AN ABSTRACT OF THE THESIS OF

Meeta Bhutani for the degree of Master of Science in Electrical and Computer Engineering presented on August 31,1998. Title: Comparison of DPCM and Subband Codec Performance in the Presence of Burst Errors.

Abstract approved : _____
Virginia L. Stonick

This thesis is a preliminary study of the relative performance of the major speech compression techniques, Differential Pulse Code Modulation (DPCM) and Subband Coding (SBC) in the presence of transmission distortion. The combined effect of the channel distortions and the channel codec including error correction is represented by bursts of bit errors. While compression is critical since bandwidth is scarce in a wireless channel, channel distortions are greater and less predictable. Little to no work has addressed the impact of channel errors on perceptual quality of speech due to the complexity of the problem. At the transmitter, the input signal is compressed to 24 kbps using either DPCM or SBC, quantized, binary encoded and transmitted over the burst error channel. The reverse process is carried out at the receiver. DPCM achieves compression by removing redundant information in successive time domain samples, while SBC uses lower resolution quantizer to encode frequency bands of lower perceptual importance. The performance of these codecs is evaluated for BERs of 0.001 and 0.05, with the burst lengths varying between 4 and 64 bits. Two different speech segments - one voiced and one unvoiced are used in testing. Performance measures include two objective tests - signal to noise ratio (SNR) & segmental SNR, and a subjective test of perceptual quality - the Mean Opinion Score (MOS). The results

obtained show that with a fixed BER and increasing burst length in bits, the total errors reduce in the decoded speech thereby improving its perceptual quality for both DPCM and SBC. Informal subjective tests also demonstrate this trend as well as indicate distortion in DPCM seemed to be less perceptually degrading than SBC.

Comparison of DPCM and Subband Codec Performance in the Presence of Burst Errors

by

Meeta Bhutani

A THESIS

submitted to

Oregon State University

in partial fulfillment of
the requirements for the
degree of

Master of Science

Presented August 31, 1998
Commencement June 1999

Master of Science thesis of Meeta Bhutani presented on August 31, 1998

APPROVED:

Redacted for Privacy

Major Professor, representing Electrical and Computer Engineering

Redacted for Privacy

Head of Department of Electrical and Computer Engineering

Redacted for Privacy

Dean of Graduate school

I understand that my thesis will become part of the permanent collection of Oregon State University libraries. My signature below authorizes release of my thesis to any reader upon request.

Redacted for Privacy

Meeta Bhutani, Author

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

LIST OF FIGURES

LIST OF FIGURES (Continued)

LIST OF FIGURES (Continued)

LIST OF TABLES

Comparison of DPCM and Subband Codec Performance in the Presence of Burst Errors

Chapter 1: Introduction

In recent years, telecommunications products have proliferated and become increasingly common in the daily lives of consumers. Wireless and mobile products and services are in increasing demand, and voice communications remains a primary and preferred means of human-to-human communication, as evidenced by the increasingly widespread use of cellular phones. Higher consumer demand for wireless products and services places more pressure on scarce resources, particularly bandwidth. As a result, service providers are anxious to take advantage of technologies that minimize bandwidth requirements while providing acceptable quality to customers. These so-called compression technologies are critical components in wireless and mobile standards, including, e.g., IS-54 [34].

In general, however, increasing compression also increases sensitivity to distortions caused during transmission. For wireless and mobile communications - as opposed to transmission over optical fiber or coaxial cable - this tradeoff results in a particular dilemma: while compression is critical since bandwidth is scarce, channel distortions are greater and less predictable. For example, changing weather, climate and terrain all can cause errors in the received signal. Some, but not all, of these errors can be corrected in the receiver. The goal in this thesis is to provide an evaluation of the relative performance of the major speech compression techniques in the presence of uncorrectable errors caused by distortion arising during the transmission process.

1.1 Background : Digital Communication Systems

Compression technologies are used in a *digital* communication system. The term digital here implies that the information signal to be transmitted - speech in this case - is discrete in both time and amplitude and thus can be represented by a sequence of bits (1's and 0's). The goal in the digital communications system is to provide a service of acceptable quality (termed toll quality in telephony) and cost to the customer at the receiving end, while reducing the bit rate required for transmission to be as low as possible.

The major components of a typical digital communications system are shown in Figure 1.1 and described briefly below. The information source is assumed to generate a relatively high quality digital message signal that has already undergone analog-to-digital (A/D) conversion. In this work, the message signal is a speech signal assumed to be of toll quality. Toll quality speech requires a minimum sampling rate of 8kHz (8000 samples per second) and the use of an 8-bit quantizer (8 bits per sample), resulting in a bit rate of 64 kbps (kilobits per second) [34].

The primary function of the source encoder is compression; i.e. to reduce the bit rate required for transmission by removing redundancy in the source, speech in this case. The source decoder reconstructs speech from the transmitted signal. Collectively encoder/decoder pairs are called codecs. The focus in this thesis is on the performance of speech compression codecs (encoder-decoder) which are part of the source encoder/decoder pair. Source encoders may also include encryption for privacy or security. Since the focus of this thesis is on compression, encryption is not used.

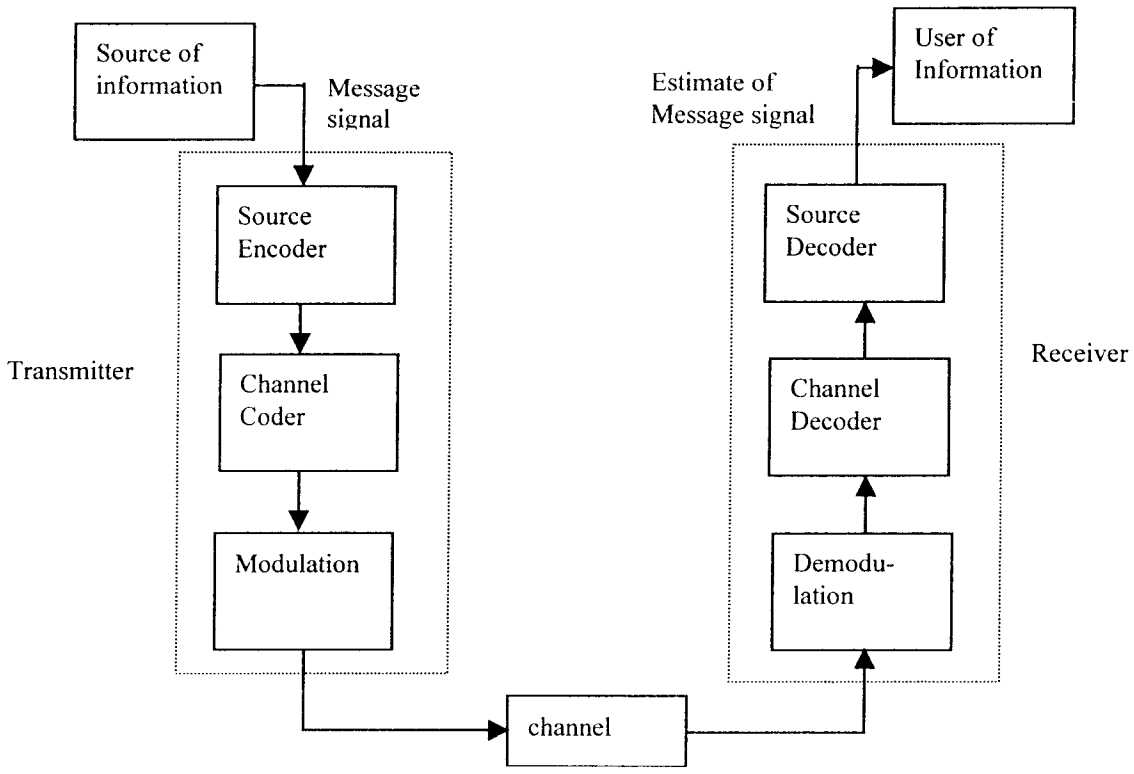


Figure 1.1: Block Diagram of a Digital Communication System

The purpose of the channel encoder is to provide robustness in the received signal to errors caused by distortion in the transmission channel. While source encoders remove redundancy to reduce the bit rate, channel encoders add some redundancy back into the signal to provide for error detection and correction at the receiver. Thus the bit rate out of the channel encoder is higher than the bit rate in, but typically still lower than the original signal bit rate. The type of channel encoding, as the name implies, depends on the type and amount of distortion expected in the channel. While there are many different types of channel codecs, the following qualitative description is generally accepted as valid. Channel codecs provide virtually perfect reconstruction of the transmitted bit sequence, unless the cumulative effect of errors causes the error correction to fail, resulting in a

burst of bit errors into the source decoder. In digital television, this effect is called the 'Cliff effect' as it results in a sudden and complete loss of picture from one of nearly perfect quality. In this thesis, the combined effect of the transmission channel, including the channel codec, is represented by bursts of bit errors as shown in Figure 1.2.

The modulator converts the bits out of the channel encoder into symbols represented by analog waveforms that are appropriate for transmission over the channel. Binary information may be encoded in different signal levels, phases, and/or frequencies. The results in this thesis are not specific to any particular type of modulator/demodulator as the cumulative effect of the channel codec, modulator/demodulator and transmission channel is reflected in the properties of bit errors into the source decoder.

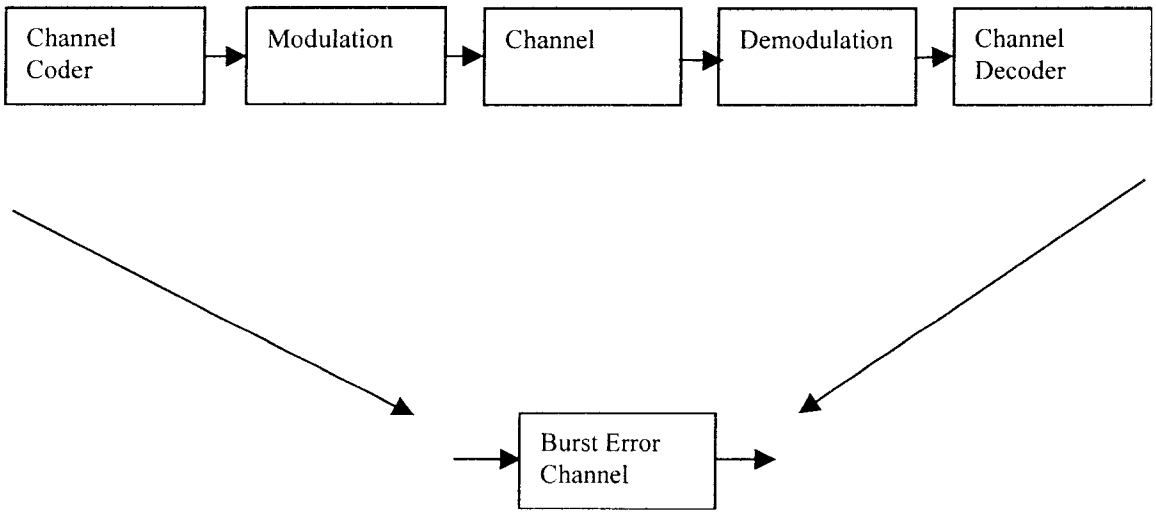


Figure 1.2: Combining of Blocks as in this thesis

1.2 Speech Compression Codec Performance

Compression is of two types: lossy and lossless. Lossless Compression takes advantage of statistical properties of the encoded signals to reduce the bit rate, as in Morse code, Hamming and Lempel-Ziev codes [11, 25]. For example, it is used for the compression of financial data where no information should be lost. Lossy compression is used for voice and video where precision is less important than perceptual quality. In this thesis, the focus is on lossy compression, which adds distortion in a controlled manner to minimize the perceptual degradation caused by reducing the bit rate. The two major types of lossy compression for speech are predictive coding, such as Differential Pulse Code Modulation (DPCM) and transform coding such as Subband Coding (SBC). Since most of the compression technologies and standards use a combination of these basic types, these two are the focus of this thesis.

A typical quantitative measure of compression codec quality is Signal-to-Quantization Noise Ratio (SQR) (which does not always correlate well with perceived quality but is still widely used). Note that the SQR evaluates performance only with respect to the loss introduced by the compression and is not impacted by transmission distortions. In contrast to source coders, channel codec performance is quantified by the number of bit errors that can be detected and corrected, or the bit error rate (BER) out given the BER in. Thus channel codec performance measures are unrelated to the quality of the speech signal output.

Little to no work has addressed the impact of channel errors on perceptual quality of speech due to the complexity of the problem. In this preliminary study, the channel and modulation have been collectively represented as bursts of errors. The channel codec

performance in this thesis has been fixed at bit error rates of 0.001 and 0.05. While typical rates out of a perfect channel codec are of the order of 10^{-20} [33], we use 0.001 and 0.05 to account for the cumulative effect of errors occurring in an overloaded channel codec. The bit error rates have been chosen so that they are significant enough to cause perceptual degradation in the final output.

There exist different objective measures of coder quality, which have the general nature of signal to noise ratio (SNR). In coding of communications signals such as speech and video, subjective measures to evaluate perceptual quality are also important. In this thesis, Signal to Noise Ratio and Segmental Signal to Noise Ratio (SEGSNR) have been used as objective measures of coder quality. Informal subjective testing has been done by calculating the Mean Opinion Score (MOS). While SNR gives the average signal to error power, segmental SNR tries to account for the impact of time varying SNR performance and so is a more suited perceptual measure. MOS is a purely subjective evaluation and does not distinguish the type of distortion. All the three are widely accepted as measures of coder quality [1, 2, 7, 14, 28, 29, 31].

