National Library of Canada
Acquisitions and Bibliographic Services Branch
395 Wellington Street
Ottawa, Ontario
K1A 0N4

Bibliothèque nationale du Canada
Direction des acquisitions et des services bibliographiques
395, rue Wellington
Ottawa (Ontario)
K1A 0N4

Your file Votre référence

Our file Notre référence

NOTICE

The quality of this microform is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

AVIS

La qualité de cette microforme dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

Canada

VARIABLE RATE SPEECH AND CHANNEL CODING FOR MOBILE COMMUNICATIONS

by

Eric Kwok Fung Yuen

B.A.Sc. Simon Fraser University, 1992

A THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF APPLIED SCIENCE
in the School
of
Engineering Science

SIMON FRASER UNIVERSITY
September, 1994

ISBN 0-612-06876-5

APPROVAL

Name: Eric Kwok Fung Yuen

Degree: Master of Applied Science

Title of thesis: Variable Rate Speech and Channel Coding for Mobile Communications

Examining Committee: Dr. John Jones, Chairman

Dr. Paul Ho
Associate Professor, Engineering Science, SFU
Senior Supervisor

Dr. Vladimir Cuperman
Professor, Engineering Science, SFU
Senior Supervisor

Dr. Jacques Vaisey
Assistant Professor, Engineering Science, SFU
Supervisor

Dr. Steve Hardy
Professor, Engineering Science, SFU
Examiner

Date Approved: September 21, 1994

PARTIAL COPYRIGHT LICENSE

Title of Thesis/Project/Extended Essay

"Variable Rate Speech and Channel Coding for Mobile Communications"

Author: _____
(signature)

K.F. (Eric) Yuen
(name)

September 21, 1994
(date)

ABSTRACT

Although the mobile communication channels are time-varying, most systems allocate the combined rate between the speech coder and error correction (channel) coder according to a nominal channel condition. This generally leads to a pessimistic design and consequently an inefficient utilization of the available resources, such as bandwidth and power. This thesis describes an adaptive coding scheme that adjusts the rate allocation between the speech and channel coder according to actual channel conditions. Two types of network controlled variable rate speech coders are considered : the embedded coders and multimode coders. Both are based on code excited linear prediction (CELP). On the other hand, the variable rate channel coders are based on the rate compatible punctured convolutional (RCPC) codes. A channel estimator is used at the receiver to track both the short term and long term fading conditions in the channel. The estimated channel state information is then used to vary the rate allocation between the speech and the channel coder, on a frame by frame basis. This is achieved by sending the estimated channel state information to the transmitter through a reverse channel. The transmitter then makes an appropriate rate adjustment. We have found that combined CELP/RCPC coders using multimode speech coder perform better than their embedded counterparts. In addition, experimental results indicate that the objective and subjective speech quality of the adaptive coders are superior than their non-adaptive counterparts operating in a Rayleigh flat fading channel. Improvements of up to 1.35 dB in segmental signal-to-noise ratio (SEGSNR) of the speech signal and up to 0.9 in informal mean opinion scores (MOS) for a combined rate of 12.8 kbit/s have been found.

ACKNOWLEDGMENTS

I would like to thank my senior supervisors, Dr. Paul Ho and Dr. Vladimir Cuperman, for their assistance and guidance throughout the course of this research. I am also thankful to all the members of the digital speech research group of Simon Fraser University for their help during the last two years. Finally, I would like to thank Edward Leung for his valuable time and patience in reading and commenting on my thesis.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

A-S	Analysis-Synthesis
A-by-S	Analysis-by-Synthesis
ACB	Adaptive Codebook
ADPCM	Adaptive Differential Pulse Code Modulation
AMPS	Advanced Mobile Phone System
ARQ	Automatic Repeat Request
ASIC	Application Specific Integrated Circuit
BCH	Bose-Chaudhuri-Hocquenghem
CCITT	International Telegraph and Telephone Consultative Committee
CDMA	Code Division Multiple Access
CELP	Code Excited Linear Prediction
CSI	Channel State Information
DPCM	Differential Pulse Code Modulation
DQPSK	Differential Quadrature Phase Shift Keying
E-TDMA	Enhanced Time Division Multiple Access
EIA	Electronic Industries Association
ESI	Error Sensitivity Information
FDMA	Frequency Division Multiple Access
FEC	Forward Error Correction
GSM	Global System for Mobile Communications
ITU-T	International Telecommunication Union
LP	Linear Prediction
LPC	Linear Predictive Coding
LSP	Line Spectrum Pairs
MOS	Mean Opinion Score
MSE	Minimum Square Error
PCM	Pulse Code Modulation
PDF	Probability Density Function
PN	Pseudo-random Noise
PRMA	Packet Reservation Multiple Access
PSD	Power Spectral Density

QAM	Quadrature Amplitude Modulation
QPSK	Quadrature Phase Shift Keying
RCPC	Rate Compatible Punctured Convolutional
RPE-LTP	Regular Pulse Excitation Long Term Prediction
SCB	Stochastic Codebook
SEGSNR	Segmental Signal to Noise Ratio
SNR	Signal-to-Noise Ratio
SQ	Scalar Quantization/Quantizer
TDMA	Time Division Multiple Access
TIA	Telecommunication Industry Association
TS-MSVQ	Tree Searched Multi-stage Vector Quantization
UEP	Unequal Error Protection
VQ	Vector Quantization/Quantizer
VSELP	Vector Sum Excited Linear Prediction
ZIR	Zero Input Response
ZSR	Zero State Response

CHAPTER 1

INTRODUCTION

In the past decade, speech transmission over mobile fading channels has become an important research area. A major concern to the communication system designers is the quality of the reconstructed speech at the receiver. Because of the severe operating condition due to channel fading, a good mobile speech communication system must maintain a given grade of reconstructed speech quality with limited bandwidth.

Cellular systems for mobile speech communications have experienced very fast growth in both North America and parts of Europe and Asia. The first generation systems are all based on analog transmissions such as the Advanced Mobile Phone System (AMPS) [1]. However, due to the need for more efficient utilization of channels, second generation systems that are based on digital transmission have been proposed and already implemented in certain areas [2, 3, 4]. Second generation systems can accommodate three to ten times more users than the first generation systems without a penalty in system performance.

One of the major components in a second generation mobile speech transmission system is the digital speech coder. In mobile radio applications, speech coders are required to have low bit rates and provide high quality speech. Many efficient speech coders have been developed for providing high quality speech at rates between 4.8 kbit/s and 16 kbit/s including code linear excited prediction (CELP) and subband coders [5]. In addition, due to severe noise disturbances in mobile channels, speech coders must be robust to channel errors as well. However, a robust speech coder still requires a channel coder to protect the encoded speech information from channel

errors. Hence, there is a great demand in robust low bit rate speech coders complemented with efficient channel coding schemes. With this motivation, we study adaptive variable rate combined speech and channel coding in which the rate allocation between the speech and channel coders can be adjusted in accordance with the actual channel conditions.

1.1 Combined Speech and Channel Coding

In most existing communication systems, the source coder (i.e., speech coder, video coder, and etc.) and the channel coder are designed separately. For example, in the ITU-T (formerly CCITT) H.261 standard for video conferencing [6], the Bose-Chaudhuri-Hocquenghem (BCH) code is used to encode the data from the video source coder disregarding the properties of the compressed data. The advantage of separating the design of source and channel coders is that it allows the channel coder to be designed independently of the actual source coder used. According to Shannon's theory [7], the source coder and the channel coder are fundamentally separable as long as they may be arbitrarily complex. In practical situations where complexity is a constraint, a degradation in performance is expected if the source and the channel coders are designed separately. Therefore, combining the designing processes of the source and the channel coders is a better approach.

Generally speaking, there are two approaches in designing combined source and channel coding systems. The first approach is to jointly optimize the source and the channel coders. The second approach is to design a channel coding strategy for the available source coder.

An example of joint optimization has been presented by Kurtenbach and Wintz for designing an optimal scalar quantizer for transmission over a binary symmetric channel [8]. It has been shown that the optimal quantizer should be designed based on both the source statistics (characterized by the source's probability density function) and the channel statistics (characterized by the channel transition probabilities). Then, in 1976, Rydbeck and Sundberg demonstrated the importance of optimizing the assignment of channel symbols to the quantizer's codebook indices [9]. It has been found that the Gray code mapping of channel symbols to codebook indices results in a more robust source coding system to channel errors and this method is often called

protection without introducing redundancy. A further step in joint source and channel coding optimization is to introduce optimal index assignment in the codebook design algorithm and to optimize the codebook for some given channel transition probabilities. Farvardin and Vaishampayan have demonstrated channel dependent vector quantizer design algorithms and the resulting quantizers are called channel optimized vector quantizers [10]. A more novel approach has been presented by Liu, Ho, and Cuperman where the decoder uses a linear combination of codebook vectors to reconstruct the original signal [11].

Although most of the joint optimization schemes have shown promising results, it is very difficult, if possible, to incorporate them into more complex and practical coders. For example, Phamdo, Farvardin, and Moriya have designed a vector quantizer for the line spectral pair (LSP) parameters which can be used in a CELP speech coder [12]. The vector quantizer was optimized based on a known channel transition probabilities matrix using a very complicated algorithm. On the other hand, the second approach for combined source and channel coding is much easier. In the second approach, the source coder should first be designed to be robust to channel errors. For example, in most of the CELP speech coders, the linear prediction coefficients are converted into LSP parameters, which are more resilient to error, before transmission. Given that there is a source coder, an appropriate forward error correcting code can be chosen and a good error control strategy can be employed to minimize the distortion in the reconstructed signal due to channel errors. For example, Cox, Hagenauer, Seshadri, and Sundberg have designed a robust subband speech coder and an error protection strategy in which speech parameters with different error sensitivities are protected by rate compatible punctured convolutional (RCPC) codes at different code rates [13]. This is known as unequal error protection (UEP).

1.2 Variable Rate Combined Speech and Channel Coding

In designing a digital communication system for transmitting speech signals, a typical design decision to be made is the partitioning of the available channel bandwidth between speech coding and channel coding. Making such a decision is relatively straight

forward if the channel is static and/or efficient utilization of the available resources is not the primary concern. For such applications, we can simply allocate the combined rate to the speech coder and the channel coder according to the "nominal" channel condition. As a matter of fact, many existing combined speech and channel coding systems are designed this way. Unfortunately, in applications such as land mobile and cellular communications where resources are limited and the channel conditions can vary over a wide range, fixed-rate allocation systems become ineffective in maintaining a given grade of service. Hence, an adaptive system which adjusts the rate allocation between the speech and channel coders is desirable for mobile communications. Adaptive rate allocation has been considered previously in [14] for a combined rate of 32 kbit/s and in [15] for a Rician fading channel. However, the former study has not considered a lower rate system, which is critical in today's mobile communications, and the latter study has not considered a combined speech and channel coding system.

This thesis demonstrates the potential benefits of adaptive variable rate combined speech and channel coding for systems operating in the mobile radio environment. Variable rate speech coders based on code excited linear prediction (CELP) and variable rate channel coders based on rate compatible punctured convolutional (RCPC) codes are considered. In this thesis, although the rates of the speech and the channel coders can be varied, their combined rates are fixed. Thus the problem at hand is equivalent to finding the optimal rate partitionings under different channel conditions. For an optimal adaptive system, the speech and channel coding rates have to be adjusted continuously, which implies that an infinite number of speech and channel coders are required. However, it is obviously impossible to implement such a system. Therefore, as a trade-off between system complexity and performance, a number of combined CELP/RCPC coders that have good performance over the typical range of channel signal-to-noise ratio (SNR) have been designed. Adaptive error control is then achieved by using the combined coder that has the best performance at a given channel SNR.

1.3 Contributions of the Thesis

The major contributions of this thesis can be summarized as follows:

1. An investigation of the performance of variable rate multimode and embedded CELP speech coders. We found that both uncoded and error protected multimode CELP speech coders perform better than their embedded counterparts.

2. The design and evaluation of adaptive variable rate combined CELP/RCPC coders operating in Rayleigh flat fading channels. We found that adaptive coders perform significantly better than their fixed rate counterparts.

1.4 Thesis Outline

In Chapter 2, various common speech coding, channel coding, modulation, and multiple access techniques are reviewed. The channel model assumed in the thesis is also described. In Chapter 3, the variable rate CELP speech coders are described in detail. Two types of variable rate CELP coders have been considered : multimode CELP and embedded CELP coders. In Chapter 4, the RCPC codes are described. A complete description of our adaptive variable rate combined CELP/RCPC coders are provided in Chapter 5, followed by an evaluation of their performance in flat fading channels in Chapter 6. We draw the conclusion and suggest some possible future works in the last chapter.

CHAPTER 2

SYSTEM OVERVIEW

The purpose of speech coding is to represent an analog speech signal by a sequence of binary digits efficiently while the goal of channel coding is to allow the binary data sequence to be reliably transmitted over a noisy channel. In this thesis, both speech and channel coders, which operate at various rates, are studied. Before having a detailed discussion of variable rate combined speech and channel coders, an overview of each major component of our system is presented in this chapter. This chapter begins with a generalized model of a mobile speech transmission system. An overview of speech coding and channel coding is presented in section 2.2 and 2.3 respectively. In section 2.4, some common digital modulation techniques and multiple access techniques are reviewed, followed by the description of a model for mobile channel.

2.1 Model of a Speech Transmission System

A generalized block diagram of a digital mobile speech transmission system is shown in Figure 2.1. The main goal of this system is to transmit the speech signal from the transmitter to the receiver as accurately as possible. In other words, the system is designed in such a way that the distortion between the original and the reconstructed speech signal is minimized.

In a digital communication system, the analog speech signal $s(t)$ is first converted to discrete form through the process of sampling and quantization. Then, data compression is performed to reduce the bandwidth required to transmit the digitized speech. This is often done by removing redundancy from the digitized speech

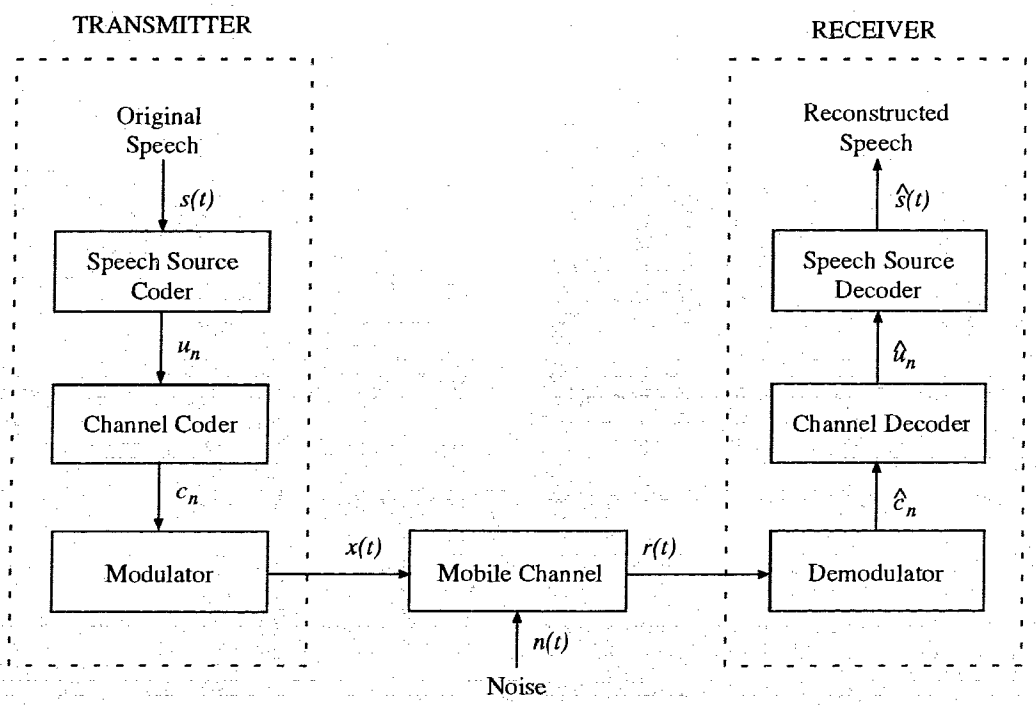


Figure 2.1: Block diagram of a mobile speech communication system

samples. The process of sampling, quantization, and data compression are often collectively known as speech coding. We shall describe some common speech coding techniques in section 2.2.

The data from the speech coder u_n is to be transmitted through a mobile channel to the receiver. In order to combat channel noise disturbances and interferences, channel coding is often required. The objective of channel coding is to transform the speech coder's output in a way such that the effects of channel impairments on the transmitted data are minimized. In other words, the role of a channel coder/decoder pair is to provide reliable communication over a noisy channel. Often, this is done by introducing redundancy to the transmitted data in a prescribed fashion. Hence, redundancy is removed in speech coding while controlled redundancy is introduced in channel coding. Some efficient channel coding techniques are described in section 2.3.

Real channels are all waveform channels and they cannot be used to transmit the sequence of digital data from the channel coder directly. A digital modulator is required to transform the digital information sequence c_n into waveforms $x(t)$ that are compatible with the characteristics of the channel.

At the receiver side, the received signals $r(t)$ are first converted into a sequence of decision values $\hat{c}_n$ through the digital demodulator. The channel and the speech decoders are then used to reconstruct the speech signal. Due to channel impairments and possible loss of information from data compression, the reconstructed signal may not be exactly the same as the original speech signal.

2.2 Speech Coding

The objective of speech coding is to efficiently represent an analog speech signal in digital form. A few reasons for an efficient digital representation of speech signals includes [16]

- Minimizing the communication channel bandwidth required for the transmission of high quality speech signals;
- Obtaining the best possible reproduction quality of speech signals over an available digital communication channel;
- Providing a concise representation of speech signals so as to reduce the amount of computations for subsequent signal processing algorithms such as data encryption.

In most communication systems, bandwidth is an expensive and/or limited resource and thus should be utilized efficiently. For example, low-bit-rate speech coders are often used in power/bandwidth limited systems such as cellular and microwave links [13, 17, 18]. Speech coding is also important in other applications including voice mail, multimedia, and secure communication systems.

Speech is an analog signal with continuity in both time and amplitude. In order to represent speech in digital form, it has to be digitized by sampling (i.e., time discretization) and then quantizing (i.e., amplitude discretization) the analog signal. Figure 2.2 shows a general block diagram of a speech coding system composed of an encoder and a decoder. The encoder digitizes the analog speech signal $s(t)$ and performs data compression. The decoder decompresses the encoded data and reconstructs an approximation $\hat{s}(t)$ of the original speech signal.

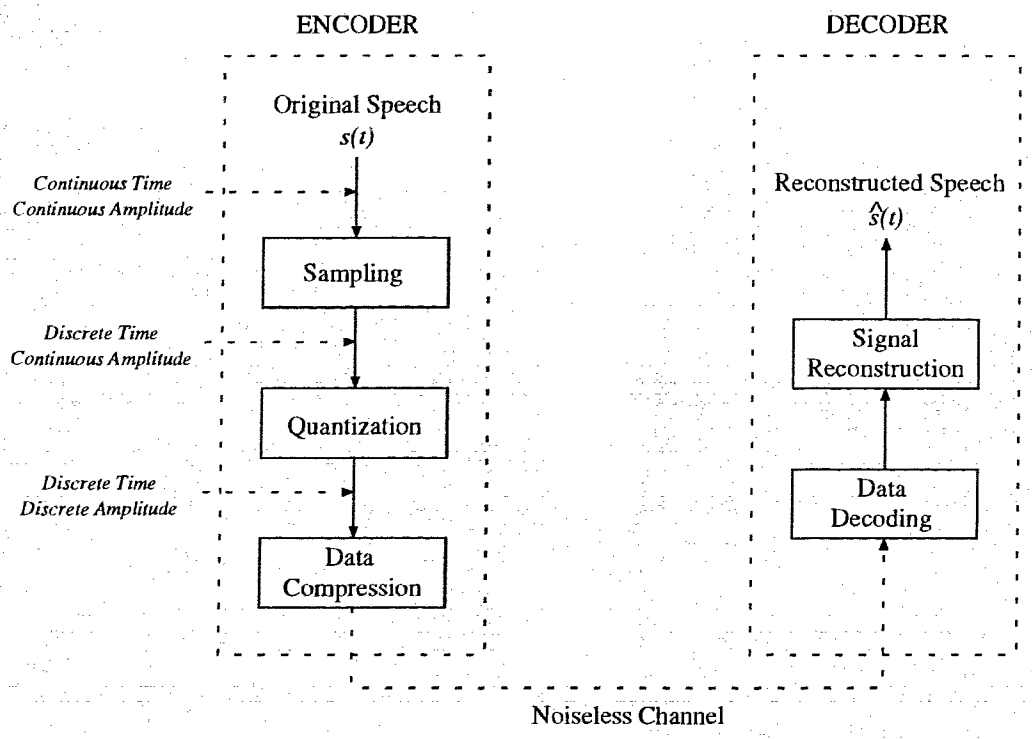


Figure 2.2: Block diagram of a speech coding system

Sampling is a lossless operation (i.e., information are preserved) as long as the conditions of the Nyquist sampling theorem are satisfied. According to the Nyquist sampling theorem [19], if an analog signal is bandlimited to B Hz, then lossless sampling can be achieved when the sampling rate is equal to or higher than $2B$ Hz. The $2B$ Hz sampling rate is often called the Nyquist rate.

In order to represent a discrete time continuous amplitude signal digitally, the amplitude of the signal has to be represented by a finite set of values. Quantization transforms each continuous valued sample into a discrete number. Quantization is inherently a lossy process which results in a loss of information because it attempts to map a continuous value into a finite set of numbers.

There are two major categories of quantization : scalar quantization (SQ) and vector quantization (VQ) [16]. A scalar quantizer observes a scalar value and selects the nearest approximating value from a predetermined finite set of numbers. More precisely, the real axis is mapped into a finite set of real numbers $C = \{y_1, y_2, \ldots, y_L\}$ where L is the size of the quantizer and C is called the codebook of the quantizer.

If the mapping and the input scalar value is denoted by $Q(\bullet)$ and x respectively, the resulting quantized value y is given by

$$y = Q(x) \qquad \text{where } y \in C. \tag{2.1}$$

y_k is chosen as the quantized value of x if it satisfies the Nearest Neighbor condition [16]

$$d(x, y_k) = \min_{y_k \in C} d(x, y_k) \tag{2.2}$$

which states that y_k is selected if the corresponding distortion $d(x, y_k)$ is minimal. For example, the square error $d(x, y_k) = (x - y_k)^2$ is one of the typical criteria. The codebook C can be designed to minimize the average quantization error if the input to the quantizer is a random process with a known probability density function (PDF) or there is a sufficiently large database representing the input random process. The Lloyd-Max algorithm can be used to design such a quantizer [20].

Vector quantization can be considered as a generalization of scalar quantization. In 1948, Claude Shannon observed that coding systems can perform better if they operate on vectors or groups of symbols rather than individual samples or symbols [7]. Instead of restricting the input and the output of the quantizer to be scalar values, a vector quantizer attempts to represent a k-th dimensional vector $\underline{x}$ by a finite set of output vectors $C = \{\underline{y}_1, \underline{y}_2, \ldots, \underline{y}_L\}$ where L is the size and C is the codebook of the vector quantizer. Similar to scalar quantization, the nearest neighbor condition is used to choose a codevector $\underline{y}_k$ that best represents $\underline{x}$ and the generalized Lloyd-Max algorithm [16] can be used to find a codebook for a particular source.

In real-world speech coding applications, a certain amount of distortion is usually allowed as long as a predetermined fidelity requirements such as subjective quality are achieved. Therefore, most of the practical speech coding systems are lossy in nature and they can be broadly classified into the following categories:

- Waveform Coding;
- Analysis-synthesis Coding;
- Analysis-by-synthesis Coding.

2.2.1 Waveform Coding

Waveform coding techniques attempt to faithfully represent and reproduce the speech signal. A digital representation of the input speech is computed which can be used to reproduce the amplitude-*vs*-time waveform as precisely as possible.

The simplest waveform coding technique is Pulse Code Modulation (PCM). This method combines sampling and uniform quantization with μ-law companding to produce digitized speech at 64 kbit/s [21]. μ-law companding is a technique for extending the dynamic range of a quantizer. However, since PCM does not reduce much redundant information in the speech signal, it is wasteful from the transmission bandwidth point of view.

Since speech signal samples are correlated, it is more efficient to quantize the difference between samples or the difference between the actual sample and the predicted sample, the so-called prediction residual [20]. The predicted sample can be computed based on the correlation characteristics of the speech signal. For example, differential pulse code modulation (DPCM) exploits the correlation between samples by quantizing the prediction residual. In DPCM, the prediction residual of an L-th order linear predictor

$$e_n = x_n - \sum_{i=1}^{L} \alpha_i x_{n-i} \tag{2.3}$$

where x_n is the n-th speech sample and $\alpha_i's$ are the prediction coefficients, is computed. Instead of quantizing the original speech sample x_n, e_n is quantized. The statistics of the speech signal actually vary slowly with time. Adaptive DPCM (ADPCM) is a DPCM coder that adapts to this slowly varying statistics by predictor adaptation, quantizer adaptation, or both [20]. ADPCM at 32 kbit/s [21] can achieve speech quality close to that of PCM at 64 kbit/s.

PCM, DPCM, and ADPCM all operate in the time domain. There are techniques that operate in the frequency domain including subband coding and transform coding. In both the subband coding and transform coding, the input signal is divided into a number of separate frequency bands and then each band is encoded separately. These techniques have the advantage that the number of bits used to encode each frequency band can be variable. The encoding accuracy of different frequency bands can then be adjusted according to their perceptual significances. For example, in speech coding, the lower frequency bands are encoded with a larger number of bits because they

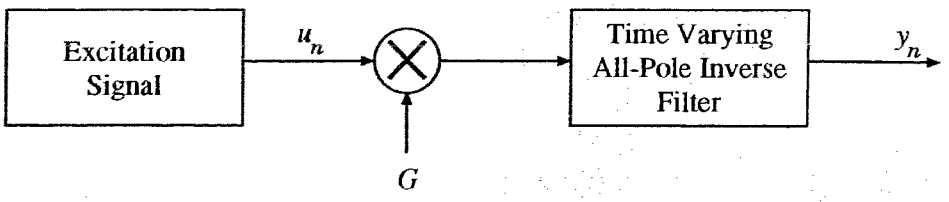


Figure 2.3: Model of the Time-varying Speech Production System

contain pitch and formant structure that must be preserved accurately. On the other hand, a small number of bits can be used to encode the higher frequency bands in which fricative and noise-like sounds occur in speech. In addition, at the cost of higher complexity, adaptive bit allocations can be employed to improve the coder's performance significantly [20].

2.2.2 Analysis-synthesis Coding

Unlike waveform coding, analysis-synthesis (A-S) coding techniques characterize the speech signal by certain perceptually important parameters and synthesize a signal that is perceptually similar to the original signal [5]. Since A-S coding systems only preserve the perceptually significant information, they usually achieve a higher data compression ratio than waveform coding systems.

Generally speaking, the human voice system can be modelled for voiced speech by a time-varying all-pole inverse filter excited by some periodic excitation signals as shown in Figure 2.3. The inverse filter is intended to model the human vocal tract and this is called the time-varying speech production model [22]. In 1974, Markel and Gray used linear prediction to derive the time-varying speech production model and derived the linear predictive coding (LPC) vocoder [23]. In an LPC vocoder, the speech signal is segmented into frames and each frame can be characterized by a small number of parameters. Instead of encoding the actual speech or error samples as in waveform coding, the parameters are encoded. Figure 2.4 shows the block diagram of an LPC vocoder. The encoder computes and quantizes the optimal linear prediction (LP) coefficients (α_i's), the gain factor (G), and the pitch period (k_p) for each speech frame. The decoder decodes the parameters and synthesizes the reconstructed speech using the time-varying speech production model.

