Codebook Excited Linear Prediction of Speech: Performance in the Presence of Channel Errors

C. C. Chu and P. Kabal

INRS-Télécommunications
3 Place du Commerce
Ile des Soeurs, Qué.
CANADA H3E 1H6

March 1988

Rapport technique de l'INRS-Télécommunications 88-10

Contents

Contents .... *i*

Figures .... *ii*

Tables .... *iii*

1. Introduction 1
2. Experiment Setup 2
   - 2.1 Channel encoding for pitch predictor information 4
3. Performance of the Prototype Design 4
4. Modifications to the Coder Design 5
   - 4.1 System reset 5
   - 4.2 Residual gain quantization 7
   - 4.3 Line spectral frequencies quantization 8
5. Performance of the Modified Design 11
   - 5.1 Performance with respect to isolated channel errors 11
   - 5.2 System reset 13
   - 5.3 Performance in the presence of i.i.d. random errors 14
   - 5.4 Performance in the presence of clustered channel errors 16
6. Conclusion 20

Appendix A. Residual gain quantizer channel error analysis 22

Appendix B. Analysis of the LSF ADPCM quantizer gain corrective measure 23

References 26

Figures

1 Block diagram of the simulated communication system ... 3

2 Segmental SNR measure, segSNR, of the performance of the coder design with respect to i.i.d. random channel errors with bit error rate Pe. ... 15

3 Log-likelihood LPC dissimilarity measure, d, of the performance of the coder design with respect to i.i.d. random channel errors with bit error rate Pe. ... 16

4 SNR and LPC dissimilarity measures of the performance of the coder design as a function of time (Test signal is TOMF8). ... 17

5 Segmental SNR measure, segSNR, of the performance of the coder design with respect to burst channel errors as a function of the degree, $\log_2 M$, of the M-sequence used for bit scrambling. ... 18

6 Log-likelihood LPC dissimilarity measure, d, of the performance of the coder design with respect to burst channel errors as a function of the degree, $\log_2 M$, of the M-sequence used for bit scrambling. ... 19

A.1 Residual gain quantization ... 22

B.1 Structure of a LSF ADPCM decoder ... 24

B.2 Approximation of $G_n^d = (G_{n-1}^d + 1)/2$ by $G_n^d = G_{n-1}^d/1.5$ ... 25

Tables

1 Statistics of the burst error sequence from JPL 17

1. Introduction

In general, there are three classes of speech coding techniques: waveform coding, source coding, and hybrid coding. The class of waveform coding techniques makes use of speech statistics in encoding. Time domain waveform coding techniques attempt to replicate input signal waveforms or waveforms derived from the input signals. Examples of time domain waveform coding are pulse code modulation (PCM) and differential pulse code modulation (DPCM). Time domain waveform coders can be very simple and easy to implement. High quality reconstructed speech is possible with higher transmission bit rates. Typical figures for PCM and simple DPCM are 64 kb/s and 32 kb/s. Lower bit rate waveform coding is possible with more complicated time domain coders or frequency domain coders such as SBC or ATC.

Low bit rate speech coding can be provided by source coders. Source coders digitize and compress speech signals by modeling, approximating and coding the parameters describing a speech production model. A requirement of source coding is a good model of speech production including differentiation between voiced and unvoiced sounds. Voiced speech is generated by exciting the vocal tract with periodic pulse trains while unvoiced speech uses a noise excitation. The parameters necessary include voiced/unvoiced decisions, the periodic excitation rates, gain factors, and parameters describing the vocal tract. A time domain example of source coder is linear predictive coding (LPC) coder. Transmission bit rates around 2 kb/s are feasible with LPC coding. However, because of inaccuracies and rigidity in the speech production model, the reconstructed speech has a synthetic quality.

Recently, many new speech coding designs based on combining aspects of waveform coding and source coding are being studied. The purpose of this work is to obtain a hybrid speech coding technique with the properties of both high reconstructed speech quality and low bit rates. Two promising techniques have been proposed. The first is called code excited linear predictive coding (CELP) [1] and the other is called multipulse speech coding [2][3]. In a CELP coder, the residual signal after linear prediction is quantized. Indices of the approximations are transmitted. Information left in the residual signals is captured and used to excite the synthesis filters in the receiver. In multi-pulse speech coding, the excitation to the synthesis model is a sequence of samples whose positions and amplitudes are coded. The major difference between the CELP and multipulse coding scheme is the absence of pitch synthesis filter in the later design. The limitation or rigidity of exciting LPC inverse filters with either a periodic pulse train or a noise signal is eliminated in either scheme, however. Good quality reconstructed signals are obtainable with either hybrid scheme with transmission bit rates between 4.8 kb/s and 16 kb/s.

The Code Excited Linear Prediction Coder, CELP, developed at this institution is a high quality medium bit rate speech coder designed on the base of residual coding technique [1]. Previous work on the CELP coder design results in a relatively light computational load due to the development of an efficient codeword selection algorithm which allows the use of a 32-codeword codebook. Reduction in transmission bit rate requirement is facilitated by coding the Line Spectral Frequency LSF representation of the LPC information. New adaptive postfiltering technique has also been considered in the design to enhance the subjective quality of the decoded speech signal. The complete design of the CELP coder has been documented in [4], [5], and [6]. As reported in [4], [5], and [6], high quality reproduced speech signal is available at the decoding end of the speech coding/decoding system when it was computer simulated in an ideal environment: noise free input background† and ideal transmission channel with zero channel error probability. This report documents the study of the effect of channel transmission errors on the performance of the CELP coder design. Modifications to the coder design which is required to increase the channel noise immunity of the coder will be considered.

This report will be organized as follows. Section 2 will briefly describe the system setup for the computer simulations and the channel noise type to be considered. The performance of the prototype design in computer simulations will be highlighted in this section as well. In section 3, the performance will be analyzed and modifications required will be described in detail. It is emphasized at this point that no major redesign of any coding algorithm will be attempted. The modifications considered in many cases are ad hoc and heuristic. They are designed to combat specific problems encountered in the simulations. Section 4 will evaluate the final performance of the modified model. Three types of channels will be considered. They are channels with isolated errors, independent and identically distributed, i.i.d., random errors with bit error probability Pe, and simulated errors for mobile radio communications respectively. Section 5 will be the section summing up the study and proposing problems remained to be solved.

2. Experiment Setup

In a real digital communication environment using a binary channel, information from the encoder is first converted into a sequence of information bits and sent into the digital channel to be transmitted to the receiving end. The receiver then interprets the received information appropriately to regenerate the quantizer output level indices to be used by the decoder to reproduce the coded input signal.

Recall that the information generated by the source coding design and interpreted by the decoder is the various quantizer output level indices. There are two ways of inserting channel transmission errors into the information stream between the encoding and decoding end of a digital communication

† Preliminary studies show that the coder performance is good when handling telephone graded input speech or white noise contaminated input speech with SNR > 20 dB.

system based on this speech coder design. Figure 1 shows the block diagram of the communication system assumed in this study. In some of the simulations to be described below, the effect of isolated channel errors to a selected piece of information is desired. For better control, the errors are imposed onto the quantizer output indices directly according to arbitrary rules designed for the selected information. The rules are designed in such a way as to maximize the effect of the impairments. On the other hand, whenever the overall performance of the coding system is of interest, an error bit sequence is exclusive-OR'ed with the information bit stream converted from the quantizer output indices. The conversion is performed according to a protocol agreed upon by both the transmitting and receiving ends of the system. This way of applying channel errors in the simulations, namely in bit level, is necessary because the information sent onto a channel is of heterogeneous nature. Effect of channel errors on the overall system performance is affected by the error protection measure employed, for example Gray code and possibly bit scrambling, and by the format of the information bit allocations. The effect of continuous error bits can better be reflected and observed in the bit level.