1.3 Prior Work/ Literature Search

Many researchers have studied various kinds of low bit rate source coders to achieve compression, such as, Differential Pulse Code Modulation, Delta Modulation (DM), Subband Coding, Code Excited Linear Prediction (CELP), Vector Sum Excited Linear Prediction (VSELP) [1, 2, 5, 8, 9, 10, 20, 21, 22]. The performance of these coders has been researched extensively for Additive White Gaussian Noise (AWGN) channels [6, 16, 18, 23, 25, 26]. Burst Error channels have been studied with emphasis on burst

error correction [14, 15, 19, 27, 32, 33]. Even though performance of low bit rate codecs has been extensively researched, to date the effect of burst errors on low bit rate speech has not been thoroughly investigated. This thesis is a preliminary study of the effects of burst errors on two different kinds of speech compression codecs. As indicated, a burst error model provides a reasonable model of the combined effect of the channel and error corrector.

1.4 The Specific Problem

The problem that has been dealt within this thesis is an exploration of the effect of burst errors, both analytically and through simulations, on the quality of encoded and decoded speech using Differential Pulse Code Modulation and Subband Coding Algorithms. Both the simulated codecs have the same bit rate of 24 kbps and reflect the two fundamental techniques used in speech compression standards. Most standards for lower bit rate typically use combination of differential and transform coding. While DPCM tries to remove the redundant information in successive time domain samples, SBC uses a lower resolution quantizer for frequency bands in which the perceptual impact is less.

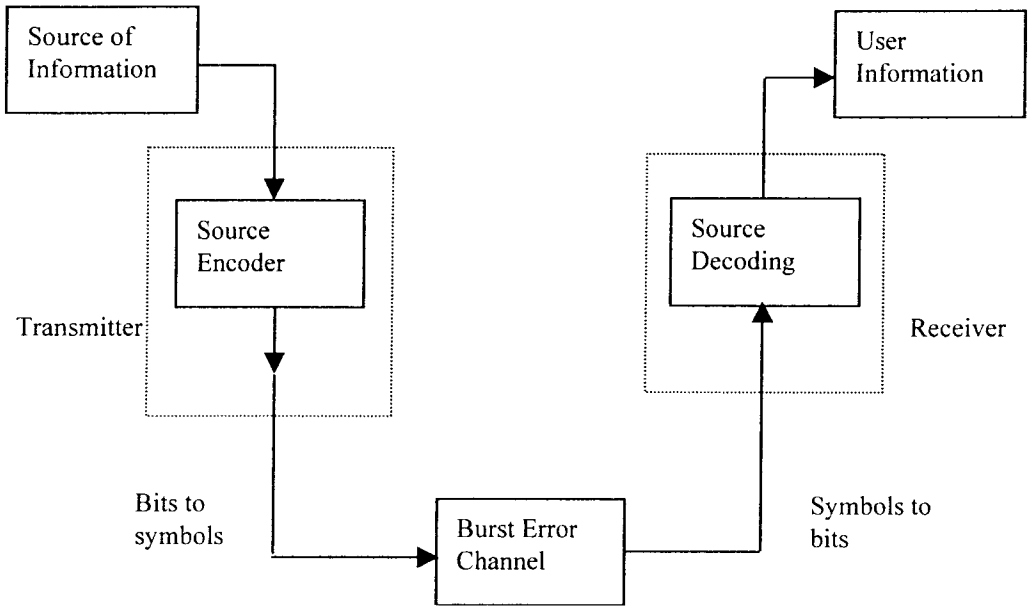


Figure 1.3: The Block Diagram as per this thesis

1.5 Thesis Organization

The organization of this thesis is as follows. The testing and development of the 24 kbps DPCM and SBC codecs are described in Chapter 2. Chapter 3 describes the burst error channel model used, and the corresponding analysis of the burst error performance of DPCM and Subband. A description of the performance measures used together with the simulation results for DPCM and SBC follow in Chapter 4. Chapter 5 summarizes results, implications, and suggestions for further research.

CHAPTER 2: Codec Simulation

2.1 The Subband Codec

2.1.1 Introduction

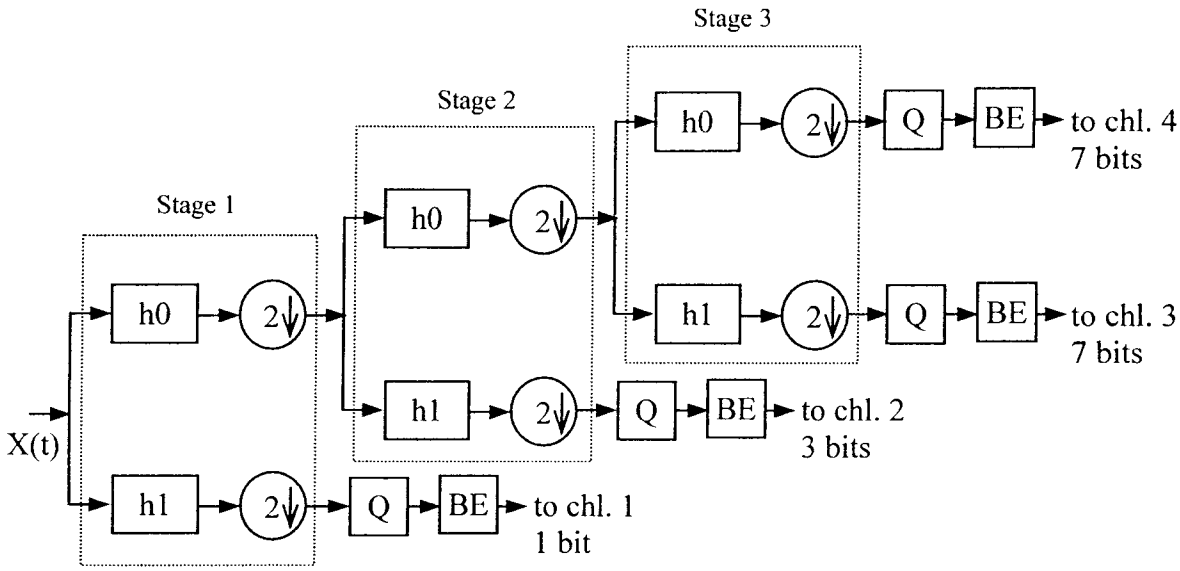
Subband Coding is a transform coding technique in which the speech signal is filtered into a number of subbands and each subband signal is separately encoded into a digital format. As with any digital encoding and compression method, the goal is to reduce the number of bits required in transmission while still preserving perceptual quality of the speech at the receiver. In SBC, this goal is achieved by using different number of bits with more quantization noise where it causes less perceptual degradation. The number of bits used in the encoding process differs for each subband signal, with more (fewer) bits assigned to subbands that are more (less) perceptually important. Since most of the speech energy is contained in the lower frequencies, the lower frequency bands are encoded using more bits than the high frequency bands. By encoding each subband individually, the quantization noise is confined within that subband. The output bit streams from each encoder are then multiplexed and transmitted.

At the receiver, demultiplexing is performed followed by decoding each subband data signal. The sampled subband signals are then combined to yield the recovered speech signal. The effect of subband coding on signal quality with respect to quantization noise and single bit channel errors has been reasonably well studied, for examples [1, 2, 9, 10].

The focus in this thesis is on the impact of burst errors - as are likely to occur out of error correction devices in wireless communications - on the quality of the received signal. In this chapter, the specific subband codec structure used here is developed and its simulation performance verified. This resulting codec will be used in the comparison studies in Chapters 3 and 4.

2.1.2 System Overview

A subband encoder comprises multiple stages as illustrated in Figure 2.1. In each stage, the input signal band is split into two equal frequency bands, comprising high and low frequencies respectively. Filters in Figure 2.1 are designated by their unit impulse response. The sampling rate at the output of each stage is halved, as indicated by the down arrow. This decimation does not result in aliasing distortion as the bandwidth of each output signal is half of the original.



h0 : Low Pass Filter Impulse Response
h1: High Pass Filter Impulse Response
Q : Quantizer
BE: Binary Encoder
chl : Channel

Figure 2.1: Subband Coder for Encoding the Speech Signal

The frequency domain representations illustrating what happens at each stage of the encoder are shown in Figure 2.2 and can be described as follows. Let the input signal be a speech signal confined to B = 4000 Hz sampled at the Nyquist rate of 8000 samples per second; i.e. Fs = 8000 in Figure 2.2. During the first filtering operation or "stage 1" in Figure 2.1, the input speech signal is split into two equal bandwidth signals: a low-pass signal in the frequency band ($0 \leq F < F_s/4$) and a high pass signal in the frequency band ($F_s/4 < F < Fs/2$) as shown in Figure 2.2(a). Next, the low-pass signal from the first stage is split into two signals having equal bandwidth: one signal compressing the lower half of frequencies in the band ($0 < F < F_s/8$) and a second signal compressing the higher

frequencies in the band ($F_s/8 < F < F_s/4$) as shown in Figure 2.2(b). In the third and final stage, the low-pass signal from the second stage is split into two equal bandwidth signals as shown in Figure 2.2(c). Thus the signal is subdivided into four frequency bands

It is important to note here that each subband filter produces Fs samples/sec - even though the bandwidth of each filter is less than the full bandwidth of the speech signal. To prevent increasing the number of samples to be transmitted above that required, the filter output is down-sampled according to the ratio of the original bandwidth B to the subband bandwidth. Note that no compression is achieved by these decimation operations.

Compression is achieved by using fewer bits to encode samples in the less perceptually important, higher frequency bands. The signal in each channel, ch_i is quantized into 2^{bi} levels (Quantizer Q), each of which is converted to b_i bits using the binary encoder (BE).

To calculate the total bit rate of the encoder consider that the bit rate per channel R_i is

$$R_i = b_i * Fs_i \tag{2.1}$$

Where b_i = Number of bits/sample & Fs_i = Number of samples per second. Then the total bit rate is

$$R = \sum_{i=1}^{L} R_i$$

Where L = number of channels, four in this case.

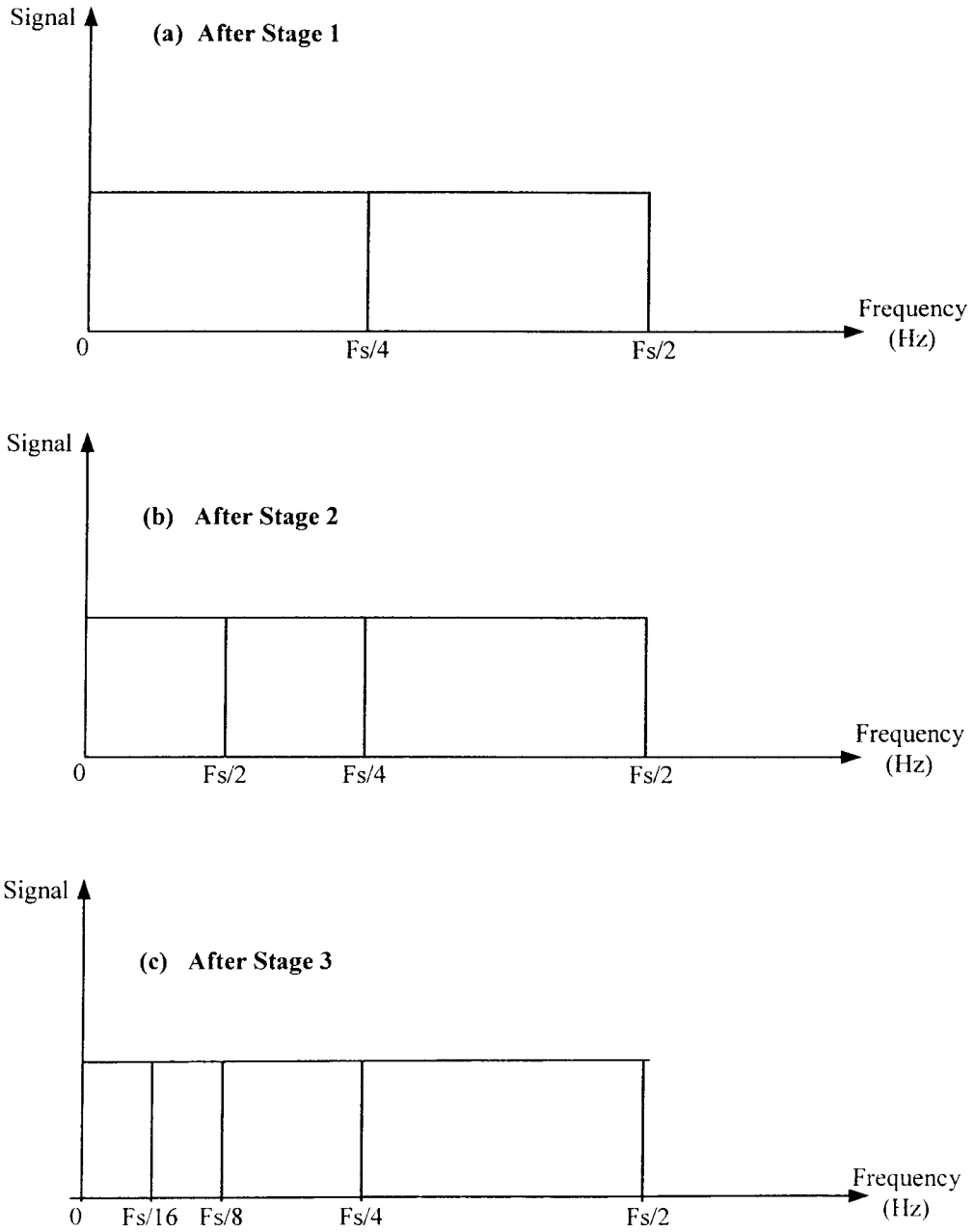


Figure 2.2: Subdivision of Signal into four frequency bands

Table 2.1 illustrates the number of bits assigned to each channel for the subband encoder used in this thesis.

