomplished using adaptive quantisers mentioned earlier. However, no attempt has been made to implement such quantisers in the present work.

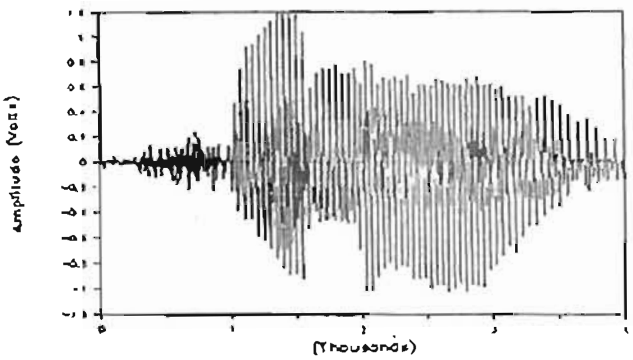


Fig. 4, Waveform of the Original Speech Signal for "Shamaal"

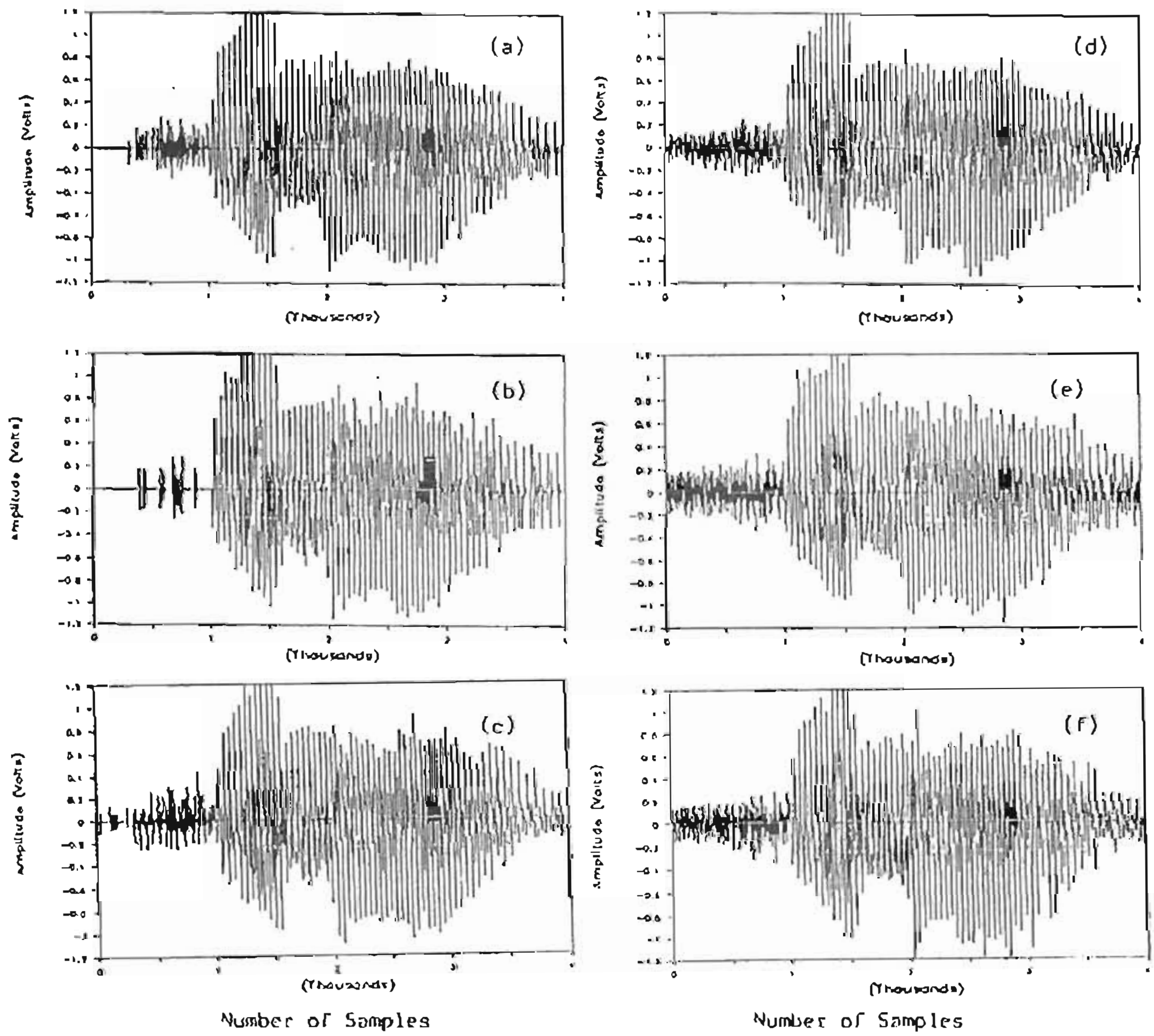


Fig. 5, Waveforms of the Reconstructed Speech for "Shamaal" from
a)Fixed Predictor & 4 bits Linear QZ b)Fixed Pred. & 3 bits Linear QZ
c)Fixed Predictor & 3 bits Gamma QZ d)Adaptive Pred. & 4 bits Linear QZ
e)Adaptive Pred. & 3 bits Linear QZ f)Adaptive Pred. & 3 bits Gamma QZ.

Fig. 6, illustrate the log magnitude of the discrete Fourier transform (obtained using 512 point FFT) as a function of normalised frequency (normalised to the sampling frequency) for two typical vowels of the original speech signals. These vowels are /aa/ in "shamaal" and /ee/ in "yameen". Figs. 7 and 8 show the log magnitude of the Fourier transform of both the reconstructed speech and prediction error signals as obtained from ADPCM and fixed DPCM systems respectively for the vowel /aa/. Figs. 9 and 10 show similar log magnitude spectral for /ee/.

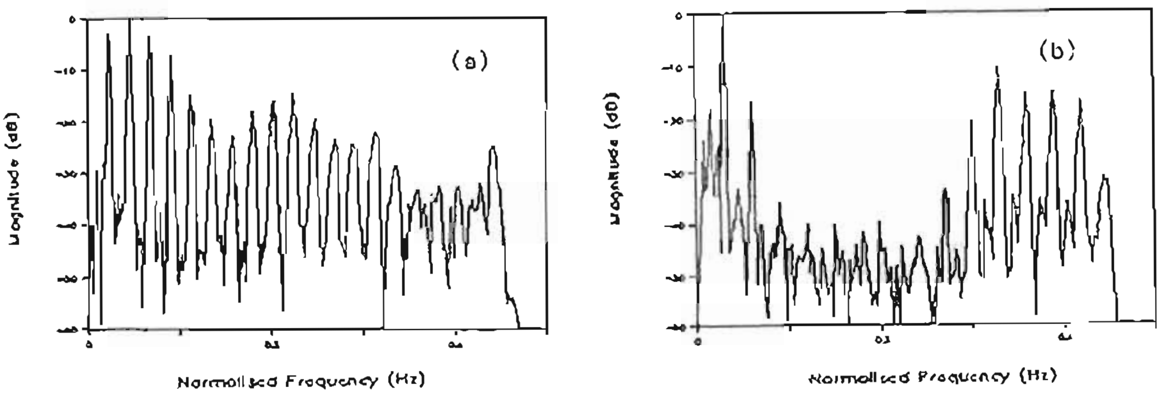


Fig. 6, Log Magnitude (dB) of the Discrete Fourier Transform Versus Normalised Frequency for Original Vowels;
a) /aa/ in "shamaal"
b) /ee/ in "yameen".

Comparing the spectrum of predictor error from ADPCM in Figs. 7 and 9 with the corresponding spectrum from fixed DPCM in Figs. 8 and 10, it can be seen that the adaptive predictor concentrates its deconvolving (redundancy removal) effort at low frequencies where the magnitude of the spectral differences between the first and subsequent formants are always high whereas the fixed predictor concentrates on the high frequencies where the differences in spectral magnitude are not significant (see Fig. 6). As a result the adaptive predictor provides more spectral flatness than the fixed predictor, i.e., the adaptive predictor removes more redundancy from the speech signal than the fixed predictor. This is due to the fact that the adaptive predictor continuously adjusts its transfer characteristics so as to match the spectral envelope of the incoming speech signal. The effect of this process was justified in the SNR results shown in tables 3, 4, and 5 where the adaptive predictor provided prediction gain (G_p) in the order of 3 to 4 dB more than the fixed predictor. Comparing the reconstructed speech spectrum from ADPCM system in Figs. 7 and 9, and the corresponding spectrum from fixed DPCM system in Figs. 8 and 10 with the spectrum of the original speech in Fig. 6, it can be concluded that:

i-The effect of quantisation noise is observed at frequencies in between the formants where the spectral density is slightly increased. However, this effect is more pronounced in the output of the fixed predictor system than that of the adaptive one (compare Fig. 9 (a), (b), and (c) with Fig. 10 (a), (b), and (c) between frequencies 0.1 and 0.3).

ii-Distortion in the harmonic line structures of the spectrum provided by the fixed DPCM system is higher than that provided by the ADPCM system.

iii-The spectral envelope provided by the ADPCM system is identical to the original in most cases, which is not the case for fixed predictor system.

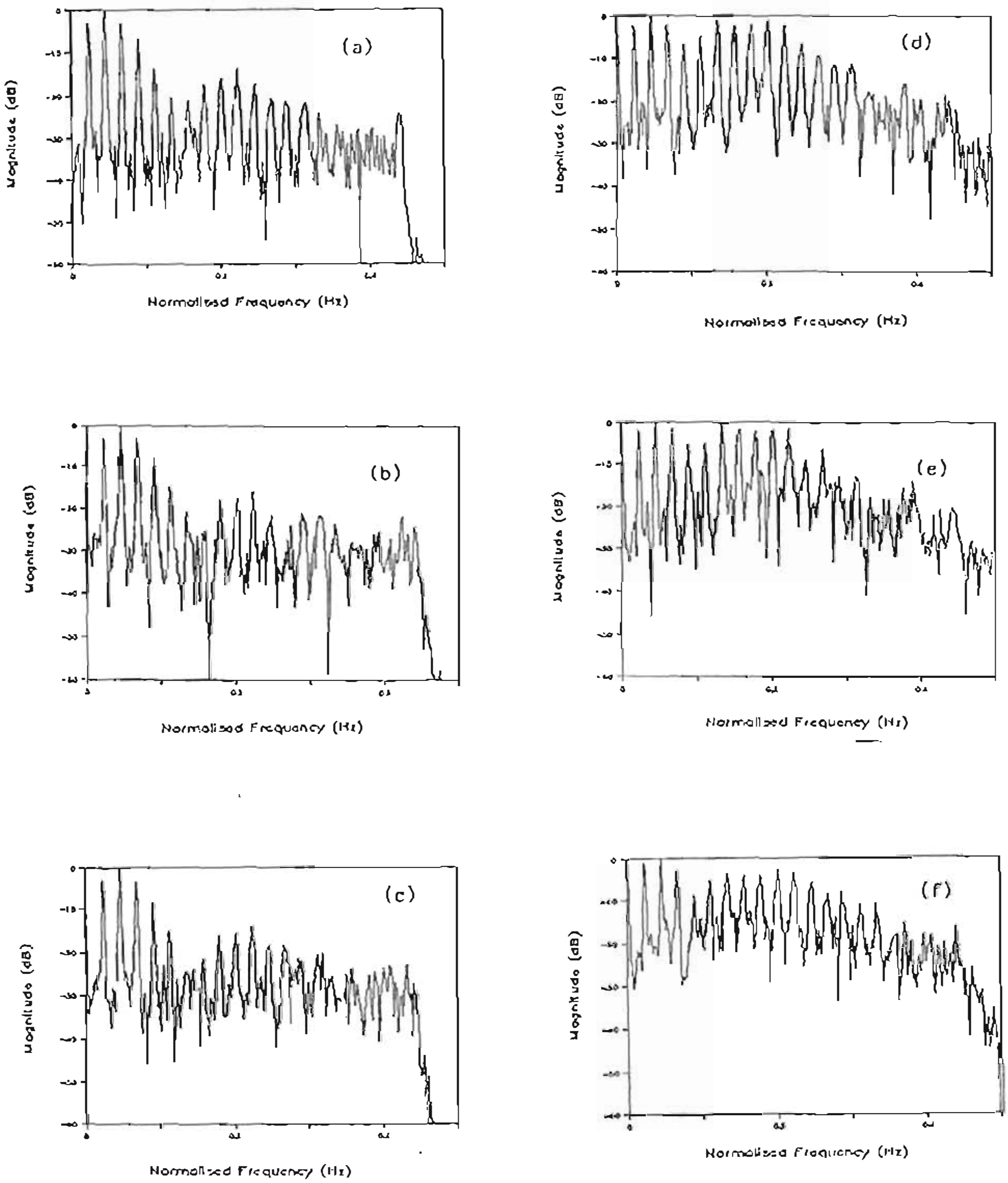


Fig. 7, Reconstructed Speech Spectrum and Quantised Prediction Error Spect. from ADPCM system with adaptive LSP Predictor for the Vowel /aa/.
a)Speech Spect. for 4-bits Lin. QZ d)Pred. Error Spect. for 4-bits Lin. QZ
b)Speech Spect. for 3-bits Lin. QZ e)Pred. Error Spect. for 3-bits Lin. QZ
c)Speech Spect. for 3-bits Gamma QZ f)Pred. Error Spect. for 3-bits Gamma QZ