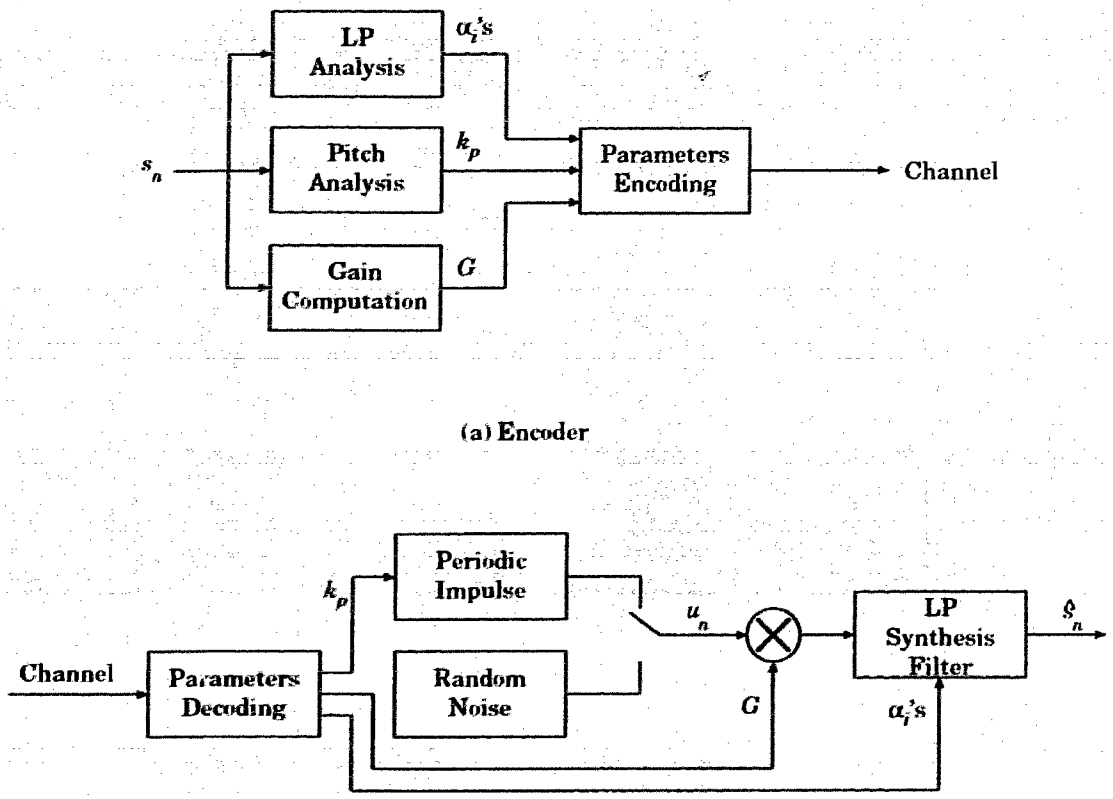


Figure 2.4: Block diagram of an LPC vocoder

In an LPC vocoder, the current speech sample is approximated by a linear combination of previous L samples. If $\hat{x}_n$ is the reconstructed speech sample at instant n and α_i's are the predictor coefficients, then

$$\hat{x}_n = e_n + \sum_{i=1}^{L} \alpha_i \hat{x}_{n-i} \tag{2.4}$$

where e_n is an uncorrelated random variable with zero mean and a variance of σ^2. The above linear predictor can also be considered as a digital filter with an input $\hat{x}_n$, an output e_n, and a system function

$$H(z) = 1 - \sum_{i=1}^{L} \alpha_i z^{-i} \tag{2.5}$$

In other words, the speech signal can be synthesized by exciting the inverse filter $1/H(z)$ with an appropriate excitation signal. Depending on whether the speech signal is voiced or unvoiced, either a sequence of periodic impulses or a sequence of white noises can be used as an excitation signal.

LPC vocoders operate at bit rates between 1.2 to 2.4 kbit/s. A 2.4 kbit/s vocoder can usually reconstruct speech with a high level of intelligibility. However, speech quality, naturalness, and speaker recognizability are all poor [21].

2.2.3 Analysis-by-synthesis Coding

Unlike analysis-synthesis coding systems, analysis-by-synthesis (A-by-S) coding systems determine the signal parameters by an optimization procedure in which synthesized signals are compared with the original signal. Therefore, a better reproduction quality from A-by-S systems can be expected. In fact, A-by-S systems can achieve toll quality at 8 kbit/s [5]. Toll quality is defined as the speech quality required in commercial telephony. In particular, code excited linear prediction (CELP) coders, which are based on the A-by-S structure, provide good reproduction quality at bit rates between 4.0 to 16.0 kbit/s.

Figure 2.5 shows a general configuration of an A-by-S encoder. The excitation codebook contains a collection of excitation vectors and the spectral codebooks contain a collection of long-term and short-term predictor coefficients. The long-term predictor, which models the far-sample correlation or the spectral fine structure of

the speech signal, is represented by

$$\frac{1}{B(z)} = \frac{1}{1 - \sum_{i=k_p-m}^{k_p+m} b_i z^{-i}} \tag{2.6}$$

where $(2m + 1)$ is the order of the long-term predictor, k_p is the pitch period of the speech signal in samples, and b_i's are the long-term predictor coefficients. On the other hand, the short-term predictor models the near-sample correlation or the speech spectral envelope and its system function is

$$\frac{1}{A(z)} = \frac{1}{1 - \sum_{i=1}^{L} \alpha_i z^{-i}} \tag{2.7}$$

where L is the order of the short-term predictor and α_i's are the short-term predictor coefficients. This filter is the same as the linear predictor described in section 2.2.2. Typical values of L are between 10 and 16 assuming an 8 kHz sampling rate [5].

One of the characteristics of an A-by-S encoder is that all the possible reconstructed signals are compared with the original signal in search for the parameters that best represent the original signal. Hence, the decoder is an integral part of the encoder (the shaded area in Figure 2.5). Specifically, for each segment $\underline{x} = (x_n, x_{n+1}, \ldots, x_{n+N})$ consisting of N consecutive speech samples, all the vectors in the excitation codebook with index i and the corresponding scale factor G are used as excitations to the predictors with spectral codebook entries (j, k). The resulting synthesized signals $\underline{y}_{i,j,k,G}$ are then compared with $\underline{x}$. Without the weighting filter, the value of the parameters (i, j, k, G) that minimize the squared error between $\underline{x}$ and $\underline{y}_{i,j,k,G}$ is the best representation of $\underline{x}$. The decoder can use codebooks identical to the encoder to synthesize the speech signal. However, better perceptual results can be obtained by exploiting the auditory masking characteristics of human hearing. The weighting filter is introduced so that the parameters (i, j, k, G) are selected based on a perceptual error criterion [5]. The weighting filter has a system function

$$W(z) = \frac{A(z)}{A(z/\gamma)} \qquad 0 \leq \gamma \leq 1 \tag{2.8}$$

and the value of γ is determined through listening test. The effect of the weighting filter is noise emphasis in high signal energy region and noise deemphasis in low signal energy region. Thus, the actual perceived noise level is reduced.

An exhaustive search of the excitation and the spectral codebooks for all possible (i, j, k, G) results in a huge computational complexity. For example, in 1985,

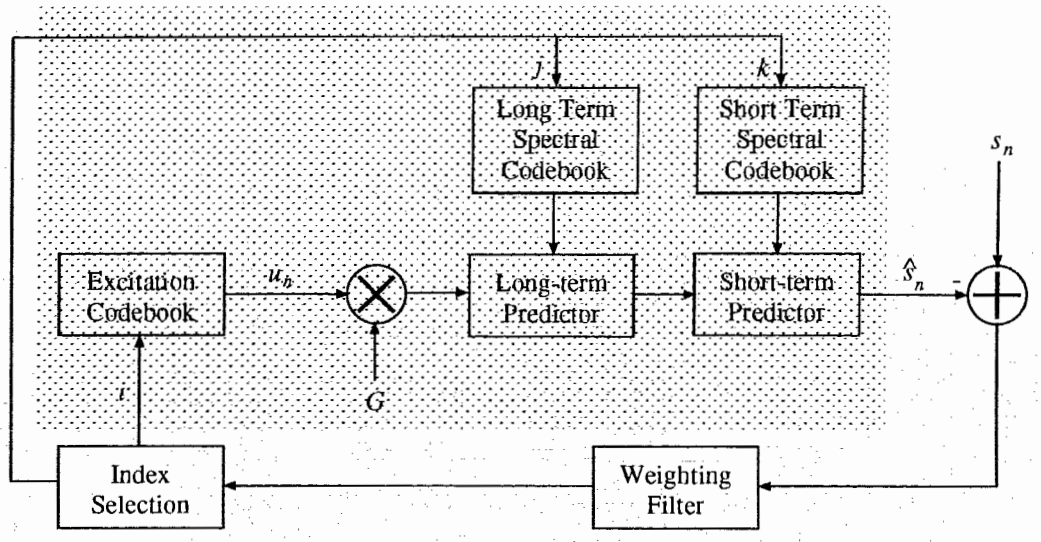


Figure 2.5: Block diagram of an analysis-by-synthesis coding system

Schroeder and Atal indicated that 125 seconds of Cray-1 supercomputer time is required to encode 1 second of a speech signal [24]. This kind of complexity prohibits a real-time implementation of A-by-S systems. Complexity reduction techniques such as ZIR-ZSR decomposition and sequential codebook search [25] can be used to reduce the computational load of the encoder.

2.2.4 Performance Measurements

Traditionally, performance of signal processing systems such as scalar and vector quantizers are measured in terms of the signal-to-noise ratio (SNR). However, for speech coding systems, the ultimate goal is to maximize the subjective quality (i.e., what the human listeners actually perceive) of the reconstructed speech. So far, it has not been possible to objectively quantify the term "speech quality". While some listeners may tolerate one kind of distortion, others may not. In addition, noisy speech which was unsatisfactory in the beginning can become acceptable after repeated listening. However, a combination of objective and subjective measures can still be used to give a good indication of the overall speech quality. In the following sections, some of the commonly used objective and subjective quality measures are presented.

Objective Speech Quality

The simplest and most commonly used objective quality criterion is the signal-to-noise ratio (SNR). If $x(t)$, $y(t)$, and $e(t) = x(t) - y(t)$ are the original signal, the reconstructed signal, and the error signal respectively, the SNR is defined as

$$\text{SNR} = 10\log_{10}(\frac{\sigma_x^2}{\sigma_e^2}) \qquad \text{dB} \tag{2.9}$$

where σ_x^2 and σ_e^2 are the variances of $x(t)$ and $e(t)$ respectively. In practice, the SNR is estimated from the original speech and the reconstructed speech samples as follows:

$$\text{SNR} = 10\log_{10}(\frac{\sum_{n=1}^{N} x_n^2}{\sum_{n=1}^{N} e_n^2}) \qquad \text{dB} \tag{2.10}$$

where N is the number of samples used in the SNR estimation.

Usually, SNR is computed over a long speech database. However, speech signal is a quasi-stationary process and the SNR measure often has a poor correlation with the human perception. This statement is particularly true when regions of poor performance mask regions of better performance. A better assessment of speech quality can be obtained by using the segmental signal-to-noise ratio (SEGSNR) defined as

$$\text{SEGSNR} = \frac{1}{K}\sum_{i=1}^{K} \text{SNR}_i \qquad \text{dB} \tag{2.11}$$

where k is the number of speech frames used for the computation of SEGSNR, and SNR_i is the SNR computed for the i-th frame of N samples with a typical value of 256. Silence frames are eliminated from the computation of SEGSNR because a small amount of noise in a silence frame can give rise to a large negative SNR. A frame is considered to be silent if its average signal power is 40 dB below the average power level of the entire speech database [5].

Subjective Speech Quality

Although SNR and SEGSNR are convenient ways to assess the quality of the reconstructed speech, they do not reliably predict the subjective speech quality. This is particular true for speech coders operating below 16 kbit/s. Alternatively, the quality of the speech coder can be evaluated by formal tests with human listeners. For

example, in the A-B preference test [26], sentence pairs processed by two different algorithms are presented to the listeners in a random fashion. The listeners then decide if they prefer sentence A or sentence B. Results of the A-B preference test are usually presented as a percentage of a preference for a particular algorithm.

On the other hand, a more precise measure is the mean opinion score (MOS) [26]. The MOS is obtained by averaging scores given by a group of untrained listeners. The listeners evaluate the speech quality on a scale of 1 (for poor quality) to 5 (for excellent quality). The reconstructed speech is considered to have toll quality if the MOS is better than 4.0. For example, the CCITT G.711 64 kbit/s PCM coder achieves a MOS of 4.25 while the CCITT G.721 32 kbit/s ADPCM coder achieves a MOS score close to 4.0 [5].

Unfortunately, subjective evaluations of speech quality or intelligibility are very time consuming and expensive. The objective measures such as SNR and SEGSNR are comparably much easier to compute. Thus, although objective quality measures do not always reliably predict the subjective quality, they are still often used for preliminary speech coder design.

2.3 Channel Coding

Channel coding is a technique for improving the reliability of digital data transmissions over a noisy channel. This is usually accomplished by introducing redundancy into the data to be transmitted. A block diagram of a channel coding system is shown in Figure 2.6. The objective of the channel coding system is to reduce the bit error (i.e., error between u_n and $\hat{u}_n$ due to e_n) probability of data transmissions over the noisy discrete channel at the cost of bandwidth expansion. The discrete channel actually composes of a modulator, a demodulator, and a waveform channel. Although the bit error probability can also be reduced by increasing the transmitter's power, it is usually more costly and introduces problems such as additional co-channel and adjacent channel interferences [27].

There exists many channel coding techniques that facilitate two basic objectives at the receiver: error detection and error correction. With error detection, the receiver can detect the presence of errors in the received data and thus an appropriate action can be taken. For example, the receiver can ask for retransmission of corrupted data.

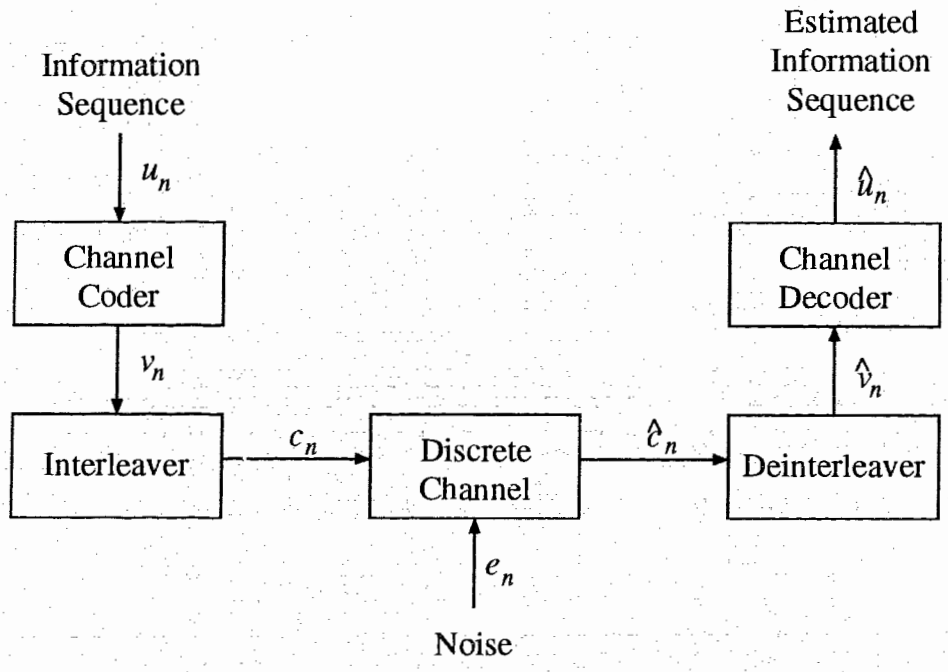


Figure 2.6: Block diagram of a channel coding system

Schemes including pure automatic repeat request (ARQ) and Hybrid ARQ operate on this principle [28, 29]. For speech transmission, delay introduced by retransmission is often unacceptable. However, with error detection, techniques such as frame substitution [30, 31, 32] and frame interpolation/extrapolation [33, 34] can be employed to improve the reconstructed speech quality in case of errors.

In this thesis, we only consider forward error correction (FEC). Compared to increasing transmitter's power, FEC is a relatively inexpensive way of improving communication reliability. Low cost implementations of FEC algorithms are readily available on custom made integrated circuits (i.e., ASICs) [35].

2.3.1 Forward Error Correction

There are two important types of forward error correction codes: linear block codes and convolutional codes [36]. In linear block codes, a block of k data bits is represented by a codeword of n bits where $k < n$. The code rate is defined as $R_c = k/n$ and it indicates the amount of redundancy introduced by the error correction code. A linear block code that accepts k bits as input and produces n output bits is often called

an (n, k) code; and there is always a unique mapping from the k data bits to the n bit codeword. Since the n bit codeword depends only on the the corresponding k data bits, linear block codes are memoryless. There are many important block codes including Hamming codes, Golay codes, Bose-Chaudhuri-Hocquenghem (BCH) codes, and Reed-Solomon codes [37, 38].

In convolutional codes, the n bit codeword depends on both the k data bits and the previous l data bits. Hence, convolutional codes have a memory of order l which is often called the constraint length of the convolutional code. In other words, the mapping from the k bit input to the n bit codeword is not unique. Similar to linear block codes, the ratio $R_c = k/n$ is called the code rate and a convolutional code with l bits of memory that accepts k bits as input and produces an n bit codeword is often called an (n, k, l) code.

For both linear block codes and convolutional codes, there are two types of decoding algorithm: hard decision and soft decision decoding [36]. If the matched filter outputs (i.e., outputs of the demodulator) are quantized into binary levels prior to channel decoding, the decoding process is called hard decision decoding. On the other hand, if a measure of reliability is provided for each received bit to the channel decoder, the decoding process is called soft decision decoding. The reliability information is often the unquantized output of the matched filter. Channel coding systems with soft decision decoding often outperforms those with hard decision decoding by 2 to 3 dB in decoding gain. However, soft decision decoding is substantially harder to implement in a block code decoder than hard decision decoding. Soft decision decoding can be accomplished much easier with convolutional codes than with block codes when the maximum likelihood Viterbi decoding algorithm is used [39, 40, 41].

2.3.2 Interleaving

Most of the well known error-correcting codes are effective when the channel errors are statistically independent. However, mobile channels often exhibit bursty error characteristics. An effective method for dealing with burst errors is to interleave the coded data in such a way that the burst errors are randomized. Thus, a code designed for independent errors can be used.

There are many interleaving/deinterleaving methods. Conceptually, the simplest

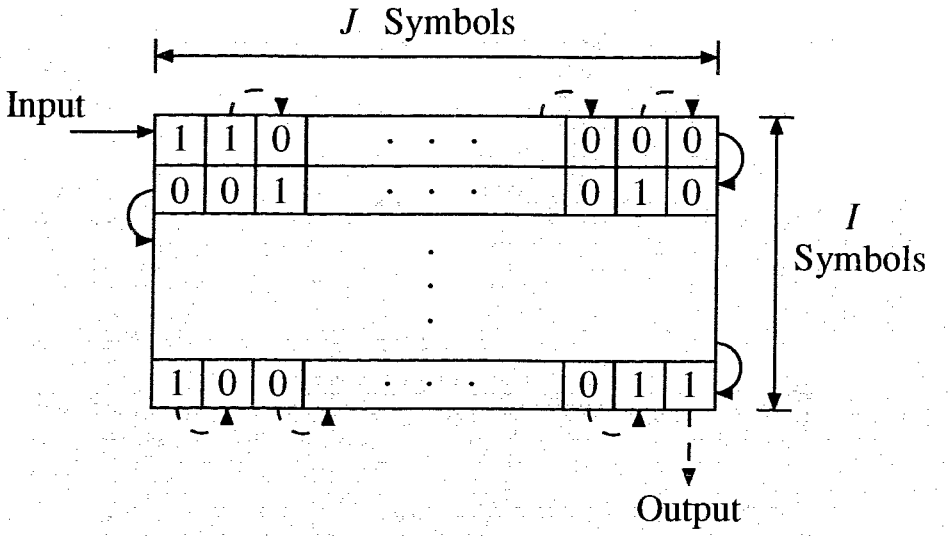


Figure 2.7: Basic concept of block interleaving

one is block interleaving. Inside a block interleaver, the interleaving buffer can be viewed as a rectangular array of storage locations. Figure 2.7 shows the matrix representation of a block interleaver with an interleaving depth of I symbols. In the interleaver, data is written into the interleaving buffer by rows. When the buffer is full (i.e., all the rows are filled up), the data is read out by columns. The deinterleaver performs the inverse operation by writing the received data into a similar buffer by columns and reading them out by rows after the buffer is full. Now, if the communication channel introduces an error burst that corrupts K data symbols where $K \leq I$, the K consecutive errors will cause only single errors after deinterleaving.

It is intuitively obvious that as I gets larger, the errors will tend to be more uncorrelated. However, the delay that results from interleaving is directly proportional to I. An ideal interleaver, which has an infinite interleaving depth, will introduce an infinite delay. For real-time speech communications, an excessive amount of end-to-end delay can cause irritating distortion in the reconstructed speech. Hence, the depth of the interleaver depends on the communication channel, the error correcting code, and the speech coding algorithm.

2.4 Modulation and Multiple Access Techniques

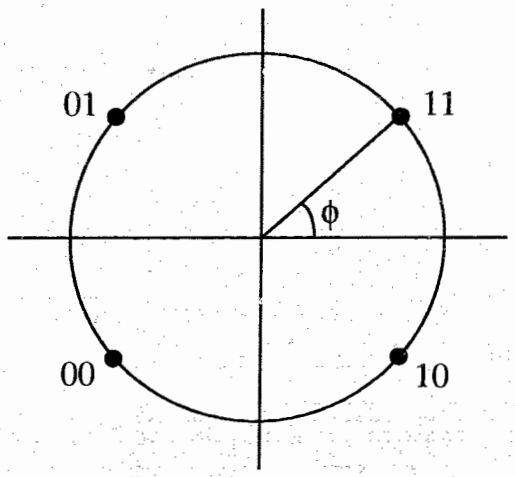
2.4.1 Digital Modulation Schemes

Digital transmission over the mobile environment has become an important research area. Since mobile channels are waveform channels, they cannot be used to transmit the sequence of digital data from the channel coder directly. On the transmitter side, a digital modulator is required to convert digital data into analog waveforms that are compatible with the characteristics of the channel. On the receiver side, a digital demodulator is required to reproduce the digital data sequence from the received waveforms.

Modulated data symbols can be detected either coherently or differentially. For example, in the North American IS-54 standard for digital cellular networks, differentially encoded and detected quadrature phase shift keying (DQPSK) is suggested as the digital modulation format [2]. The advantage of differential detection is that a coherent phase reference at the receiver is not required. Therefore, DQPSK is very robust to the random phase fluctuations introduced by a mobile channel since channel statistics change very little between adjacent data symbols.

On the contrary, although an elaborate method for estimating the carrier phase is required, the performance of coherent detection is significantly better than differential detection. For example, DQPSK is approximately 3 dB poorer in performance than coherent QPSK. In practice, very reliable phase estimates can be obtained by using the pilot symbol technique. Known symbols are inserted into modulated data sequences periodically. By comparing the received known symbols with their true values, the receiver can estimate the channel variations and establish a phase reference for coherent detection. It has been shown that pilot symbol assisted modulation outperforms differential detection under most conditions [42].

In this thesis, we have decided to use coherent QPSK and assumed that the phase distortion can be perfectly estimated at the receiver. The QPSK constellation is shown in Figure 2.8. Theoretically, higher level modulation with better bandwidth efficiency such as 16 level quadrature amplitude modulation (QAM) can be used to improve the system throughput. However, 16-QAM requires too much power which prohibits its use on hand-held portable units [43].



Data Bit		Phase
x[i]	x[i+1]	ϕ
0	0	$-3\pi/4$
0	1	$3\pi/4$
1	0	$-\pi/4$
1	1	$\pi/4$

Figure 2.8: QPSK signal constellation

2.4.2 Multiple Access Methods

In a wireless digital radio network, the multiple access method has a significant impact on how the network is to be organized. Therefore, a multiple access scheme is one of the most important design factors to be considered. For digital systems, there are three basic multiple access schemes, namely the frequency division multiple access (FDMA), the time division multiple access (TDMA), and the code division multiple access (CDMA). In theory, it does not matter whether the spectrum is divided into frequencies, time slots, or codes. However, one may be better suited in a certain communication media than another. For example, FDMA is presently being used in digital residential cordless phone systems, mainly because of its narrowband nature which does not require any anti-multipath equalization [44]. On the other hand, the North American IS-54 [2] and the European GSM digital cellular standard [4] both adopt TDMA as their multiple access scheme. A competing digital cellular system based on CDMA has recently been approved by the Telecommunication Industry Association (TIA) as the North American interim standard IS-95 [3]. In addition, the third generation wireless systems are expected to provide services in voice, video, and data transmission [45]. This leads to new requirements for the design of multiple access schemes such as variable rate transmissions. In this section, all three multiple access schemes are briefly described.

In FDMA, a given frequency band is divided into many frequency channels. Each

user is assigned a separate non-overlapping frequency channel. At the base station, signals from all the users are combined through a common power amplifier. At each individual mobile user, the signal is upconverted to the assigned frequency band. FDMA can be used as a multiple access scheme for both digital and analog communications. Figure 2.9 shows a possible format of FDMA. FDMA is adopted by the advanced mobile phone standard (AMPS), which is the North American first generation wireless system. However, in order to implement variable rate transmissions, FDMA systems will require changes in the radio channel bandwidth dynamically.

In TDMA, there is only a single *narrowband* radio channel with an aggregate transmission rate of R bit/s. The channel is divided into N time slots in which each time slot can allow a bit rate of R/N bit/s. Then, with a multiplexer, a maximum of N users can share the same channel by using their designated time slots. Figure 2.10 shows the operation of a TDMA system. Work has been done to incorporate variable rate services in TDMA systems including the enhanced TDMA (E-TDMA) [46] and the packet reservation multiple access (PRMA) [47]. Both the E-TDMA and the PRMA assign TDMA time slots dynamically to users on request.

In CDMA, there is only a single *wideband* radio channel. Signals from N users are modulated with N different high frequency pseudo-random noise (PN) sequences which are orthogonal (i.e., uncorrelated) to each other. Signals from different users are combined using a common amplifier. In the receiver, since signals from different users are uncorrelated, the signal from the desired user can be obtained by correlating the received signal with the appropriate PN sequence. Figure 2.11 shows the operation of a CDMA system. A variable rate CDMA digital cellular system has been developed and demonstrated by Qualcomm, Inc. [48] and it has been approved by TIA as an alternative interim standard for the second generation wireless systems.

In this thesis, we have assumed a TDMA multiple access scheme similar to the one specified in the IS-54. The total transmission rate of a single TDMA channel is assumed to be 48 kbit/s, in which 20% is used for transmission overhead. Hence, a total of 38.4 kbit/s is available for error protected speech data. The duration of a single TDMA frame is 40 ms and each frame is divided into 12 time slots. We have assumed that each TDMA channel can accommodate either 3 or 4 users. If there are 3 users, each user is assigned 4 time slots, yielding a user transmission rate of 12.8 kbit/s. On the other hand, if there are 4 users, each user is assigned 3 time slots,

yielding a user transmission rate of 9.6 kbit/s. Figure 2.12 indicates the arrangement of user data for the two different user transmission rates.