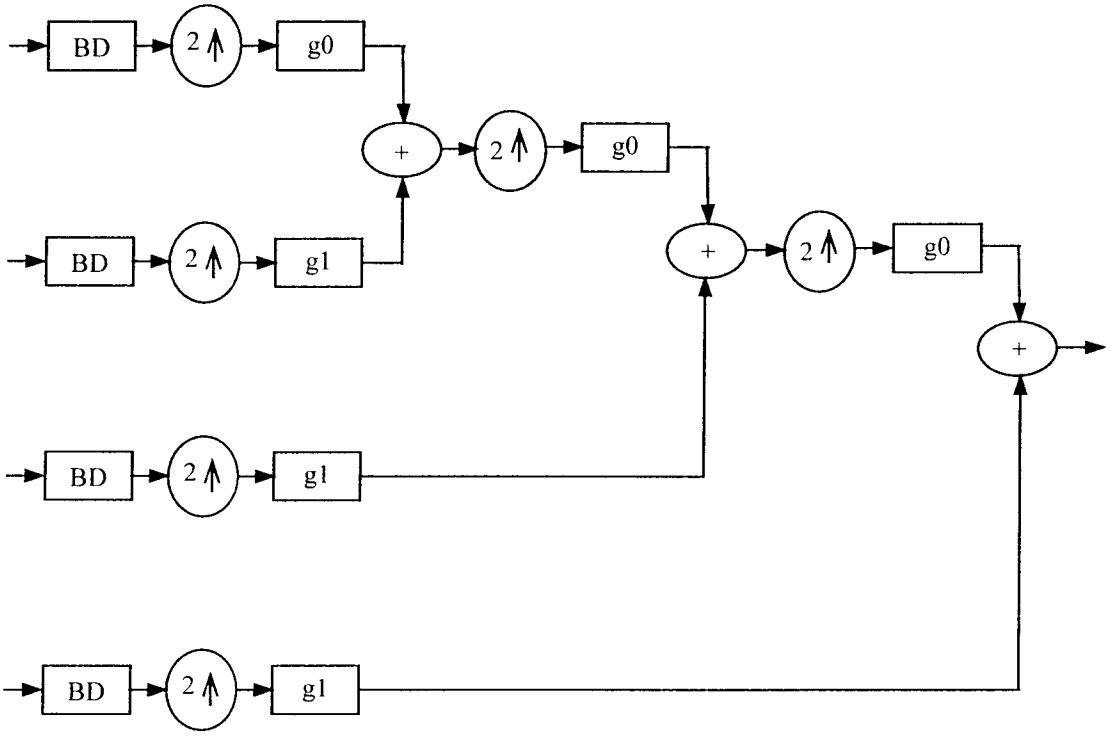
Channel Number	Freq. Band (Hz)	Fs(Hz)	No. of bits/sample	Bit Rate per channel
1	2000-4000	4000	1	4 kbps
2	1000-2000	2000	3	6 kbps
3	500-1000	1000	7	7 kbps
4	0-500	1000	7	7 kbps

Table 2.1: Channel bit assignments

Thus the total bit rate can be computed as the sum of bit rates per channel resulting in a bit rate out of the encoder of 24 kbps. After quantization and binary encoding, the information from each channel is multiplexed together into frames. Each frame comprises of 4 samples from channel 1 (1 bit per sample), 2 samples from channel 2 (3 bits per sample), 1 sample from channel 3 (7 bits per sample), 1 sample from channel 4 (7 bits per sample).Thus [1 1 1 1 2 2 3 4] is the composition of the frame where the numbers denote samples from the given channel number. Each frame comprises 24 information bits. Frames are transmitted at 1000 frames per second, yielding the expected 24 kbps.

At the receiver, the aim is to reconstruct the original speech signal from the subband signal with minimal distortion for a given transmission bit rate. Figure 2.3 shows the decoding for the subband encoded speech signal, which is basically the reverse of the encoding process. The binary decoder (BD) converts bits back into sample values, typically using a look up table. Up sampling, denoted by up arrows, is used to convert the

signals back to 8 kHz speech in stages. Filters are denoted by their impulse responses and are used to filter noise and aliasing distortion.



BD : Binary Decoder
g0 : Reconstruction low-pass filter impulse response
g1 : Reconstruction high-pass filter impulse response

Figure 2.3: Decoding of Subband Encoded Signal

It is important to note that the decimation and interpolation processes can result in aliasing distortion, but can be avoided by careful design of the filters h0(n), h1(n), g0(n), g1(n).

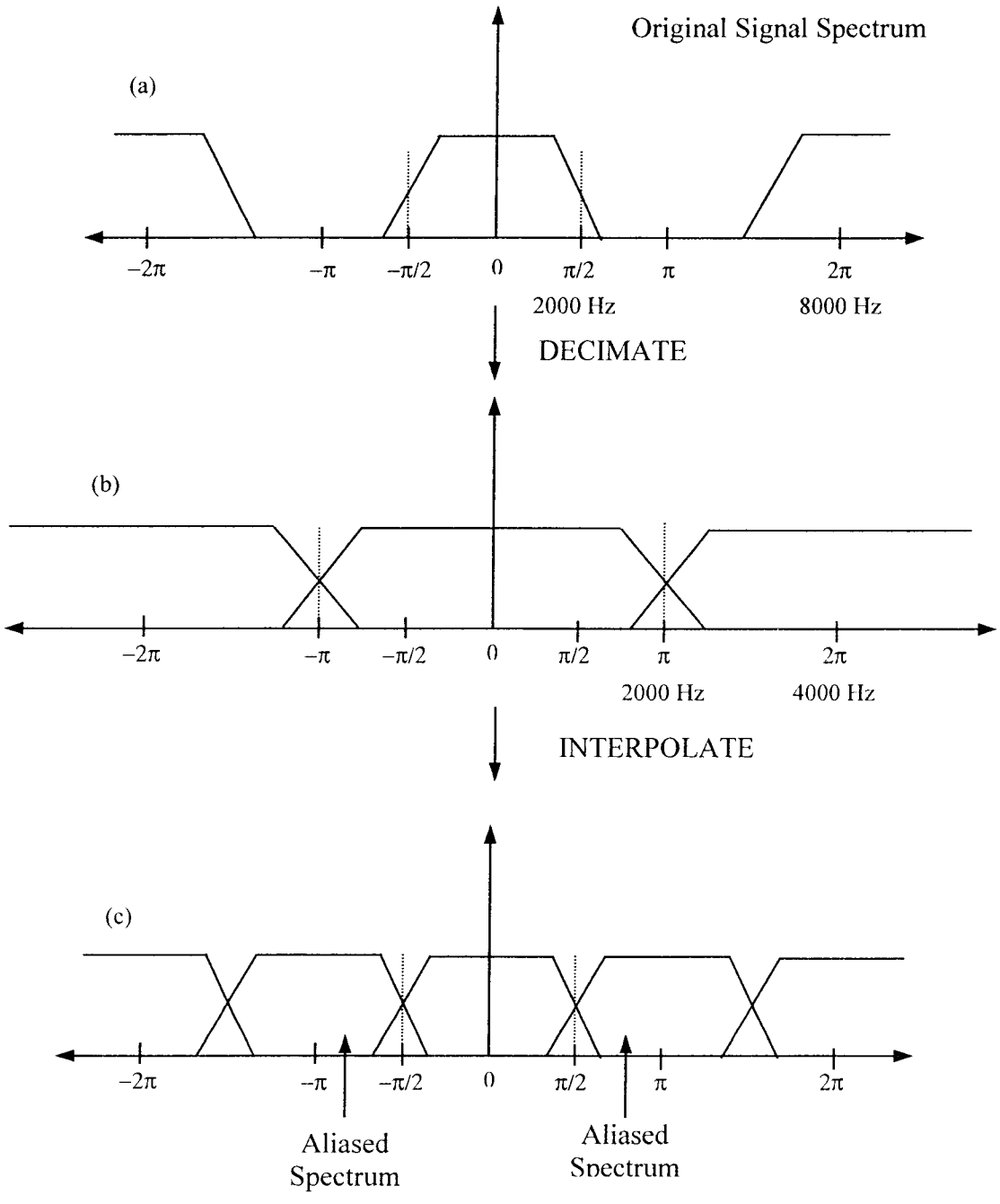


Figure 2.4: Decimation & Interpolation Process: original spectrum(a); decimation(b); interpolation(c).

Figure 2.4 illustrates the impact of decimation and interpolation process in the frequency domain and the aliasing resulting from it. Consider an original signal sampled at 8000 kHz having the spectrum illustrated in Figure 2.4(a). For DT filters, π corresponds to Fs/2 = 4 kHz and thus $\pi/2$ corresponds to 2 kHz. After decimation, the spectrum of the original signal appears to stretch as shown in Figure 2.4(b) as π now corresponds to Fs/2 = 2 kHz and $\pi/2$ to 1 kHz. After interpolation, the spectrum returns its original shape, but now has distortion, called aliasing, which needs to be rejected by passing it through an appropriate reconstruction filter. Quadrature mirror filters provide near zero aliasing and perfect reconstruction. These filters are described in detail in the next section.

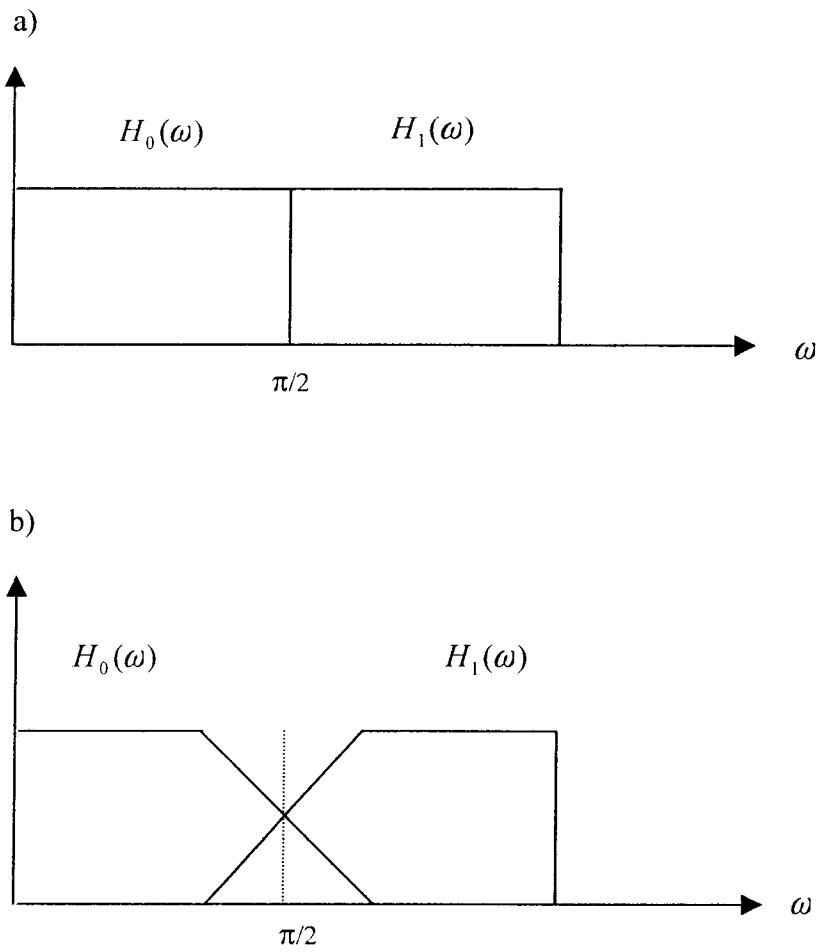


Figure 2.5: Ideal(a) and QMF(b) filters

2.1.3 Subband Filters: Quadrature Mirror Filters

Filter design is particularly important in achieving good performance in subband coding as aliasing resulting from decimation of the subband signals must be negligible. The frequency response for Ideal Filters (also known as Rectangular Perfect Reconstruction and Brick Wall) is shown in Figure 2.5(a) and is not physically unrealizable. Practical filters have non-zero transition bands, which can lead to aliasing.

A solution to the aliasing problem is to design quadrature mirror filters (QMF), to eliminate aliasing. QMF filters have important frequency response characteristics similar to those shown in Figure 2.5(b). The sum of the filter frequency responses, H0(w) + H1(w) is nearly flat. Thus if a signal is filtered by H0(w) and H1(w), the sum of the resulting output signals results in the original signal, i.e.

$$|Y(\omega)| = |(H_0(\omega) + (H_1(\omega)||X(\omega)|$$
$$|(H_0(\omega) + (H_1(\omega)| = 1$$

2.1.3.1 Quadrature Mirror Filters (QMF)

The basic building block in applications of QMF is the two channel QMF bank as shown in Figure 2.6. This 2 channel QMF system is used below to explain how QMF filters are designed to prevent aliasing distortion. Note that this corresponds to a 1 stage subband encoder. This system is called a multi-rate digital filter structure that employs two decimators in the signal analysis section and two interpolators in the signal synthesis section. Let the impulse responses for lowpass and high pass filters in the analysis section be $h_0(n)$ and $h_1(n)$, respectively. Similarly, let the impulse responses of the lowpass and high pass filters in the synthesis section be $g_0(n)$ and $g_1(n)$, respectively.