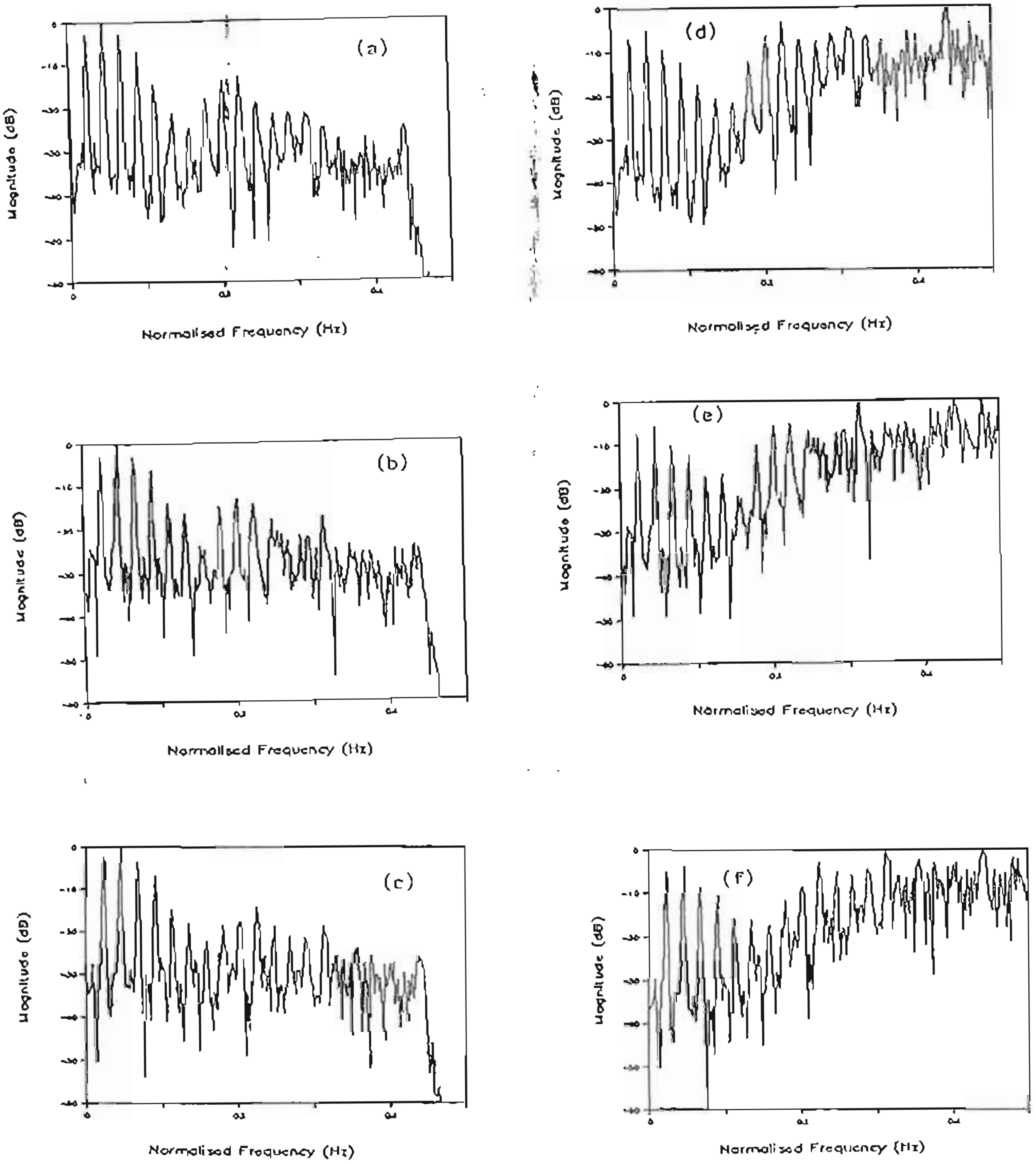


Fig. 8, Reconstructed Speech Spectrum and Quantised Prediction Error Spectrum from DPCM system with Fixed Predictor for the vowel /aa/.
a)Speech Spect. for 4-bits Lin. Qz d)Pred. Error Spect. for 4-bits Lin. QZ
b)Speech Spect. for 3-bits Lin. QZ e)Pred. Error Spect. for 3-bits Lin. QZ
c)Speech Spect. for 3-bits Gamma QZ f)Pred. Error Spect. for 3-bits Gamma QZ