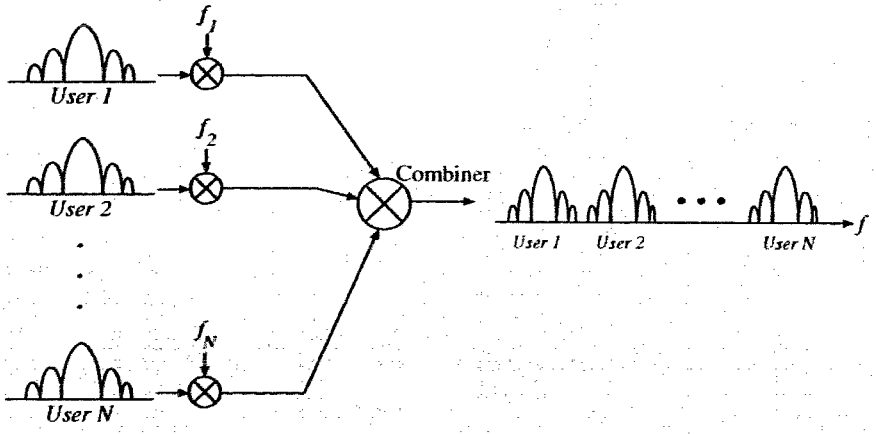


Figure 2.9: Block diagram of a FDMA system

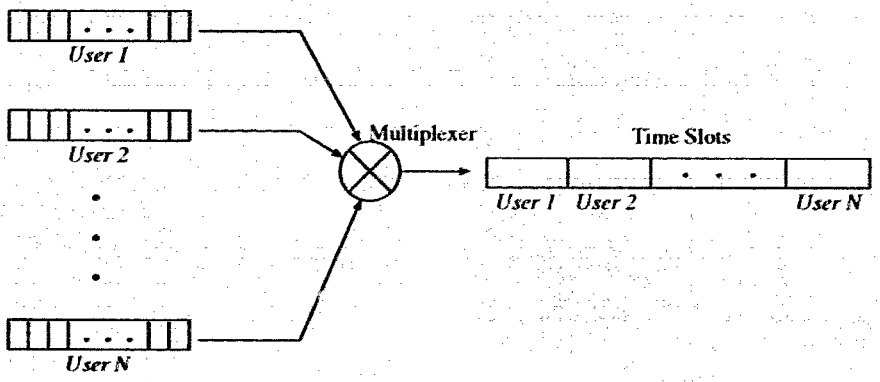


Figure 2.10: Block diagram of a TDMA system

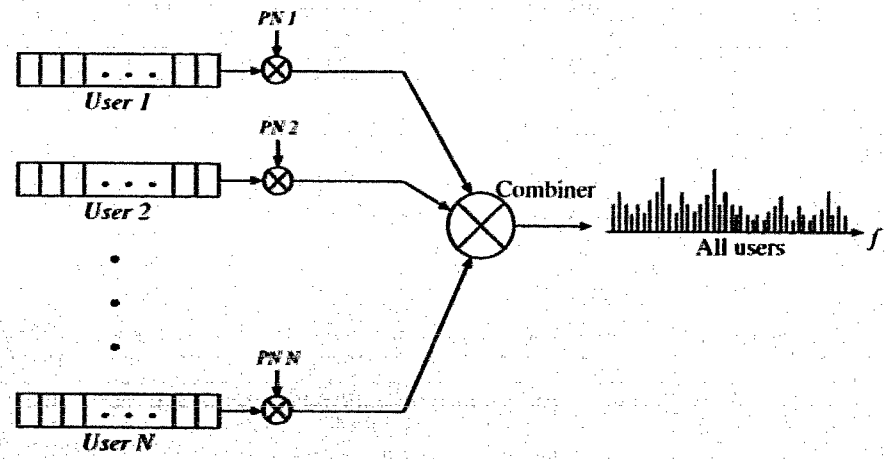
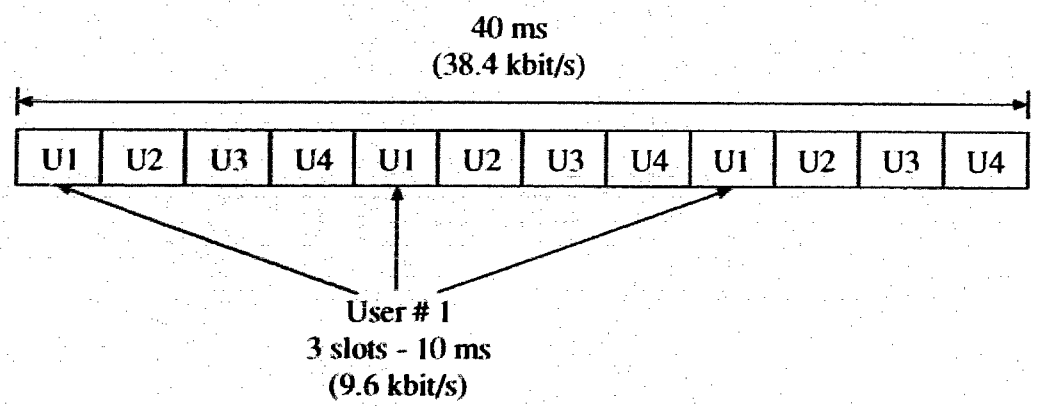
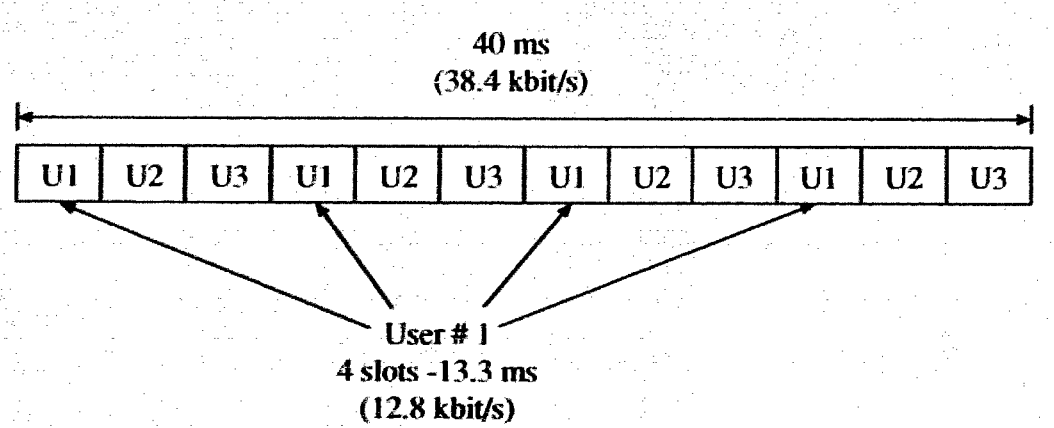


Figure 2.11: Block diagram of a CDMA system



(a) 4 Users per TDMA Channel



(b) 3 Users per TDMA Channel

Figure 2.12: Frame structure of the multiple access scheme

2.5 Model of the Transmission Channel

Typical transmission channels include telephone lines, mobile radio links, microwave links, and satellite links. These channels are subject to various kinds of noise disturbances. For example, on a mobile channel, radio signals are subject to significant propagation-path loss and fading due to shadowing and multipath effects [27]. In addition, man-made disturbances such as noise from a vehicle's electrical ignition system and interference from adjacent channels in a cellular network are not uncommon. The overall effect of the noise disturbances is to introduce an additive and a multiplicative noise component to the transmitted signal. In this section, the mobile fading channel which consists of short-term Rayleigh fading and long-term Log-normal fading is briefly described. Baseband representations are used without loss of generality.

2.5.1 Fading Channel

In the mobile radio environment, the system performance is a function of the radio frequency, vehicular speed, channel bandwidth and geographical location of the vehicle. For example, there are usually a large number of diffuse scatters and/or reflectors which move randomly relative to each other. This results in a multipath fading signal (i.e., short-term fading). In addition, different propagation paths can experience different time delays and signal attenuations. If the time delay spread is small compared to one symbol duration, the short-term fading is frequency non-selective (also called flat fading) [36]. Otherwise, a large delay spread causes frequency selective fading. In this thesis, we only consider flat fading. On the other hand, the average received signal level decreases as the mobile unit moves away from the base station and depends on the terrain configuration between the mobile unit and the base station (i.e., long-term fading).

In general, the received signal is subject to both short-term and long-term fading effects. The baseband equivalent of the received signal can be written as

$$r(t) = d(t)s(t) + n(t) \tag{2.12}$$

where $d(t)$ is the multiplicative distortion (also the received signal envelope) and $n(t)$ is the additive complex Gaussian noise with zero mean and a power spectral density

(PSD) of N_o. Since we assume perfect coherent detection, $d(t)$ is assumed to be a real value random process without loss of generality. $d(t)$ can be further decomposed as

$$d(t) = d_s(t)d_l(t) \tag{2.13}$$

where $d_s(t)$ is a Rayleigh process modelling the short term fading and $d_l(t)$ is a Log-normal process modelling the long term fading.

After receiver matched filtering and sampling, the output of the QPSK coherent demodulator is

$$r_k = d_k s_k + n_k \tag{2.14}$$

where d_k is the fading envelope, s_k is the complex QPSK symbol, and n_k is the white Gaussian noise.

Short-term Fading

Short term fading is caused by the multipath phenomenon as mentioned earlier. The fading amplitude $d_s(t)$, which is the magnitude of the complex short-term fading gain, can be modelled by the Rayleigh fading process [27]. The complex fading gain has a normalized autocorrelation function

$$\rho_\tau = J_0(2\pi f_D \tau) \tag{2.15}$$

where $J_0(\bullet)$ is the modified Bessel function of the first kind and order zero, and f_D is the maximum Doppler frequency (also called the fade rate). The fade rate f_D can be expressed in terms of the vehicle speed v and the carrier frequency f_c as

$$f_D = \frac{v}{c} f_c \tag{2.16}$$

where $c = 3.0 \times 10^8$ is the speed of light. In addition, the normalized fade rate is defined as $f_D T$ where T is duration of a symbol.

Long-term Fading

Long term fading is caused by the propagation path loss and the variation of terrain configuration between the base station and the mobile unit. Instead of representing

the fading process in time, we can also express it in terms of the distance x from the base station. In that case,

$$d(x) = d_s(x)d_l(x) \tag{2.17}$$

Empirical results indicate that the long term fading process $d_l(x)$ follows a Log-normal distribution [27]. That means the PDF of $D_l(x)$, where $D_l(x)$ is the power of $d_l(x)$ expressed in dB, is

$$p(D_l) = \frac{1}{\sqrt{2\pi\sigma^2}} exp\{-\frac{(D_l - P_x)^2}{2\sigma^2}\} \tag{2.18}$$

where P_x is the average power (expressed in dB) of the received signal at a distance x away from the base station. Previous studies have shown that P_x in linear scale follows the inverse 4-th power law which states that

$$P_x = P_0 - 40\log(x) \tag{2.19}$$

where P_0 is the power at a reference distance close to the transmitting antenna but in the far-field. [27].

2.6 Summary

In this chapter, we gave a brief introduction to speech coding, channel coding, and digital modulation. We also described the multiple access scheme and the model of mobile channels considered in this thesis.

CHAPTER 3

VARIABLE RATE SPEECH CODING

In the design of conventional speech coding systems, it is often assumed that the speech coder operates at a fixed rate. However, since speech is non-stationary and the channel capacity is sometimes time-variant, speech coders that operate at variable rates are desirable. In this chapter, code excited linear prediction (CELP) coders based on the analysis-by-synthesis structure are described. The basic structure of our CELP speech coder is first described, followed by a description of our variable rate CELP speech coders.

3.1 CELP algorithm

In the last decade, the development of digital mobile communication systems have progressed significantly. In these systems, voice is still the dominant service provided. Digital transmission of speech can provide a more secure and bandwidth efficient service with good reproduction quality. In particular, code excited linear prediction (CELP) is considered to be a powerful technique in producing high-quality speech at bit rates in the range from 4.0 to 16.0 kbit/s. In fact, the North American IS-54 digital cellular standard employs the vector sum excited linear prediction (VSELP) coder, which is a variant of CELP [2]. On the other hand, the European GSM digital cellular standard adopts the regular-pulse-excitation long term prediction (RPE-LTP) coder, which is one of the many realizations of the analysis-by-synthesis structure [4].

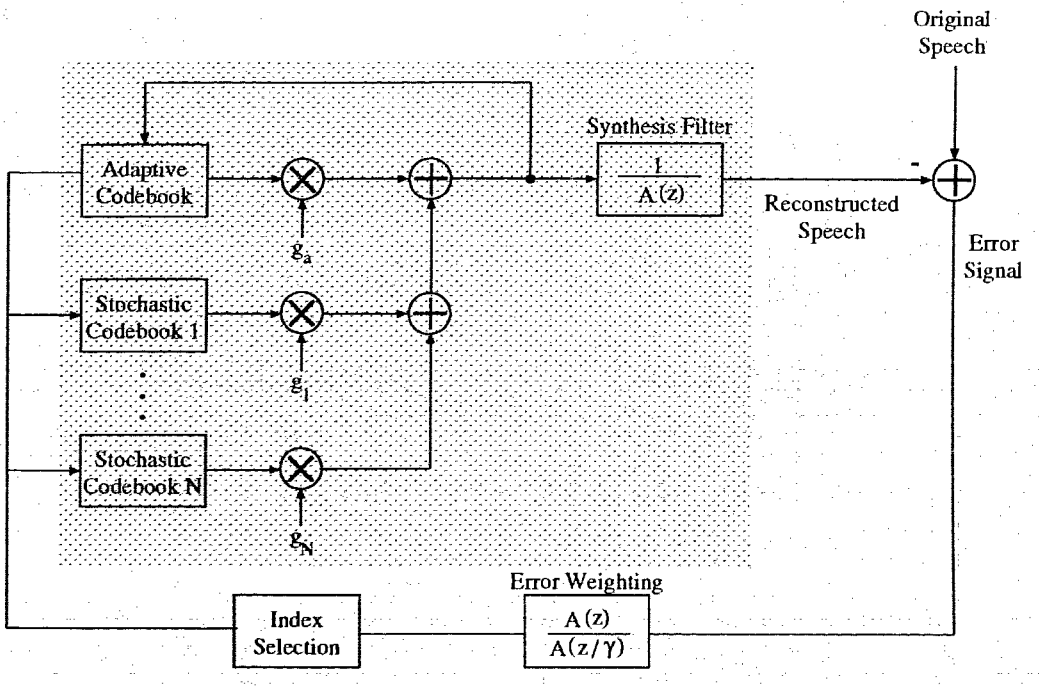


Figure 3.1: Block diagram of a CELP speech coder

The objective of this section is to describe the basic structure of the CELP coder employed in this thesis. A block diagram of a generic CELP encoder structure is shown in Figure 3.1. As mentioned in Section 2.2.3, the CELP speech coder is based on the analysis-by-synthesis structure. One major characteristic of an A-by-S system is that every possible synthesized signal is compared to the original signal in order to obtain the parameters that best represent the original signal. As a result, the decoder, which is shown as the shaded area in Figure 3.1, is an integral part of the encoder.

The reconstructed speech is synthesized by filtering an excitation signal with a synthesis filter based on the short-term linear prediction (LP) coefficients. LP analysis is performed on the original speech signal to determine the LP coefficients. The excitation signal is formed by the weighted sum of the contributions from the adaptive codebook and stochastic codebooks. The adaptive codebook represents the periodicity of the speech signal, which is a realization of the long-term pitch filter. The stochastic codebooks contain random excitation vectors. The codebook indices and the scale factors are determined by an analysis-by-synthesis procedure as described in Section 2.2.3.

The CELP encoder shown in Figure 3.1 is characterized by high computational

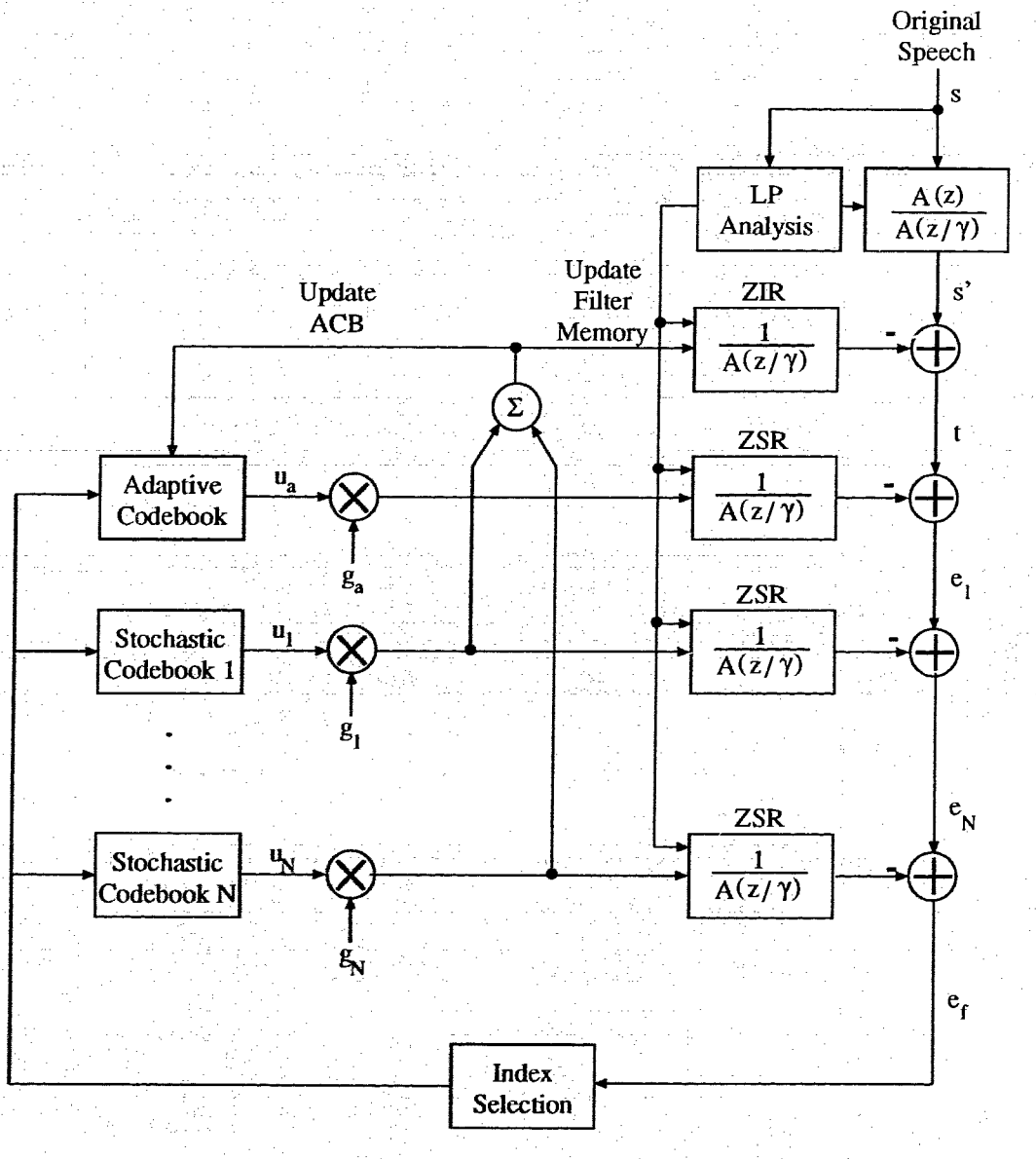


Figure 3.2: Block diagram of a complexity reduced CELP speech coder

complexity. A complexity reduced version is shown in Figure 3.2. This complexity reduced version has three major differences from the high complexity one: combining the synthesis filter and the perceptual weighting filter; decomposing the synthesis filter output into the zero state response (ZSR) and zero input response (ZIR); and searching the codebooks sequentially.

In the complexity reduced coder, instead of filtering the error signal with the weighting filter

$$W(z) = \frac{A(z)}{A(z/\gamma)}, \tag{3.1}$$

both the original speech signal and the excitations are filtered by $W(z)$. The combined synthesis and weighting filter has a transfer function

$$H_c(z) = \frac{1}{A(z)} \frac{A(z)}{A(z/\gamma)} = \frac{1}{A(z/\gamma)}. \tag{3.2}$$

In addition, the combined filter response is separated into the zero input response (ZIR) and zero-state response (ZSR), a technique known as ZIR-ZSR decomposition [25]. To further reduce the complexity of the coder, the best codebook vectors and the corresponding gain values are determined sequentially. Typically, the best adaptive codebook vector is determined first, followed by the stochastic codebooks. However, this sequential search method is suboptimal. A simple improvement can be achieved by re-optimizing the gain values after the ACB and the SCB vectors are determined [5].

The input speech is processed on a frame by frame basis. The CELP encoder divides the input speech samples into a series of analysis frames which are further divided into a number of subframes. LP analysis is performed once every analysis *frame* while the excitation parameters are determined once every *subframe*. In the following sections, we describe in detail the various major components of the complexity reduced CELP encoder.

3.1.1 Linear Prediction Analysis and Quantization

For every analysis frame, an L-th order LP analysis is performed to obtain the optimal LP coefficients, which are used to construct the L-th order short-term synthesis filter. The autocorrelation method is used to obtain the LP coefficients [22]. If $r_i(\tau)$ denotes the estimate of the autocorrelation function of the i-th analysis frame at lag τ and

α_j denotes the j-th LP coefficient, the following system of equations can be used to solve for the optimal LP coefficients:

$$\sum_{j=1}^{L} \alpha_j r_i(|j-k|) = r_i(k) \tag{3.3}$$

for $k = 1, 2, \ldots, L$. This is a system of L linear equations with L unknowns (i.e., α_j's) which are called the Wiener-Hopf or Yule-Walker equations. This system of equations can be solved efficiently using the Levinson-Durbin algorithm [16].

Before quantization and transmission, the LP coefficients α_j's are converted into line spectral pair (LSP) parameters. The LSP parameters have better quantization properties than LP coefficients [50]. Tree searched multi-stage vector quantization (TS-MSVQ) is used to quantize the LSP parameters [51]. TS-MSVQ is a novel technique to reduce the complexity of the quantization procedure.

In order to ensure a smooth transition between adjacent sets of LP coefficients, the coefficients are interpolated in the LSP domain. Linear interpolation is done every subframe as follows:

$$\mathrm{LSP}_i(j) = (1 - \frac{i}{N_{sub}})\mathrm{LSP}_p(j) + \frac{i}{N_{sub}}\mathrm{LSP}_c(j) \tag{3.4}$$

where N_{sub} is the total number of subframes in an analysis frame
$\mathrm{LSP}_i(j)$ is the j-th LSP for the i-th subframe
$\mathrm{LSP}_p(j)$ is the j-th LSP of the previous analysis frame
$\mathrm{LSP}_c(j)$ is the j-th LSP of the current analysis frame.

3.1.2 Adaptive and Stochastic Codebooks Search

As described earlier, the excitation signal is formed by the sum of the scaled contributions from the adaptive codebook (ACB) and stochastic codebooks (SCBs). In the following sections, the structure of the ACB and SCBs are described, followed by a description on best codevector/gain selection.

Adaptive Codebook

The adaptive codebook is used to realize a single tap closed-loop pitch predictor. ACB is composed of past excitation vectors and uses the pitch delay k_p as an index in the

codebook. This approach was introduced by Singhal and Atal [52] and then further developed by Rose and Barnwell [53]. For a given k_p, the ACB generates excitation samples of the form

$$\underline{u}_a(n) = g_a \underline{u}_p(n - k_p) \tag{3.5}$$

where $\underline{u}_a$ is the ACB excitation, g_a is the optimal gain value for the given k_p, and $\underline{u}_p$ is the past excitation. Typical values of k_p ranges from 20 to 147 samples [5].

Stochastic Codebook

In a full complexity CELP encoder, the stochastic codebook(s) can be generated using a random Gaussian number generator with zero mean and unit variance. However, in our complexity reduced system, a special form of SCB is used containing sparse and overlapped shift by -2 codevectors with ternary-valued samples (i.e., -1, 0, 1). Figure 3.3 shows the structure of the SCB. As indicated in the diagram, if the j-th SCB vector is defined as $\underline{u}_j$, then the $(j-1)$-th vector is defined recursively as

$$\underline{u}_{j-1}(i) = \underline{u}_j(i-2) \tag{3.6}$$

for $i = 3, \ldots, k$ where k is the dimension of the excitation vector. $\underline{u}_{j-1}(1)$ and $\underline{u}_{j-1}(2)$ are found by shifting in the next two samples of the codebook. The sparsity of the codebook is defined as the percentage of zero entries in the codevectors. In our complexity reduced system, the sparsity is 77%. The SCB can be generated in three steps:

- Generate an SCB using a random Gaussian number generator;
- Center clip the SCB (i.e., set the value of an entry to zero if its magnitude is smaller than a threshold) with a threshold determined by the desired sparsity of the SCB;
- Quantize the SCB to three levels : -1, 0, and 1.

The special SCB structure allows end point correction as explained below and thus reduces the computational load significantly [54]. If the filter's ZSR $\underline{y}_j$ due to an excitation $\underline{u}_j$ has been determined, $\underline{y}_{j-1}$ due to $\underline{u}_{j-1}$ can be found by first shifting $\underline{y}_j$ by two as follows:

$$\underline{y}_{j-1}(i) = \underline{y}_j(i-2) \tag{3.7}$$

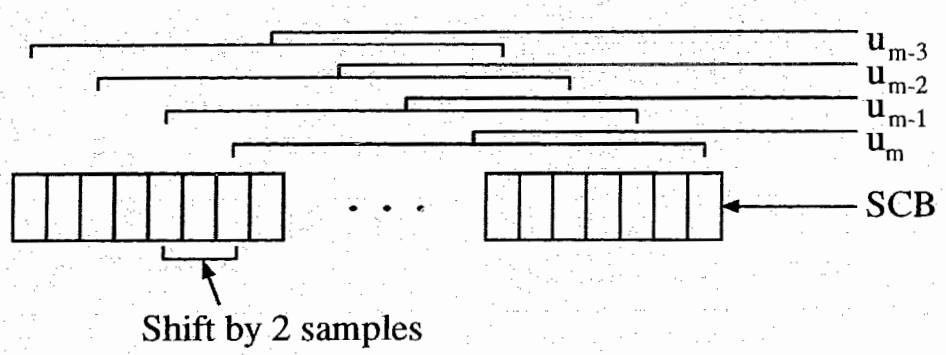


Figure 3.3: Structure of an overlapped shift stochastic codebook

for $i = 3, \ldots, k$, and setting $\underline{y}_{j-1}(1)$ and $\underline{y}_{j-1}(2)$ to zero. The filter's correct ZSR is then obtained by convolving only the new excitation samples $\underline{u}_{j-1}(1)$ and $\underline{u}_{j-1}(2)$ with the filter's impulse response and adding the results to $\underline{y}_{j-1}$ found in (3.7). Hence, with end point correction, only two convolutions are required for each codebook vector (except the first vector). In addition, because of the sparsity of the SCB, $\underline{u}_{j-1}(1)$ and $\underline{u}_{j-1}(2)$ are zero most of the time. In those cases, no convolutions are required. Therefore, a significant amount of computations are saved.

Computation of Best Excitation

As mentioned earlier, the best excitation vectors are determined sequentially. Usually, the best ACB vector is determined first, followed by the SCBs. For the ACB, the best excitation is determined as follows.

Since $H_c(z)$ is a linear filter, its output response $\underline{y}^i$ as a result of an ACB excitation vector $\underline{u}_a^i$ can be separated into the ZSR and ZIR

$$\underline{y}^i = \underline{y}_{zir} + g_a^i \underline{y}_{zsr}^i \tag{3.8}$$

where $\underline{y}_{zir}$ is the ZIR response, $\underline{y}_{zsr}^i$ is the ZSR response due to the excitation $\underline{u}_a^i$, and g_a^i is the corresponding gain. The ZIR is the output of the synthesis filter with zero input while the ZSR is the output of the synthesis filter with an excitation and zero filter memory (i.e., zero initial conditions). Thus, the ZIR is independent of i while each $\underline{u}_a^i$ will result in a different ZSR.

If we define an $M \times M$ matrix H as

$$H = \begin{bmatrix} h(1) & 0 & \dots & 0 \\ h(2) & h(1) & \dots & 0 \\ \dots & \dots & \dots & 0 \\ h(M) & h(M-1) & \dots & h(1) \end{bmatrix} \tag{3.9}$$

where M and $h(n)$ are the order and the impulse response of the synthesis filter respectively, (3.8) can be written as

$$\underline{y}^i = \underline{y}_{zir} + g_a^i H \underline{u}_a^i. \tag{3.10}$$

If we define a vector $\underline{t} = \underline{s}' - \underline{y}_{zir}$ in which $\underline{s}'$ is a subframe of speech samples filtered by $W(z)$, the optimization problem becomes finding the $\underline{u}_a^i$ and g_a^i that minimizes the square error

$$\epsilon = \|\underline{t} - g_a^i H \underline{u}_a^i\|^2. \tag{3.11}$$

By differentiating (3.11) with respect to g_a^i and equating the resulting expression to zero, the gain value for a particular codebook vector can be found to be

$$g_a^i = \frac{\underline{t}^T H u_a^i}{\|H u_a^i\|^2} \tag{3.12}$$

and (3.11) becomes

$$\epsilon = \|\underline{t}\|^2 - \frac{(\underline{t}^T H u_a^i)^2}{\|H u_a^i\|^2}. \tag{3.13}$$

The first term in the above equation does not depend on the index i. Hence, the optimization problem is equivalent to maximizing

$$\epsilon = \frac{(\underline{t}^T H u_a^i)^2}{\|H u_a^i\|^2}. \tag{3.14}$$

The $g_a^i \underline{u}_a^i$ that maximizes (3.14) is the best excitation.

After the best ACB excitation is determined, a new target vector $\underline{e}_1 = \underline{t} - g_a^{opt} H \underline{u}_a^{opt}$ is used for selecting the first stage best SCB excitation. The latter SCB stages' best excitations are then obtained in the same fashion. After all the ACB and SCB best excitations are determined, the gain values are reoptimized to obtain a smaller ϵ. For example, if there are one ACB and one SCB, the encoder's objective is to minimize

$$\epsilon = \|\underline{t} - g_a^{opt} H \underline{u}_a^{opt} - g_1^{opt} H \underline{u}_1^{opt}\|^2. \tag{3.15}$$

By equating the partial derivatives of (3.15) with respect to g_a^{opt} and g_1^{opt} to zero, a linear system of two equations with two unknowns is obtained. By solving the system of equations, the gains that minimize (3.15) can be obtained.

3.1.3 Gain Quantization

After the codebook gains are obtained, they need to be quantized. In order to reduce the complexity, we use a shape-gain/multistage vector quantizer (VQ) to jointly quantize the codebook gains. In a shape-gain VQ, the source vector $\underline{x}$ is normalized by a gain value g_{norm} to obtain a normalized source vector $\underline{x}_{norm}$. Thus, the problem becomes quantizing the scalar gain value g_{norm} and the shape vector $\underline{x}_{norm}$. The operation of a shape-gain VQ is shown in Figure 3.4. The reason for using a shape-gain VQ is that the same pattern of variation in a vector may recur with a wide variety of gain values. This technique enables the VQ to handle a source with a wide dynamic range without having a huge VQ codebook.

The VQ for quantizing the normalized codebook gain vectors is implemented using a multistage VQ. In a multistage VQ, the quantization task is divided into successive stages. The first stage performs a relatively crude quantization of the source vector. Then, the second stage quantizer operates on the error vector between the original and the quantized first stage vector and so on. The reconstructed vector is the sum of the quantized outputs of all the stages. The operation of a two-stage VQ is illustrated in Figure 3.5.