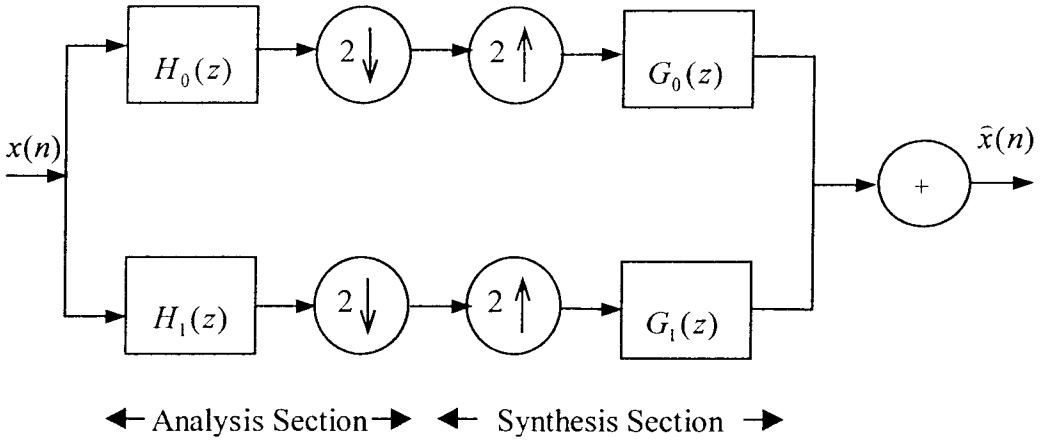


Figure 2.6: The Two Channel QMF Filter Bank

Then the Fourier transforms of the signals at the outputs of the two decimators are

$$X_{a0}(\omega) = \frac{1}{2}\left[X\left(\frac{\omega}{2}\right)H_0\left(\frac{\omega}{2}\right) + X\left(\frac{\omega-2\pi}{2}\right)H_0\left(\frac{\omega-2\pi}{2}\right)\right]$$
$$X_{a1}(\omega) = \frac{1}{2}\left[X\left(\frac{\omega}{2}\right)H_1\left(\frac{\omega}{2}\right) + X\left(\frac{\omega-2\pi}{2}\right)H_1\left(\frac{\omega-2\pi}{2}\right)\right]$$

Let $X_{s0}(\omega)$ and $X_{s1}(\omega)$ represent the two inputs to the synthesis section, then the spectrum $X(\omega)$ of the output signal is simply

$$X(\omega) = X_s\ (2\omega)G\ (\omega) + X_{s1}(2\omega)G_1(\omega)$$

If there is no noise, then the analysis and synthesis filters are so connected such that

$$X_{a0}(\omega) = X_{s0}(\omega)$$
$$X_{a1}(\omega) = X_{s1}(\omega)$$

In this case,

$$X(\omega)=\frac{1}{2}[H_0(\omega)G_0(\omega)+H_1(\omega)G_1(\omega)]X(\omega)+ \\ \frac{1}{2}[H_0(\omega-\pi)G_0(\omega)+H_1(\omega-\pi)G_1(\omega)]X(\omega-\pi) \tag{2.2}$$

Where the first term represents the desired signal output from the QMF bank, and the second term represents the effect of aliasing. To eliminate aliasing, the term $[H_0(\omega-\pi)G_0(\omega)+H_1(\omega-\pi)G_1(\omega)]$ in equation 2.2 should be zero, which can be accomplished by selecting

$$G_0(\omega)=H_1(\omega-\pi) \\ G_1(\omega)=-H_0(\omega-\pi) \tag{2.3}$$

If $H_0(\omega)$ is a lowpass filter and $H_1(\omega)$ is a mirror image high pass filter, as shown in Figure 2.5(b), then they can be expressed as

$$H_0(\omega)=H(\omega) \\ H_1(\omega)=H(\omega-\pi)$$

where $H(\omega)$ is the frequency response of a lowpass filter. In time domain the corresponding relations are

$$h_0(n)=h(n) \\ h_1(n)=(-1)^n h(n) \tag{2.4}$$

Thus $H_0(\omega)$ and $H_1(\omega)$ have mirror image symmetry about the frequency π/2. Also

$$G_0(\omega)=2H(\omega) \\ G_1(\omega)=-2H(\omega-\pi) \tag{2.5}$$

In time domain these relations become

$$g_0(n)=2h(n) \\ g_1(n)=-2(-1)^n h(n)$$

The scale factor 2 here results due to the interpolation factor used to normalize the overall frequency response of the QMF. With this choice of the filter characteristics, the aliasing component vanishes. Thus the aliasing resulting from decimation in the analysis section of the QMF bank is perfectly canceled by the image signal spectrum that arises due to interpolation. The two-channel QMF thus behaves as a linear, time-invariant system.

2.1.3.2 QMF Filter Design

Since the QMF filters are critical in subband coding, the design of the QMF filters used here is detailed below. The following steps describe the operations for designing a QMF filter, where the term half-band filter implies a filter with a cut-off frequency half of the original signal bandwidth.

Step 1: Design a linear-phase FIR half-band filter of length 2N-1 such that

Pass Band Frequency $\omega_p = 0.8 * \pi/D$

Stop Band Frequency = $\pi - \omega_p$

Stop Band Attenuation < -90 dB

Ripple < 0.0001 dB

Step 2: Construct an all-positive magnitude half-band filter from the filter obtained from Step 1.

Step 3: Compute the zeros, z_i, of the filter designed in Step 2 (MATLAB's tf2zp function)

Step 4: Construct $h_0(n)$ by using only the zeros having magnitude less than 1, i.e., $|z_i| \leq 1$.

The filters designed for this thesis are described below.

Step 1:

To achieve perfect reconstruction, a linear-phase FIR half-band filter of length 2N-1 is designed first. A half-band filter is defined as a zero-phase FIR filter whose impulse response satisfies the condition

$b(2n) =$ constant for $n \neq 0$

$b(2n) = 0$ otherwise

Hence all even numbered samples are zero except at $n = 0$. The zero phase requirement implies that $b(n) = b(-n)$.

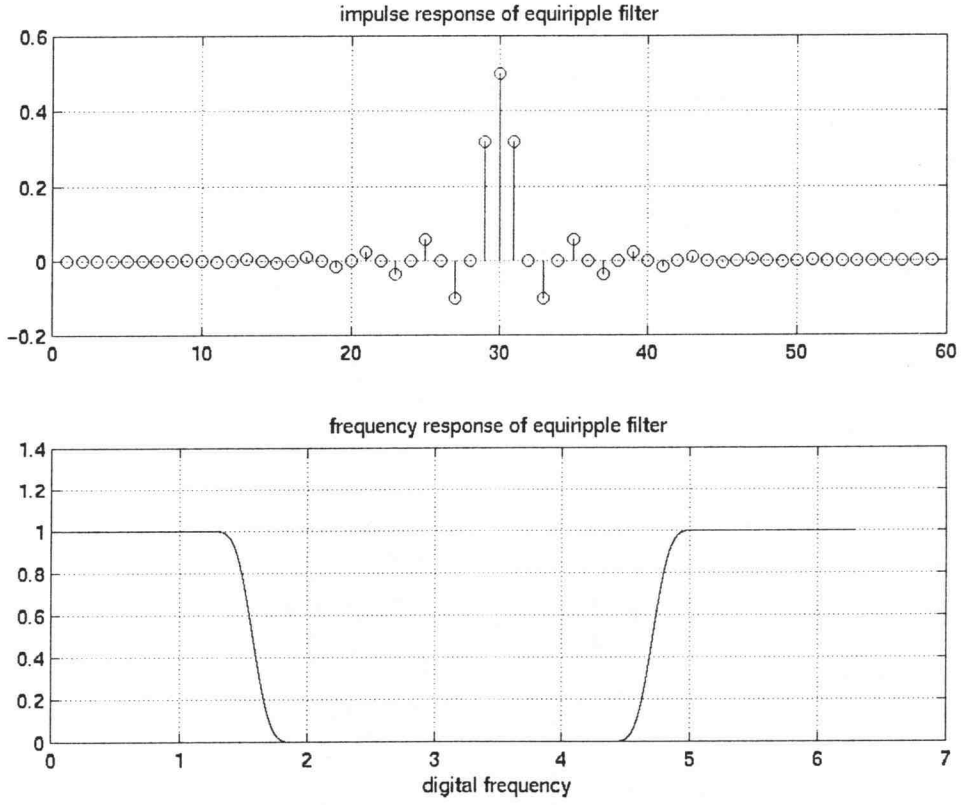


Figure 2.7: Impulse Response of Equiripple Filter

An equiripple filter of length 59 (N = 30) satisfied the required specifications. Figure 2.7 shows the resulting filter designed here. Note that the filter $B(\omega)$ satisfies the condition $B(\omega) + B(\pi - \omega)$ is equal to a constant for all frequencies.

Step 2:

Next, an all-positive half band filter $B_+(\omega)$ is constructed from $B(\omega)$ with the response

$$B_+(\omega) = B(\omega) + Ke^{-j\omega(N-1)}$$

where K is a constant. This filter is called all positive because its magnitude response is now positive at all frequencies.

Step 3:

Since the frequency response of $B_+(\omega)$ is nonnegative, it can be spectrally factored as

$$B_+(z) = H(z)H(z^{-1})z^{-(N-1)}$$

or

$$B_+(\omega) = |H(\omega)|^2 e^{-j\omega(N-1)}$$

where $H(\omega)$ is the frequency response of an FIR filter of length N(=30) with real coefficients.

Step 4:

Aliasing can be prevented by choosing $H_1(z), G_0(z)$, and $G_1(z)$ as follows

$$H_0(z) = H(z)$$
$$H_1(z) = -z^{-(N-1)} H_0(-z^{-1})$$
$$G_0(z) = z^{-(N-1)} H_0(z^{-1})$$
$$G_1(z) = z^{-(N-1)} H_1(z^{-1}) = -H_0(-z)$$

Figures 2.8, 2.9, 2.10, and 2.11 show the above filters designed using the method described above. As can be seen in Figure 2.9, the magnitude in the pass-band is twice that of $h_0(n)$. Recall the effect of decimation followed by interpolation is to decrease the magnitude by 2, resulting in the need for the gain term in the reconstruction filter. This gain occurs for the high-pass reconstruction filter as well as shown in Figure 2.11.

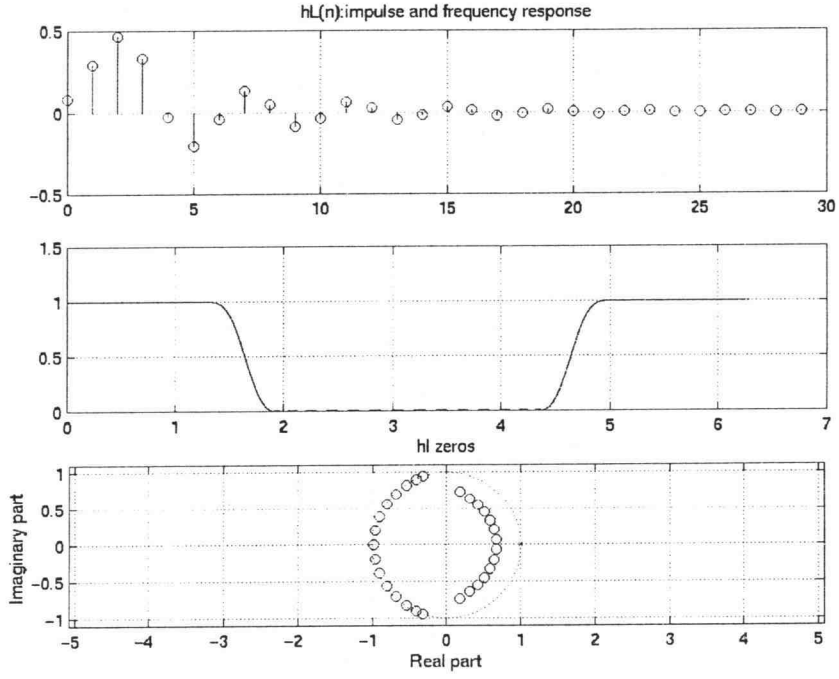


Figure 2.8: Impulse and Frequency Response for low pass filter $h_0(n)$

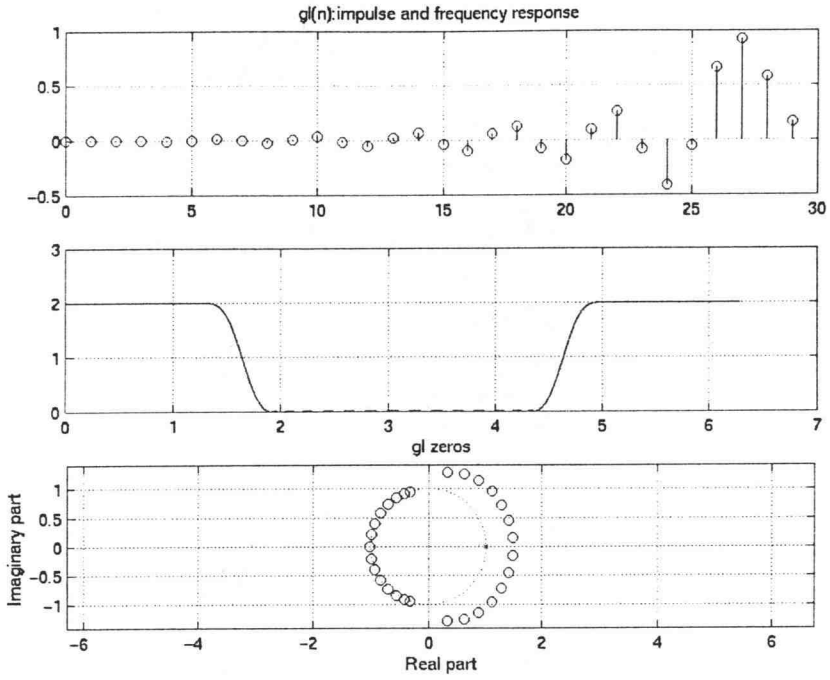


Figure 2.9: Impulse and Frequency Response for Reconstruction low pass filter $g_0(n)$