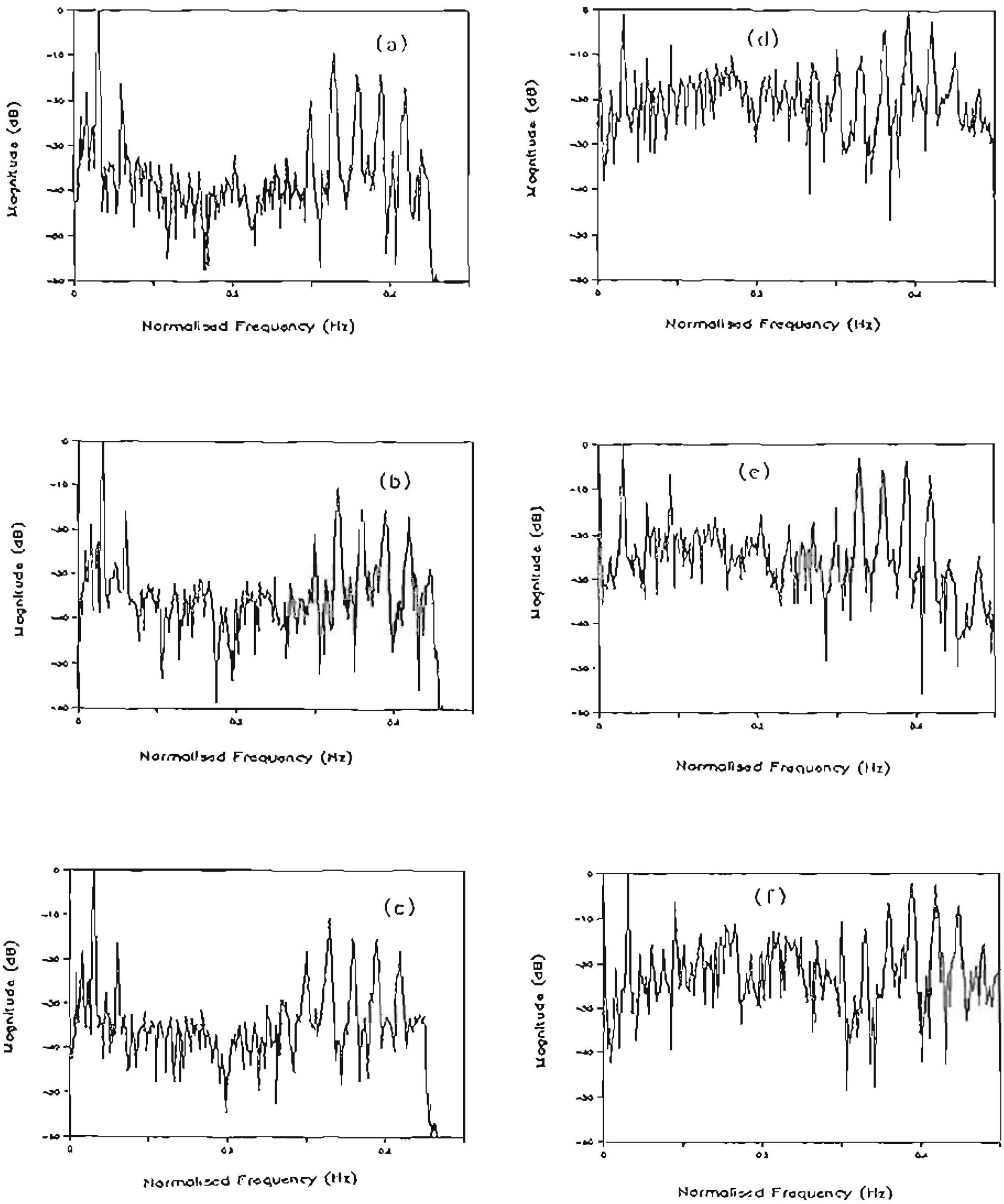


Fig. 9,Reconstructed Speech Spectrum and Quantised Prediction Error Spectrum from ADPCM system with Adaptive Predictor for the Vowel /ee/.
a)Speech Spect. for 4-bits Lin. QZ d)Pred. Error Spect. for 4-bits Lin QZ
b)Speech Spect. for 3-bits Lin. QZ e)Pred. Error Spect. for 3-bits Lin QZ
c)Speech Spect. for 3-bits Gamma QZ f)Pred. Error Spect. for 3-bits Gamma QZ