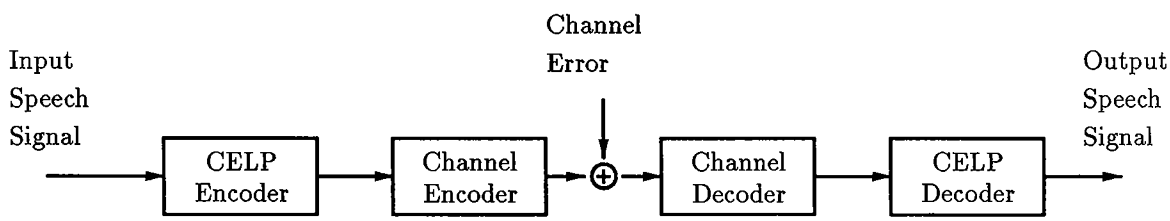

Fig. 1 Block diagram of the simulated communication system

In the prototype encoder design, seven types of information are generated and transmitted. They are the line spectral frequencies, LSF's, for every second frames, LSF interpolation factors for the other frames, pitch predictor tap coefficients and delays, magnitude of the residual gains, sign of the residual gains, and residual codewords for the subframes †. The format of information bit allocation is defined in this order. The formant predictor information for every two frames is followed by the pitch and residual information of each subframe contained in this pair of frames.

Many channel error protection schemes have been designed to ensure the information transmitted by one side of a communication system to be received with minimum damage. It was found that it is more advantageous to invest additional bits for error protection purpose than for the improvement of an encoder design [7]. Due to strict limitation to transmission rate requirement, no redundant error protection scheme will be considered in this study. However, Gray code is applied to

† With a sampling frequency of 8 kHz, a frame is defined as a block of 120 samples and a subframe 40 samples.

indices-to-bits conversion and vice versa. The effect of bit scrambling as an error protection measure will be studied also when fading channels such as in multipath transmission are considered.

2.1 Channel encoding for pitch predictor information

It was reported in [6] that 10 bits are used to jointly code 73 possible predictor delays and 14 possible predictor tap coefficient quantizer outputs. Special care is required in the indices-to-bits conversion process for the two pieces of information so as to minimize the effect of perturbation in one information on the other.

It is noted that

$$73 \cdot 14 = 64 \cdot 14 + 9 \cdot 14.$$

Accordingly, the collection of 10-bit codes designed for the joint pitch information coding is divided into two sections such that 64 delay values and 14 tap coefficients can be coded independently using 6-bit and 4-bit Gray codes, respectively. However, because of the odd numbers involved, the other 9 delay values and 14 tap coefficients have to be coded together using 10-bit strings. Therefore an error in a code in the first collection would affect either the predictor delay or the tap coefficient only whereas an error inflicted in a codeword in the second collection could result in incorrect decoding of both pieces of information. To minimize the probability of experiencing the later effect, the codes in the second collection are assigned to 9 delay values which occur less often.

3. Performance of the Prototype Design

With the system setup described as above, the performance of the prototype design is evaluated. For the computer simulations, a set of eight audio speech utterances from the data base is used. They are

1. PIPM8: (male) The pipe began to rust while new.
2. DOUG3: (male) It's easy to tell the depth of a well.
3. CATM8: (male) Cats and dogs each hate the other.
4. CANM8: (male) The red canoe is gone.
5. TOMF8: (female) Tom's birthday is in June.
6. OAKF8: (female) Oak is strong and also gives shade.
7. CATF8: (female) Cats and dogs each hate the other.
8. THVF8: (female) Thieves who rob friends deserve jail.

Each one of these utterances is recorded under well controlled recording conditions and therefore has high input signal-to-noise ratio, SNR.

Preliminary studies show that the prototype design is extremely sensitive to channel error perturbations. The effect of i.i.d. random channel errors even with a very low bit error rate can be disastrous. Line spectral frequencies crossovers are found in many frames in the decoder output. System instability which is manifested by continuous synthesized signal energy saturation is found in every trial.

4. Modifications to the Coder Design

In view of the poor performance of the prototype design of the CELP coder in the presence of channel transmission errors, certain modifications to the design are considered. The objective of these measures is to increase the robustness of the coder to channel error perturbations. Specifically, the sensitivity of the coder to channel noise has to be reduced and the process of "forgetting" the presence of past channel errors has to be speeded up. While system reset is designed as to the later effect, changes in certain parameter coding designs are also necessary. In this section, the design of a system reset mechanism and the change in parameter coding schemes will be detailed.

4.1 System reset

System reset is an effective way of system error control. Its primary usage is to help the system reconfigure itself, isolate the effect of channel errors and forget the presence of perturbations in the decoder due to past channel transmission errors. This section gives the design of a system reset mechanism for the system as shown in Fig. 1.

In a communication system or at least in this speech coder design, system reset has to be carried out in both the encoder and decoder synchronously. An identical system reset mechanism must be designed for the two ends of the communication channel such that the decoder can assume a new system configuration identical to the one in the encoder. Otherwise, received information will be incorrectly interpreted, and undesired result is generated. In the sequel, system reset refers to re-initialization of all elements with memory. These components are the synthesis filters and the DPCM coders for LSF and residual gain quantization. Simulations show that re-initialization of memories of the parameter coding devices only could lead to undesired result.

Because of the absence of error detection bits, it is not possible to determine whether the received signal in the decoder has been corrupted by channel errors or not unless the effect of the errors is too severe. For example, when crossovers occur in the received LSF's, the presence of channel transmission errors can be ensured. It is then obvious that system reset cannot be carried out only immediately after the perturbation of channel errors. Furthermore, the encoder cannot foresees the appearances of channel errors. As a convention, the reset mechanism in either end of the channel may trigger a reset only in silence or intervals with low energy content. By doing so, synthesized speeches will not be interrupted in a speech segment resulting in distortions due to changes in system status. Based on an energy measure which is available to both the encoder and decoder, system reset can be carried out.

Design of the system reset mechanism basically involves selection of the energy measure and an appropriate threshold with respect to which a decision is made. Among energy measures available in the coder design, energy of the synthesized speech signal is the only one truly reflecting the energy content of a speech signal. However, any good measure for system reset should not be affected by channel errors, especially instability of the synthesis filters. Because energy of the synthesized speech signal could be affected by instability of the LPC and the pitch synthesis filters, an energy measure defined as the sum of squared residual gains of the past N subframes of samples is chosen instead. Although this measure does not faithfully show the energy contour of the synthesized utterance, it was found to be a good indicator of the absence of signal energy. When the result of such an energy measure is lower than a chosen threshold, the subframe in concern is classified as silence and a system reset is triggered. Two basic control parameters then are of interest. They are the silence detection threshold and the number of subframes N over which the squared residual gains are summed.

These two parameters have to be chosen so that system reset can be performed without causing perceptible distortion. In addition to avoiding reset in a speech segment of whatever amount of energy content, empirical results show that a certain time difference should also be allocated between a system reset and the onset of the next speech segment. Otherwise, undesired components such as clicks may be generated. In view of the energy fade in and fade out characteristics, the value of N should therefore be large to enable the mechanism to ignore short pauses between adjacent words and the threshold be not too low so as to start a reset early. However, if the value of N is too large, no silence will be detected. A large N also allows more accumulated error in residual gain to be included to alter the correct decision. A relatively low threshold is also required to avoid system reset in the middle of weak speech segments.

In various computer simulations, different combinations of N and threshold are examined for the high SNR input sentences. The following combination is found to give the most favorable synthesized speech signal with least perceptual distortions both in the presence and absence of channel errors[◇].

$$N = 6$$

$$\text{SilenceThreshold} = 150$$

A third control parameter is also devised to control the maximum rate of system reset. It is found that undesired perceptual effect can be observed if system reset is performed continuously or with a high rate in silence intervals. Under the same simulation condition, a distance equivalent to about 60 subframes of samples is found appropriate to separate two adjacent resets.

A problem with system reset is that a decision cannot be made based on look-ahead information. Without introducing extra delay, there is no way to determine how far the onset of the next speech segment is away from a reset. In general, the subjective quality of synthesized output generated with system reset and in the absence of channel errors has extraneous components in the form of clicks and noise. In some cases, the reproduced signal sounds slightly distorted around where a reset takes place. Other than these, the new synthesized output sounds identical to that generated without reset. More information on the effect of system reset on system performance will be given in Section 5.2.

4.2 Residual gain quantization

In the prototype of the CELP coder design, the sign and magnitude of the residual gain are coded separately. A single bit is allocated to the sign of a residual gain, and four bits are used to code its magnitude based on the DPCM coding scheme. In the presence of channel transmission errors, a perturbation on a received sign bit effectively changes the phase of the received block of residual signal selected from the codebook. The magnitude of the residual gain factor is not affected. Furthermore, such an error is local to the subframe in concern. Error propagation into subsequent residual gain decoding is not expected. On the other hand, error propagation in differential decoding of the residual gain magnitude can result in unpleasant effect. In the CELP coder design, a time invariant predictor of unity tap coefficient is used in the DPCM coding of the residual gain magnitude. It was shown using an idle channel, input signal energy being zero, that the effect of a single bit of channel error on the DPCM quantizer output did result in decoding problem lasting for many frames [†]. A leakage factor less than 0.98 is required if fast relief is desired.