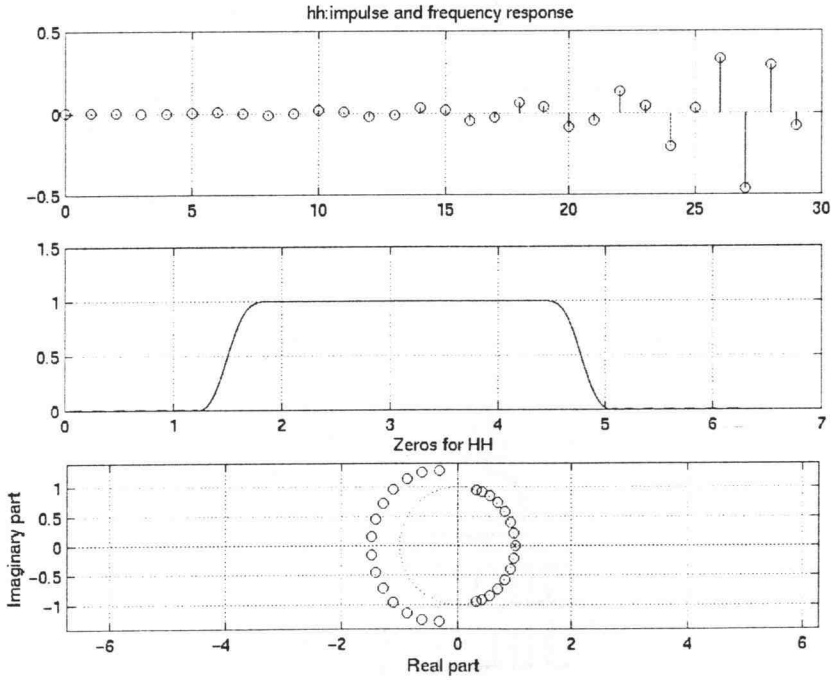


Figure 2.10: Impulse and Frequency Response for High pass filter $h_1(n)$

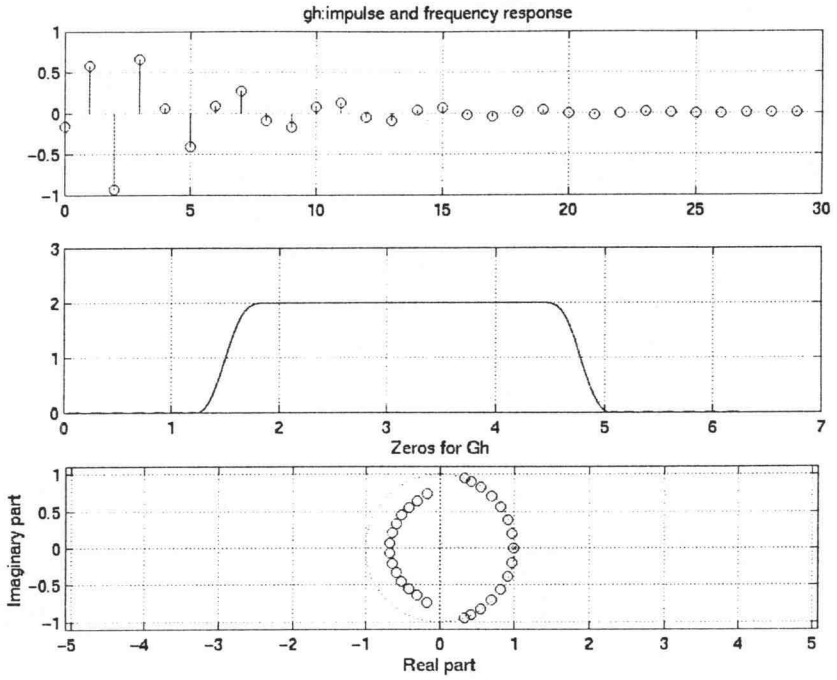


Figure 2.11: Impulse and Frequency Response for reconstruction filter $g_1(n)$

2.1.3.3 Performance Verification of a Two Channel QMF Bank

Before we can put these filters in our actual codec, it is important to find out if they are working properly, i. e., to ensure that there is minimal aliasing resulting from decimation and interpolation. In order to test the above filters, consider the two channel QMF bank as shown in Figure 2.6. If input to this filter bank is an impulse, then after passing through the set of filters it should get perfectly reconstructed at the other end and there should be no aliasing. Figure 2.12 shows the impulse input to this bank. This impulse has a height of 1 and is padded with zeros.

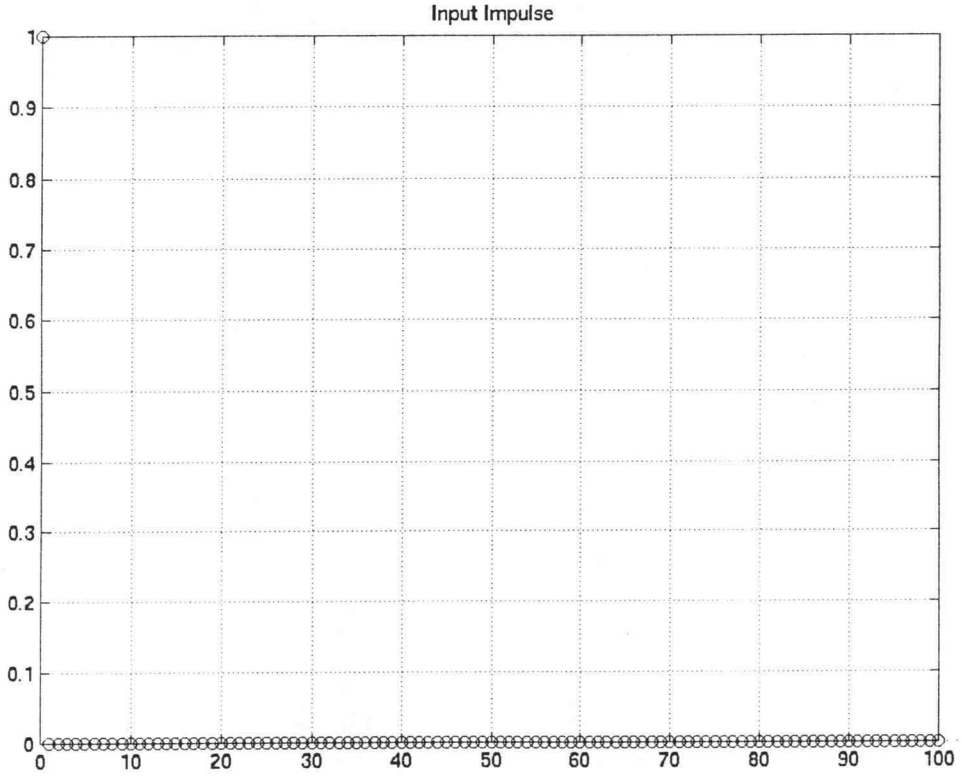


Figure 2.12: Impulse Input to the 2-Channel Filter Bank

Figure 2.13 shows this impulse after it has been convolved with the low pass and high pass filters in the analysis section before decimation. This action gives the impulse response of the filters as shown in the Figure 2.13.

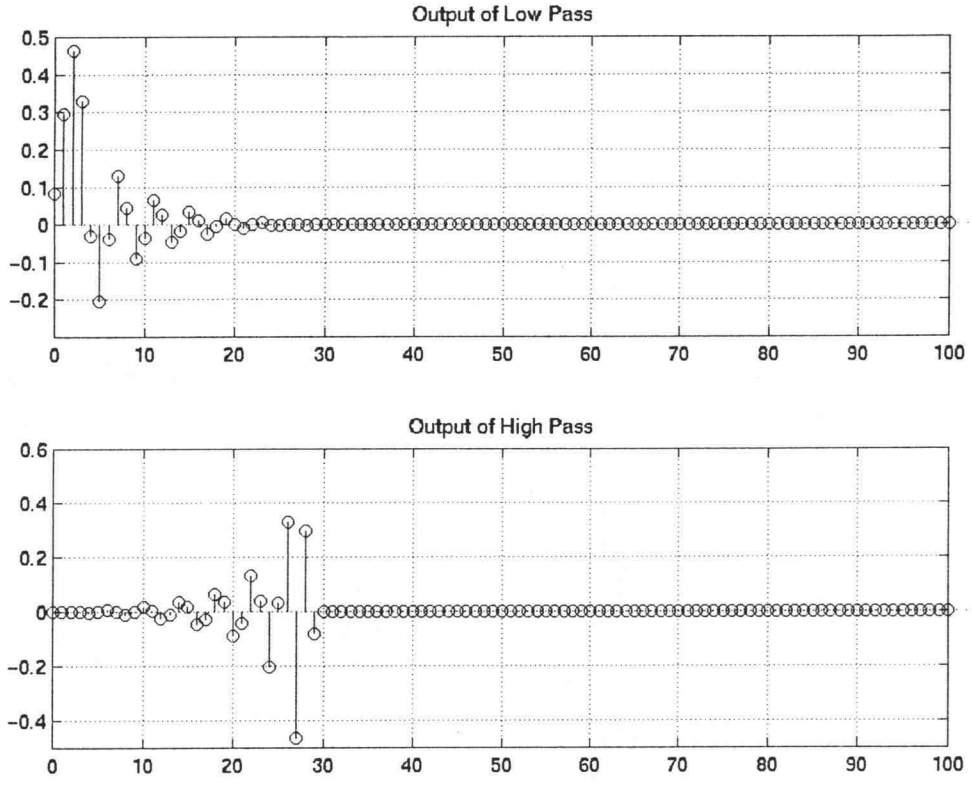


Figure 2.13: Output after the Analysis Section

Next after going through the process of decimation and interpolation and then filtering through the low and high pass reconstruction filters, we obtain the response as shown in Figure 2.14.

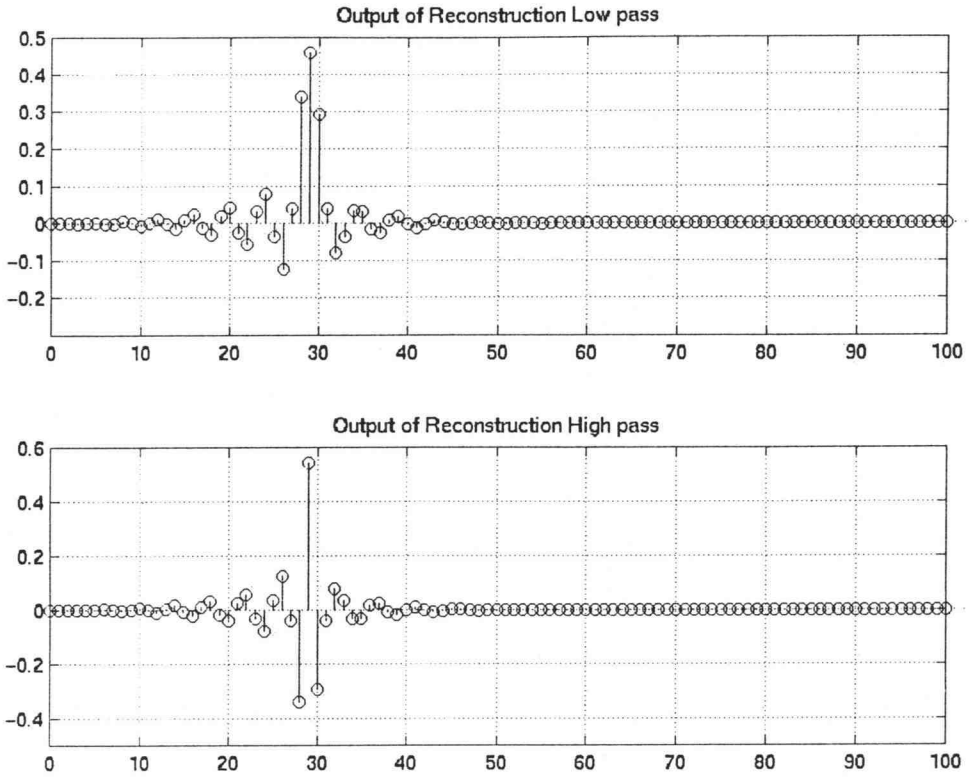


Figure 2.14: Output after the Synthesis Section

The outputs from the reconstruction filters are added together to get the final result which as can be seen in Figure 2.15 is the original delayed impulse due to the delay in the filters. Thus, it was possible to reconstruct back the input at the synthesis side after passing through the filter bank. We know now that our filters work and can now be used to build the complete subband codec.