Assuming that there is one ACB and one SCB in the CELP encoder, the codebook gain vector can be defined as $\underline{G} = \{g_a^{opt}, g_1^{opt}\}$. The corresponding normalization factor for each component of $\underline{G}$ is defined as [55]

$$\begin{aligned} g_{norm} &= \frac{\|\underline{t}\|^2}{(\sqrt{N_s})(\|\underline{u}\|^2)(g_{filter})} \\ &= \frac{g_{frame}}{(\|\underline{u}\|^2)(g_{filter})} \end{aligned} \tag{3.16}$$

where $\underline{t}$ and $\underline{u}$ are the target vector and the excitation vector defined in section 3.1.2 respectively, N_s is the number of samples in a subframe, and g_{filter} is the gain of the synthesis filter. Since the values of all the parameters in (3.16) except g_{frame} are known in both the encoder and the decoder, only g_{frame} is needed to be quantized and transmitted. The g_{frame} is quantized in the logarithmic domain by a scalar quantizer.

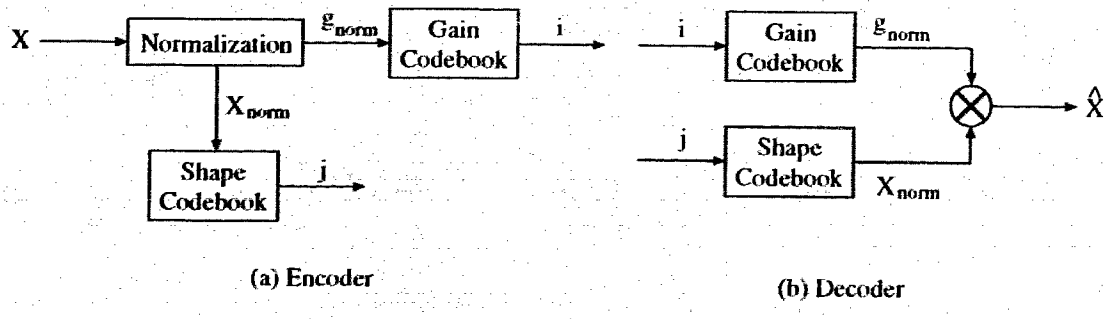


Figure 3.4: Block diagram of a shape-gain VQ

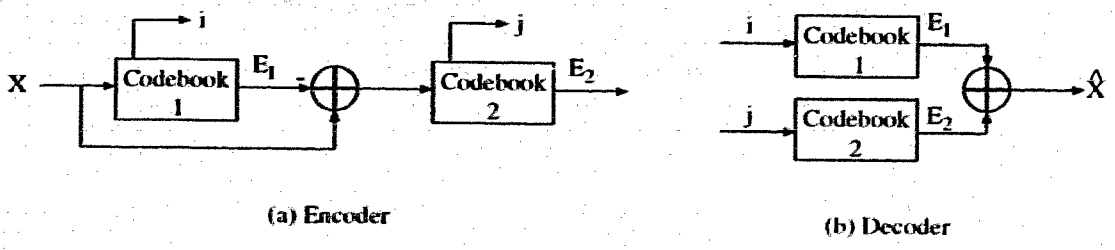


Figure 3.5: Block diagram of a two stage VQ

Then, the normalized gain vector $\underline{G}_{norm}$ is quantized by the multistage shape gain quantizer. In our complexity reduced system, the maximum number of stages in the shape-gain quantizer is two. However, complexity can be further reduced by having more stages. The reconstructed vector (i.e., found by summing the quantized output of the first and second stage codebooks) that best represents $\underline{G}_{norm}$ is found using an $M-L$ search technique. Starting with the first stage codebook, the M codevectors which achieve the lowest errors using the minimum-square-error (MSE) criterion

$$\epsilon = \|\underline{G}_{norm} - \underline{\hat{G}}_i\|^2 \tag{3.17}$$

in which $\underline{\hat{G}}_i$'s are the codevectors, are selected and M error vectors are computed. The second stage codebook is searched M times. For each codebook search, the L codevectors which achieve the lowest MSEs are selected. In the end, there are a total of $(M \times L)$ vectors formed by summing the M codevectors from the first stage with the corresponding L codevectors from the second stage. The best representation of the $\underline{G}_{norm}$ is selected among the $(M \times L)$ vectors using the error criterion

$$\epsilon = \|\underline{t} - \hat{g}_a^i H \underline{u}_a^{opt} - \hat{g}_1^i H \underline{u}_1^{opt}\|^2 \tag{3.18}$$

where the $\underline{u}$'s are the best ACB/SCB excitations and $\{\hat{g}_a^i, \hat{g}_1^i\}$ is the i-th vector among the $(M \times L)$ vectors.

3.1.4 Updating ACB and Filter Memory

After the best excitation is determined, the speech signal can be synthesized with the following parameters

$$(\underline{u}_a^{opt}, \underline{u}_1^{opt}, \hat{g}_{frame}, \hat{g}_a^{opt}, \hat{g}_1^{opt}) \tag{3.19}$$

where $\underline{u}_a^{opt}$ and $\underline{u}_1^{opt}$ are the best ACB and SCB excitation vectors

$\hat{g}_{frame}$ is the quantized value of g_{frame}

$\hat{g}_a^{opt}$ and $\hat{g}_1^{opt}$ are the quantized value of the ACB and SCB gains.

The best excitation is then given by

$$\underline{u}^{opt} = \frac{\hat{g}_{frame}}{g_{filter}} \left(\frac{\hat{g}_a^{opt}}{\|\underline{u}_a^{opt}\|^2} \underline{u}_a^{opt} + \frac{\hat{g}_1^{opt}}{\|\underline{u}_1^{opt}\|^2} \underline{u}_1^{opt} \right). \tag{3.20}$$

The ACB memory is updated by replacing the old ACB memory with (3.20). In addition, (3.20) is used to excite the synthesis filter for updating the filter's memory.

3.2 Variable Rate CELP Coders

In the design of conventional speech coding systems, it is often assumed that the speech coder operates at fixed rate. However, this approach does not exploit two important characteristics of speech signals: speech pauses and large variance in the short-term entropy of the speech signal due to its non-stationary characteristic. In normal telephone conversations, it has been found that about 60% of the time is spent in silence [56]. These intermittent and entropy-varying properties can often be utilized to design speech communication systems, such as digital cellular networks, more efficiently. Moreover, in certain applications, the channel capacity is actually variable, such as in packet networks. Hence, speech coders that operate at variable rates are desirable.

Variable rate speech coders can be divided into two distinct categories:

- *Source-controlled* variable rate speech coders, where the coding rate is a function of the time-varying activities of the speech signal;
- *Network-controlled* variable rate speech coders, where the coding rate is determined by an external control signal, such as one generated by the communication network in response to the network traffic condition.

There are two types of network-controlled variable rate coders: multimode and embedded variable rate coders. For multimode coders, a different mode of encoding is performed for each bit rate option. Each encoding mode may have different bit allocations or even entirely distinct coding algorithms. In embedded coders, a single coding algorithm generates a fixed-rate data stream from which one of the several reduced rate data signals can be extracted by a simple bit-dropping procedure. Thus, lower rate signals are embedded in the higher rate data bit stream.

In this thesis, our interest is in network-controlled variable rate speech coders. Based on CELP, we have constructed both a multimode and an embedded variable rate coder. They are described in detail in the following sections.

3.2.1 Multimode CELP coder

In this thesis, we have built a multimode variable rate CELP coder based on a modular CELP coder that can be adjusted to operate at 12.8, 9.6, 8.0, and 5.0 kbit/s. Different bit rates are obtained by using different bit allocations. However, the short term filter and adaptive codebook remain the same regardless what the bit allocation is. Figure 3.6 shows the structure of the multimode variable rate coder. An external signal is sent to the bit rate control unit and then an appropriate coder is selected. The bit allocations of the multimode coder at different bit rates are shown in Table 3.1 to Table 3.4.

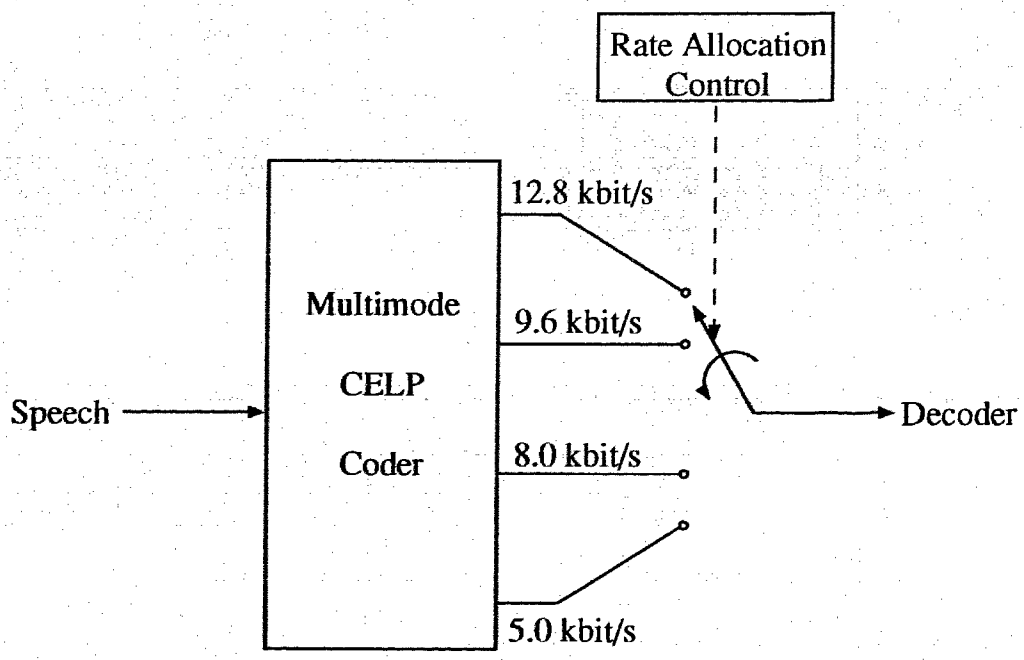


Figure 3.6: Block diagram of the multimode CELP coder

Table 3.1: 12.8 kbit/s CELP (multimode) configuration

	LP/g_{frame}	ACB/SCB	Gain VQ
Update	20 ms	20/4 = 5 ms	
Parameters	10 LSPs/ 1 gain	1 pitch for ACB (pitch: 20 to 147) 6 SCB stages (5,5,5,4,5,5 bits)	1 ACB gain 6 SCB gain
Method	Autocorrelation/ Scalar Quantization	Full Search	*1st gain VQ:* ACB + 3 SCB gains 2 stages (5,5 bits) *2nd gain VQ* 3 SCB gains 2 stages (4,6 bits)
Bits Per Frame	LP: 24 g_{frame}: 8	ACB index: 28 SCB indices: 116	1st gain VQ: 40 2nd gain VQ: 40
Rate	1600	7200	4400
TOTAL = 12800 bit/s			

Table 3.2: 9.6 kbit/s CELP (multimode) configuration

	LP/g_{frame}	ACB/SCB	Gain VQ
Update	20 ms	20/4 = 5 ms	
Parameters	10 LSPs/ 1 gain	1 pitch for ACB (pitch: 20 to 147) 4 SCB stages (5,5,5,5 bits)	1 ACB gain 4 SCB gain
Method	Autocorrelation/ Scalar Quantization	Full Search	*1st gain VQ:* ACB + 4 SCB gains 2 stages (7,6 bits)
Bits Per Frame	LP: 24 g_{frame}: 8	ACB index: 28 SCB indices: 80	1st gain VQ: 52
Rate	1600	5400	2600
TOTAL = 9600 bit/s			

Table 3.3: 8.0 kbit/s CELP (multimode) configuration

	LP/g_{frame}	ACB/SCB	Gain VQ
Update	20 ms	20/4 = 5 ms	
Parameters	10 LSPs/ 1 gain	1 pitch for ACB (pitch: 20 to 147) 3 SCB stages (5,5,5 bits)	1 ACB gain 3 SCB gain
Method	Autocorrelation/ Scalar Quantization	Full Search	*1st gain VQ:* ACB + 3 SCB gains 2 stages (5,5 bits)
Bits Per Frame	LP: 24 g_{frame}: 8	ACB index: 28 SCB indices: 60	1st gain VQ: 40
Rate	1600	4400	2000
TOTAL = 8000 bit/s			

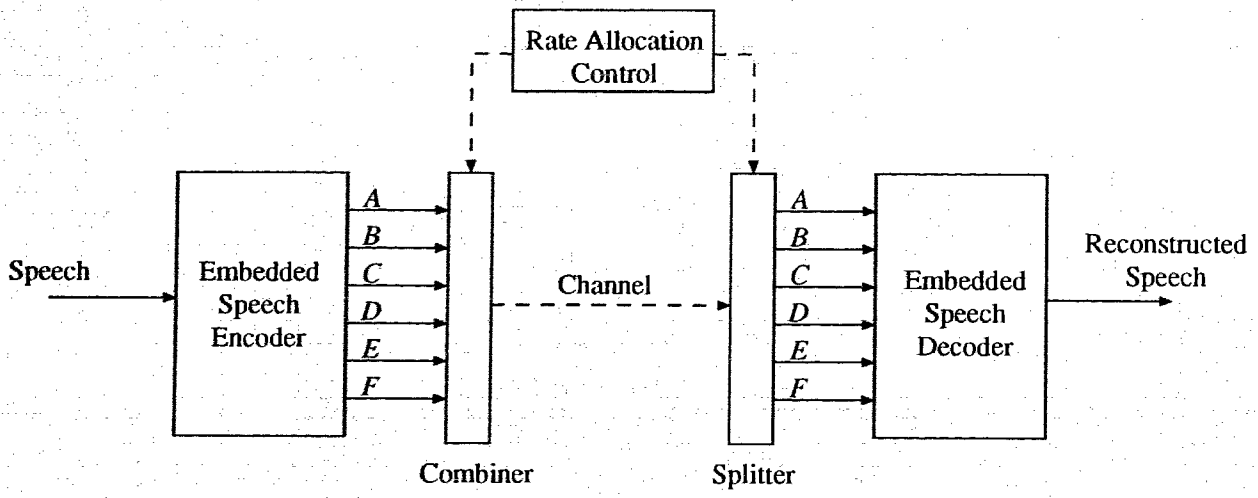
Table 3.4: 5.0 kbit/s CELP (multimode) configuration

	LP/g_{frame}	ACB/SCB	Gain VQ
Update	20 ms	20/4 = 5 ms	
Parameters	10 LSPs/ 1 gain	1 pitch for ACB (pitch: 20 to 147) 1 SCB stage (5 bits)	1 ACB gain 1 SCB gain
Method	Autocorrelation/ Scalar Quantization	Full Search	*1st gain VQ:* ACB + 1 SCB gain 1 stages (5 bits)
Bits Per Frame	LP: 24 g_{frame}: 8	ACB index: 28 SCB indices: 20	1st gain VQ: 20
Rate	1600	2400	1000
TOTAL = 5000 bit/s			

3.2.2 Embedded CELP coder

In this thesis, we have constructed an embedded variable rate CELP speech coder that can operate at 12.8, 9.6, 8.0, 7.0, 6.0, and 5.0 kbit/s. The embedded coder generates a fixed rate data stream at 12.8 kbit/s. Lower rate data streams can be obtained by dropping some bits from the fixed rate data stream. In other words, a lower rate data stream is embedded into a higher rate data stream. Therefore, a communication system with an embedded speech coder can drop bits without informing the encoder. For example, in case of a network overload, the network can decide to reduce the network load by dropping some of the bits. This is illustrated in Figure 3.7.

The basic structure of the CELP coders used in the multimode and embedded coders are identical. In the multimode coder, all the ACB and SCB excitations are used for updating the ACB memory and the short-term filter memory. On the other hand, in the embedded coder, only the excitations required for obtaining the lowest rate encoded speech (i.e., 5.0 kbit/s) are used for updating the memories. Thus, although the embedded encoder always encodes the speech signal at 12.8 kbit/s, the decoder's memories will not be different from those in the encoder as long as the 5.0 kbit/s data are received. The receiver can enhance the quality of the reconstructed speech by using information in additional to those in the 5.0 kbit/s data stream. The bit allocations of the embedded coder at different bit rates are shown in Table 3.5 to Table 3.10.



Available Data	Speech Coding Rate
A	5.0 kbit/s
A,B	6.0 kbit/s
A,B,C	7.0 kbit/s
A,B,C,D	8.0 kbit/s
A,B,C,D,E	9.6 kbit/s
A,B,C,D,E,F	12.8 kbit/s

Figure 3.7: Block diagram of the embedded CELP coder

Table 3.5: Bit allocation of the 12.8 kbit/s embedded CELP coder

	LP/g_{frame}	ACB/SCB	Gain VQ
Update	20 ms	20/4 = 5 ms	
Parameters	10 LSPs/ 1 gain	1 pitch for ACB (pitch: 20 to 147) 6 SCB stages (5,5,5,4,5,5 bits)	1 ACB gain 6 SCB gain
Method	Autocorrelation/ Scalar Quantization	Full Search	*1st gain VQ:* ACB + 3 SCB gains 2 stages (5,5 bits) *2nd gain VQ* 3 SCB gains 2 stages (4,6 bits)
Bits Per Frame	LP: 24 g_{frame}: 8	ACB index: 28 SCB indices: 116	1st gain VQ: 40 2nd gain VQ: 40
Rate	1600	7200	4400
TOTAL = 12800 bit/s			

Table 3.6: Bit allocation of the 9.6 kbit/s embedded CELP coder

	LP/g_{frame}	ACB/SCB	Gain VQ
Update	20 ms	20/4 = 5 ms	
Parameters	10 LSPs/ 1 gain	1 pitch for ACB (pitch: 20 to 147) 4 SCB stages (5,5,5,4 bits)	1 ACB gain 4 SCB gain
Method	Autocorrelation/ Scalar Quantization	Full Search	*1st gain VQ:* ACB + 3 SCB gains 2 stages (5,5 bits) *2nd gain VQ* 1 SCB gain 1 stage (4 bits)
Bits Per Frame	LP: 24 g_{frame}: 8	ACB index: 28 SCB indices: 80	1st gain VQ: 40 2nd gain VQ: 16
Rate	1600	5200	2800
TOTAL = 9600 bit/s			

Table 3.7: Bit allocation of the 8.0 kbit/s embedded CELP coder

	LP/g_{frame}	ACB/SCB	Gain VQ
Update	20 ms	20/4 = 5 ms	
Parameters	10 LSPs/ 1 gain	1 pitch for ACB (pitch: 20 to 147) 3 SCB stages (5,5,5 bits)	1 ACB gain 3 SCB gain
Method	Autocorrelation/ Scalar Quantization	Full Search	*1st gain VQ:* ACB + 3 SCB gains 2 stages (5,5 bits)
Bits Per Frame	LP: 24 g_{frame}: 8	ACB index: 28 SCB indices: 60	1st gain VQ: 40
Rate	1600	4400	2000
TOTAL = 8000 bit/s			

Table 3.8: Bit allocation of the 7.0 kbit/s embedded CELP coder

	LP/g_{frame}	ACB/SCB	Gain VQ
Update	20 ms	20/4 = 5 ms	
Parameters	10 LSPs/ 1 gain	1 pitch for ACB (pitch: 20 to 147) 2 SCB stages (5,5 bits)	1 ACB gain 2 SCB gain
Method	Autocorrelation/ Scalar Quantization	Full Search	*1st gain VQ:* ACB + 2 SCB gains 2 stages (5,5 bits)
Bits Per Frame	LP: 24 g_{frame}: 8	ACB index: 28 SCB indices: 40	1st gain VQ: 40
Rate	1600	3400	2000
TOTAL = 7000 bit/s			

Table 3.9: Bit allocation of the 6.0 kbit/s embedded CELP coder

	LP/g_{frame}	ACB/SCB	Gain VQ
Update	20 ms	20/4 = 5 ms	
Parameters	10 LSPs/ 1 gain	1 pitch for ACB (pitch: 20 to 147) 1 SCB stage (5 bits)	1 ACB gain 1 SCB gain
Method	Autocorrelation/ Scalar Quantization	Full Search	*1st gain VQ:* ACB + 1 SCB gain 2 stages (5,5 bits)
Bits Per Frame	LP: 24 g_{frame}: 8	ACB index: 28 SCB indices: 60	1st gain VQ: 40
Rate	1600	2400	2000
TOTAL = 6000 bit/s			

Table 3.10: Bit allocation of the 5.0 kbit/s embedded CELP coder

	LP/g_{frame}	ACB/SCB	Gain VQ
Update	20 ms	20/4 = 5 ms	
Parameters	10 LSPs/ 1 gain	1 pitch for ACB (pitch: 20 to 147) 1 SCB stage (5 bits)	1 ACB gain 1 SCB gain
Method	Autocorrelation/ Scalar Quantization	Full Search	*1st gain VQ:* ACB + 1 SCB gain 1 stage (5 bits)
Bits Per Frame	LP: 24 g_{frame}: 8	ACB index: 28 SCB indices: 20	1st gain VQ: 20
Rate	1600	2400	1000
TOTAL = 5000 bit/s			

3.3 Summary

In this chapter, we described in details the various major components of the CELP speech coders that we have studied in the course of this thesis. We described the structure of the CELP speech coder which includes the short term filter, the adaptive codebook, and the stochastic codebook. The methods used in obtaining and quantizing the speech parameters were also discussed. We also described the structure of our variable rate multimode and embedded CELP speech coders.

CHAPTER 4

VARIABLE RATE CHANNEL CODING

For most low rate speech coders, speech quality degrades significantly due to transmission errors. For mobile communications, the transmission channel is characterized by severe disturbances. In this thesis, forward error correction (FEC) is considered as the channel protection scheme. In particular, rate compatible punctured convolutional (RCPC) codes with block interleaving are adopted. In this chapter, the general convolutional code and RCPC codes are described. Then, the performance of the RCPC codes at various code rates under flat fading channel is presented.

4.1 Convolutional Codes

In recent years, convolutional codes with Viterbi decoding have become one of the most widely used forward error correction techniques. Convolutional codes are well known as good random error correcting codes. Together with interleaving, they can be used for channel coding in mobile communications. In addition, soft decision decoding can be easily incorporated into the Viterbi decoding algorithm. Potentially, the decoding gain can be increased by approximately 3 dB, as compared to hard decision decoding [37, 38]. This is why convolutional codes with Viterbi decoding are widely used for error protection in various applications [40, 13, 17].

A convolutional code can be specified by a state diagram or a trellis diagram. Figure 4.1 shows the state diagram of a simple rate 1/2 convolutional code with a

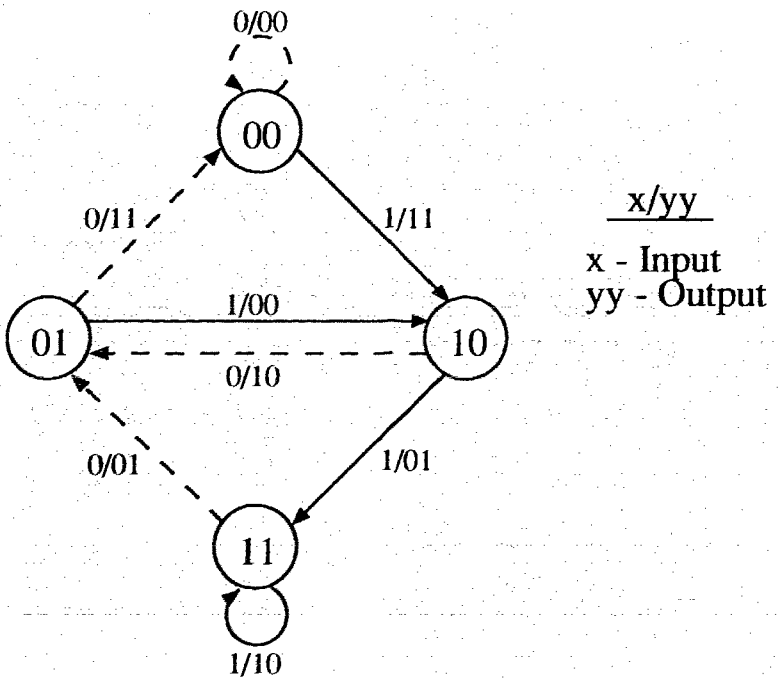


Figure 4.1: State diagram of a rate 1/2 convolutional code

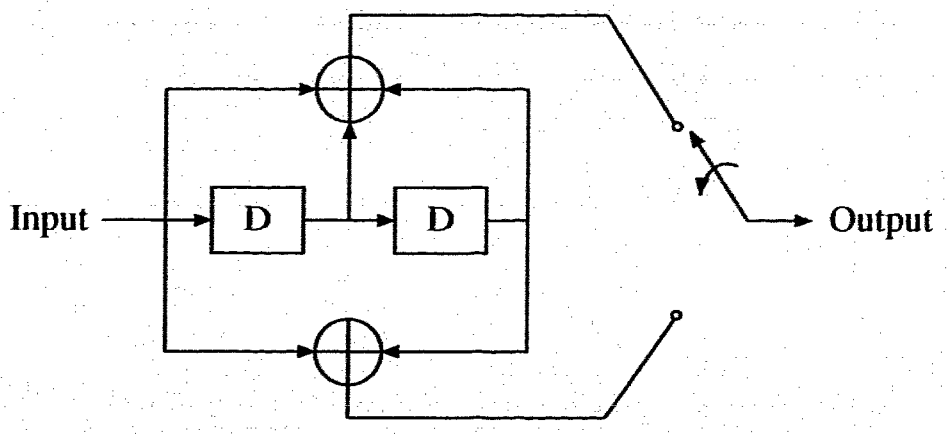


Figure 4.2: Block diagram of a rate 1/2 convolutional code encoder

constraint length of 2. The corresponding encoder is also shown in Figure 4.2. At the beginning of encoding, the convolutional encoder always start at state 00. Then, depending on the input bit, the next state will be either state 00 or state 10, with outputs 00 or 11 respectively, and so on. At the end of a data frame, the encoder is reset to the initial state 00 by encoding a number of zeros. In other words, the trellis or the state diagram is terminated at the state 00. The number of zeros required is equal to the constraint length of the convolutional code.

A convolutional code can also be completely described by its generator matrix. The generator matrix specifies the connections between the delay elements and the modulo-2 adders of the encoder. For the convolutional code encoder shown in Figure 4.2, the generator matrix is

$$G = \left\{ \begin{array}{ccc} 1 & 1 & 1 \\ 1 & 0 & 1 \end{array} \right\} \tag{4.1}$$

where a "1" in the matrix means a connection while a "0" means no connection.

4.2 Rate Compatible Punctured Convolutional Codes

The design of an error protection scheme usually consists of selecting a fixed channel code with a certain complexity and correction capability. The same code is then used for all data at any instant. However, it is sometimes desirable to have different degrees of error protection for different kinds of data and at different instances. For example, in a PCM coder, the effect of an error in the sign bit is more damaging than an error in the least significant magnitude bit [21]. The amount of error protection that is needed is called the error sensitivity information (ESI). In order to exploit the ESI of the data, a channel coder that is capable of operating at different code rates is required. In addition, the characteristics of a mobile channel are extremely time varying. This time variation can be measured at the receiver, thus yielding the channel state information (CSI). A flexible and adaptive speech transmission system should be designed using both the ESI and CSI to the maximum benefit.

Since most of the speech coding schemes generate blocks of encoded speech, it

seems natural to use a block channel coding scheme. Variable rate block codes have been studied in literature following the work of Masnick and Wolf [57]. However, these block codes cannot easily accommodate broadly varying unequal error protection needs within one codeword. They are also not easily decodable if soft decision decoding is employed. In a mobile fading channel, considerable coding gain can be achieved by soft decision decoding. Therefore, Hagenauer proposed the use of rate compatible punctured convolutional (RCPC) codes [41]. The RCPC codes are easily decodable by the Viterbi algorithm in which soft decision decoding can easily be incorporated. In addition, only one encoder/decoder is required even if the code rate changes several times within a speech frame. With RCPC codes, Cox has built a speech transmission system in which different speech bits that have different error sensitivities are protected by RCPC codes at different code rates. This technique is known as unequal error protection (UEP) and a gain of 5 dB in channel SNR (comparing to uniform error protection) has been found [13]. Hence, we suggest the use of RCPC codes as our variable rate channel coding scheme.

In our study, both variable rate channel coding in the speech frame level and the system level have been considered. In the speech frame level, encoded data is categorized into a few classes and different classes of data are protected by RCPC codes at different code rates (i.e., UEP). In the system level, the bit allocation between speech coding and channel coding is adjusted according to the mobile channel condition.