The DPCM coder for the residual gain magnitude is designed again in an iterative approach. Specifically, the DPCM quantizer breakpoints and output levels are sought. To start off the iteration,

◇ Given that the energy of each subframe of residual signal in the codebook has unit variance, the value of the silence threshold selected corresponds to the total energy of the scaled residual signal over $N = 6$ subframes or 240 samples. The average energy of a 240-sample block of the eight input sentences used in the experiment is 36258.78.

† Appendix A shows the analysis of error propagation effect of the combined residual gain coding scheme.

the DPCM coder is given the quantizer designed for the prototype design. The predictor tap coefficient ρ is equal to the first order auto-correlation of the unquantized residual gain magnitudes selected during the analysis stage of the residual coding process ‡. The predictor tap coefficient is found to be

$$\rho = 0.87$$

During each iteration, a sequence of best residual quantizer input samples, differential residual, is generated. A new quantizer is designed based on the histogram of this sequence by minimizing the mean squared quantization error. The process is stopped when there is little change in the shape of the histogram resulted from the iterations. The final quantizer design after all iterations is obtained by selecting the best dynamic range from the final histogram and minimizing the mean squared quantization error based on the histogram clipped at the chosen bounds.

4.3 Line spectral frequencies quantization

It is observed that the magnitude of the single tap pitch predictor tap coefficient very often is larger than unity. This implies unstable pitch synthesis filter realization according to Schur-Cohn's stability test. Nevertheless, no adverse effect is observed due to pitch synthesis filter instability. Most of the problems experienced in the presence of channel transmission errors can be attributed to the coding scheme for the odd line spectral frequencies. Among all the parameters generated by the prototype encoder, the odd LSF's are most sensitive to perturbation due to channel transmission errors. As mentioned in the last section on the performance of the coder, instability of LPC synthesis filters in the decoding stage of the communication system renders the system designed useless in the presence of channel transmission errors. In the sequel of this subsection, the odd LSF ADPCM coder design will be examined and necessary modifications will be detailed.

It is recalled that leakage was introduced in the DPCM coders for the odd line spectral frequencies in the prototype coder design. However, it was noticed from the LSF parameter traces that once a transmission error is made on an odd LSF ADPCM quantizer output index, the LSF decoded stays away from its correct path with little signs of converging back to it. The effect usually persists until another error occurs or a system reset is triggered. In the light of this observation the ADPCM predictor tap coefficients should be chosen again with more leakage.

Each ADPCM time invariant predictor tap coefficient for an odd LSF coding was chosen to be the first order auto-correlation value of the mean compensated ADPCM input [6]. The new coefficients are also obtained as such from the analysis stage of the CELP coder, which uses the newly designed residual gain quantizer, but in the absence of system reset. With the value of β for

‡ The coder design with no system reset is used in the iterative computation.

the mean estimation set to 0.5, the following auto-correlation values are obtained for the 1st, 3rd, 5th, 7th, and 9th LSF ADPCM predictors respectively

$a_1 = 0.277$

$a_3 = 0.330$

$a_5 = 0.429$

$a_7 = 0.349$

$a_9 = 0.514$

Even though heavy leakage as computed above is applied to the differential odd LSF coding, the system design still does not perform satisfactorily. Further experimental results show that the problem lies with the backward adaptive nature of each time varying LSF DPCM coder. The effect of channel transmission error on the DPCM decoder input mean estimations and quantizer gains can be carried on. If these parameters are set time invariant, the effect of channel error can be nullified very quickly together with the predictor tap coefficients selected.

4.3.1 Odd LSF DPCM input mean estimations and quantizer gains

Among the two sets of parameters: LSF ADPCM decoder input mean estimations and quantizer gains, one or both sets has to be made time invariant to improve the channel error immunity of the coder design. Nevertheless, the improvement has to be achieved at the expense of synthesized speech quality. If only the mean estimation process is turned off, the synthesized output is of lower quality than it used to be. When the gain adaptation process is turned off instead, more obvious distortions are introduced in the form of chirps and clicks at speech/silence transitions [6]. Very unnatural synthesized speech is obtained if both parameters are made constant.

Analyses show that when an error is made in the transmission of an odd LSF ADPCM quantizer output index, the decoded LSF has an error term which is a function of the received quantizer output, the received quantizer output error, the quantizer gain and its error at that frame. For the subsequent frames, this error term in the decoded LSF is passed on with a scaling factor β which was defined for the ADPCM mean estimation process [6]. Although this scaling factor β was chosen equal to 0.5, less than unity, the magnitude of this error term also depends on the other parameters mentioned.Error propagation due to this term alone can be deleterious if it is larger than unity. Because the negative effect due to time invariance of mean estimations is less than that due to constant quantizer gain and that the error term, thus the potential deleterious effect, mentioned above can be cancelled with $\beta = 0$, it is decided that the ADPCM mean estimation process will be turned off.

As for the ADPCM gain adaptation algorithm, no major modification is considered so as to preserve the general perceptual quality of the synthesized speech signal. A small change is required, however. Experimental results show that subjective quality of the reproduced utterances is reduced to a relatively large degree when a LSF ADPCM quantizer gain factor is perturbed as it is being amplified with a multiplication factor larger than unity by the gain adaptation mechanism. The perturbation in this situation is amplified as well. To reduce the effect of error amplification, scaling factors for amplifying all LSF ADPCM quantizer gains are reduced. It is observed from the parameter traces that there are not many occasions in which the gains are increased. Most of the time, a gain drops from its reset value or fluctuates with a tendency to drop. The values of the amplification factors can therefore be reduced with little significant effect on the general performance of the LSF coders. An amplification factor of 1.2 is chosen empirically to replace the previously selected value 1.5 which was used for all except the 9^{th} odd LSF ADPCM quantizer gain adaptation 6. As expected, computer simultions show no degradation in the general synthesized speech quality resulted after the change. However, the change reduces the negative effect of perturbation to the quantizer gain. For example, it gives the system more time to attenuate the error before it causes instability.

4.3.2 LSF crossovers

In the presence of channel transmission errors, it is not unreasonable to expect LSF crossovers in the receiving end of the system because of the independence of ADPCM coders for the odd LSF's. In this section, an ad-hoc procedure is designed to handle LSF crossovers and to prevent subsequent realization of unstable formant synthesis filters. Once an odd LSF crossover is detected, the procedure applies brute force based on heuristic to correct the problem. The even LSF's are then decoded accordingly. It is noted that this measure will not guarantee realization of a new and trouble free LPC synthesis filter, but only mitigate potential disastrous effect.

Assume that a crossover between two or more odd LSF's is confirmed. It is then necessary to identify the odd LSF's which strays from its normal course and crosses over the other odd LSF's. In the case when an odd LSF is found crossing over more than one other odd LSF's, this LSF is easily identified as the one incorrectly decoded. If the crossover is between only two odd LSF's, then the one which is further away from its other odd LSF neighbor is considered as the one going stray†. Once the LSF in trouble is located, it is pulled away from the crossover point towards its other neighbor by a distance which is a faction of the distance between its two neighboring odd LSF's under the assumption that these two neighbors are correctly decoded. The identification and rectification processes are proceeded in this manner for all the odd LSF's in ascending order. Because a relative distance is applied, this measure will not introduce new crossovers. From the parameter traces of corrupted LSF's in contrast to their clean counterparts, the factions are determined to be

1/2 for LSF 1

1/3 for LSF 3

† The lower and upper bounds are considered as two LSF's for the measure.

1/4 for LSF 5

1/5 for LSF 7 and

1/6 for LSF 9.

This arbitrary setting takes into account the larger sensitivity of the low order LSF's to channel errors.

The measure described above just prevents the LPC synthesis filter from becoming unstable at the current frame. It is understood that the quantizer gains for the ADPCM decoders affected are contaminated as well. Further measure such as unscaling of the quantizer gains should be taken to reduce the probability of having crossovers in the coming decoding operations. Specifically, a unscaling factor

$$\begin{cases} 1.5 & \text{or} \\ 1/1.5 & \end{cases}$$

has been selected by taking the quantizer gain amplification/attenuation factors into account. Depending on the crossover pattern and the signs of inputs to the affected ADPCM decoders, a scheme has been designed to apply either 1.5 or 1/1.5 to the quantizer gains. The analysis of this design can be found in Appendix B.