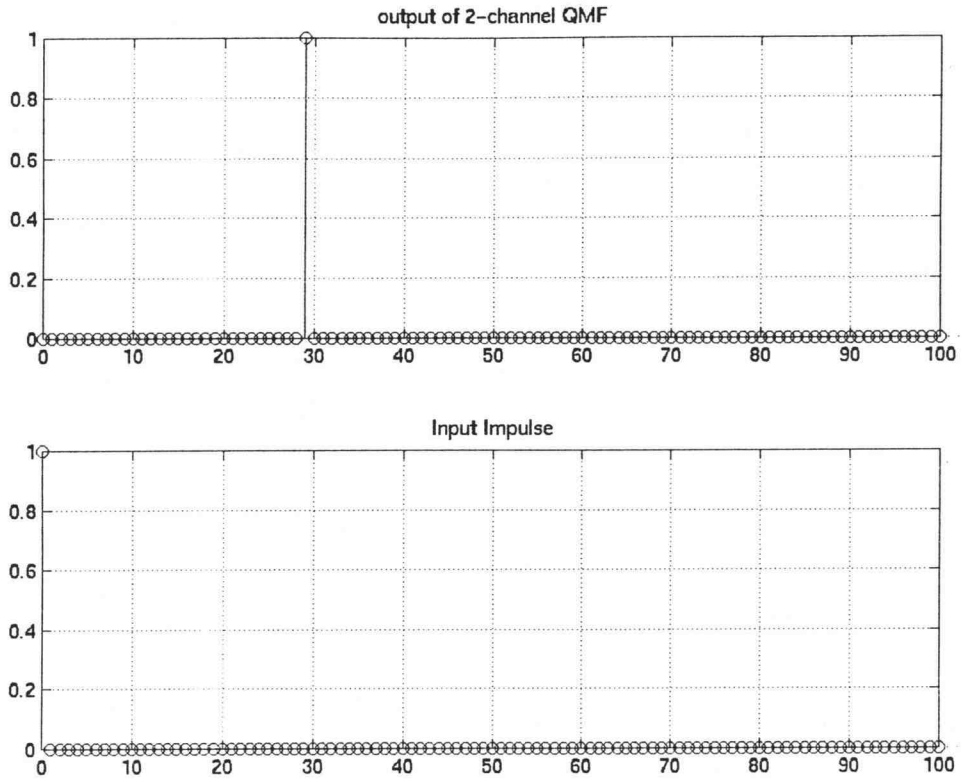


Figure 2.15: QMF Output and Input

2.1.4 MATLAB Implementation of a Subband Codec

The 24 kbps subband codec illustrated in Figure 2.1 has been implemented in MATLAB. The speech segment is the statement "We were away a year ago". As can be seen from the frequency spectrum, most of the energy is in the lower frequency bands. The peakiness of the spectrum is indicative of voiced speech

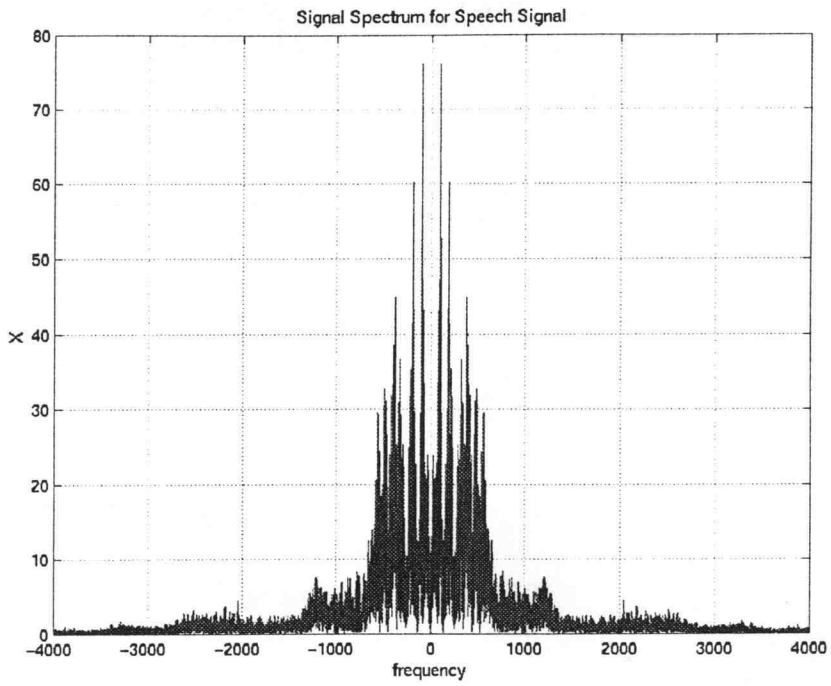


Figure 2.16: FFT of the input speech segment

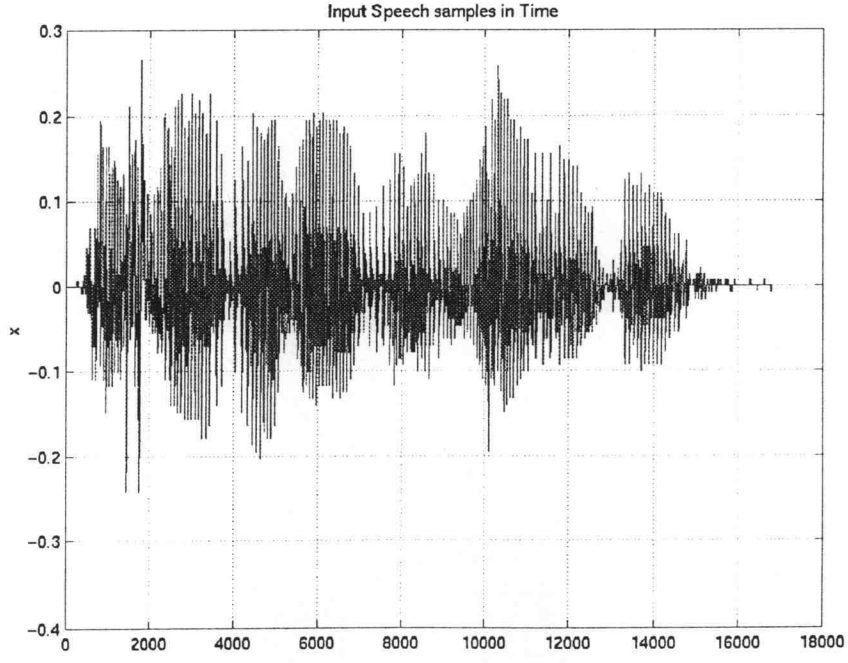


Figure 2.17: Time Domain Plot of Input speech signal

Figure 2.18 shows the output signal spectrum after the first set of low pass and high pass operations, respectively, in the first stage. Figure 2.19 shows the resulting signal spectrum after decimation by 2 and filtering by the second set of low pass and high pass filters. As can be seen from Figure 2.18, around 2000 Hz there is an overlap of the signal for both the filters which is due to the overlapping transition bands of the low and high pass QMF filters

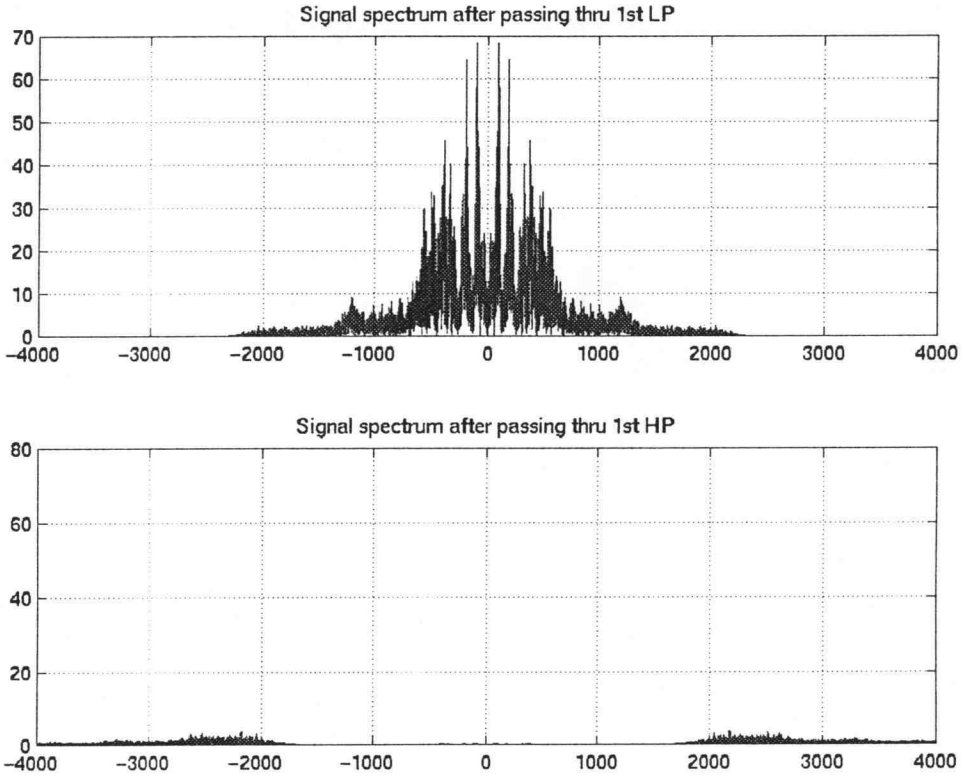


Figure 2.18: Signal Spectrum after passing through the first LP and HP Filters

Also, note that the signal power in the high frequency band, as shown in Figure 2.18, is very low. For the 24 kbps codec simulated here, this high frequency information

is encoded using only one bit. The outputs of the filters in the second stage are shown in Figure 2.19.

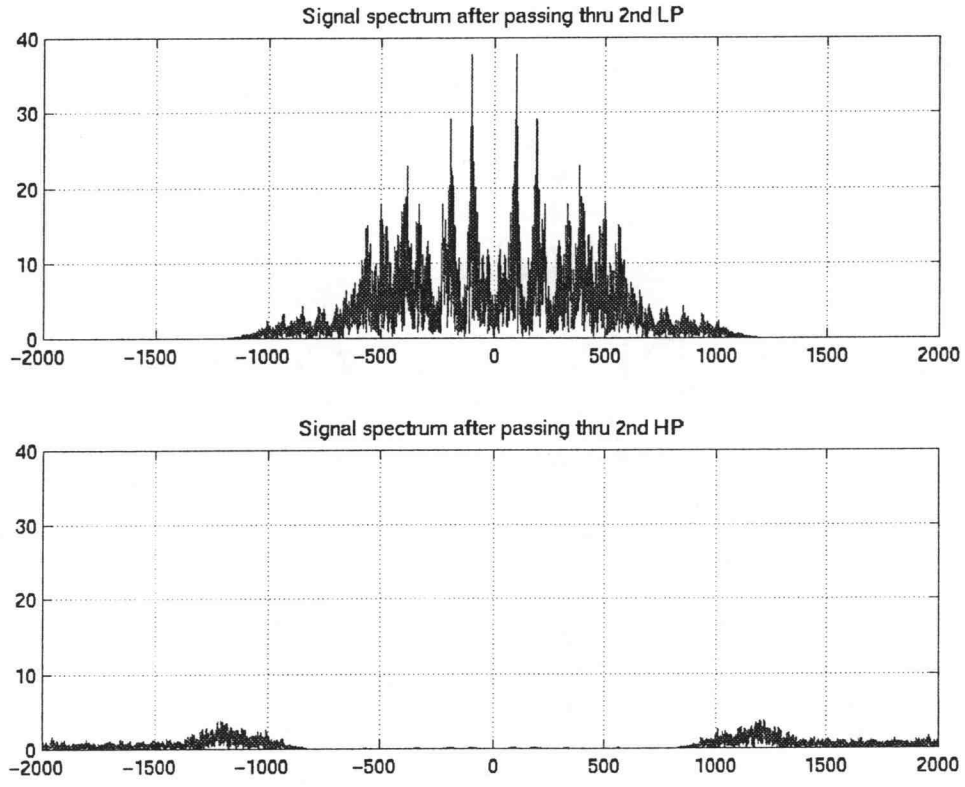


Figure 2.19: Signal Spectrum after passing through the 2nd set of LP and HP Filters

The signal energy in the high frequency band of 1000 - 2000 Hz shown in Figure 2.19 is encoded using three bits. As was seen before, most of the signal energy is still concentrated in the lower frequency regions.

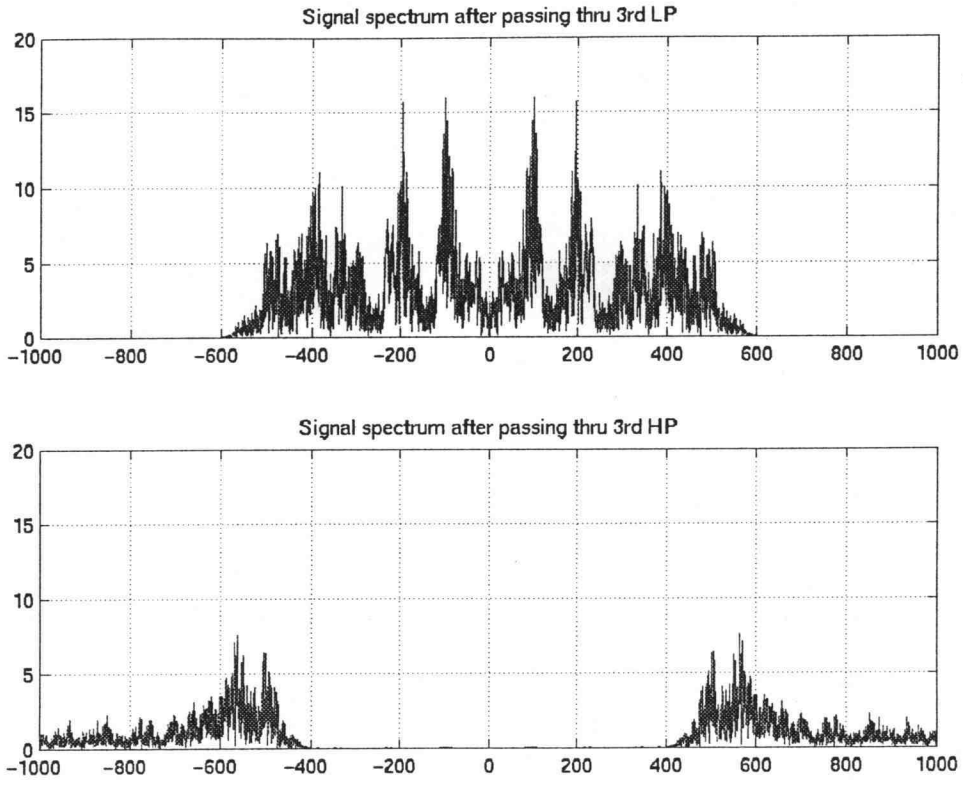


Figure 2.20: Signal Spectrum after passing through the 3rd set of LP and HP filters

The outputs of the filters in the third stage of the codec after decimation are shown in Figure 2.20. The speech information in the regions from 500 - 1000 Hz and 0 - 500 Hz each is encoded using 7 bits.

At the decoder the speech signal is reconstructed. Figure 2.21 shows the reconstruction of the speech waveform after the first and second stages of the receiver as defined in Figure 2.3.