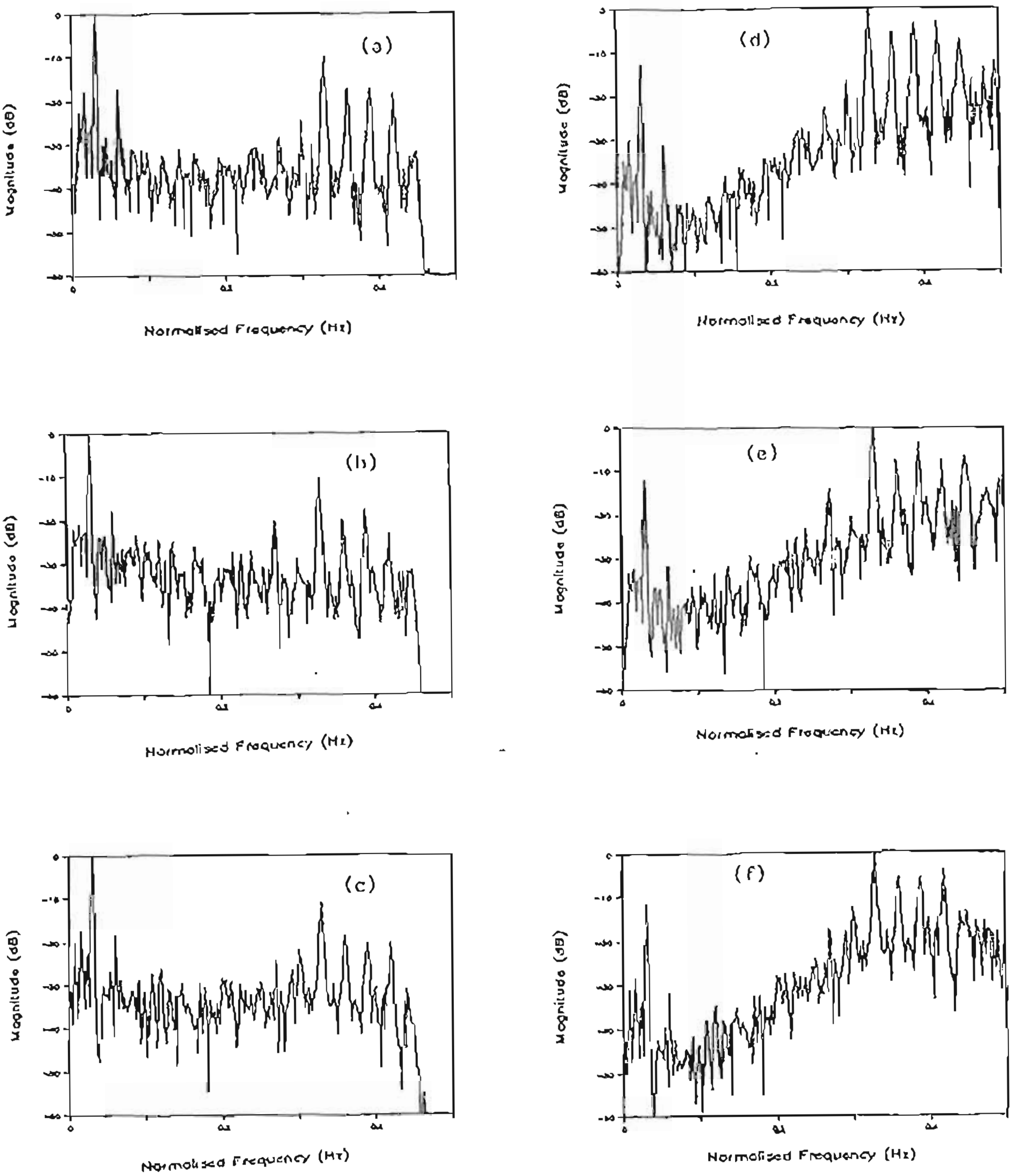


Fig.10, Reconstructed Speech Spectrum & Quantised Prediction Error Spectrum from DPCM System with Fixed Predictor for the Vowel /ee/.
a)Speech Spect. for 4-bits Lin. QZ d)Pred. Error Spect. for 4-bits Lin. QZ
b)Speech Spect. for 3-bits Lin. QZ e)Pred. Error Spect. for 3-bits Lin. QZ
c)Speech Spect. for 3-bits Gamma QZ f)Pred. Error Spect. for 3-bits Gamma QZ

For listening tests, the reconstructed speech as well as the original speech prepared earlier were passed through a 12-bits linear D/A converter and 4th. order Butterworth lowpass filter with its -3 dB point chosen at 3.5 KHz. Several informal listening tests have been conducted to assess the quality of the reconstructed speech from both the ADPCM and fixed DPCM systems. Although, all listeners judged the reconstructed speech as of high quality and retain its naturalness, they prefered the reconstructed speech from ADPCM system over that of the fixed DPCM system all the time. Evenmore, they could not differentiate between the output of the fixed DPCM system using 4-bits/sample linear quantiser and the output from the ADPCM system using 3-bits/sample Gamma quantiser. In other words, it is found that the quality of the received speech from ADPCM system at 24 Kbits/Sec. is identical to that received from fixed DPCM system at 32 Kbits/Sec.
Note: All results obtained above could have been obtained at bit rates between 19.2 to 25.6 Kbits/Sec. (instead of 24 to 32 Kbits/Sec.) if the original speech is bandlimited to 3.2 KHz and sampled at 6.4 KHz.

The final series of experiments were carried out to study the effects of bit error rate on the performance of the ADPCM system. In these experiments, the quantised prediction error was perturbed with different bit error rates before applying it as input to the predictor (see Fig. 2). The results have shown that error rates as high as 10^{-3} do not produce any noticeable degradation in the output speech from the ADPCM system with adaptive LSP predictor. However, this error rate of 10^{-3} turns the DPCM system with fixed predictor of the same order into divergence. The ADPCM system can function with error rates up to $5x10^{-2}$ and still produce speech that is marginally intelligible. Moreover, bit error rates are more severe on systems using linear quantisers than on those using nonlinear quantisers.

CONCLUSIONS

An improved system for speech digitisation using adaptive differential pulse code modulation (ADPCM) is introduced. The system uses an adaptive LSP 4th. order predictor, linear and nonlinear quantisers to achieve a 3 to 4 dB increase in SNR over DPCM system with fixed optimum predictor of the same order. This increase can be used to improve speech quality at moderate data rates on the order of 24 to 32 Kbits/Sec. or to retain the same quality and reduce the data rate below 20 Kbits/Sec. The latter alternative permits the use of narrow-band channels. The system provided high quality natural speech at the bit rate specified and produced intelligible speech at bit error rate as high as 5 %.