It is noted that the stabilization routine described in this section is an a posteriori and heuristic measure. Also, the routine is invoked only when the error has severe effect on the system. Although the measure shows sign of improving the performance of the system with respect to channel transmission errors, it is not capable of rectifying the problem completely.

5. Performance of the Modified Design

In the last section, several modifications to the prototype coder design were described in addition to the design of system reset. These modifications are to the odd LSF ADPCM coding/decoding, and residual gain quantization. The overall performance of the modified system will be examined in this section in the presence of channel transmission errors. The effect of each modification on the overall performance will not be considered, however.

5.1 Performance with respect to isolated channel errors

In order to understand how each parameter reacts to the presence of channel errors, an isolated error is applied to each type of received quantizer output indices except for the residual codeword indices and the odd LSF interpolation factor quantization indices. Specifically, the parameters to be studied are the LSF's, residual gain, and pitch predictor tap coefficient and delay. It is believed that the effects of errors in residual codeword indices and LSF interpolation factors on system performance are relatively harmless; therefore they are not considered. The experiment is repeated for each parameter with the error hit applied in different selected locations of the utterances: silence, voiced, and unvoiced segments.

Experimental results show that an error typically can alter the characteristics of the speech reproduction model. However, the disastrous effect encountered in the prototype design is not observed again. In most cases, the perceptual impact of channel errors is acceptable and even negligible so long as there is no saturation or rapid increase in synthesized signal energy. Also, unlike in a sample-by-sample coder the effect of channel errors observed in this design is smeared over subframes of samples. The error is in the form of unnaturalness such as being throaty and muffle and sometimes mispronunciation is detected. In some cases, slight bongs, clicks, and flutterings of short duration may be observed. The worst situation arises when the LPC synthesis filter becomes or tend to be unstable. Energy of the synthesized output signal increases rapidly in this case producing a very annoying perceptual distortion. The cause of this abnormal behavior can be due to direct error hit on the odd LSF ADPCM quantizer output indices, perturbation to the odd LSF ADPCM quantizer gain through backward adaptation or malfunction of system reset. Perturbations to other parameters have relatively slight or little perceptual impact.

Typically, each one of this kind of undesired effect: be it negligible or annoying lasts with no prolonged error propagation and affects the reproduction of a single word only. This is so regardless of the location of corruption in the utterance and severity of the effect. Because of the use of different quantizer designs for different information, different parameters have different sensitivity to direct/indirect error perturbations and different error propagation properties. Specifically, errors on the codeword selection, sign of the residual gain, pitch information, and even numbered LSF's are local to the point of error hits. Except for the even numbered LSF's, they are not affected by errors inflicted to other information of the system either. The other parameters: magnitude of the residual gain and odd numbered LSF's, however, are coded using DPCM scheme with or without backward adaptation. Error propagation is expected in the decoding of these parameters. Furthermore, the backward DPCM quantizer gain adaptation opens a way allowing perturbation to other system parameters to have influence on all line spectral frequencies decoding. Because of the lack of redundant error protection measures, any perturbation to the system configuration can therefore directly or indirectly causes error propagation in all ADPCM decoders. With the modifications in force, obvious improvement is achieved. Experimental results show that all LSF's are back or tend to be back in track after a perturbation regardless of the point of occurrence. The perceptual effect of errors is very localized although it usually takes approximately 20 frames for a heavily perturbed LSF or 20 subframes for a perturbed residual gain to physically return to its normal path completely. As a result of this, the rest of the synthesized utterances after recuperation sound as if no perturbation has occurred. Experimental results also show that this self-tracking capability is present even in the absence of system reset.

5.2 System reset

System reset is the first mechanism devised in this study to control error propagation due to channel transmission errors. This measure was designed to restart the complete speech coding system at carefully selected locations of the signal so as to force the system to " forget" the presence of any possible channel errors received in the decoding end of the communication system. In this subsection, the performance of the system reset mechanism is to be evaluated. A final decision will be made as to whether the system reset design is really necessary.

The system reset mechanism was designed as a function of three control parameters: silence threshold, window width for residual energy computation and minimum recuperating time required between two successive system resets. These control parameters were chosen empirically to allow the system reset mechanism to meet the objective with minimal interference or distortion introduced to the operation of the coder. The design does perform satisfactorily in computer simulations. Normally, operations in both the encoding and decoding ends of the system are in perfect synchronization after a system reset. The effect of any channel error is completely nullified. However, certain shortcomings are found with the design.

Perfect synchronization is required all the time. It was realized that the residual energy measure could be affected by channel transmission errors in residual gains. A sequence of residual gains over N subframes was used to minimize the effect of error in a single residual gain on the decision. Nevertheless, the decision can still be altered in the decoder because of error propagation in the residual gain DPCM decoder. This kind of error can usually be found if a channel error is inflicted on the residual gain information in silence segments of the signal in which small fluctuations in residual gains are enough to send the energy measure across the decision threshold. Result shows that the general perceptual quality of the decoded signal is not affected. However at where system reset has errors, loud clicks, musical noise or bongs appear. These extraneous components do not appear to last or propagate. In the worst case, system instability is observed and the perceptual impact is large although short in duration.

Another negative feature of the system reset design is its dependence on input speech signal to the encoder. Specifically, the control parameters should be adjusted to the signal-to-noise ratio, SNR, of the input signal. The setting of the control parameters used in the simulations assumes input signal of high SNR value. However, background noise can be excessive in real working environment. As more white noise is added to the input speech signal, the prediction residual contains a lower percentage of unpredictable speech components. The residual gain values computed become less informative to distinguish the presence of speeches from silence. The resulting energy measure may not be reliable at all if the SNR level drops too low †. In retrospect, correct system reset decisions may be difficult to obtain when the input signal energy level is low also. This is so even if the input is free from background noise.

Recall from the last section that the modified coding system is very good in automatic retracking. After a channel transmission error is added to the received information, the perturbed LSF's and other parameters converge back to the pathes assumed in the absence of error in a few frames. This is found so before the next system reset is triggered or even when the complete system reset mechanism is turned off ‡. In view of this self-tracking capability, system reset is found not necessary. Furthermore, the stringent conditions required by the system reset design make it an even undesirable measure.

5.3 Performance in the presence of i.i.d. random errors

In this section, the performance of the modified coder design with system reset turned off and in the presence of random errors in the channel is evaluated. The quantized information obtained from the analysis of each utterance is sent through a channel with bit error probability Pe and decoded for the reproduction of the utterance. Each synthesized sentence are then compared to the corresponding one obtained in the absence of channel errors, $Pe = 0$.

The objective measures adopted here are the segmental signal-to-noise ratio, segSNR, computed for blocks of 15 ms long and the LPC dissimilarity measure proposed by Itakura [8] [9]. The segSNR measure for each utterance is defined as the segmental channel error free reproduced signal energy to channel noise energy ratio. The LPC dissimilarity measure, d, quantifies the effect of channel noise on the decoded formant structure of the speech reproduction model. The value of d increases as perturbation on the formant structure becomes more pronounced ◇. Figure 2 and 3 respectively show the segSNR and the averaged d measures averaged over all utterances as functions of Pe. Figure 4 shows the time domain behavior of the channel error free reproduced signal energy to channel noise energy ratio, SNR, and the d measure for one of the utterance processed.

Informal subjective listening test was performed to evaluate the performance of the coder at $Pe = 10^{-4}, 5 \cdot 10^{-4}, 10^{-3}$, $5 \cdot 10^{-3}$, and 10^{-2}. In general, the coder performs well at $Pe < 5 \cdot 10^{-3}$. AsPe increases, distortions such as flutters, clicks, hoarseness, or muffling can be observed.

---

† Unsatisfactory performance can be obtained with SNR lower than 20 dB.

‡ In the absence of system reset, the whole system may not be able to completely go back to the normal mode. Very slight off tracking can still occur at sparse locations in the future after an error hit. No undesired perceptual impact is observed due to this residual effect.

◇ It was reported that when $d > 0.3$, the LPC dissimilarity is significant. However, it was also pointed out by Jayant [9] that this absolute value of d does not faithfully reflect the true subjective quality. One should pay attention to the relative assessments as well.