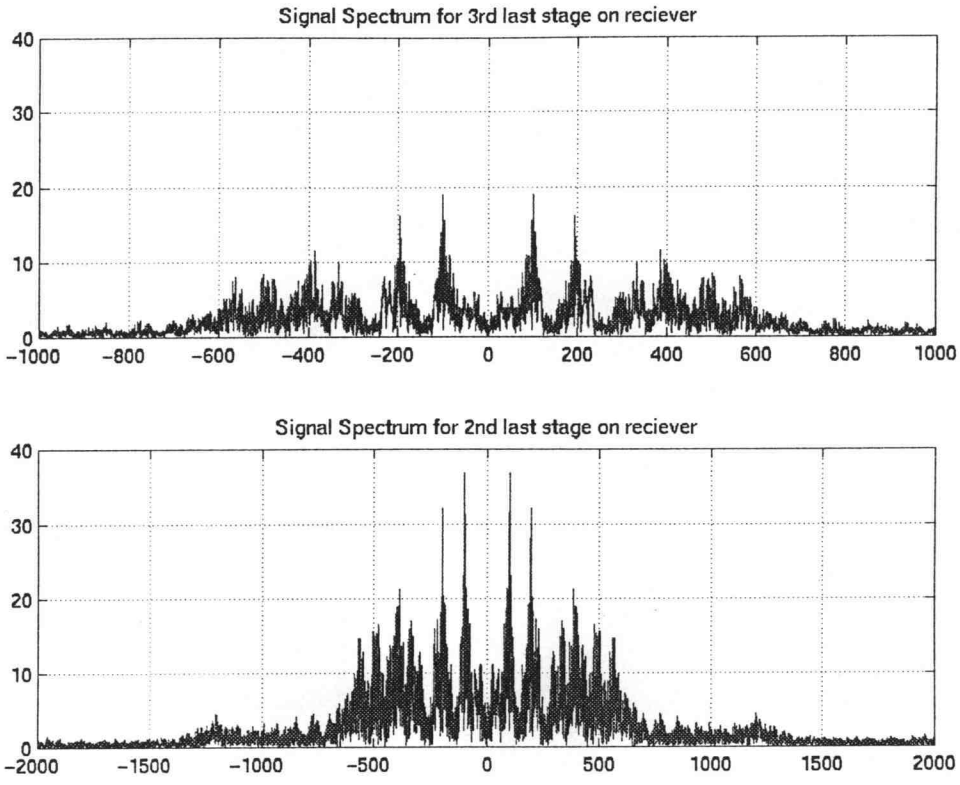


Figure 2.21: Reconstructed speech signal at 3rd and 2nd last stages at the receiver

Recall that reconstruction includes interpolation and recombination of information from the different subbands. Figure 2.22 shows the signal after the final reconstruction at the last stage in the receiver. Note that after each stage the signal spectrum more closely resembles the original speech spectrum.

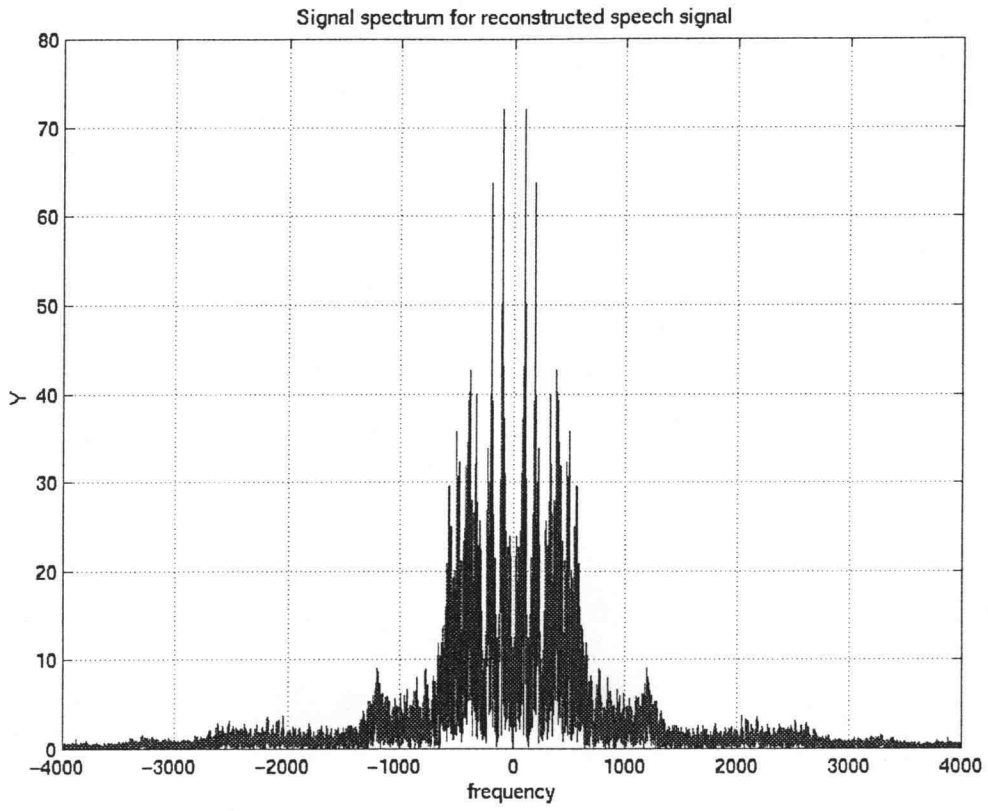


Figure 2.22 : Speech Spectrum for the reconstructed speech at the receiver

Looking at the above graph and comparing it visually with the actual speech spectrum shows very little difference, again validating the subband codec simulation. The reconstructed speech signal in the time domain is shown in Figure 2.23 and is similar to the actual speech plot. The 24 kbps speech sounds very similar to the original recorded at 64 kbps. The signal to noise ratio is 11.5655 dB which agrees with SQR of 12 dB in standard text [2].

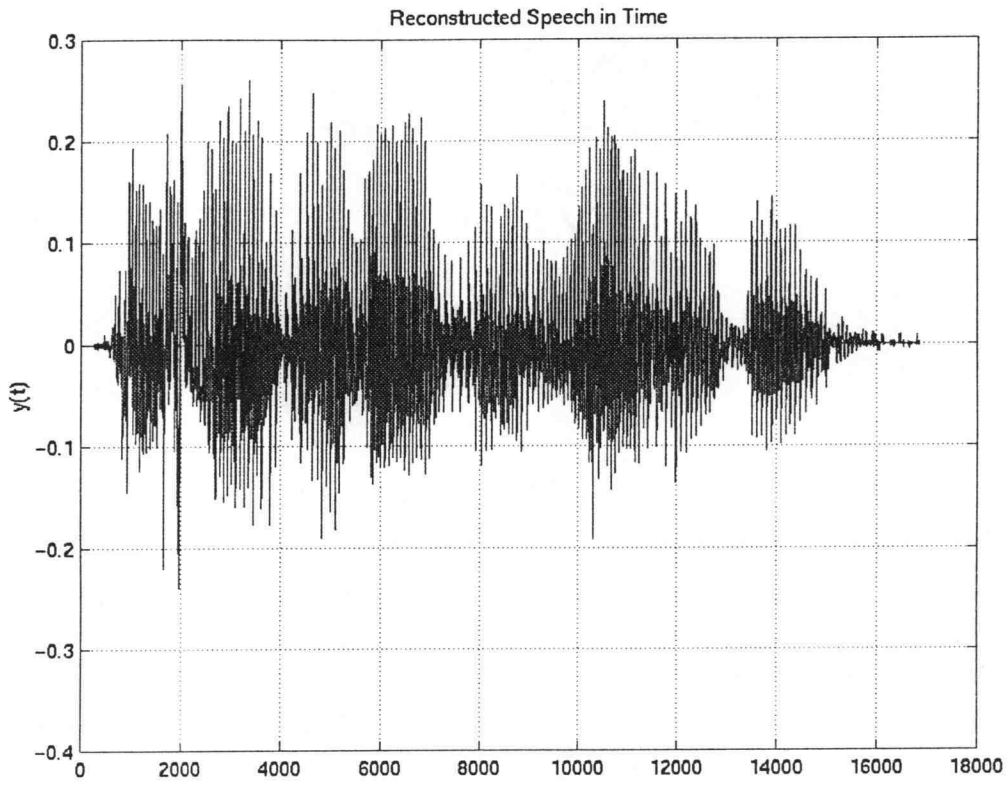


Figure 2.23: Reconstructed Signal in Time

2.1.5 Analytical SQR for Subband Coding

In subband coding, each subband waveform $x_k(t)$ is sampled at a rate f_{sk} and encoded using N_k bits per subband sample. The original speech signal has a sampling frequency of 8000 samples/sec. From the equation (2.1), the transmission rate in SBC can be computed by summing the bit rates needed to code individual subbands:

$$I = \sum_{k=1}^{M} f_{sk} N_k \text{ bits/sec} \qquad (2.6)$$

To simplify the analysis, we assume non-overlapping subbands. In this case, there is no correlation between signals in adjacent subbands. Thus the total signal variance σ^2_x is simply the sum of the subband variances σ^2_{xk} i.e.

$$\sigma^2_x = \sum_{k=1}^{4} \sigma^2_{x_k}$$

Intuitively, recall that the input signal variance σ^2_x is equal to the area under the power spectral density. Similarly σ^2_{x1} and σ^2_{x2} are equal to the areas under the power spectral density curves (PSD) for each subband. Since the subbands are non-overlapping, it is clear that the total area is the sum of the subband PSD areas.

$$\sigma^2_x = \sigma^2_{x1} + \sigma^2_{x2}$$

Extending this analysis for all the bands and assuming ideal filters, we can now say that the individual variances σ^2_{rk} of subband reconstruction errors for each band can be added to obtain the variance σ^2_{rSBC} of the signal reconstruction error:

$$\sigma^2_{rSBC} = \sum_{k=1}^{M} \sigma^2_{rk}$$

The reconstruction error variance of a conventional full-band PCM coder, with a bit rate equal to the average bit rate N bits/sample is given by σ^2_{rpcm}. Then there exists a gain G_{SBC} which is the SQR improvement due to subband coding and is given as:

$$G_{SBC} = \frac{\sigma^2_{rPCM}}{\sigma^2_{rSBC}}$$

$$SQR_{SBC}(dB) = SQR_{PCM}(dB) + 10 \log G_{SBC} \qquad (2.7)$$

2.1.6 Comparison of Simulated and Analytical SQR Measurements

To verify the analysis, the simulated and analytical SQR results are compared. The analytical SQR is calculated as the sum of the reconstruction errors in each channel. The simulated SQR has been calculated using the following formula :

$$\mathrm{SQR}_{\mathrm{SBC}}(\mathrm{dB}) = 10 * \log_0 \frac{\sigma^2{}_x}{\sigma^2{}_r} \tag{2.8}$$

where $\sigma^2{}_x$ is the total signal power and $\sigma^2{}_r$ is the total reconstruction error power.

Table 2.2 illustrates the reconstruction errors in each band and the total reconstruction error calculated from the difference between the original and the reconstructed speech at the receiver.

$\sigma^2{}_{r1}$	$\sigma^2{}_{r2}$	$\sigma^2{}_{r3}$	$\sigma^2{}_{r4}$	$\sum_{k=1}^{4} \sigma^2{}_{rk}$	$\sigma^2{}_r$
6.4791e-5	2.1027e-5	5.2398e-4	2.3699e-4	8.4679e-4	1.8940e-4

Table 2.2: Reconstruction errors in each band

The total reconstruction error power calculation assumes ideal filters and non-overlapping subbands. The actual filters have overlapping transition bands and thus the sum of the individual reconstruction error powers is greater than the simulated end to end reconstruction error, although the values are very close.

The SQR values corresponding to the simulated and analytical total reconstruction error are given in table 2.3. The simulated SQR is computed by taking the ratio of the signal power to the overall reconstruction error power, equation (2.8). The analytical SQR is computed by computing the gain term and then using equation (2.7). Since no

other distortions are introduced, the reconstructed signal should only contain the quantization noise. Again the simulated and analytical results are not identical but are very close, validating the analytical assumption. These results are very close to the SQR of 12 dB typically assumed for 24 kbps SBC [2].

SQR simulated(dB)	SQR PCM(dB)	Gain G_{SBC}(dB)	SQR analytical(dB)
11.5655	9.3146	2.0255	11.3401

Table 2.3: Analytical and simulated SQR for Subband

2.2 DPCM

2.2.1 Introduction

The term Pulse Code Modulation (PCM) refers to analog to digital conversion by sampling and quantization. The standard uncompressed 64 kbps speech is a PCM signal. PCM is robust to channel interference and is easily converted back to the analog speech signal. Data compression is used to remove the redundancy present in a PCM signal and thereby reduce the bit rate of the transmitted data without serious degradation in signal quality.

Since speech signals sampled at 8 kHz do not change in value rapidly from one sample to the next, a sample can be predicted with reasonable accuracy from previous samples. Compression can be achieved by transmitting the difference between the signal and its predicted value rather than the signal itself. Differential Pulse Code Modulation (DPCM) uses this idea to achieve compression.

The goal in this thesis is to achieve speech coding using a 24 kbps DPCM codec, transmit it over a bursty channel and observe the effect at the receiving end. This section describes the development of a 24 kbps DPCM codec simulation.

2.2.2 System Overview

In Differential Pulse Code Modulation, difference between the input sample and a prediction value is transmitted, rather than on the sample itself. The difference signal can be quantized using fewer bits than required for the original signal, resulting in

compression. Coding methods using this prediction idea are called predictive coding methods.

2.2.2.1 Predictive Coding for Compression

Accurate prediction requires a good model. Speech can be modeled as the output of a linear system comprising all poles (an AR model). At the transmitted end, an inverse model is used. The parameters of this system are time varying, but can be viewed as fixed for each utterance of about 20 msec. For AR models, the optimal inverse model can be computed by using *linear prediction.* The predicted value is a weighted sum of past values and the model parameters are computed by minimizing the power in the difference between the actual and predicted signals.