Reducing the bit rate even further by including adaptive quantiser as well as adaptive LSP predictor requires further study and may be a topic for future research.

REFERENCES

1-Flanagan, J.L., "Focal Points in Speech Communication Research," IEEE Trans. Comm., Vol. COM-19, pp.1006-1015, Dec. 1971.

2-Flanagan, J.L.,Schroeder, M.R., Atal, B.S., Crochiere, R.E., Jayant, N.S., and Tribolet, J.M., "Speech Coding," IEEE Trans. Comm., Vol. COM-27, No.4, pp.710-736, April 1979.

3-Rabiner, L.R., and Schafer, R.W., "Digital Processing of Speech Signals," Prentice-Hall Inc., Englewood Cliffs, N.J. 1978.

4-Atal, B.S., and Schroeder, M.R.,"Adaptive Predictive Coding of Speech Signals," Bell Syst. Tech. J., pp.1973-1986, 1970.

5-Zaki, F.W., "A Study of Adaptive Predictive Coding for Speech Signals," Ph.D. Thesis, Liverpool University, U.K., 1981.

6-Gibson, J.D., Jones, S.K., and Melsa, J.L., "Sequentially Adaptive Prediction and Coding of Speech Signals," IEEE Trans. Comm., Vol. COM-22, pp.1789-1797, Nov. 1974.

7-Widrow, B., "Adaptive Filters I: Fundamentals," Stanford Electronics Labs., Stanford, Calif., Rep. SEL-66-126, Dec. 1966.

8-Lucky, R.W., "Adaptive Redundancy Removal in Data Transmission," Bell Syst. Tech. J., pp.549-573, April 1968.

9-Moye, L.S., "Self-Adaptive Filter Predictive-Coding," Proc. of the Int. Zurich Seminar on Integrated Systems for Speech, Video and Data Communication, Zurich, 15-17 March 1972.

10-Moye, L.S., "Study of the Self-Adaptive Filter Deconvolver in Digital Speech Transmission Systems," Final Rep., Stand. Telecomm. Labs., Feb. 1973.

11-Sage, A.P., and Melsa, J.L., "Estimation Theory with Applications to Communications and Control," McGraw-Hill, N.Y., 1971.

12-Max, J., "Quantising for Minimum Distortion," IRE Trans. Inform. Theory, Vol. IT-6, pp.7-12, March 1960.

13-Paez, M.D., and Glisson, T.H., "Minimum Mean Square Error Quantisation in Speech, PCM and DPCM Systems," IEEE Trans. Comm., Vol. COM-20, pp.225-230, April 1972.

14-Cohn, D.L., and Melsa, J.L., "The Residual Encoder: An Improved ADPCM System for Speech," IEEE Trans. Comm., Vol. COM-23, pp.935-941, Sept. 1975.

15-Gibson, J.D., "Sequentially Adaptive Backward Prediction In ADPCM Speech Coders," IEEE Trans. Comm., Vol. COM-26, pp.145-150, Jan. 1978.

16-Gibson, J.D., Berglund, V.P., and Sauter, L.C., "Kalman Backward Adaptive Predictor Coefficient Identification in ADPCM with PCQ," IEEE Trans. Comm., Vol. COM-28, pp.361-371, March 1980.

17-Jayant, N.S., "Adaptive Quantisation with a One-Word Memory," Bell Syst. Tech. J., Vol. 52, pp.1119-1144, Sept. 1973.

18-Zaki, F.W., "A New Adaptive Line Spectral Pair Filter for Linear Prediction of Speech," Mansoura Engineering Journal (MEJ), Vol.15, No.1, pp.E1-E13, June 1990.

19-McDonald, R.A., "Signal-to-Noise and Idle Channel Performance of Differential Pulse Code Modulation Systems- Particular Applications to voice Signals," Bell Syst. Tech. J., Vol. 45, pp.1123-1151, Sept. 1966.

20-Zaki, F.W. , " Learning Characteristics of a New Adaptive Line Spectral Pair Filters," to be Published at the Signal Processing Journal, Vol. 25 Issue 3, 1991.