These distortions becomes obvious with $Pe = 5 \cdot 10^{-3}$ and severe with $Pe = 10^{-2}$ or higher. Although cracklings due to instability appear occasionally, the perceptual effect is not too annoying. While some syntheiszed sentences remain intelligible at $Pe = 10^{-2}$, the presence of large amount of extraneous components and severe distortion to the speech reproduction model at $Pe = 10^{-2}$ make some synthesized utterances very annoying and even barely intelligible.

An informal test is also performed to compare the impact of random channel errors on the modified CELP design and on a 32 Kbits/s ADPCM coder. This ADPCM coder design is conformed to the CCITT specifications. The major difference is that full precision rather than sign algorithm is used in the prediction coefficients adaptation in this ADPCM coder design. The comparison of perceptual impact on the performance of the two coders is carried out with random errors at $Pe = 10^{-3}$. It is found that the perceptual impact of the errors is larger in the ADPCM coder. Rapid trains of crackling are observed throughout all decoded sentences from the ADPCM while errors with the previously described nature are perceived at only a few isolated locations in the CELP decoded sentences. Although the 32 Kbit/s ADPCM coder has excellent general performance compared to the 4.8 Kbit/s CELP coder design, the perceptual impact of channel errors on its performance is more pronounced.

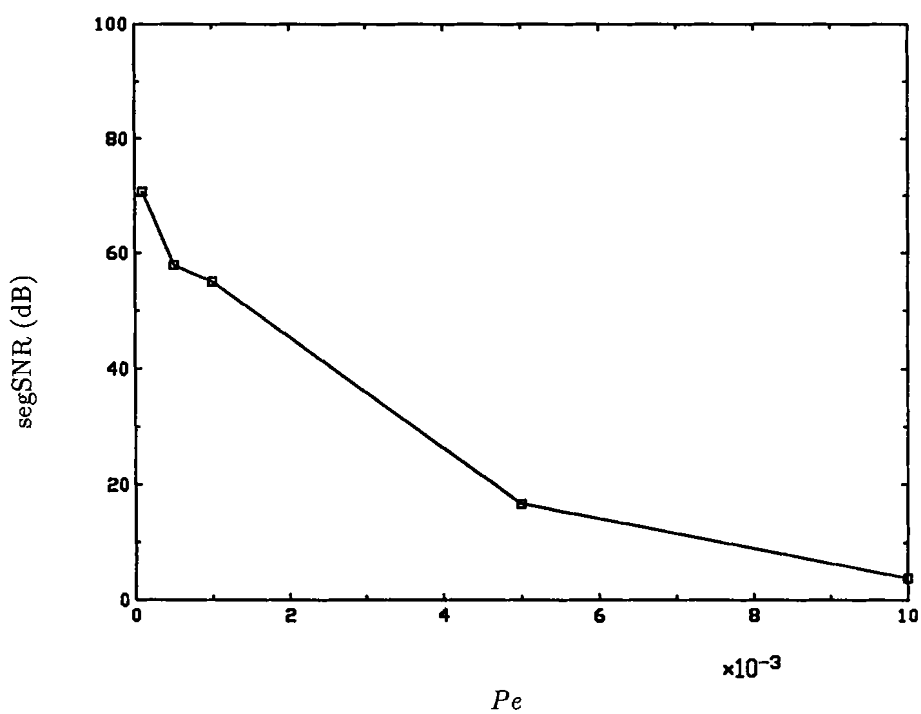

Fig. 2 Segmental SNR measure, segSNR, of the performance of the coder design with respect to i.i.d. random channel errors with bit error rate Pe.

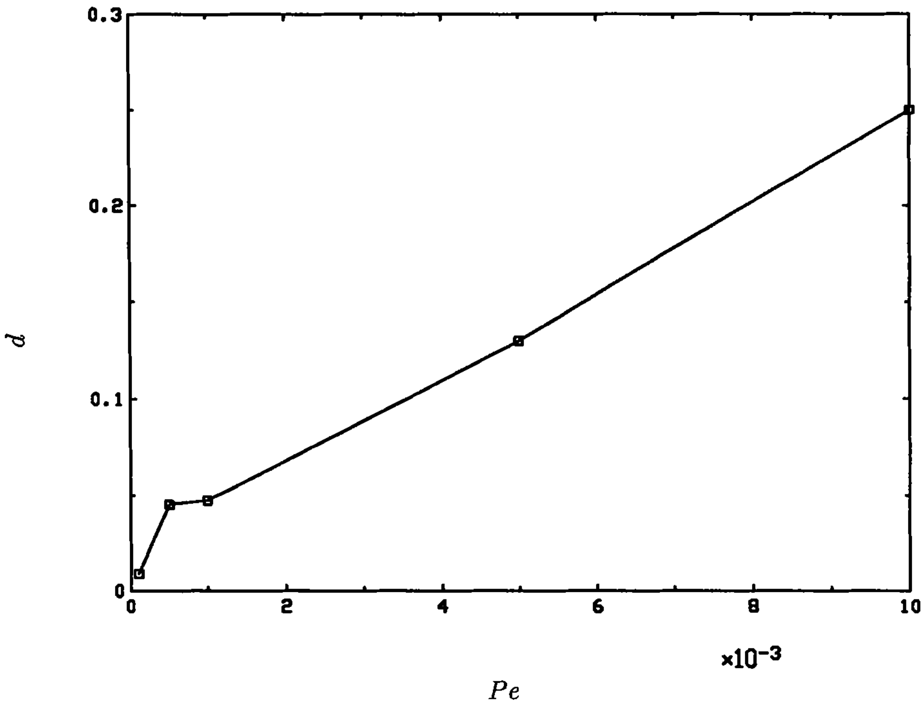

Fig. 3 Log-likelihood LPC dissimilarity measure, d, of the performance of the coder design with respect to i.i.d. random channel errors with bit error rate Pe.

5.4 Performance in the presence of clustered channel errors

One of the possible application of digital speech communication technology is to mobile radio telephony. An advantage is that it allows the realization of all digital signaling and information links. A more appealing advantage is that a digital speech coder is more robust than an analog system to multipath fading which is often encountered in mobile radio telephony [7]. This section evaluates the performance of the modified coding system in the presence of multipath fading or in the digital domain burst channel errors.

The burst error sequence used in this study is generated with a MSAT channel simulator in Jet Propulsion Laboratory, JPL *. The averaged bit error rate computed for the complete sequence of one million bits are 0.0011. As defined by Jayant [7], the mean and median length of contiguous 0's, error free subsequences, in the error bit stream, I_{aver} and I_{med}, are computed. Similarly, D_{aver} and D_{med} which stand for the mean and median length of contiguous 1's in the error bit stream are calculated for the burst error sequence. It is noticed from Table 1 that I_{aver} is extremely large as compared to I_{med}. This fact indicates the presence of some extreme long error free subsequences and many short subsequences of contiguous 0's in the burst error bit stream. The ratio of D_{aver} to

* This error sequence is provided for this study by Dr. Allen Gersho of University of California, Santa Barbara.

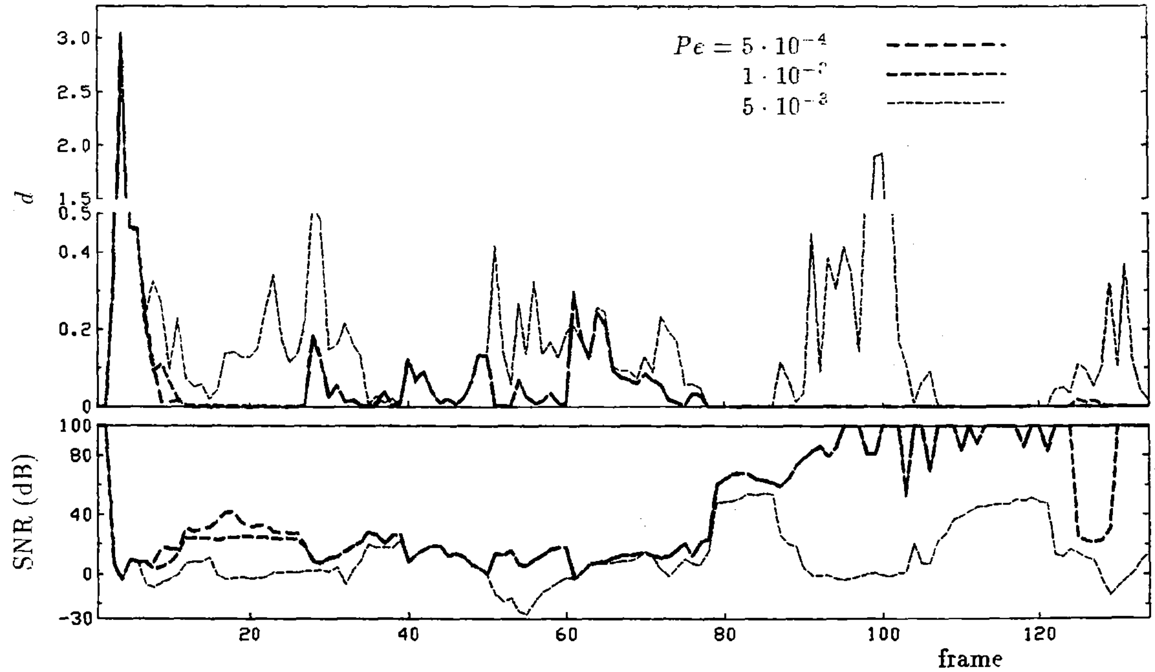

Fig. 4 SNR and LPC dissimilarity measures of the performance of the coder design as a function of time (Test signal is TOMF8).