4.2.1 Constructing RCPC Codes

Puncturing is a method of obtaining higher rate convolutional codes from a lower rate mother code [37]. For a rate $1/n$ mother code with constraint length l, the generator matrix is

$$G = \begin{Bmatrix} g_{10} & g_{11} & \cdots & g_{1l} \\ g_{20} & g_{21} & \cdots & g_{2l} \\ \vdots & \vdots & \vdots & \vdots \\ g_{n0} & g_{n1} & \cdots & g_{nl} \end{Bmatrix} \tag{4.2}$$

where g_{ij}'s are the coefficients that specify the connections in the encoder. There is also a puncturing matrix

$$a(q) = \left\{ \begin{array}{cccc} a_{10} & a_{11} & \dots & a_{1p} \\ a_{20} & a_{21} & \dots & a_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n0} & a_{n1} & \dots & a_{np} \end{array} \right\} \tag{4.3}$$

of size $n \times p$ associated with each punctured code. In (4.3), p is called the puncturing period and a_{ij}'s specify the puncturing rules. If $a_{ij} = 0$, the bit at that position in the encoded code sequence is punctured. Otherwise, the bit is transmitted.

If m positions in the puncturing matrix are zeros, we obtain a punctured convolutional code of rate

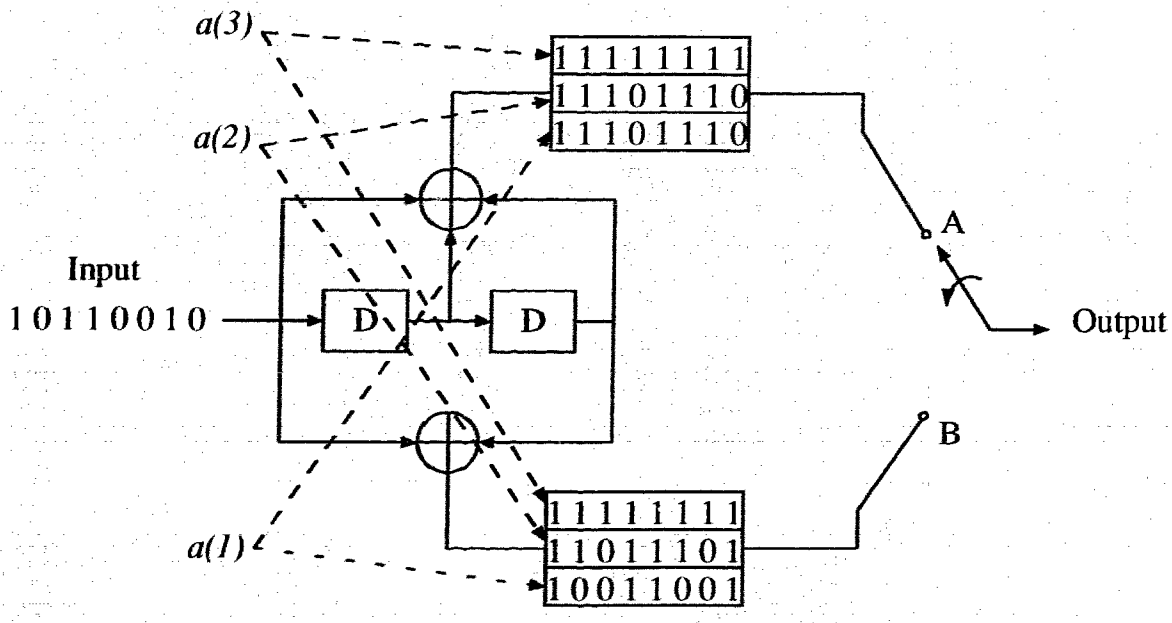
$$R_q = \frac{p}{np - m} \tag{4.4}$$

where q specifies the puncturing matrix $a(q)$ used in obtaining the punctured code. Thus, we can obtain a family of higher rate codes based on a lower rate mother code. Figure 4.3 shows an encoder that constructs RCPC codes at three different code rates. The encoder is based on a rate 1/2 mother code. In order to obtain a rate 2/3 code, 2 out of every 8 output bits are punctured. That means for 4 input bits, there are 6 output bits, which is the definition of a rate 2/3 code. A rate 4/5 code can also be obtained in a similar fashion. It should be noted that the encoder uses the same shift register for all rates. Only the multiplexer rule is changed, and the multiplexer functions according to the puncturing rule $a(q)$.

The parameter q can be changed during encoding as long as the decoder uses the same sequence of puncturing rules for proper demultiplexing. It also describes the order of puncturing. It is advisable that the family of RCPC codes follow the so-called *rate compatibility restriction* [41] which states that if $a_{ij}(q_o) = 1$, then

$$a_{ij} = 1 \qquad \text{for all } q \geq q_o \geq 1; \tag{4.5}$$

and the value of q is always increasing. That means if $n(q)$ denotes the number of 1's in $a(q)$, then $n(1) \leq n(2) \leq n(3)$ and so on, which implies that the encoder always change from a higher rate code to a lower rate code within a data frame. At the end of the frame, the code is terminated by encoding a number of zeros (see section 4.1).



a(q)	Rate	Output
a(3)	1/2	A -> 1 1 0 0 0 1 1 0 B -> 1 0 0 1 1 1 1 0
a(2)	2/3	A -> 1 1 0 - 0 1 1 - B -> 1 0 - 1 1 1 - 0
a(1)	4/5	A -> 1 1 0 - 0 1 1 - B -> 1 - - 1 1 - - 0

- means no output

Figure 4.3: Operation of a punctured convolutional encoder with $p = 8$

Table 4.1: Rate-compatible punctured convolutional codes

RCPC Code $l = 4, p = 8$	Generator Matrix $G = \begin{Bmatrix} 1 & 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 & 1 \end{Bmatrix}$			
Code Rate	1/3	1/2	2/3	4/5
q	4	3	2	1
Puncturing Matrix $a(q)$	1111 1111 1111 1111 1111 1111	1111 1111 1111 1111 0000 0000	1111 1111 1010 1010 0000 0000	1111 1111 1000 1000 0000 0000
Free Distance	11	7	4	3

In this thesis, we use the family of codes suggested by Hagenauer [41] and some of the codes are listed in Table 4.1. The mother code here is a rate 1/3 convolutional code with the generator matrix as shown in Table 4.1. The constraint length is 4 and the puncturing period is 8. The free distance of the codes are also shown in Table 4.1. The free distance is defined as the minimum Hamming distance between the all-zero path and any other path. The larger the free distance, the better the code is.

4.2.2 Decoding RCPC codes

At the receiver, the Viterbi algorithm is used for decoding. The basic Viterbi decoding algorithm can be found in many books that cover convolutional codes [37, 38, 36]. In the Viterbi algorithm, the path metric for a given path in the trellis is

$$\begin{aligned} m_p &= \sum_{j=1}^{D}\sum_{i=1}^{n}\beta_{ij} \\ &= \sum_{j=1}^{D}\sum_{i=1}^{n}a_{ij}b_{ij}d_{ij} \end{aligned} \tag{4.6}$$

where D is the decoding depth

n is the number of bits at each trellis branch

β_{ij} is the metric of the i-th bit at the j-th trellis branch

$a_{ij} \in \{0, 1\}$ is the corresponding entry in the puncturing matrix

$b_{ij} \in \{+1, -1\}$ is the i-th code bit at the j-th trellis branch

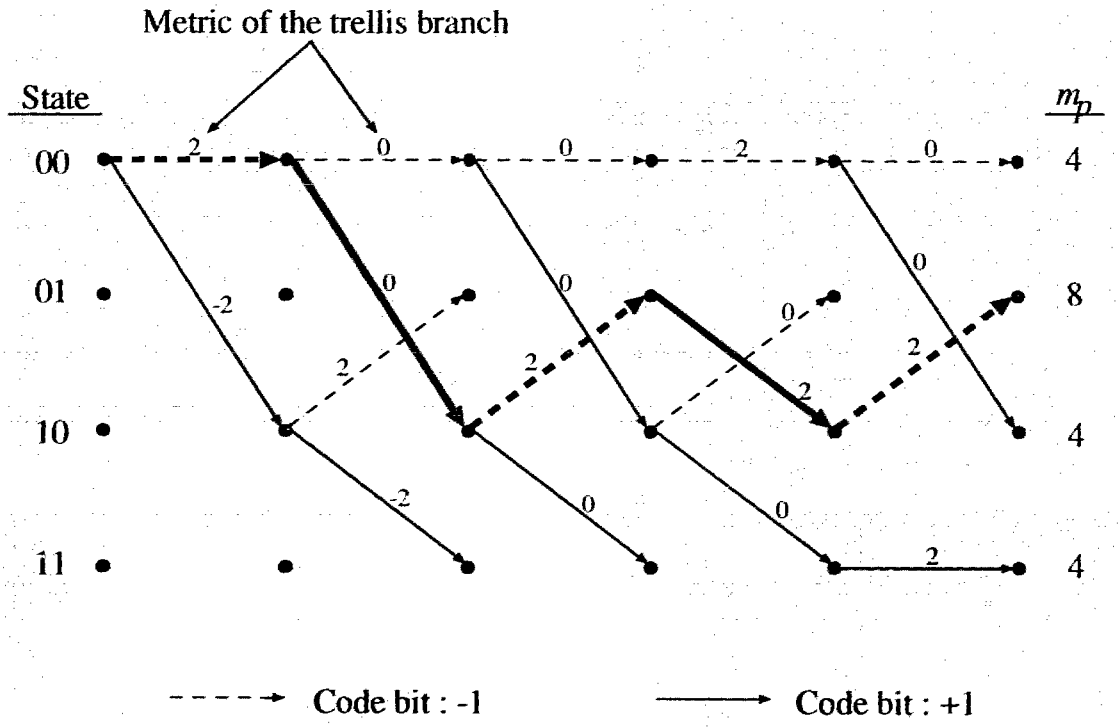
d_{ij} is the demodulator output for the i-th code bit at the j-th trellis branch.

For hard decision decoding, $d_{ij} \in \{+1, -1\}$ which corresponds to $\{1, 0\}$ in binary number representation. For soft decision decoding, d_{ij} is the unquantized output of the demodulator, which is a real number. The path that has the maximum m_p is the maximum likelihood path and the code bits along that path are considered to be the best estimate of the transmitted data. Figure 4.4 illustrates the Viterbi algorithm for decoding a simple sequence. The data are encoded by the rate 1/2 encoder as shown in Figure 4.2. The darkened path in Figure 4.2 is the maximum likelihood path. In order to demonstrate the error correcting capability of the code, an error is inserted in the fourth bit of the coded data. It can be seen that the correct data is decoded.

Since the modulation format is QPSK (see Chapter 2), two data bits are represented by a single channel symbol. With a typical QPSK constellation, we can separate the effect of the real and imaginary part of the complex modulation signal. Suppose two data bits $\{x_{k1}, x_{k2}\}$ are modulated into a single QPSK symbol. If the received symbol is r_k, the output of the demodulator is

$$d_{k1} = \mathrm{Re}\{r_k\} \tag{4.7}$$

$$d_{k2} = \mathrm{Im}\{r_k\} \tag{4.8}$$



Original Data	-1, +1, -1, +1, -1
Coded Data	-1, -1, +1, +1, +1, -1, -1, -1, +1, -1
Received Data	-1, -1, +1, -1, +1, -1, -1, -1, +1, -1
Decoded Data	-1, +1, -1, +1, -1

Single bit error

Figure 4.4: An illustration of the Viterbi algorithm

for soft decision decoding and

$$d_{k1} = \text{sign}(\text{Re}\{r_k\}) \tag{4.9}$$

$$d_{k2} = \text{sign}(\text{Im}\{r_k\}) \tag{4.10}$$

, where sign(•) gives the sign of the argument, for hard decision decoding.

4.2.3 Interleaving Strategy

Since RCPC codes are random error correcting codes, their performance degrade when errors happen in bursts. Therefore, an interleaver is required to decorrelate the errors. Strictly speaking, an interleaver with an infinite depth (i.e., ideal interleaving) guarantees the errors to be independent. In reality, Fung has found that an interleaving depth roughly equivalent to the duration of a quarter fade cycle is sufficient to produce the same effect as ideal interleaving [58]. A fade cycle is defined as the reciprocal of the normalized fade rate $f_D T$.

Suppose the vehicle is traveling at a speed of 90 km/h. For a 900 MHz carrier and an aggregate transmission rate of 48 kbit/s, $f_D T$ is about 0.003. Then, a quarter of the fading cycle is $0.25/f_D T \approx 83$ symbols, which is equivalent to 166 bits. A block interleaver with an interleaving depth of 166 bits will introduce an intolerable delay for speech transmissions.

As mentioned in Chapter 2, the transmission rate per user is either 9.6 or 12.8 kbit/s. Since each frame is 40 ms long, there are 384 bits and 512 bits per frame for a transmission rate of 9.6 kbit/s and 12.8 kbit/s respectively. We have chosen an interleaving depth of 16 bits (8 symbols) such that the interleaving depths are identical in all bit rates. Thus, the interleaving buffer for a transmission rate of 9.6 kbit/s and 12.8 kbit/s are of size 16×24 and 16×32 respectively.

4.2.4 Performance of RCPC Codes in Rayleigh Flat Fading Channels

We have examined the bit error performance of the RCPC codes (listed in Table 4.1) with an interleaving depth of 16 bits and they are shown in Figure 4.5. A full precision soft decision Viterbi decoder with a decoding depth (i.e., D in (4.6)) of 50 bits has

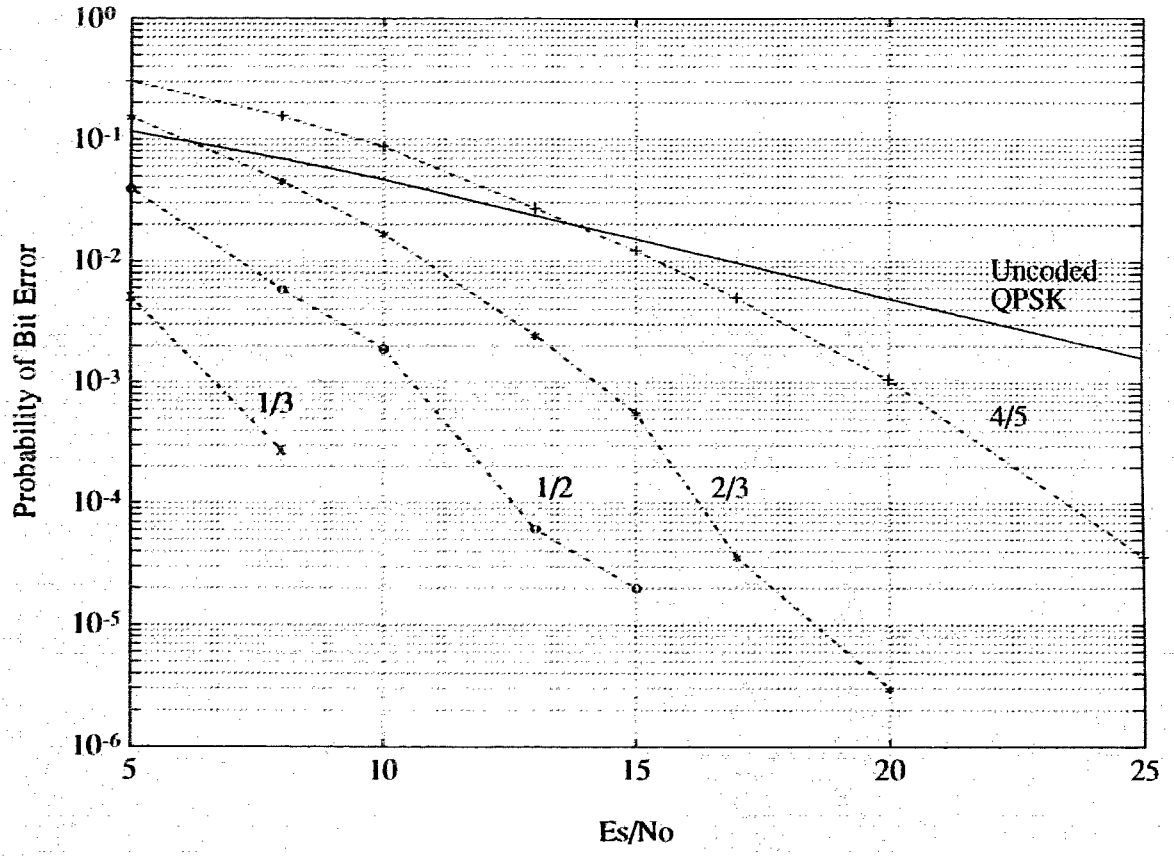


Figure 4.5: Bit error performance of RCPC codes in flat fading channel

been used. Hong has found that the decoding depth should be at least ten times of the constraint length of the RCPC codes [17]. QPSK modulation with perfect phase recovery and a normalized fade rate $f_D T$ of 0.003 have been assumed.

4.3 Summary

In this chapter, we discussed the basic theory behind encoding and decoding RCPC codes. We described the RCPC codes and the interleaving strategy used in this thesis. We also evaluated the performance of the RCPC codes with QPSK modulation in Rayleigh flat fading channel. We found that rate 1/2 and 1/3 RCPC codes have good error performance for the range of channel SNR that are of interest in mobile communications.

CHAPTER 5

VARIABLE RATE SPEECH AND CHANNEL CODING

In the previous chapters, we discussed the basics of the speech and channel coding techniques used in this study. In this chapter, we will show the design procedure of variable rate combined speech and channel coders. In this study, both variable rate coding in the speech frame level and the system level have been considered. Section 5.1 describes variable rate coding in the speech frame level by means of unequal error protection. Section 5.2 describes variable rate coding in the system level by means of adaptive assignment of speech and channel coding rates.

5.1 Combined Speech and Channel Coding

5.1.1 System Description

As discussed in Chapter 3, a CELP coder can produce good quality speech at rates as low as 4 kbit/s in the absence of channel errors. However, the reconstructed speech quality degrades dramatically in the presence of channel errors. In particular, mobile radio channels are very noisy due to fading. The impact of channel fading can cause a very high bit error rate which results in a speech communication system that is of unacceptable quality. Hence, an efficient forward error correction system is required to alleviate channel errors.

Our experiments indicate that some output bits of our CELP coders are very

sensitive to channel errors, while the others are not. A bit is sensitive to channel errors if a transmission error in that particular bit causes a large degradation in the reconstructed speech quality. An efficient error control system is to protect the output of the CELP coder with an unequal error protection scheme in which different speech bits that have different sensitivities are protected by error correcting codes at different code rates [13].

Figure 5.1 shows the block diagram of our combined speech and channel coder. In the transmitter, the speech is first processed by the CELP encoder on a frame by frame basis. The data within a speech frame is then classified into three different groups. The first two groups, which contain data that are most sensitive to channel errors, are protected by RCPC codes at different code rates. The remaining group, which contain data that are least sensitive to channel errors, are left unprotected. The above arrangements are necessary in order to maximizing the effectiveness of channel coding with a limited channel bandwidth. The error protected speech is then interleaved and sent to the receiver where various decoding operations are performed to reconstruct the speech signals.

After evaluating the performance of the RCPC codes under a Rayleigh fading channel (see Figure 4.5), we have decided to use the rate 1/3 and rate 2/3 codes. The rate 1/3 code can provide very good error protection to the most sensitive bits in the CELP coder output even at very low channel signal-to-noise ratio (SNR). On the other hand, the rate 2/3 code can provide ample protection for the less sensitive bits at moderate channel SNR.

5.1.2 Evaluation of Bit Error Sensitivity

As explained in Chapter 3, in a CELP encoder, the speech signal is analyzed frame by frame and represented by the linear prediction coefficients, the adaptive and the stochastic codebook excitations, and their corresponding gain values. These parameters are either scalar quantized or vector quantized, and the codebook indices of the quantizers are transmitted to the decoder through a physical channel with noise disturbances.

On a clean channel, the transmitted and received indices are identical. However, on a noisy channel, the received indices may be different from the transmitted ones.

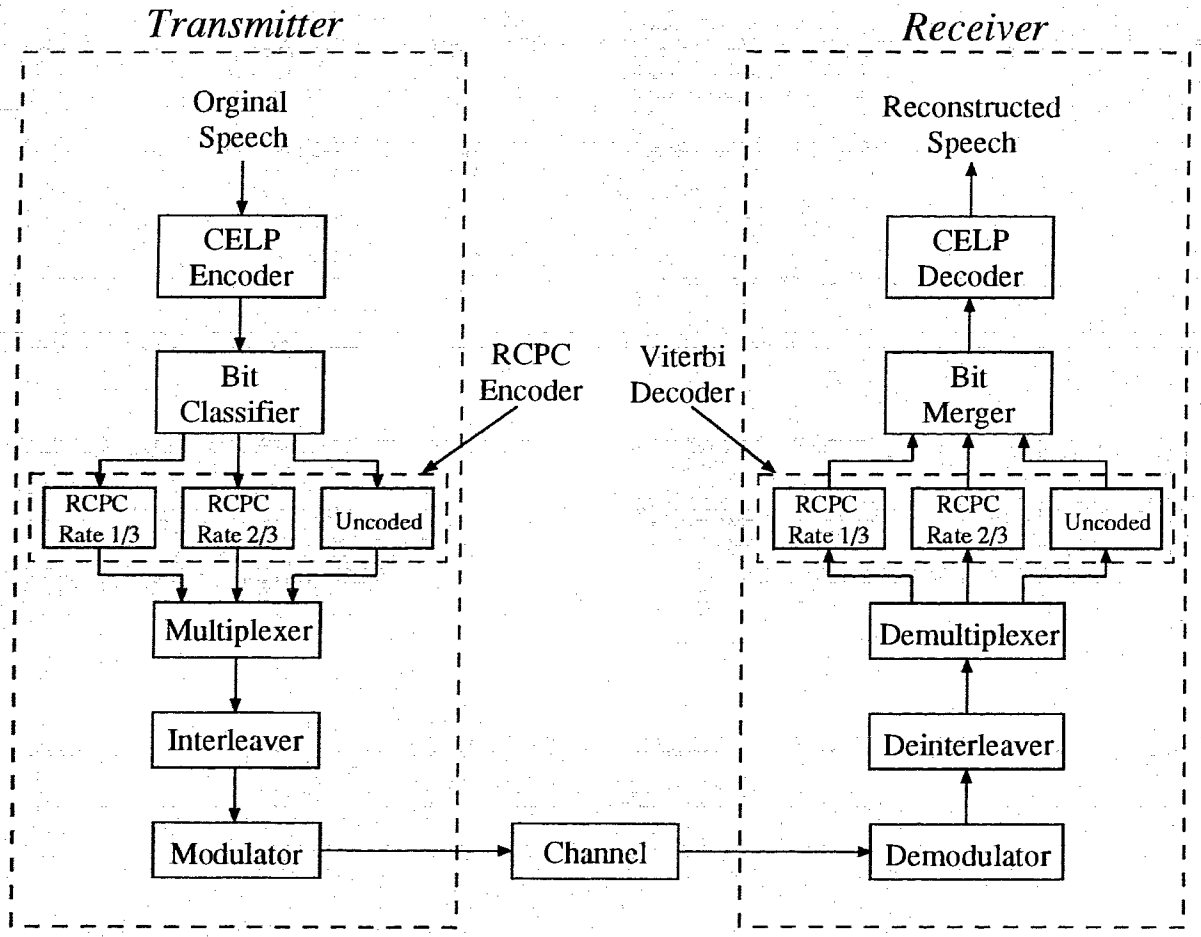


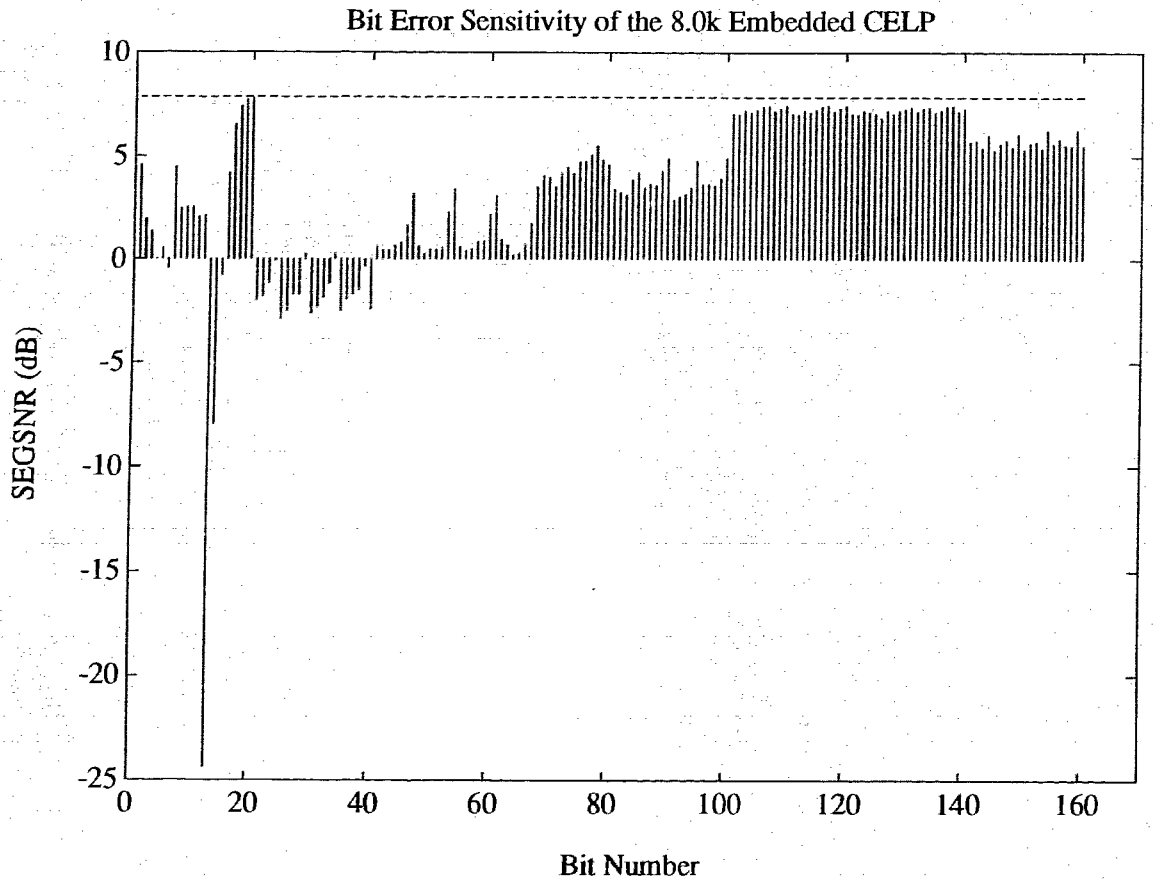
Figure 5.1: Block diagram of the combined speech and channel coder

Consequently, the parameters used to synthesize the speech signal at the decoder are different from the chosen ones at the encoder. This can cause a large degradation in the quality of the reconstructed speech.

We have observed that errors occur at different indices and at different positions within an index have different levels of influence on the quality of the reconstructed speech. That is, some bits in the CELP coder output are more sensitive to channel errors than the others. As mentioned earlier, unequal error protection (UEP) can be utilized to protect the CELP coder's output from channel errors efficiently. In order to apply UEP, the error sensitivity information (ESI) of our different CELP coders are evaluated. In this section, the ESI of our embedded CELP operating at 8.0 kbit/s coder is shown. However, the procedure in obtaining the ESI is identical to every coders and the ESI of the remaining coders are shown in Appendix A.

The error sensitivity of a particular bit within a CELP speech frame is determined by introducing an error in that particular bit and keeping the remaining bits error-free. The segmental signal-to-noise ratio (SEGSNR) between the original speech and the reconstructed speech is then measured. The resulting SEGSNR can be used as an indication of the error sensitivity of that particular bit. As mentioned in Chapter 3, the SEGSNR is a better indication of the objective quality of the reconstructed speech signal than the normal SNR criterion.

Figure 5.2 indicates the bit error sensitivity for different bits of a speech frame from the 8.0 kbit/s embedded CELP coder. The ESI test has been done on 625 frames of speech spoken by both male and female speakers. More sensitive bits have lower SEGSNG. For the 8.0 kbit/s embedded coder, there are 160 bits per frame and the correspondence between the bit number and the encoder parameter is also shown in Figure 5.2. From the graph, it can be observed that the bits representing the various gain values and the adaptive codebook indices are more sensitive to errors than the other parameters.



Parameters	Bit Numbers
LSP	1 - 12, 69 - 80
ACB	41 - 68
SCB	81 - 140
g_{frame}	13 - 20
GainVQ	21 - 40, 141 - 160

Figure 5.2: Error sensitivity of the 8.0 kbit/s embedded CELP coder

5.1.3 Optimal Code Rate Allocation

The objective of this section is to discuss the issue of determining the optimal channel code rate allocation (i.e., UEP) strategies for the CELP speech coders. A full search method has been used to obtain the optimal code rate allocations using the SEGSNR criterion.