Error Type	I_{aver}	D_{aver}	I_{med}	D_{med}
Burst	1531.27	1.75	1.13	0.855

Table 1 Statistics of the burst error sequence from JPL

(D_{aver}+ I_{aver}) gives the average bit error rate $Pe = 0.0011$ of the sequence.

Although significant robustness to channel noise can be achieved with explicit transmission of coder side information, no redundant error protection measure such as time diversity coding is attempted. Instead, bit scrambling technique is used so as to alter the statistics of channel errors. Specifically, parallel bit scrambling performed in a block by block fashion is employed in an attempt to randomize the burst channel errors. Locations of the information bits within a block of length M are reorganized according to the pattern of a degree $\log_2 M$ maximal-length feedback shift register sequence (M-sequence) [10]. Inverse operations are carried out in the channel decoder as shown in Fig. 1 to descramble the received information. The effect of the parallel bit scrambling/descrambling operations is to alter the clustered nature of the burst errors. As opposed to the sequential scrambling/descrambling operations where a continuous input bit stream to a shift register scrambler is transformed, an error bit will not stimulate generation of a sequence of error bits in the descrambler. The disadvantage of the parallel operation is the introduction of delay. The bigger the block size is, the more delay is involved.

A single long speech utterance is formed by concaternating together the eight utterances from the database. The resulting information stream has about 89,500 bits. Instead of scrambling this bit stream and descrambling the corrupted bit stream later in the decoder, this information bit sequence is exclusive-OR'ed with a sequence of scrambled error sequence. The received binary digits are decoded and used to synthesize the composite signal directly.

The segSNR and d measures as defined in the last section are used here as well. Figures 5 and 6 show the performance as functions of the bit scrambling block size M.

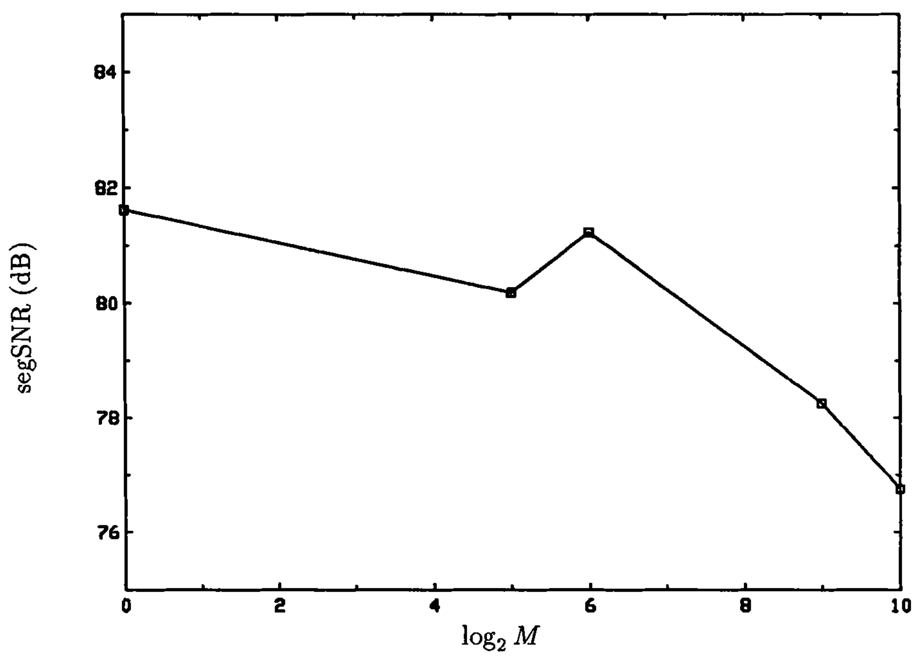

Fig. 5 Segmental SNR measure, segSNR, of the performance of the coder design with respect to burst channel errors as a function of the degree, $\log_2 M$, of the M-sequence used for bit scrambling.

It is noted from Fig. 5 and Fig. 6 that the objective performance of the design degrades and converges to that obtained for random channel errors of rate $Pe = 0.001$ as the bit scrambling block size M increases. Because of the extreme length of some error free intervals (> 1500), it is true that bit scrambling based on M-sequences of degree $\log_2 M \leq 10$ cannot effectively reallocate error bits into error free regions and randomize the clustered errors. However, the information suggested by Fig. 5 and Fig. 6 is believed to be typical and further results can be deduced based on it.

Without bit scrambling, distortions due to clustered channel errors are confined to certain parameters or subframes. As a M-sequence of higher degree is used, many more frames are affected by at least one error bit. It was found that about five times more subframes of the composite signal

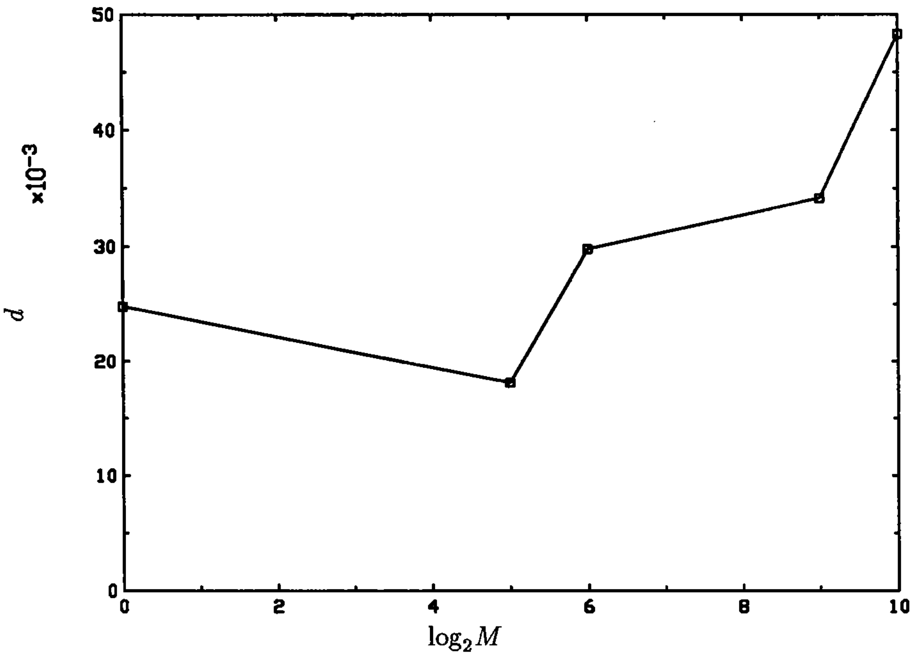

Fig. 6 Log-likelihood LPC dissimilarity measure, d, of the performance of the coder design with respect to burst channel errors as a function of the degree, $\log_2 M$, of the M-sequence used for bit scrambling.

are corrupted by 90 channel errors if bit scrambling with a M-sequence of degree 10 is used. In general, little difference is found in the perceptual impact on the decoding of a parameter due to a single error bit and due to a contiguous sequence of error bits. It is therefore subjectively favorable to have distortions confined to as few parameters as possible without regard to how many bits in the bit streams of these parameters are affected, and have the effect limited in time and space. In this respect, clustered error is more preferrable to i.i.d. random errors which tends to corrupt every frame of a signal independently.

Informal subjective listening tests show that the perceptual impact due to burst errors with an average error rate $Pe = 0.001$ is very small Most of the errors have negligible effect. Obvious perceptual distortions were found on only three isolated words: 'cat', 'friends', and 'rust' throughout the reproduced composite signal even though the number of subframes affected increases with M*. Subjectively, these distortions vary with the degree $\log_2 M$ of the M-sequences. However, no consistent relationship can be concluded between M and the distortions. The rest of the composite signal sound as good as if no errors were present. This result is consistent with the observation performed with random errors. Therefore the coder design is expected to be able to perform well in the presence of even worst channel conditions.

---

* The M-sequences used are of degree ≤ 10.

According to the objective performance of the coder in the presence of clustered errors, it is concluded that the performance of the coder obtained for i.i.d. random errors gives the lower bound of the channel error performance. A real channel with burst error pattern is more favorable than that with random errors of the same average bit error rate to this speech coder design.

6. Conclusion

The prototype coder design is extreme sensitive to transmission channel errors. Most of the serious problems are due to backward quantizer gain adaptation in odd LSF ADPCM coders. Modifications to all DPCM coder designs and other parts of the CELP coder design are considered to minimize the effect of channel errors and to help stabilize operations of the CELP decoder. System reset was also studied. However, it was found to be unnecessary and even undesired after its effect and requirements are examined. Because of strict limitations to transmission bit rate requirement, no redundant information is utilized for channel error protection purpose.

Although degradation in synthesized speech quality is expected due to the modifications in force, little loss of general perceptual quality of reproduced speech signal is observed. In general, effect of channel errors is smeared over subframes of samples. It alters characteristics of speech reproduction model and thus pronounciation of certain utterances. Nevertheless, the perceptual impact is small. The most serious effect of channel errors is observed when the decoding system becomes or tends to be unstable. Such results are not observed ofter, however.

Computer simulations show that clustered channel errors are more favorable to i.i.d. random errors to the CELP coder design. Only three isolated words are found distorted perceptually by clustered channel errors of average bit error rate $\approx 10^{-3}$ in decoding a 11 seconds long utterance. In the presence of random channel errors, the coder design is capable of coding and decoding speech signal with high quality at bit error rate $Pe \leq 5 \cdot 10^{-3}$. Annoying to unacceptable result is obtained at higher bit error rate. At $Pe = 10^{-3}$, the 4.8 Kbit/s CELP coder design is found to be preferred to a 32 Kbit/s ADPCM waveform coder in terms of perceptual impact of channel errors. It is concluded that the modified design has very good channel noise immunity and should be a good candidate in mobile radio telephony and similar applications.

In general, it is not known which information received by the decoder is inflicted by channel errors whenever an error occurs in transmission. Only from the crossover pattern can it be guessed which LSF is in trouble and the nature of the problem. At this stage, there is nothing further can be done to dramatically improving the channel performance of the coder without any major redesigns. However, the ADPCM quantizer breakpoint and output level distributions can be redesigned. In the expense of a few more bits, error detection can be implemented for the pitch and residual signal information transmission. When a channel error is detected in the received pitch and residual signal information of a subframe, the corrupted information may be substituted for by the received information of the previous subframe. This scheme may be feasible in view of the short subframe size and high correlation of this information. Based on the performance of the coder, error detection is not recommended for LSF information transmission. With limited error detection information, the decoder will not be able to determine which LSF is corrupted. Replacing the complete set of LSF by that from the previous frames would disrupt the continuity of operations of all the LSF DPCM decoders. Even if the receiver is capable of isolating the error to a particular LSF received and leaving other LSF DPCM decoder intact in the correction procedure, the LSF information from the previous two frames may not be useful in the substitution because of the low LSF update rate. The impact of disrupting the operations of the DPCM decoders and substituting information from the previous frames for the current corrupted LSF can be as bad if not worse than the impact of the channel error itself. Nevertheless, the exact effect of any operation based on error detection information, especially on the operation of the residual gain and LSF DPCM decoders, remains to be investigated. If higher transmission bit rate can be afforded, the LSF ADPCM quantizers should be made forward adaptive with the quantizer gain information encoded and transmitted explicitly to the receiver.

Appendix A. Residual gain quantizer channel error analysis

This appendix briefly analyzes the effect of channel errors inflicted in the residual gain magnitude quantizer output indices.

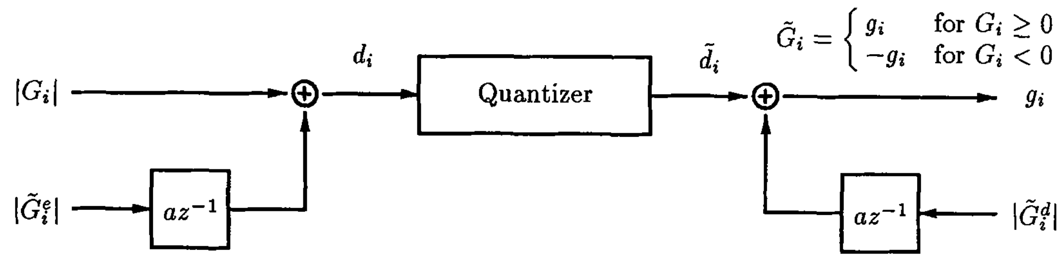

Fig. A.1 Residual gain quantization

In the absence of channel error, ideal channel,

$$
\begin{aligned}
|\tilde{G}_i^e| &= |\tilde{G}_i^d| \\
d_i &= |G_i| - a|\tilde{G}_{i-1}^e| \\
\tilde{d}_i &= d_i + q_i \\
g_i &= \tilde{d}_i + a|\tilde{G}_{i-1}^d| \\
&= |G_i| - a|\tilde{G}_{i-1}^e| + q_i + a|\tilde{G}_{i-1}^d| \\
&= |G_i| + q_i.
\end{aligned}
$$

Therefore in the absence of channel errors

$$
\tilde{G}_i = \begin{cases} G_i + q_i & \text{for } G_i \geq 0 \\ G_i - q_i & \text{for } G_i < 0 \end{cases}
$$

Assume a real channel and the first channel error p_i on the DPCM quantizer output indices appearing at time i, then

$$
\begin{aligned}
d_i &= |G_i| - a|\tilde{G}_{i-1}^e| \\
\tilde{d}_i &= d_i + q_i + p_i \\
g_i &= \tilde{d}_i + a|\tilde{G}_{i-1}^d| \\
&= |G_i| - a|\tilde{G}_{i-1}^e| + q_i + p_i + a|\tilde{G}_{i-1}^d| \\
&= |G_i| + q_i + p_i
\end{aligned}
$$

$\Longrightarrow$

$$
\tilde{G}_i^d = \begin{cases} G_i + q_i + p_i & \text{for } G_i \geq 0 \\ G_i - q_i - p_i & \text{for } G_i < 0 \end{cases}
$$

or

$$
\tilde{G}_i^d = \begin{cases} \tilde{G}_i^e + p_i & \text{for } G_i \geq 0 \\ \tilde{G}_i^e - p_i & \text{for } G_i < 0 \end{cases}
$$

By triangular inequality, this result implies

$$|\tilde{G}_i^d| \le |\tilde{G}_i^e| + |p_i|$$

At time $i+1$,

$$d_{i+1} = |G_{i+1}| - a|\tilde{G}_i^e|$$

with

$$\tilde{G}_i^e = \begin{cases} G_i + q_i & \text{for } G_i \ge 0 \\ G_i - q_i & \text{for } G_i < 0 \end{cases}$$

then

$$\begin{aligned} \tilde{d}_{i+1} &= d_{i+1} + q_{i+1} + p_{i+1} \\ g_{i+1} &= \tilde{d}_{i+1} + a|\tilde{G}_i^d| \\ &= |G_{i+1}| - a|\tilde{G}_i^e| + q_{i+1} + p_{i+1} + a|\tilde{G}_i^d| \\ &\le |G_{i+1}| + q_{i+1} + |p_{i+1}| + a|p_i| \end{aligned}$$

The decoded residual gain at time $i+1$ is

$$\tilde{G}_{i+1}^d \begin{cases} \le G_{i+1} + q_{i+1} + |p_{i+1}| + a|p_i| & \text{for } G_{i+1} \ge 0 \\ \ge G_{i+1} - q_{i+1} - |p_{i+1}| - a|p_i| & \text{for } G_{i+1} < 0 \end{cases}$$

or

$$\tilde{G}_{i+1}^d \begin{cases} \le \tilde{G}_{i+1}^e + |p_{i+1}| + a|p_i| & \text{for } G_{i+1} \ge 0 \\ \ge \tilde{G}_{i+1}^e - |p_{i+1}| - a|p_i| & \text{for } G_{i+1} < 0 \end{cases}$$