Recall that the total bit rate for combined speech and channel coding is either 12.8 kbit/s or 9.6 kbit/s. Table 5.1 and 5.2 summarize the information about the different combined speech and channel coders evaluated in this thesis. For example, for the 8.0 kbit/s embedded CELP coder, there are 160 bits in each speech frame of 20 ms. For a combined speech and channel coding rate of 12.8 kbit/s, there are 256 bits altogether in each error protected speech frame, of which 96 bits are used for error protection. We arrange the 96 redundant bits among the 160 speech output bits according to their bit error sensitivities. The data within a speech frame is arranged in an order as shown in Figure 5.3. The leftmost bit is the least sensitive bit, and the rightmost bit is the most sensitive bit. The error sensitivity level of the bits from left to right is increasing.

We have considered 2 levels of error protection in this thesis. The data within a CELP speech frame is classified into 3 groups, in which two of them are protected by the rate 1/2 and rate 2/3 RCPC codes and the remaining one is left unprotected. More levels of error protection could be considered. However, it will lead to an intractable amount of computation time for performing the optimal code rate allocation search.

Our objective is to find out the code rate allocation that optimizes the performance of a combined CELP/RCPC coder under the SEGSNR criterion. A direct way of doing this is through an exhaustive search of all the possible code rate allocations. Since there are only three groups of data, it is plausible to perform a full search.

Table 5.1: Configurations of combined embedded speech and channel coders

Coder	Combined Rate [bit/s]	Speech Coding Rate [bit/s]	Channel Coding Rate [bit/s]
1	12800	12800	0
2	12800	8000	4800
3	12800	5000	7800
4	9600	9600	0
5	9600	8000	1600
6	9600	7000	2600
7	9600	6000	3600
8	9600	5000	4600

Table 5.2: Configurations of combined multimode speech and channel coders

Coder	Combined Rate [bit/s]	Speech Coding Rate [bit/s]	Channel Coding Rate [bit/s]
9	12800	12800	0
10	12800	9600	3200
11	12800	8000	4800
12	12800	5000	7800
13	9600	9600	0
14	9600	8000	1600
15	9600	5000	4600

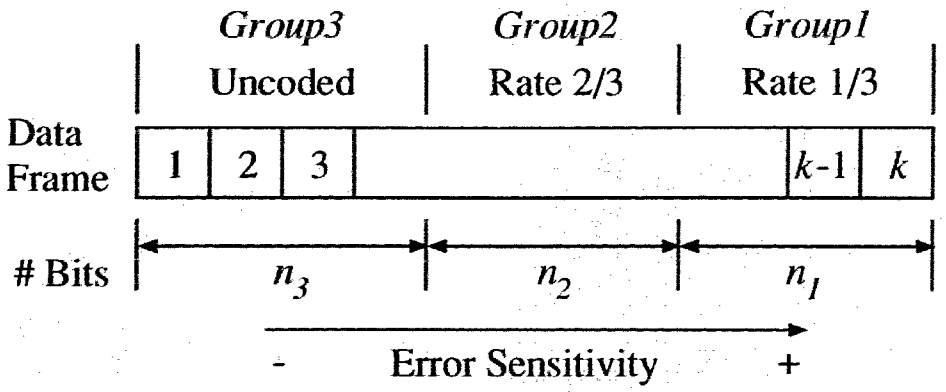


Figure 5.3: Grouping of information bits in each frame according to their error sensitivities

5.1.4 Optimal Code Rate Search

As shown in Figure 5.3, n_1, n_2, and n_3 are the number of bits assigned to each group and they are all positive integers. Assuming that the bit rate of the speech coder is 8.0 kbit/s, the following equation

$$n_1 + n_2 + n_3 = 160 \tag{5.1}$$

holds true. If the bit rate of the combined speech and channel coder is 12.8 kbit/s, the number of redundant bits is 96, out of which 6 bits are used for terminating the trellis at the end of a frame (i.e., 4 input bits coded by a rate 2/3 code results in 6 output bits). Since n_3 is uncoded, n_2 is protected by the rate 2/3 code, and n_1 is protected by the rate 1/3 code, the number of redundant bits can be represented by

$$\begin{aligned} 2n_1 + 0.5n_2 + 6 &\leq 96 \\ n_2 &\leq 180 - 4n_1. \end{aligned} \tag{5.2}$$

From (5.1) and (5.2), we have

$$n_3 \geq 3n_1 - 20. \tag{5.3}$$

(5.2) and (5.3) indicate that both n_2 and n_3 are functions of n_1. We can then obtain the ranges of n_1, n_2, and n_3 to be:

$$7 \leq n_1 \leq 45$$

$$0 \leq n_2 \leq 152$$

$$1 \leq n_3 \leq 115$$

and the number of possible code rate allocations is 39. Results of the optimal code rate allocation searches are shown in Table 5.3. Although the code rate allocation is a function of the channel SNR, we found that it is insensitive to changes in channel SNR. Therefore, we assumed the channel SNR to be 23 dB and obtained the results in Table 5.3.

Table 5.3: Result of the optimal code rate searches

Coder	Number of Searches	Results		
		n_1	n_2	n_3
2	39	31	56	73
3	8	67	32	1
5	14	3	40	117
6	24	7	64	69
7	30	13	80	27
8	19	25	72	3
10	30	17	48	127
11	39	32	52	76
12	8	67	32	1
14	14	4	36	120
15	19	25	72	3

5.2 Variable Rate Combined Speech and Channel Coding

5.2.1 System Description

Figure 5.4 shows the block diagram of our adaptive variable rate combined speech and channel coding system. In our system, adaptive error control is achieved by using the combined speech and channel coder that has the best performance at a given channel signal-to-noise ratio (SNR). Different combined speech and channel coders that have good performance over the typical range of channel SNR have been designed as described in Section 5.1.

After the CELP and RCPC encoders, the encoded binary digits are used to modulate a carrier signal using the QPSK modulation format (see Section 2.4). The modulated signal is transmitted over a mobile radio channel modelled as a Rayleigh flat fading channel with a superimposed long-term fading (see Section 2.5). That means the transmitted signal is subjected to both random amplitude and phase distortions that vary at a rate proportional to the vehicle speed. As mentioned in Section 2.4, we have assumed that both the amplitude and phase distortions introduced by the channel can be perfectly estimated at the receiver.

At the receiver, a channel state estimator is required. In addition to estimating the carrier phase, the channel estimator also estimates the long term fading statistics of the channel. Estimation of the long term statistics are sent periodically by the receiver to the transmitter through the reverse channel. Those information are required for selecting an appropriate combined CELP/RCPC coder for the prevailing channel SNR. Given that the transmitter has selected the appropriate coder, the receiver can decode the received information by selecting the correct decoder. The information contained in the header of an error protected speech frame is used to identify the appropriate decoder at the receiver.

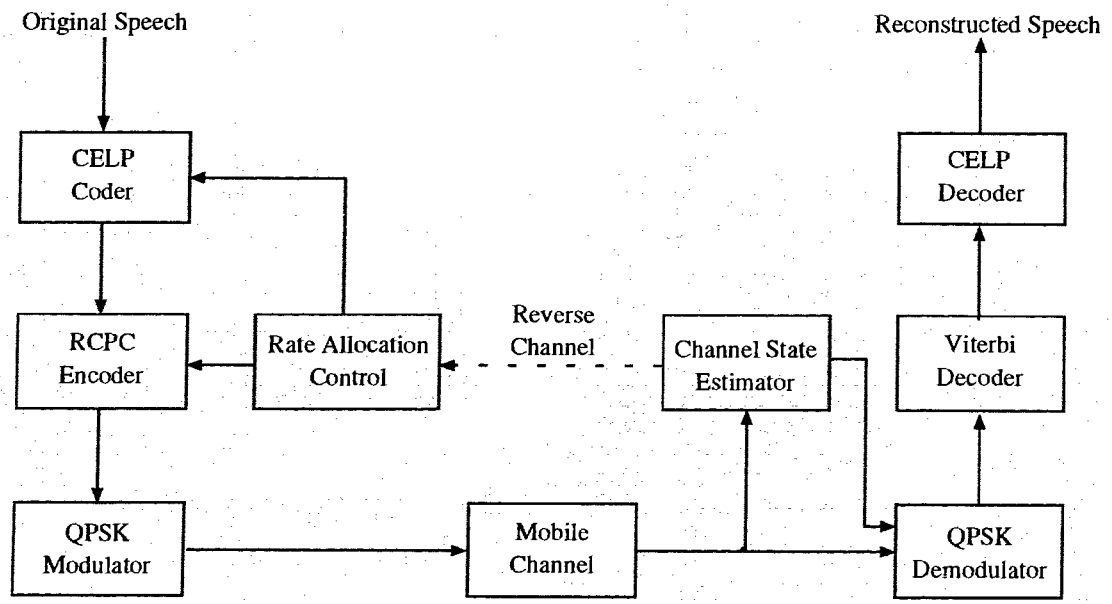


Figure 5.4: Block diagram of the variable rate system with feedback channel

5.2.2 Channel State Estimation

Before demodulation can take place at the receiver, the channel state estimator in Fig. 5.4 will first have to estimate the received signal envelope $d(t)$ in (2.13) or equivalently $d(x)$ in (2.17). For this purpose, we assume that pilot symbols are inserted periodically into the transmitted data stream. Since these symbols are known to the receiver, it allows the latter to estimate the fading envelope by comparing the values of the transmitted and the received pilot symbols. As shown in [42], very reliable estimates can be obtained this way.

As pointed out earlier, the fading envelope $d(x)$ is the product of the short term fading process $d_s(x)$, where it is assumed to be Rayleigh distributed with a mean of unity, and the long term fading process $d_l(x)$. The long term fading process can be estimated from $d(x)$ (which under the ideal condition, is the output of the pilot symbol assisted estimator) according to the following [59] :

$$\begin{aligned} \hat{d}_l(x) &= \frac{1}{2\zeta}\int_{x-\zeta}^{x+\zeta} d_l(y)d_s(y)\,dy \\ &= d_l(x)\frac{1}{2\zeta}\int_{x-\zeta}^{x+\zeta} d_s(y)\,dy \end{aligned} \tag{5.4}$$

where $\hat{d}_l(x)$ is the estimate of $d_l(x)$ and 2ζ is the window size used in the averaging

process. The above averager operates on the principle that $d_l(x)$ remains more or less constant over the span of the window.

If the length ζ is chosen properly, the estimate long term fading process $\hat{d}_l(x)$ will approach $d_l(x)$. This implies that

$$\frac{1}{2\zeta}\int_{x-\zeta}^{x+\zeta} d_s(y)\,dy = 1. \tag{5.5}$$

We can determine the length ζ that satisfies (5.5) by computing the estimation variance

$$\sigma^2_{\hat{d}_l} = E\{\hat{d}_l^2(x)\} - E\{\hat{d}_l(x)\}^2. \tag{5.6}$$

In [59], Lee has provided a detailed description of the procedure used to obtain (5.6) with different values of ζ. He has found that a length of between 20λ and 40λ, where λ is the wavelength of the carrier, is the proper length for averaging the fading envelope.

5.2.3 Protocol for Coder Selection

Once $d_l(x)$ is estimated, the receiver transmits the estimate to the transmitter over the reverse channel. Upon receiving the estimate, the transmitter selects the most appropriate combined CELP/RCPC coder. In the header of every error protected speech frame, there is a field indicating the coder used to encode this particular frame. Since the maximum number of combined coders at each combined rate is 5, 3 bits are sufficient to indicate the selected coder. The receiver then uses this information to select the appropriate decoder. The suggested protocol is shown in Figure 5.5.

As shown in section 5.2.2, the proper window size for estimating the long term statistics (i.e., 2ζ) is between 20 to 40 wavelengths. For a 900MHz carrier, the wavelength is approximately 0.33 meter. Therefore, the range of 2ζ is from 6.66 meters to 13.2 meters. For a vehicle speeds of 90 km/h, the above values correspond to an estimation interval of between 0.27 and 0.54 s. Since there are 25 frames/s in our variable rate system, a proper combined CELP/RCPC coder should be selected every 7 to 13 frames.

5.2.4 Simulation Model for Channel Condition

In order to compare the performance of our adaptive variable rate coders with their non-adaptive counterparts, a profile of variations in channel condition (i.e., changes

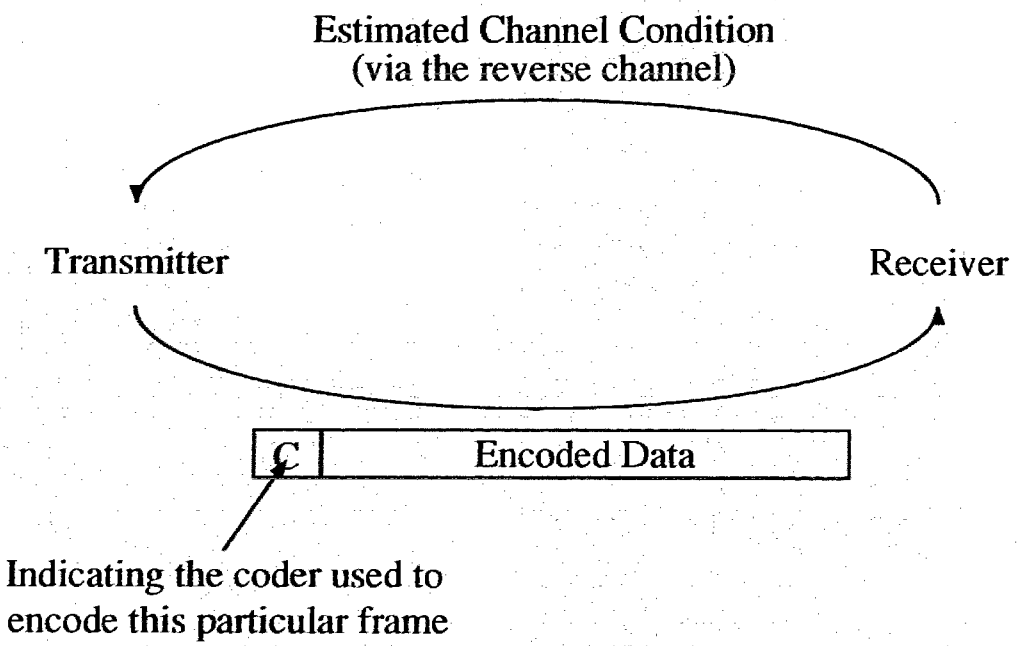


Figure 5.5: Suggested protocol for coder selection

in channel SNR over time) is required. Hence, we have derived a simple model for simulating the variations. The simulated channel variations were then used to adjust the speech/channel coding bit allocation of our adaptive variable rate coders.

As shown in Figure 5.6, we have assumed a circular coverage area with a radius R and an inner circle with a radius A that contains no users. If the users are uniformly distributed within the cell, the fraction of users within the i-th ring is

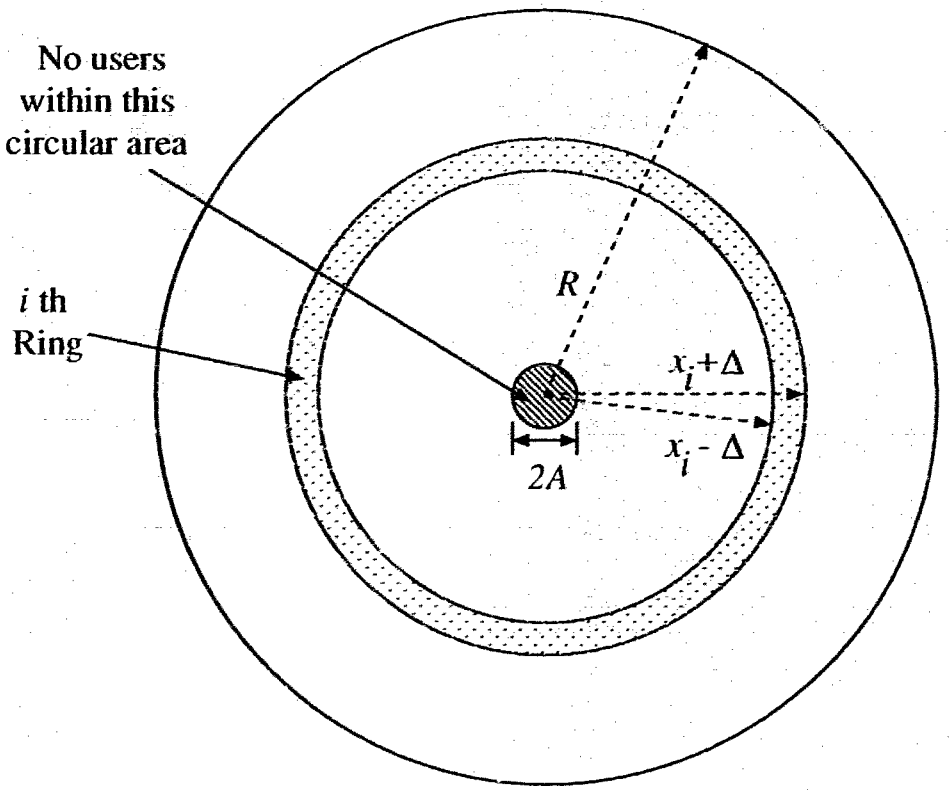
$$f_i = \frac{4\Delta x_i}{(R^2 - A^2)} \tag{5.7}$$

In accordance with the commonly used inverse 4th power law [27], the power at a distance x away from the transmitter is inversely proportional to x^4. Then, the signal power at distance x away from the base station is

$$P(x) = P(A) - 40(\log(x) - \log(A)) \quad x \geq A \tag{5.8}$$

where $P(\bullet)$, in dB, is the received power at a certain distance from the base station. Let the power $P(x)$ in linear scale be $L(x)$. The channel SNR as a function of x is then equivalent to

$$\frac{E_s}{N_o}(x) = \frac{\sigma_{d_s}^2}{N_o B} L(x) \tag{5.9}$$



Value of parameters in the simulation model

R	A	$P(A)$	σ
10 km	1 km	50 dB	8 dB

Figure 5.6: Model for simulating the variations in channel SNR

where $\sigma_{d_s}^2$ is the variance of the short term fading process and is normalized to unity, N_o is the power spectral density of the channel AWGN, and B is the equivalent noise bandwidth. For every point x in the coverage area, we assume that $L(x)$ is log-normal distributed with a standard deviation of σ dB. Then, the probability density function (PDF) of the channel SNR in dB is given by

$$p(SNR) = \sum_{i=1}^{N_{ring}} f_i \frac{1}{\sqrt{2\pi\sigma^2}} exp\{-\frac{(SNR - P(x_i))^2}{2\sigma^2}\} \tag{5.10}$$

where $N_{ring} = (R - A)/(2\Delta)$ is the number of rings within the cell.

For example, if the parameters shown in Figure 5.6 are assumed, the simulation model results in approximately an overall log-normal distribution (for the entire coverage area) with a mean of 22.8 dB and a standard deviation of 12.9 dB. Values in Figure 5.6 are chosen in a way such that the mean of the resulting PDF is approximately 23 dB, which is the typical SNR of interest for mobile communications.

5.3 Summary

In this chapter, we described in details the design procedure of combined CELP/RCPC coders. We described the method in obtaining the optimal RCPC code rate allocation strategies for the CELP speech coders, and thus obtained sets of combined CELP/RCPC coder with a combined rate of 12.8 and 9.6 kbit/s. We also described the various components of our adaptive variable rate combined CELP/RCPC systems. In order to evaluate the performance of our variable rate combined coders, we have derived a simulation model for channel SNR. It has been found that the channel variations using this model approximately follow a log-normal distribution.

CHAPTER 6

EXPERIMENTAL RESULTS

In last chapter, we described the design procedure of combined CELP/RCPC coders as well as the structure of the adaptive variable rate combined CELP/RCPC system. In this chapter, we will evaluate the performance of various combined CELP/RCPC coders operating at 12.8 and 9.6 kbit/s under the Rayleigh flat fading channel. Then, the procedure for obtaining the adaptive coders and their performance will be shown.

6.1 Combined Speech and Channel Coders

6.1.1 Performance of Combined Speech and Channel Coders

With the optimal code rate allocations as shown in Table 5.3, we have evaluated the performance of different combined CELP/RCPC coders over different channel SNR. Table 6.1 summarizes the simulation parameters and the simulation results are shown in Figure 6.1 to 6.4. In Figure 6.1 to 6.4, the vertical scale is the segmental signal-to-noise ratio (SEGSNR) which indicates the quality of the reconstructed speech. The SEGSNR values were averaged over 625 frames of speech with both male and female speakers. The horizontal scale is the channel SNR.

As shown in Figure 6.1, the coder that has the best performance above a channel SNR of 25 dB is coder #1, which is the uncoded 12.8 kbit/s embedded speech coder. However, its performance degrades drastically as the channel SNR decreases. On the other hand, coder #3 , which is the 5.0 kbit/s embedded CELP coder with 7.8 kbit/s of error protection, still performs reasonably well below a channel SNR of 10

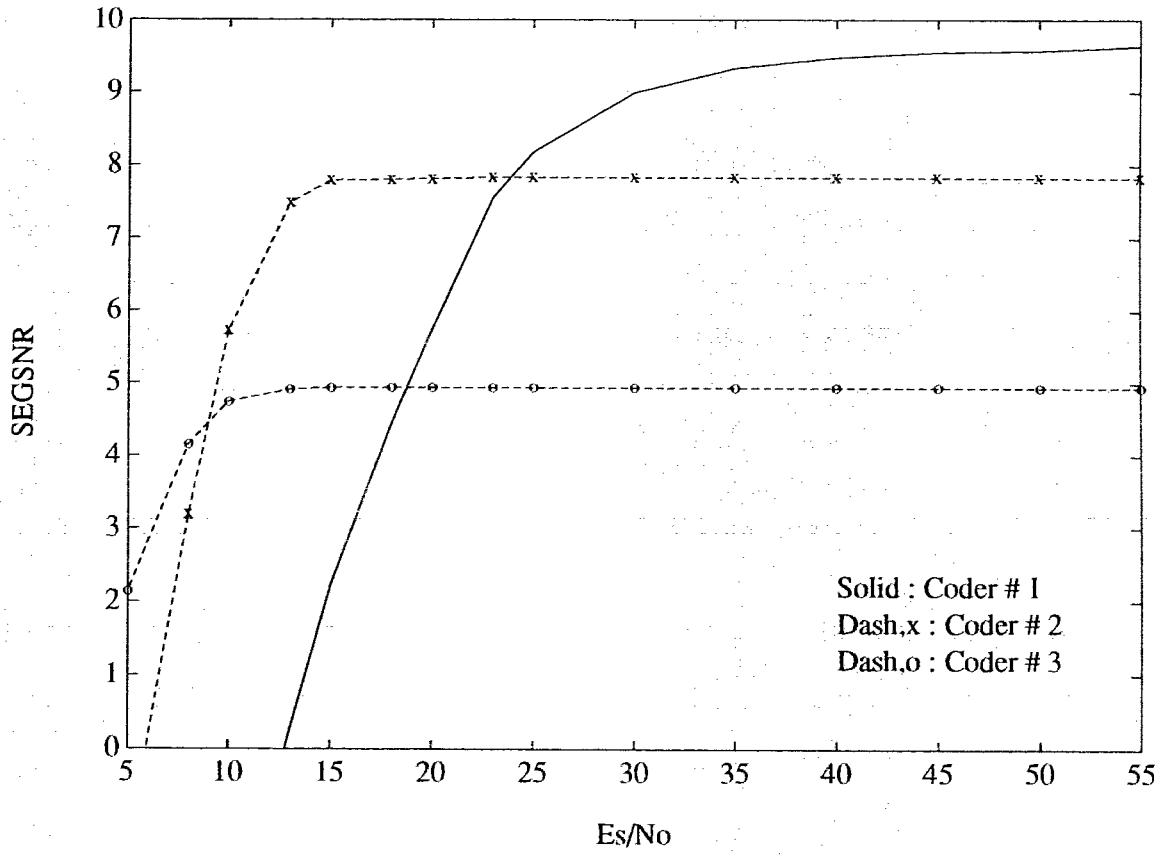
dB. However, its performance at higher channel SNR is significant worse than coder #1. For example, at a channel SNR of 40 dB, coder #1 outperforms coder #3 by about 4.5 dB in SEGSNR which indicates a significant difference in the quality of the reconstructed speech.

If the channel SNR varies over a wide range, using coder #1 will result in a speech communication system with unacceptable quality at low channel SNR. On the other hand, using coder #3 will result in a speech communication system that reconstructs speech with quality worse than necessary most of the time. The same scenario can also be observed in Figure 6.2, 6.3, and 6.4.

In addition, the simulation results indicate that for the same transmission rate and rate partitioning between speech and channel coding, combined coders using multimode speech coders are better than those using embedded speech coders. This result is expected because of the constrained characteristics of the embedded coders.

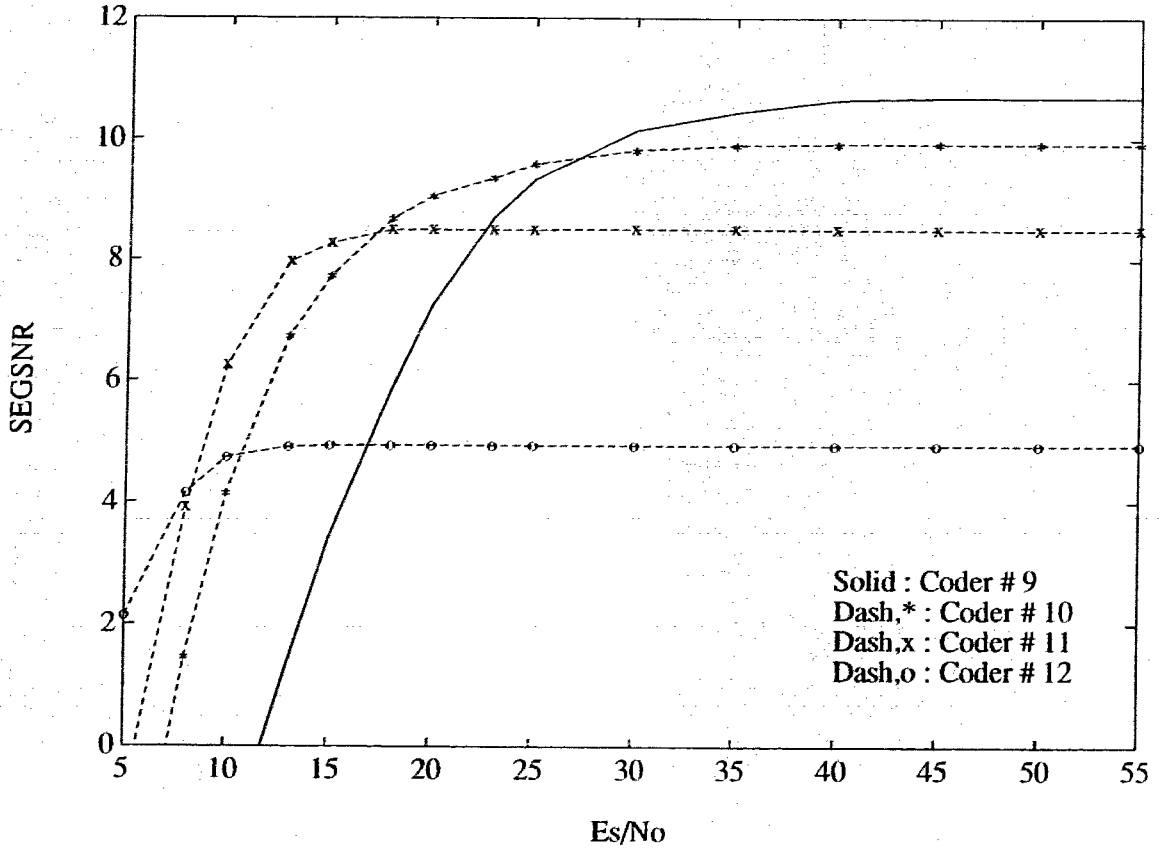
Table 6.1: Summary of simulation parameters

Parameter	Value
Modulation Method	QPSK
Carrier Frequency	900 MHz
Vehicle Speed	90 km/h
Aggregate Transmission Rate	48 kbit/s
Normalized Fade Rate	0.003
Interleaving Depth	16 bits