Therefore, the effect of channel errors on the residual gain magnitude quantization can be generalized by the following relation

$$\tilde{G}_i^d = f(\tilde{G}_i^e, \sum_{j=0}^{i} a^{i-j}|p_j|)$$

Appendix B. Analysis of the LSF ADPCM quantizer gain corrective measure

Let

G_n^e be a LSF DPCM quantizer gain calculated by the encoder at time n,

G_n^d be the corrupted version of G_n^e computed by the decoder,

G be the initialized/reset value at time 0, and

$a_1, ..., a_j, ..., a_n$ be the scaling factors at time $j = 1, 2, ...,$ and n

respectively. Assume that the adaptation process for the ADPCM quantizer gain is perturbed at time i. Instead of a_i, a scaling factor a_i' is then used in the computation of G_n^d. Further assume that $a_{i+1}, ..., a_n$ are not affected by this perturbation. Then

$$G_n^e = Ga_1a_2...a_{i-1}a_ia_{i+1}...a_n$$

and

$$G_n^d = Ga_1a_2...a_{i-1}a_i'a_{i+1}...a_n$$

This implies that

$$G_n^d = G_n^e \frac{a_i'}{a_i}$$

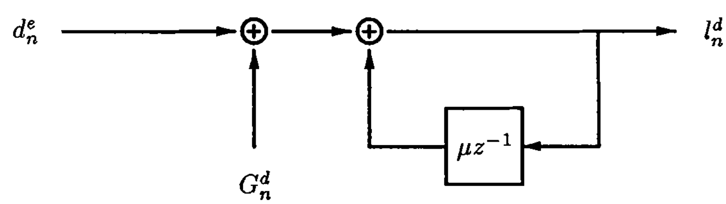

Fig. B.1 Structure of a LSF ADPCM decoder

Assuming no error in the LSF ADPCM quantizer output index received by the ADPCM decoder as shown in Fig. B.1, then

$$\begin{aligned} l_n^d &= \mu l_{n-1}^d + G_n^d d_n^e \\ &\approx \mu l_{n-1}^e + G_n^d d_n^e \qquad \text{provided} \quad \mu < 1 \\ l_n^d &\approx \mu l_{n-1}^e + G_n^e d_n^e \frac{a_i'}{a_i} \end{aligned}$$

It is noted that both l_{n-1}^e and l_n^d are both positive quantities. Depending on the crossover pattern detected and the sign of the quantizer output received by the ADPCM decoder, four cases can be deduced for the magnitude of the ratio $\frac{a_i'}{a_i}$.

1. l_n^d too high, d_n^e positive $\Longrightarrow \frac{a_i'}{a_i} > 1$
2. l_n^d too high, d_n^e negative $\Longrightarrow \frac{a_i'}{a_i} < 1$
3. l_n^d too low, d_n^e positive $\Longrightarrow \frac{a_i'}{a_i} < 1$
4. l_n^d too low, d_n^e negative $\Longrightarrow \frac{a_i'}{a_i} > 1$

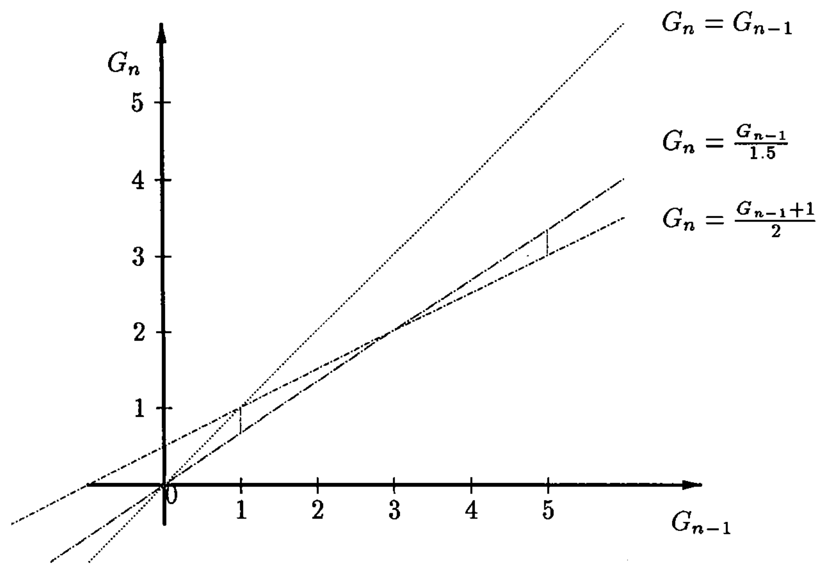

Fig. B.2 Approximation of $G_n^d = (G_{n-1}^d + 1)/2$ by $G_n^d = G_{n-1}^d/1.5$

Therefore G_n^d can be corrected by checking the crossover pattern, the sign of the received ADPCM decoder input and then applying the right unscaling factor

$$\frac{a_i}{a_i'}$$

It was reported in [6] that G_n^d is updated from G_{n-1}^d by applying the scaling factor $a_n = 1.2, 1.0$, or $1/1.5$ or the function $G_n^d = (G_{n-1}^d + 1)/2$. If the function $G_n^d = (G_{n-1}^d + 1)/2$ is approximated by $G_n^d = G_{n-1}^d/1.5$, the error is as shown in the filled region in Fig. B.2 and this error is small for $G_{n-1}^d < 5$ †. With this approximation, the scaling factors can be generalized as

$$a_i, a_i' = 1.2, 1.0, \text{or } 1/1.5$$

and

$$\frac{a_i}{a_i'} = 1.8, 1.5, 1.2, 1.0, \frac{1}{1.2}, \frac{1}{1.5}, \text{or} \frac{1}{1.8}$$

Because the actual values of both a_i and a_i' are not known, the scaling factor sought in this analysis for quantizer gain error correction is chosen to be

$$\begin{cases} 1.5 & \text{or} \\ 1/1.5 & \end{cases}$$

† The normal range of the value of a quantizer gain is [1.0,5.0].

References

1. M. Schroeder and B.S. Atal, "Code Excited Linear Prediction (CELP): High quality speech at very low bit rates", *Proc. Int. Conf. on Acoustics, Speech and Signal Proc.*, Paper 25.1, March 1985.
2. B.S. Atal and J.R. Remde, "A new model of LPC excitation for producing natural-sounding speech at low bit rates", *Proc. Int. Conf. on Acoustics, Speech and Signal Proc.*, pp. 614–617, May 1982.
3. S. Singhal and B.S. Atal, "Improving performance of multi-pulse LPC coders at low bit rates", *Proc. Int. Conf. on Acoustics, Speech and Signal Proc.*, Vol. 1, Paper 1.3, March 1984.
4. P. Kabal, "Code Excited Linear Prediction Coding of Speech at 4.8 kb/s", *INRS-Telecommunications Technical Report 87-36*, July 1987.
5. J.-L. Moncet and P. Kabal, "Codeword Selection for CELP Coders", *INRS-Telecommunications Technical Report 87-35*, July 1987.
6. C.C. Chu and P. Kabal, "Coding of LPC Parameters for Low Bit Rate Speech Coders", *INRS-Telecommunications Technical Report 87-19*, March 1987.
7. N.S. Jayant, "Step-Size transmitting differential coders for mobile telephony", *The Bell System Technical Journal*, pp. 1557–1581, Vol. 54, No. 9, November 1975.
8. F. Itakura, "Minimum prediction residual principle applied to speech recognition", *IEEE Trans. Acoust., Speech and Signal Processing*, vol. ASSP-23, pp. 67–72, Feb. 1975.
9. M.R. Sambur and N.S. Jayant, "LPC analysis/synthesis from speech inputs containing quantizing noise or additive white noise", *IEEE Trans. Acoust., Speech and Signal Processing*, pp. 488–494, Vol. ASSP-24, No. 6, Dec. 1976.
10. D.V. Sarwate and M.B. Pursley, "Crosscorrelation properties of pseudorandom and related sequences", *IEEE Proceedings*, pp. 593–619, Vol. 68, No. 5, May 1980.