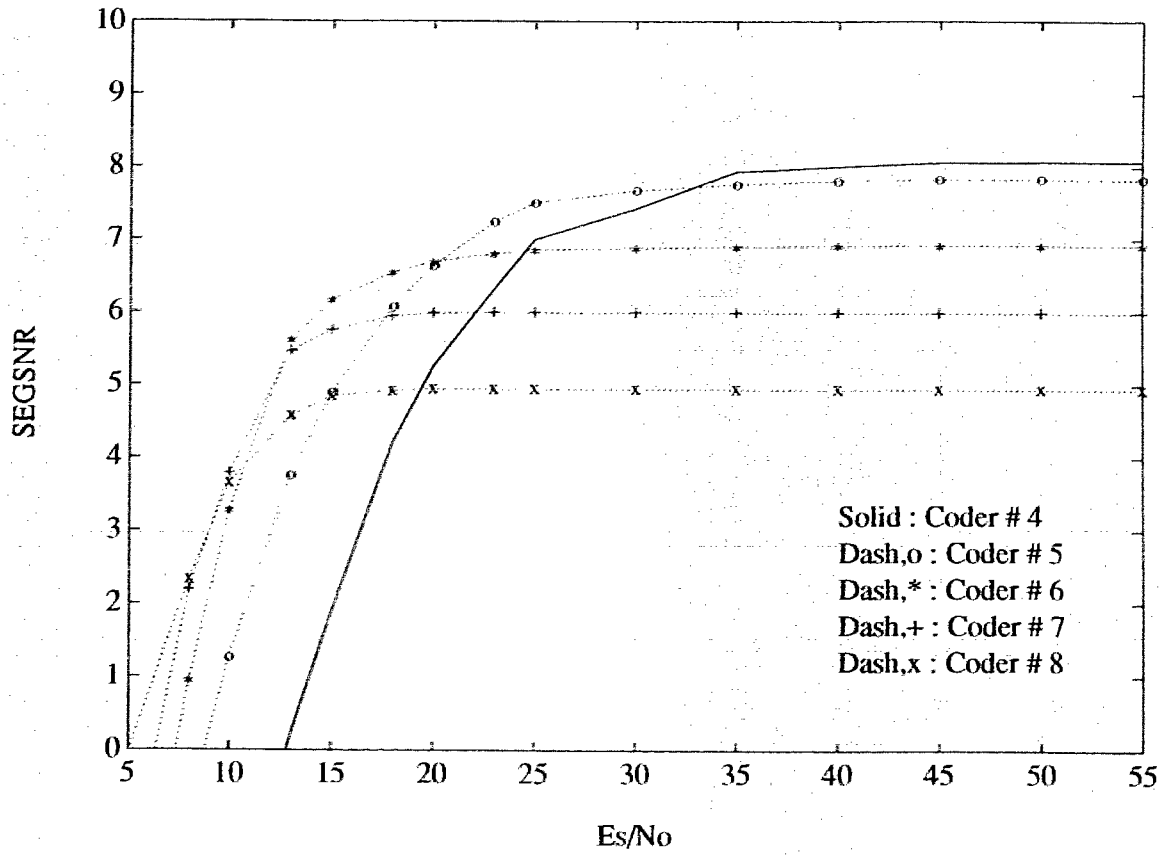
Coder	Speech Coding Rate [bit/s]	Channel Coding Rate [bit/s]
1	12800	0
2	8000	4800
3	5000	7800

Figure 6.1: Performance of combined embedded CELP/RCPC coders for combined rate = 12.8 kbit/s



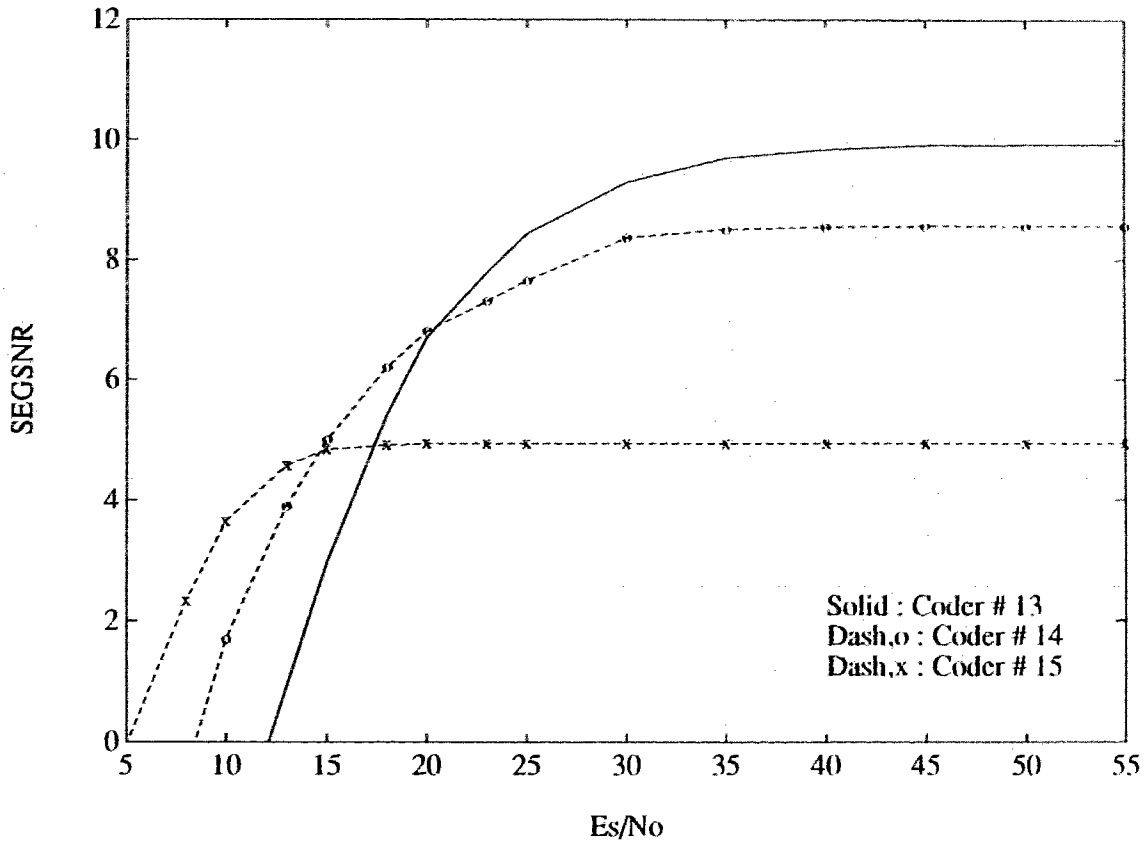
Coder	Speech Coding Rate [bit/s]	Channel Coding Rate [bit/s]
9	12800	0
10	9600	3200
11	8000	4800
12	5000	7800

Figure 6.2: Performance of combined multimode CELP/RCPC coders for combined rate = 12.8 kbit/s



Coder	Speech Coding Rate [bit/s]	Channel Coding Rate [bit/s]
4	9600	0
5	8000	1600
6	7000	2600
7	6000	3600
8	5000	4600

Figure 6.3: Performance of combined embedded CELP/RCPC coders for combined rate = 9.6 kbit/s



Coder	Speech Coding Rate [bit/s]	Channel Coding Rate [bit/s]
13	9600	0
14	8000	1600
15	5000	4600

Figure 6.4: Performance of combined multimode CELP/RCPC coders for combined rate = 9.6 kbit/s

6.2 Variable Rate Speech and Channel Coders

6.2.1 Selection of the Appropriate Combined Coder

Figure 6.5 is identical to Figure 6.1 except that a curve indicating the ideal performance of an adaptive variable rate system is shown. As shown in the graph, an adaptive system outperforms its non-adaptive counterparts provided that it can always select the combined CELP/RCPC coder that has the best performance at a given channel SNR.

In order to decide the appropriate combined coder to be used at a given channel SNR, we have to inspect the performance of all the combined coders over the typical range of channel SNR, which are shown in Figure 6.1 to 6.4. After close inspection, we have obtained the ranges of channel SNR in which the combined coders have the best performance and the results are shown in Table 6.2 to 6.5. The values in Table 6.2 to 6.5 correspond to different crossover points of the performance curves shown in Figure 6.1 to 6.4. This means that rate change is triggered by crossing thresholds.

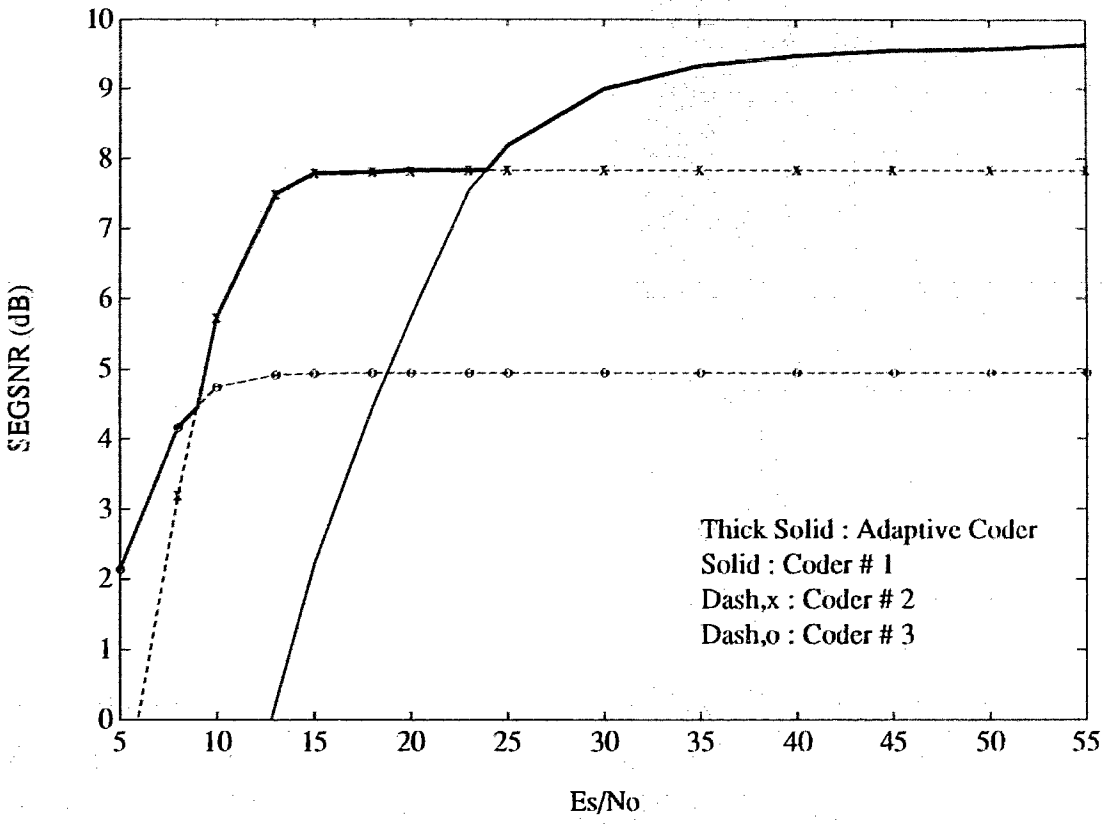


Figure 6.5: Ideal Performance of an adaptive coder operating at 12.8 kbit/s

Table 6.2: Configuration of the 12.8 kbit/s embedded adaptive coder

Coder	SNR from	SNR to
3	$-\infty$	9.0 dB
2	9.0 dB	24.0 dB
1	24.0 dB	$+\infty$ dB

Table 6.3: Configuration of the 12.8 kbit/s multimode adaptive coder

Coder	SNR from	SNR to
12	$-\infty$	8.3 dB
11	8.3 dB	17.4 dB
10	17.4 dB	27.5 dB
9	27.5 dB	$+\infty$

Table 6.4: Configuration of the 9.6 kbit/s embedded adaptive coder

Coder	SNR from	SNR to
8	$-\infty$	9.7 dB
7	9.7 dB	12.1 dB
6	12.1 dB	20.6 dB
5	20.6 dB	33.1 dB
4	33.1 dB	$+\infty$

Table 6.5: Configuration of the 9.6 kbit/s multimode adaptive coder

Coder	SNR from	SNR to
15	$-\infty$	14.7 dB
14	14.7 dB	20.6 dB
13	20.6 dB	$+\infty$

6.2.2 Performance of Variable Rate Combined Speech and Channel Coders

In order to evaluate the performance of our adaptive variable rate coders, a profile of channel SNR variations over time is required. With the channel SNR model described in section 5.2.4, an average channel SNR value is generated at random for each frame. This random SNR is then used to control the variance of the short term fading process and to select the most appropriate combined coder to encode the frame of speech. The corresponding received signal is decoded and the quality of the reconstructed speech is evaluated in terms of the SEGSNR and the mean opinion score (MOS). A scale of 1 (for poor quality) to 5 (for excellent quality) is used for MOS evaluation and the reconstructed speech is considered to have toll quality if the MOS is better than 4.0.

In our channel SNR model, the average channel SNR values are generated independently. However, in reality, consecutive samples of the long term fading component (i.e., the average channel SNR) are highly correlated. It can be easily shown that the SEGSNR is unaffected by this simulation model. On the other hand, the subjective MOS would be different if the average SNR in successive frames are correlated. Thus a proper way to interpret our MOS results is to view them as results generated under the idealistic condition of infinite interleaving at the frame level.

Table 6.6 compares the performance of adaptive and non-adaptive coders operated at various combined rates and with the simulation parameters shown in Table 6.1. The average channel SNR experienced by each frame is randomly generated. The SEGSNR values were averaged over 625 frames of speech spoken by both male and female speakers. The MOS were obtained through an informal MOS test conducted with 10 untrained listeners. As shown in Table 6.6, an improvement of up to 1.35 dB in SEGSNR and 0.9 in MOS (for a combined rate of 12.8 kbit/s using combined embedded CELP/RCPC coders) is obtained when adaptive coding is employed.

In addition, Table 6.6 indicates that combined coders with embedded CELP coder are much inferior than those with multimode CELP coder. For example, at a combined rate of 12.8 kbit/s, the adaptive coder with multimode CELP coder outperforms the coder with embedded CELP coder by 0.77 dB in SEGSNR and 0.4 in MOS. The degradation in performance suggests that the CELP structure may not be suitable for implementing embedded coders.

Table 6.6: Performance of coders obtained with the channel SNR simulation model

Coder(s) Used	Speech Coder	Combined Rate [bit/s]	Type	SEGSNR [dB]	MOS
1,2,3	Embedded	12800	Adaptive	7.40	3.3
2	Embedded	12800	Fixed	6.05	2.4
9,10,11,12	Multimode	12800	Adaptive	8.11	3.7
11	Multimode	12800	Fixed	6.89	2.9
4,5,6,7,8	Embedded	9600	Adaptive	5.58	2.2
5	Embedded	9600	Fixed	4.42	2.0
13,14,15	Multimode	9600	Adaptive	6.54	2.5
14	Multimode	9600	Fixed	5.13	2.3

6.3 Summary

In this chapter, the performance of various combined CELP/RCPC coders were shown. We found that both uncoded and error protected multimode coders have better performance than their embedded counterparts. In addition, we also described the procedure in obtaining the adaptive variable rate combined CELP/RCPC coders and evaluated their performance with the channel SNR simulation model described in Chapter 5. We found that the adaptive coders perform significantly better than their fixed rate counterparts. This is particularly true for coders operating at 12.8 kbit/s. An improvement of 1.35 dB in SEGSNR and 0.9 in MOS was obtained with the adaptive coder operating at 12.8 kbit/s.

CHAPTER 7

CONCLUSIONS AND SUGGESTIONS FURTHER WORK

7.1 Conclusions

This thesis has considered variable rate combined speech and channel coding over mobile radio channels. The speech coders are capable of operating at different rates and they are all based on code excited linear prediction (CELP). In particular, two different types of variable rate CELP coder have been considered : multimode and embedded variable rate coders. For channel coding, we have adopted the rate compatible punctured convolutional (RCPC) codes with block interleaving. A single RCPC encoder/decoder can encode/decode RCPC codes at different code rates. Since the code rate can be changed within a single frame, unequal error protection (UEP) is employed to effectively protect the speech data from channel errors. With the variable rate CELP and RCPC coders, we have proposed four adaptive variable rate systems that operate at combined rate of 9.6 kbit/s and 12.8 kbit/s. The adaptive systems can adjust the bit rate partitioning between speech and channel coding dynamically in accordance with the channel condition. Two of these systems employ multimode CELP coder while the remaining two systems employ embedded CELP coder.

We found that adaptive coders based on multimode CELP coders are better than

their embedded counterparts. However, the structure of the embedded CELP coders allows an efficient implementation of variable rate coder and easy rate control in network overload situations. In this thesis, the performance of the adaptive coders have been shown to be substantially better than the non-adaptive coders. Improvements of up to 1.35 dB in SEGSNR and 0.9 in MOS for a combined rate of 12.8 kbit/s have been found. These results confirm that adaptive coding is a promising technique that should be considered in the next generation of mobile speech communication systems in order to increase robustness to channel errors and channel utilization.

7.2 Suggestions for Future Work

Some suggestions for further work are as follows:

1. The performance of the adaptive coders in this thesis are evaluated under a Rayleigh flat fading channel. Further tests of the system performance in a frequency selective fading channel can be considered.

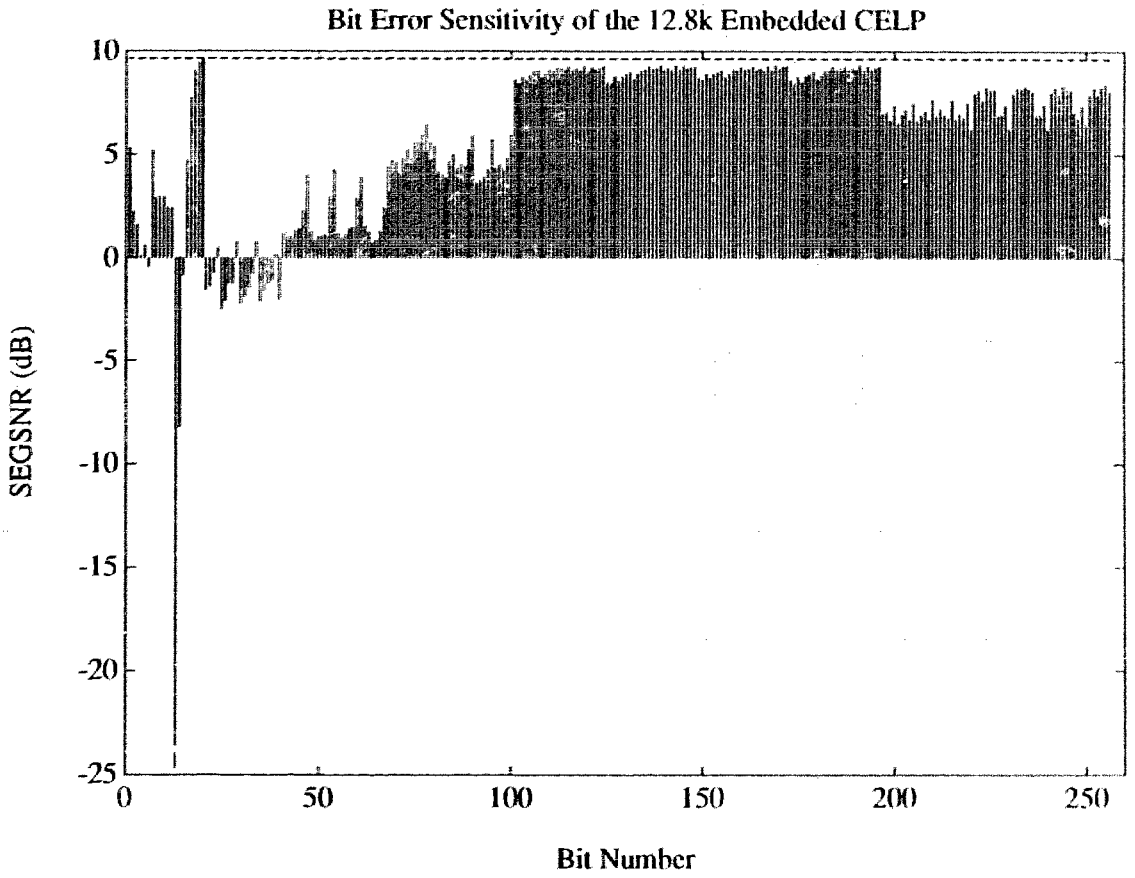
2. In this thesis, the combined rates are fixed to be either 9.6 or 12.8 kbit/s. Systems that are capable of varying the speech and channel coding rates without a restricted combined rate can be evaluated. They may be useful if variable rate transmission is allowed such as in a CDMA network.

3. Error concealment techniques for speech coders can be considered in order to improve the performance of combined coders under deep fades, which can result in multiple frame losses.

4. In light of the recent advancement in CDMA and power control techniques, the effect of power control on the performance of joint speech and channel coders can be considered.

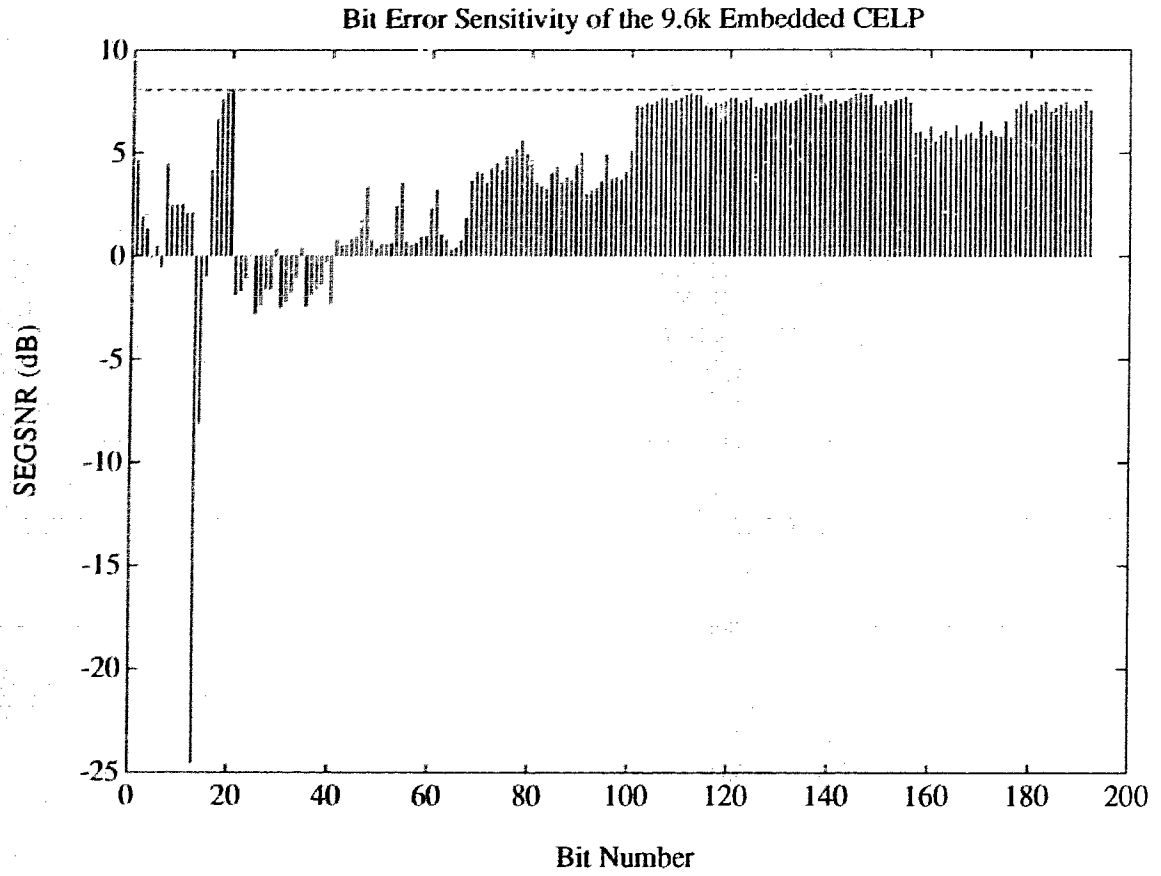
Appendix A

Bit Error Sensitivity of CELP Speech Coders



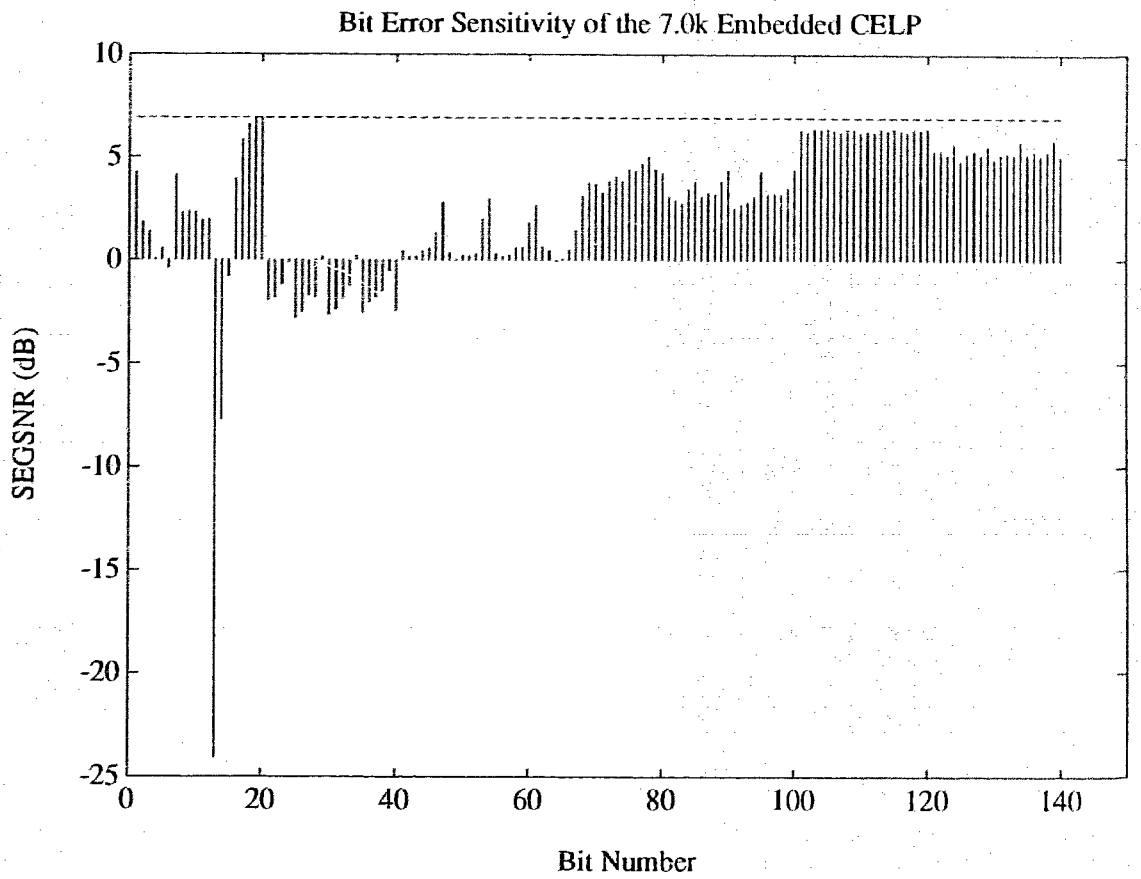
Parameters	Bit Numbers
LSP	1 - 12, 69 - 80
ACB	41 - 68
SCB	81 - 196
g_{frame}	13 - 20
GainVQ	21 - 40, 197 - 256

Figure A.1: Error sensitivity of the 12.8 kbit/s embedded CELP coder



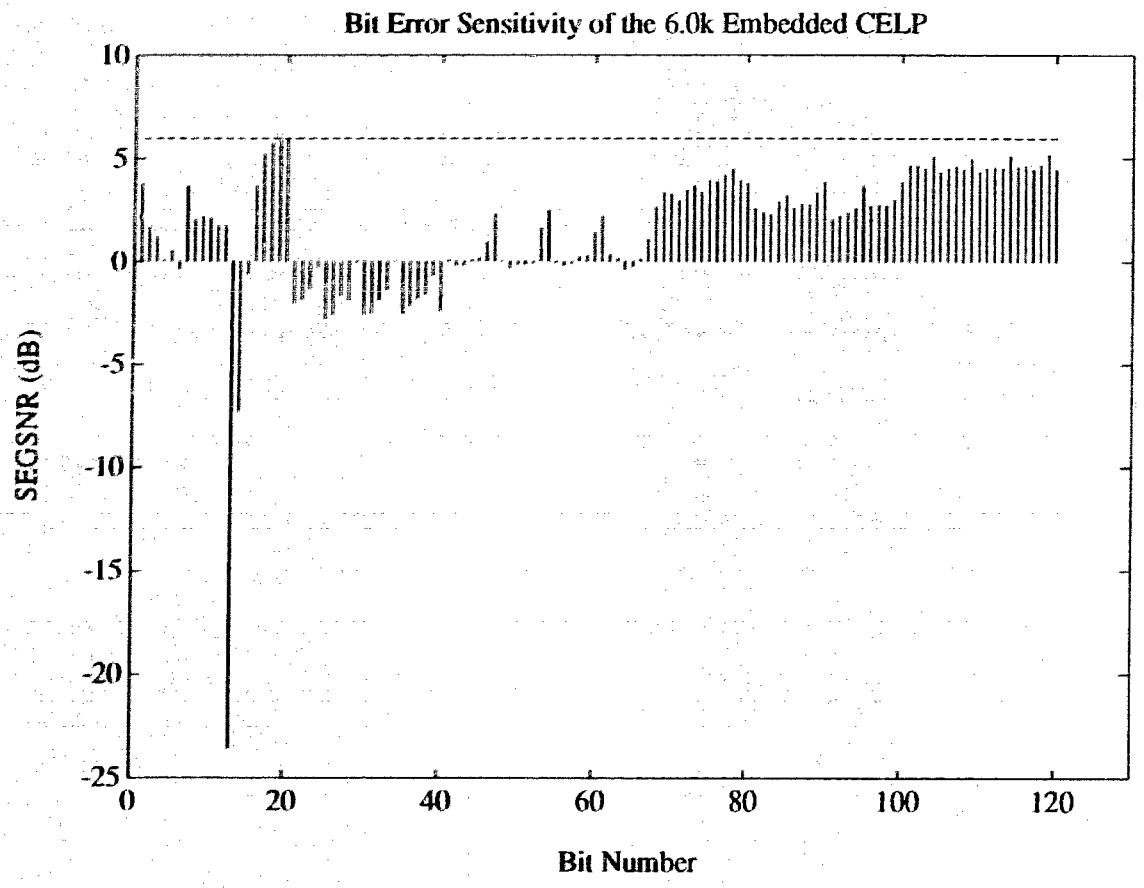
Parameters	Bit Numbers
LSP	1 - 12, 69 - 80
ACB	41 - 68
SCB	81 - 156
g_{frame}	13 - 20
GainVQ	21 - 40, 157 - 192

Figure A.2: Error sensitivity of the 9.6 kbit/s embedded CELP coder



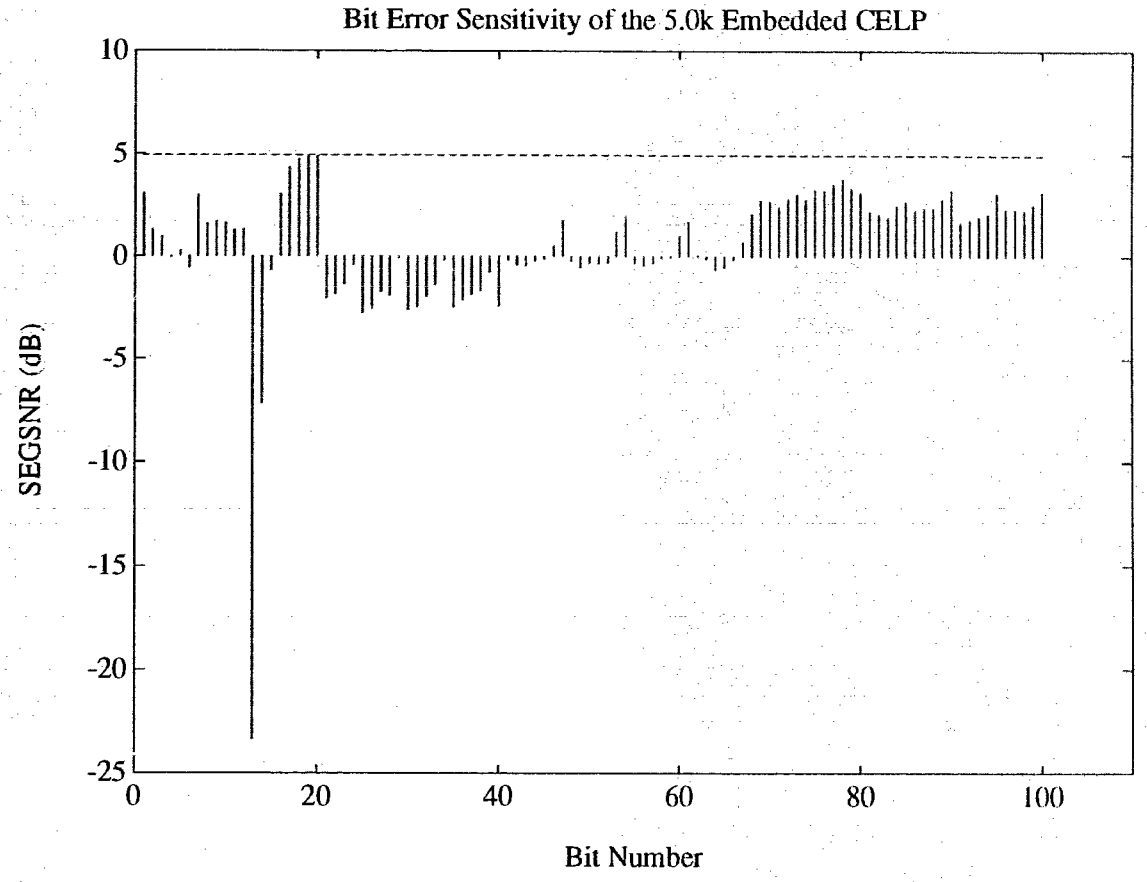
Parameters	Bit Numbers
LSP	1 - 12, 69 - 80
ACB	41 - 68
SCB	81 - 120
g_{frame}	13 - 20
GainVQ	21 - 40, 121 - 140

Figure A.3: Error sensitivity of the 7.0 kbit/s embedded CELP coder



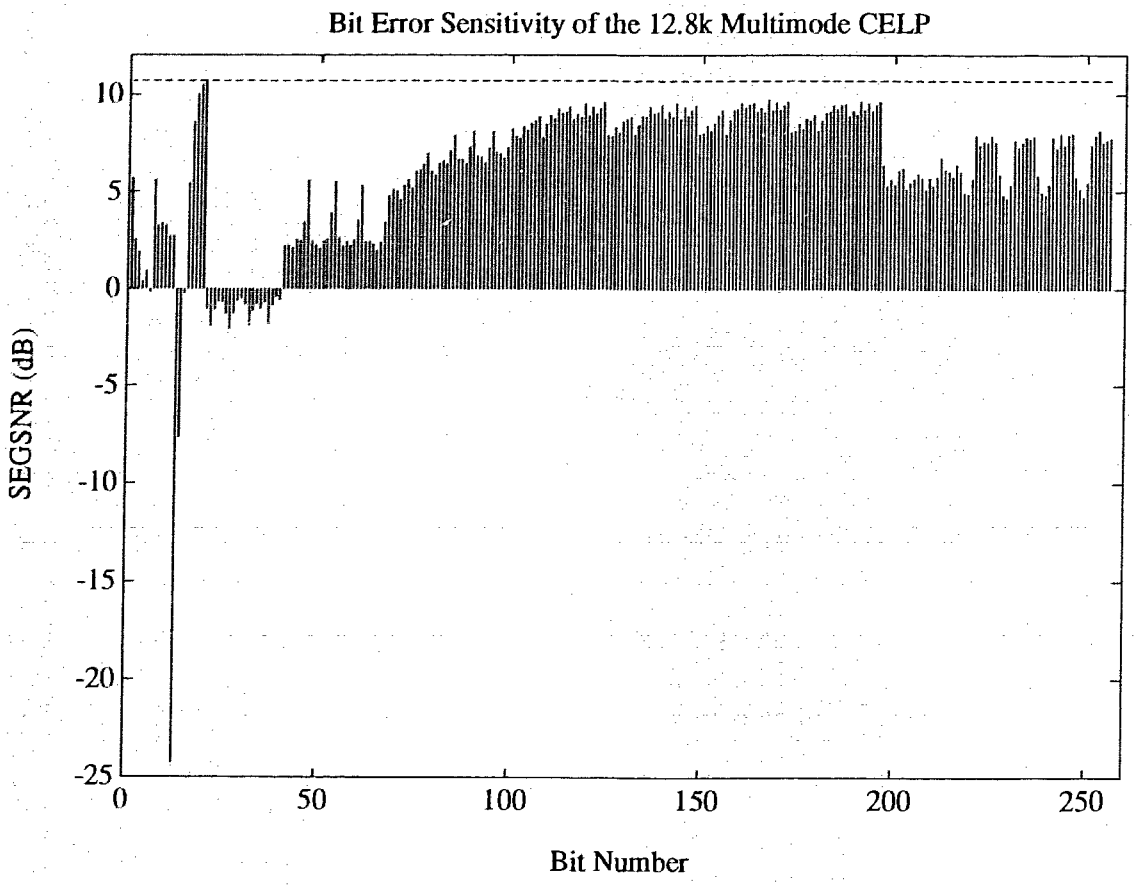
Parameters	Bit Numbers
LSP	1 - 12, 69 - 80
ACB	41 - 68
SCB	81 - 100
g_{frame}	13 - 20
GainVQ	21 - 40, 101 - 120

Figure A.4: Error sensitivity of the 6.0 kbit/s embedded CELP coder



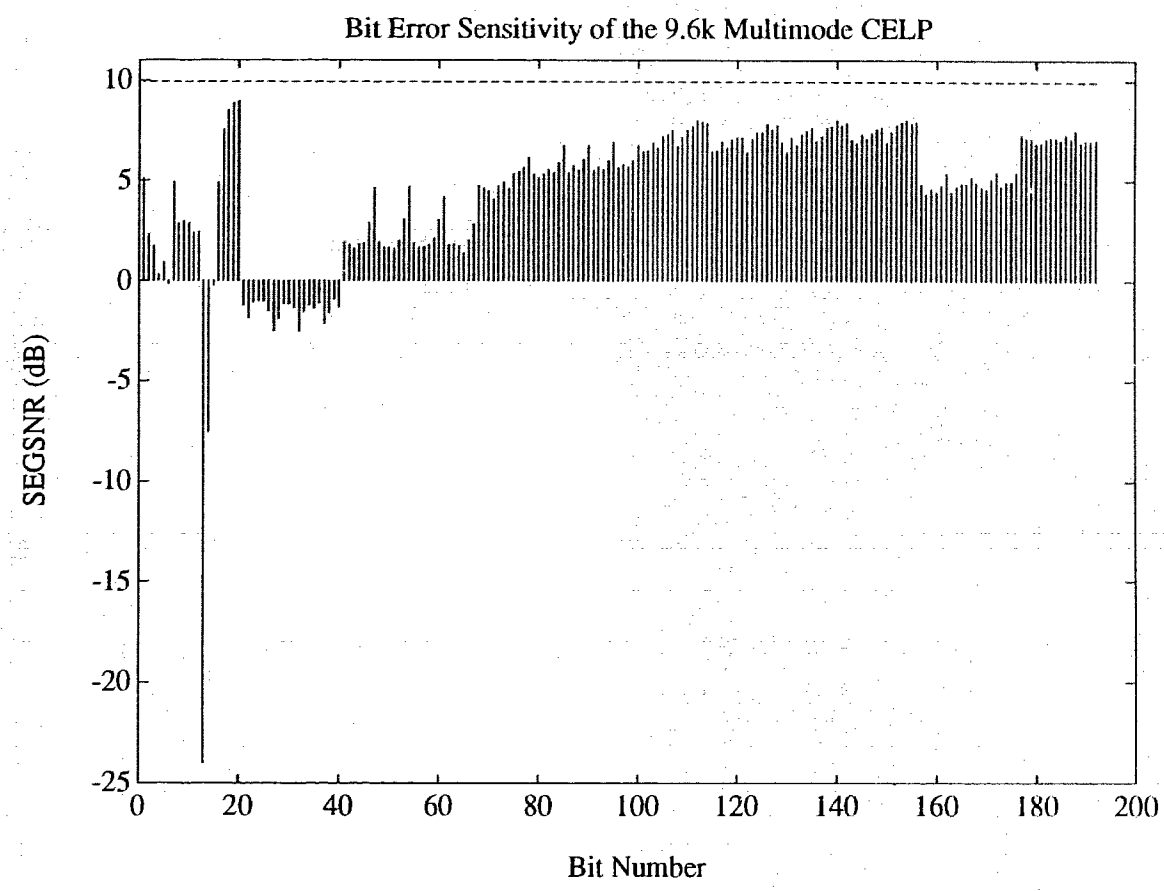
Parameters	Bit Numbers
LSP	1 - 12, 69 - 80
ACB	41 - 68
SCB	81 - 100
g_{frame}	13 - 20
GainVQ	21 - 40

Figure A.5: Error sensitivity of the 5.0 kbit/s embedded CELP coder



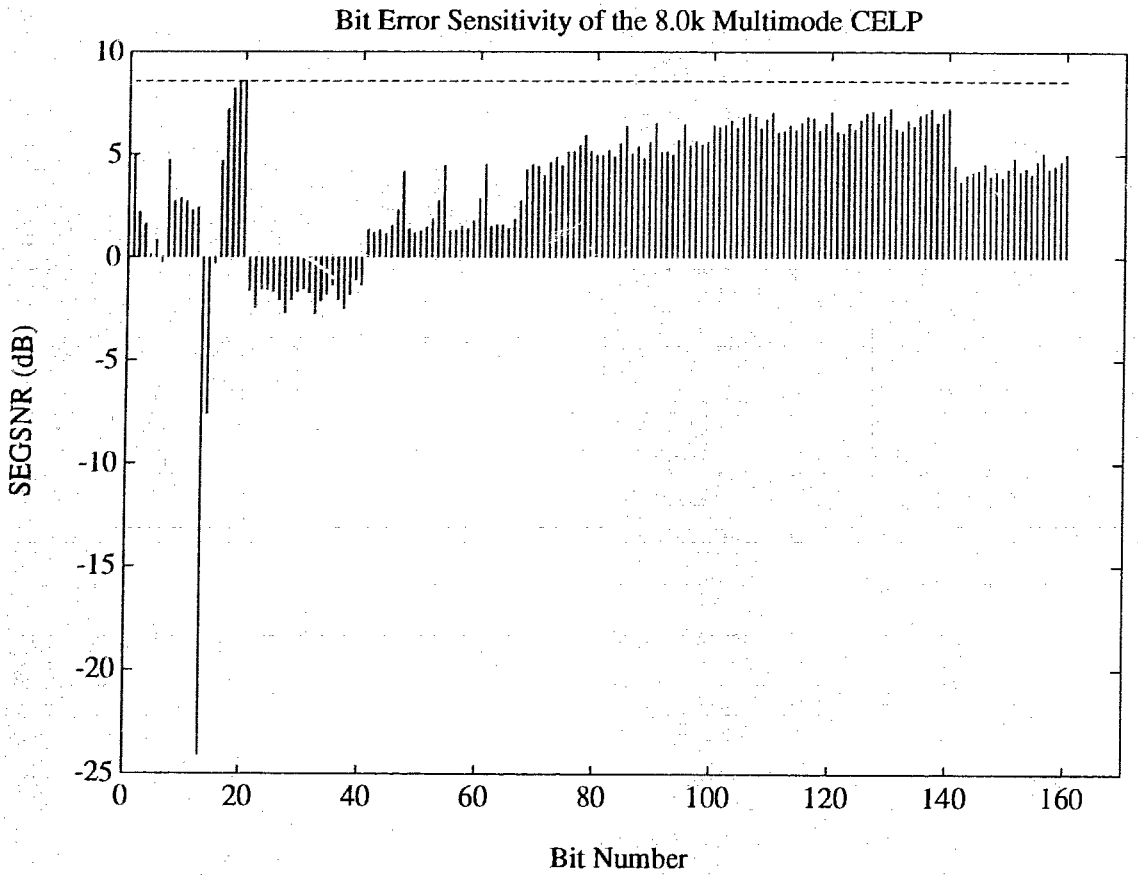
Parameters	Bit Numbers
LSP	1 - 12, 69 - 80
ACB	41 - 68
SCB	81 - 196
g_{frame}	13 - 20
GainVQ	21 - 40, 197 - 256

Figure A.6: Error sensitivity of the 12.8 kbit/s multimode CELP coder



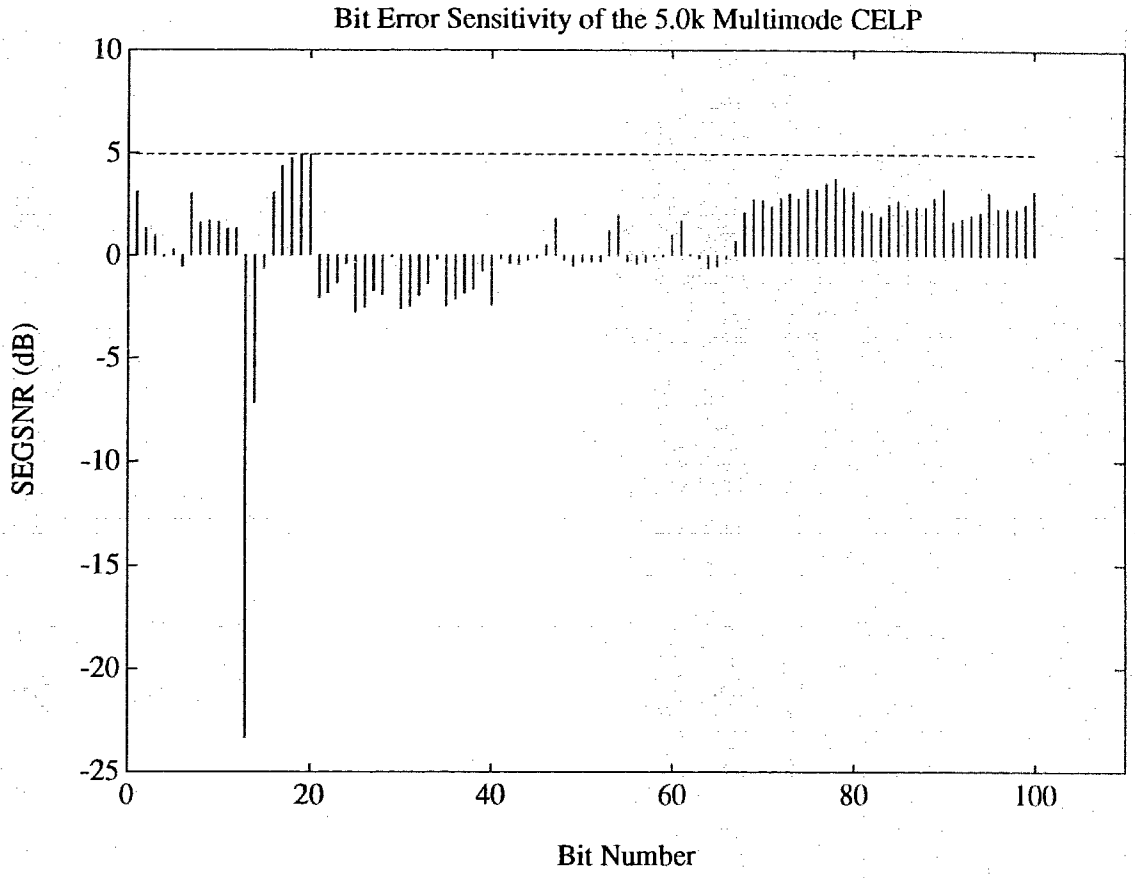
Parameters	Bit Numbers
LSP	1 - 12, 77 - 88
ACB	49 - 76
SCB	89 - 168
g_{frame}	13 - 20
GainVQ	21 - 48, 169 - 192

Figure A.7: Error sensitivity of the 9.6 kbit/s multimode CELP coder



Parameters	Bit Numbers
LSP	1 - 12, 69 - 80
ACB	41 - 68
SCB	81 - 140
g_{frame}	13 - 20
GainVQ	21 - 40, 141 - 160

Figure A.8: Error sensitivity of the 8.0 kbit/s multimode CELP coder



Parameters	Bit Numbers
LSP	1 - 12, 69 - 80
ACB	41 - 68
SCB	81 - 100
g_{frame}	13 - 20
GainVQ	21 - 40

Figure A.9: Error sensitivity of the 5.0 kbit/s multimode CELP coder

REFERENCES

[1] EIA/TIA, "EIA/TIA-553 mobile station - land station compatibility specifications," 1989.

[2] EIA/TIA, "IS-54 cellular system : dual-mode subscriber equipment - network equipment compatibility specification," Dec. 1989.

[3] EIA/TIA, "IS-95 mobile station - base station compatibility standard for dual-mode wideband spread spectrum cellular system," July 1993.

[4] ETSI/GSM, "Recommendation 06.10 : European digital cellular telecommunication system," Mar. 1992.

[5] V. Cuperman, "Speech coding," in *Advances in Electronics and Electron Physics*, vol. 82, Academic Press, Inc., 1991.

[6] CCITT Draft Recommendation H.261, *Video Codec for audio visual services at* $p \times 64$ *kbit/s*, Mar. 1990.

[7] C. Shannon, "A mathematical theory of communications," *The Bell System Technical Journal*, vol. 27, pp. 379–423, 1948.

[8] A. Kurtenbach and P. Wintz, "Quantization for nosiy channels," *IEEE Transactions on Communications*, 1969.

[9] N. Rydbeck and C. Sundberg, "Analysis of digital errors in nonlinear PCM," *IEEE Transactions on Communications*, vol. 24, pp. 59–65, Jan. 1976.

[10] N. Farvardin and V. Vaishampayan, "On the performance and complexity of channel-optimized vector quantizers," *IEEE Transactions on Information Theory*, pp. 155–160, Jan. 1991.

[11] V. Cuperman, F. Liu, and P. Ho, "Robust vector quantization for noisy channels using soft decision and sequential decoding," *To be published in IEE Journal*, 1994.

[12] N. Phamdo, N. Farvardin, and T. Moriya, "Combined source-channel coding of LSP parameters using multi-stage vector quantization," in *Speech and Audio Coding for Wireless and Network Applications* (B. Atal, V. Cuperman, and A. Gersho, eds.), Kluwer Academic Publishers, 1993.

[13] R. Cox, J. Hagenauer, N. Seshadri, and C. Sundberg, "Subband speech coding and matched convolutional channel coding for mobile radio channels," *IEEE Transactions on Signal Processing*, vol. 39, pp. 1717–1731, Aug. 1991.

[14] D. Goodman and C.-E. Sundberg, "Combined source and channel coding for variable-bit-rate speech transmission," *The Bell System Technical Journal*, vol. 62, pp. 2017–2036, Sept. 1983.

[15] B. Vucetic, "An adaptive coding scheme for time-varying channels," *IEEE Transactions on Communications*, vol. 39, pp. 653–663, May 1991.

[16] A. Gersho and R. Gray, *Vector Quantization and Signal Compression.* Kluwer Academic Publishers, 1992.

[17] S. Hong, "Combined speech and channel coding for mobile radio applications," Master's thesis, Simon Fraser University, 1993.

[18] W. Gardner, P. Jacobs, and C. Lee, "QCELP: A variable rate speech coder for CDMA digital cellular," in *Speech and Audio Coding for Wireless and Network Applications* (B. Atal, V. Cuperman, and A. Gersho, eds.), Kluwer Academic Publishers, 1993.

[19] A. Oppenheim and R. Schafter, *Discrete-time Signal Processing.* Prentice-Hall, 1989.

[20] N. Jayant and P. Noll, *Digital Coding of Waveforms.* Prentice-Hall, 1984.

[21] N. Jayant, "Signal compression: technology targets and research directions," *IEEE Journal on Selected Areas in Communications*, vol. 10, pp. 796–818, June 1992.

[22] P. Papamichalis, *Practical Approaches in Speech Coding.* Prentice-Hall, Inc., 1987.

[23] J. Markel and A. Gray, "A linear prediction vocoder simulation based on autocorrelation method," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 23, no. 2, pp. 124–134, 1974.

[24] M. Schroeder and B. Atal, "Code excited linear prediction (CELP) : high quality speech at very low bit rate," in *Proceedings ICASSP'85*, pp. 937–940, 1985.

[25] G. Davidson and A. Gersho, "Complexity reduction methods for vector excitation," in *Proceedings ICASSP'86*, 1986.

[26] S. Dimolitsas, "Subjective assessment methods for the measurement of digital speech coder quality," in *Speech and Audio Coding for Wireless and Network Applications* (B. Atal, V. Cuperman, and A. Gersho, eds.), Kluwer Academic Publishers, 1993.

[27] W. Lee, *Mobile Communications Design Fundamentals.* John Wiley and Sons, Inc., 2 ed., 1993.

[28] S. Lin, D. Costello, and M. Miller, "Automatic repeat request error control schemes," *IEEE Communications Magazine*, vol. 22, no. 12, pp. 5–17, 1984.

[29] Q.Yang and V. Bhargava, "Delay and coding gain analysis of a truncated Type II Hybrid ARQ protocols," *IEEE Transactions on Vehicular Technology*, vol. 42, pp. 22–31, Feb. 93.

[30] D. Sereno, "Frame substitution and adaptive post-filtering in speech coding," in *EuroSpeech'92*, pp. 595–598, 1992.

[31] M. Yong, "Study of voice packet reconstruction methods applied to CELP speech coding," in *Proceedings ICASSP'92*, pp. 125–128, 1992.

[32] H. Su and P. Mermelstein, "Improving the speech quality of cellular mobile systems under heavy fading," in *Proceedings ICASSP'92*, pp. 121–124, 1992.

[33] T. Leung, "Voice frame reconstruction methods for CELP speech coders," Master's thesis, Carleton University, 1993.

[34] A. Husain and V. Cuperman, "Classification and spectral extrapolation based packet reconstruction for low-delay speech coding," in *Submission to Globecom'94*, 1994.

[35] S. Atungsiri, A. Kondoz, and B. Evans, "Error control for low-bit-rate speech communication systems," *IEE Proceedings-I*, vol. 140, pp. 97–103, Apr. 1993.

[36] J. Proakis, *Digital Communications*. McGraw-Hill, 1989.

[37] G. Clark and J. Cain, *Error correction coding for digital communications*. Plenum Press, 1981.

[38] S. Lin and D. Costello, *Error control coding: fundamentals and applications*. Prentice-Hall, 1983.

[39] J. Hagenauer, "Forward error correction coding for fading compensation in mobile satellite channels," *IEEE Journal on Selected Areas in Communications*, vol. 5, pp. 215–225, Feb. 1987.

[40] J. Hagenauer, "Rate-compatible punctured convolutional codes (RCPC codes) and their applications," *IEEE Transactions on Communications*, vol. 36, pp. 389–400, Apr. 1988.

[41] J. Hagenauer, N. Seshadri, and C. Sundberg, "The performance of rate compatible punctured convolutional codes for digital mobile radio," *IEEE Transactions on Communications*, vol. 38, pp. 966–980, June 1990.

[42] J. Cavers, "An analysis of pilot symbol assisted modulation for rayleigh fading channels," *IEEE Transactions on Vehicular Technology*, vol. 40, pp. 686–693, Nov. 1991.

[43] J. Chuang, "The effects of delay spread on 2-PSK, 4-PSK, 8-PSK, and 16-QAM in a portable radio environment," *IEEE Transactions on Vehicular Technology*, vol. 38, pp. 43–45, May 1989.

[44] P. Mermelstein, A. Jalali, and H. Leib, "Integrated services on wireless multiple access networks," in *Proceedings ICC'93*, pp. 863–867, 1993.

[45] D. Raychaudhuri and N. Wilson, "Multimedia transport in next-generation personal communication networks," in *Proceedings ICC'93*, pp. 858–862, 1993.

[46] A. Gersho and E. Paksoy, "An overview of variable rate speech coding for cellular networks," in *Speech and Audio Coding for Wireless and Network Applications* (B. Atal, V. Cuperman, and A. Gersho, eds.), Kluwer Academic Publishers, 1993.

[47] D. Goodman, "Cellular packet communications," *IEEE Transactions on Communications*, vol. 38, pp. 1272–1280, Aug. 1990.

[48] Qualcomm, Inc., "Proposed EIA/TIA interim standard - wideband spread spectrum digital cellular system dual-mode mobile station - base station compatibility standard," *Submitted to the TIA TR45.5 Subcommittee*, Apr. 1992.

[49] M. Jeruchim, P. Balaban, and K. Shanmugan, *Simulation of communication systems.* plenum press: New York, 1992.

[50] P. Kabul and R. Ramachandran, "The computation of Line Spectral frequencies using Chebyshev polynomials," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 34, pp. 1419–1426, Dec. 1986.

[51] B. Bhattacharya, W. LeBlanc, S. Mahmoud, and V. Cuperman, "Tree-searched multistage vector quantization of LPC parameters for 4 kb/s speech coding," in *Proceedings ICASSP'92*, 1992.

[52] S. Singhal and B. Atal, "Improving the performance of multipulse coders at low bit rates," in *Proceedings ICASSP'84*, pp. 1.3.1–1.3.4, 1984.

[53] R. Rose and T. Barnwell, "Design and performance of an analysis by synthesis class of predictive speech coders," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 38, pp. 1489–1503, 90.

[54] W. Loo, "Real time implementation of an 8.0/16.0 kbit/sec vector quantized code excited linear prediction speech coding algorithm using the TMS 320C51 digital signal processor," Master's thesis, Simon Fraser University, 1993.

[55] P. Lupini, "TN-001v4: vector quantization of gains in CELP." Technical notes, Aug. 1993.

[56] P. Brady, "A technique for investigating on-off patterns of speech," *The Bell System Technical Journal*, vol. 44, pp. 1–22, 65.

[57] B. Masnick and J. Wolf, "On unequal error protection codes," *IEEE Transactions on Information Theory*, vol. 13, pp. 600–607, Oct. 1967.

[58] D. Fung, "Error performance of PSK signals transmitted over correlated Rayleigh fading channels," Master's thesis, Simon Fraser University, 1991.

[59] W. Lee, "Estimate of local average power of a mobile radio signal," *IEEE Transactions on Vehicular Technology*, vol. 34, pp. 22–27, Feb. 1985.