The goal in linear prediction is to create a filter that models the speech production process. If we sample a speech signal at a high enough rate, we can " predict" the next sample from the previous ones. Let $x(n)$ be the discrete time unquantized input signal, $x_p(n)$ be the prediction of it, Thus a speech sample can be approximated as a linear combination of past speech samples i.e.,

$$x_p(n) = \sum_{k=1}^{N} \alpha_k x(n-k)$$
$$e(n) = x(n) - x_p(n)$$

where α_k are the linear prediction coefficients ; *e(n)* is the difference signal and is called the prediction error.

The predicted value is thus the output of the prediction filter, which is a finite impulse response filter (FIR) whose system function is

$$P(z) = \sum_{k=1}^{p} \alpha_k z^{-k}$$

and whose input is the signal $x(n)$. Compression is achieved by transmitting the error signal instead of actual input signal and using fewer bits(quantizer levels). Even though fewer bits are used, the quantizer noise remains small as the error signal has a much smaller dynamic range than the input signal. The signal is reconstructed at the receiver. A Linear Predictive Coder and Decoder are shown in Figure 2.24

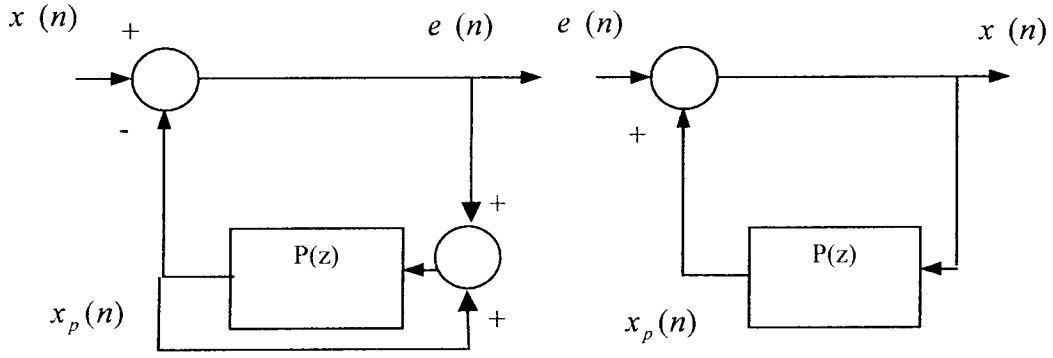


Figure 2.24 : Predictive Coding for Compression

The optimum prediction coefficients α_k are defined uniquely as the minimization of the squared differences (over a finite interval) between the actual speech samples and the linearly predicted ones: minimum mean square error. Methods like the Levinson-Durbin algorithm have been used to obtain these coefficients efficiently [1].

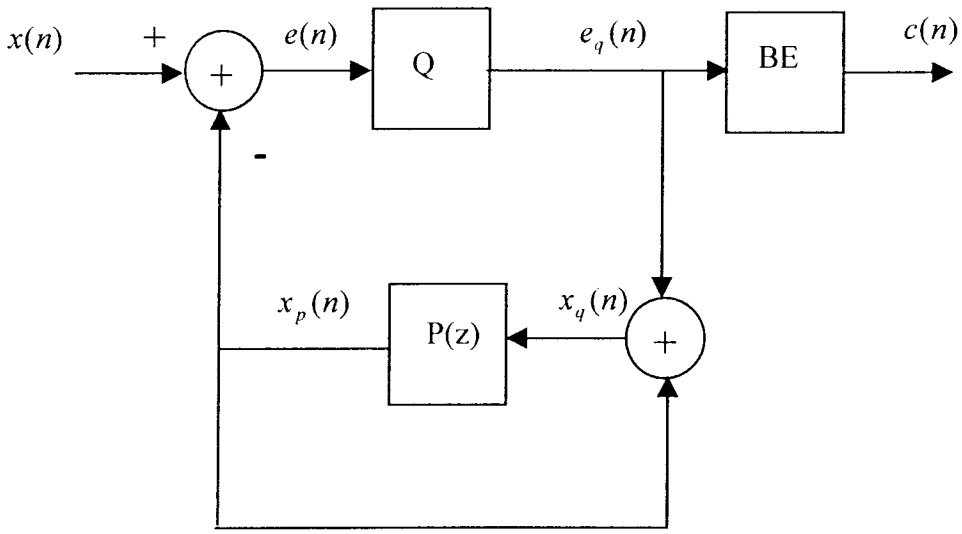
Note that the reconstruction filter transfer function

$$H(z) = \frac{1}{1 - \sum_{k=1}^{p} \alpha_k z^{-k}}$$

is an all-pole filter, i.e. an IIR filter, and the reconstructed signal is

$$x(n) = e(n) + \sum_{k=1}^{p} \alpha_k x(n-k)$$

However, now if the error out of the transmitter as shown in Figure 2.24 is quantized, then the receiver uses this quantized error as input to the system. In contrast, the transmitter used the *unquantized error* as input to the predictor. The additional error introduced at the input to the receiver is passed through the IIR filter *H(z)*, resulting is an accumulation of errors in the reconstructed speech. DPCM avoids this situation by inserting a quantizer in the loop at the transmitter as shown in Figure 2.25.



Q: Quantizer
BE: Binary Encoder
P(z): Predictor

Figure 2.25: DPCM Encoder

2.2.2.2 DPCM Encoder & Decoder

Figure 2.25 depicts the encoder for DPCM. Here, $x_q(n)$ is obtained from $x\ (n)$ and the quantized error $e_q(n)$. Thus the predicted value is computed as

$$x_p(n) = \sum_{k=1}^{p} \alpha_k x_q(n-k) \tag{2.9}$$

The important point to note here is that the input to the transmitter's predictor is the same as the input to the receiver's predictor (in the absence of noise). The predictor order p used in this thesis is 10, which is generally considered to provide a reasonable estimate for male speech.

Let $q(n)$ be the quantization error, then the quantized difference signal is

$$e_q(n) = e(n) + q(n) \tag{2.10}$$

As can be seen in the Figure 2.25, the quantized difference signal $e_q(n)$ is added to the predicted value $x\ (n)$ to produce the prediction filter input,

$$x_q(n) = x_p(n) + e_q(n) \tag{2.11}$$

Substituting for $e_q(n)$ from equation (2.10) in the above equation (2.11) results in:

$$x_q(n) = x_p(n) + e(n) + q(n) \tag{2.12}$$

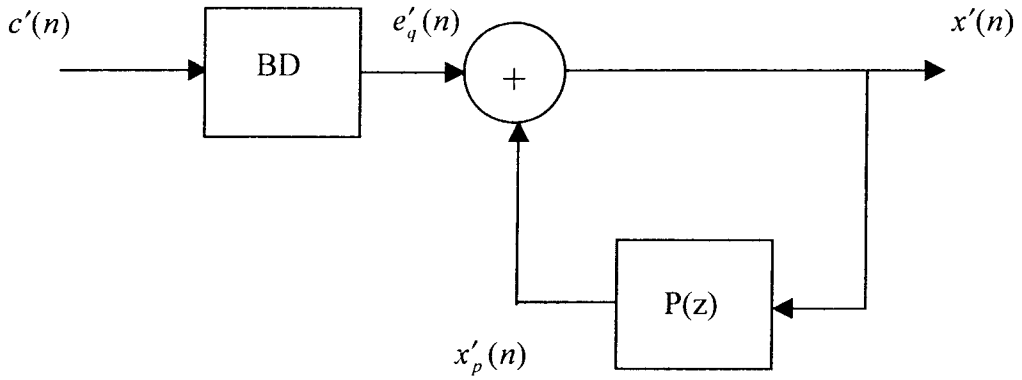
However, since $e(n) = x(n) - x_p(n)$, substituting this expression into (2.12) results in the following expression for the quantizer:

$$x_q(n) = x_p(n) + x(n) - x_p(n) + q(n) = x(n) + q(n) \tag{2.13}$$

Thus from equation 2.13 it can be seen that, independent of the properties of the predictor system P(z), the quantized signal at the prediction filter input differs from the original signal $x(n)$ only by the quantization error $q(n)$.

Thus if the prediction is good, the variance of the prediction error $e(n)$ will be smaller than the variance of *x(n)*. A quantizer with fewer levels can be used to produce a quantization error with a smaller variance than would be possible if the input signal were quantized directly as in standard PCM.

The quantizer used here is a 3 bit fixed, uniform quantizer. Better performance can be obtained by using adaptive quantization, which is responsive to changing levels and spectrum of input speech signal. But here since the goal in this thesis is a preliminary analysis, a simple uniform quantizer has been used. Figure 2.26 depicts the decoder, which reconstructs back the transmitted signal at the receiving end. Comparing this figure to the decoder in Figure 2.24, the only difference is the Binary Decoder (BD) present in Figure 2.26. The quantized version of the original input signal is reconstructed using the same prediction filter as used in the transmitter. In the absence of channel noise, the binary encoded signal at the receiver input is same as the binary encoded signal at the transmitter output. Thus the corresponding receiver output is equal to $x_q(n)$ which differs from the original input *x(n)* only by the current quantization error $q(n)$.



BD: Binary Decoder
P(z): Linear Predictor

Figure 2.26: DPCM Decoder

2.2.3 Analytical Signal to Quantization Ratio

The signal-to-quantization noise ratio is defined as

$$SQR = \frac{E\left[x^2(n)\right]}{E\left[q^2(n)\right]} = \frac{\sigma_x^{\,2}}{\sigma_q^{\,2}} \tag{2.14}$$

which can be written as

$$SQR = \frac{\sigma_x^{\,2}}{\sigma_e^{\,2}} \frac{\sigma_e^{\,2}}{\sigma_q^{\,2}} = Gp * SNR_Q \tag{2.15}$$

where

$$SNR_Q = \frac{\sigma_e^{\,2}}{\sigma_q^{\,2}}$$

is the signal-to-quantizing noise ratio of the quantizer, and the quantity

$$Gp = \frac{\sigma_x^2}{\sigma_e^2}$$

is defined as the processing gain due to the differential configuration. The quantity Gp, when greater than unity, represents the gain in signal-to-noise ratio that is due to the differential quantization scheme.

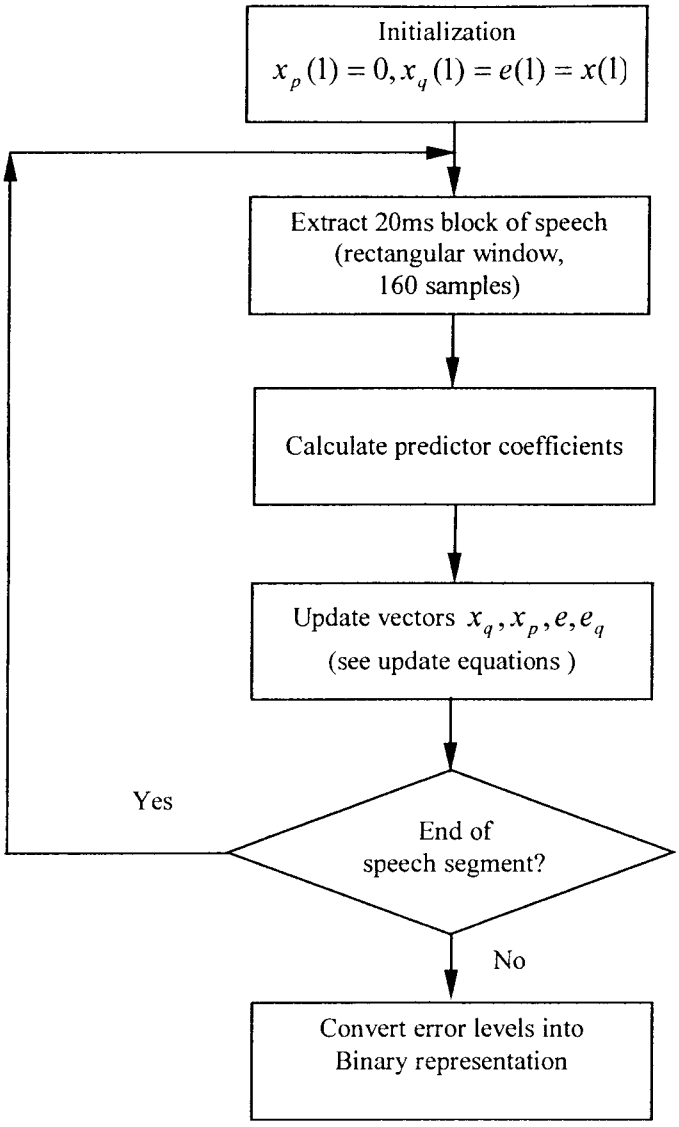
For a given message signal, the variance σ_x^2 is fixed, so that Gp is maximized by minimizing the variance σ_e^2 of the prediction error. The analytical expression for SQR for DPCM can be written in terms of the SQR for PCM together with the gain term can be expressed in dB as:

$$SQR_{DPCM}\,(dB) = SQR_{PCM}\,(dB) + 10\log Gp \qquad (2.16)$$

2.2.4 MATLAB Implementation

2.2.4.1 Implementation Flow Chart

The following Figure 2.27 shows an implementation flow chart for a DPCM coder. The difference signal has been binary encoded using 3 bits/sample and @ 8000 samples/sec, resulting in a bit rate of 24 Kbps. The speech sample used here to verify performance is the same as that used for subband containing predominantly voiced sounds, "We were away a year ago".



UPDATE EQUATIONS

$$x_p(j) = \sum_{k=1}^{10} \alpha_k x_q(j-k)$$

$$e(j) = x(j) - x_p(j)$$

$$e_q(j) = Q[e(j)]$$

Figure 2.27: Implementation Flow Chart

2.2.4.2 Histogram for Input and Difference Signals

A histogram plot depicts the number of samples per quantizer level, providing a coarse estimate of the probability density function. Figure 2.28 shows the histogram for the input signal, and Figure 2.29 is the histogram plot for the prediction error which is actually transmitted over the channel.

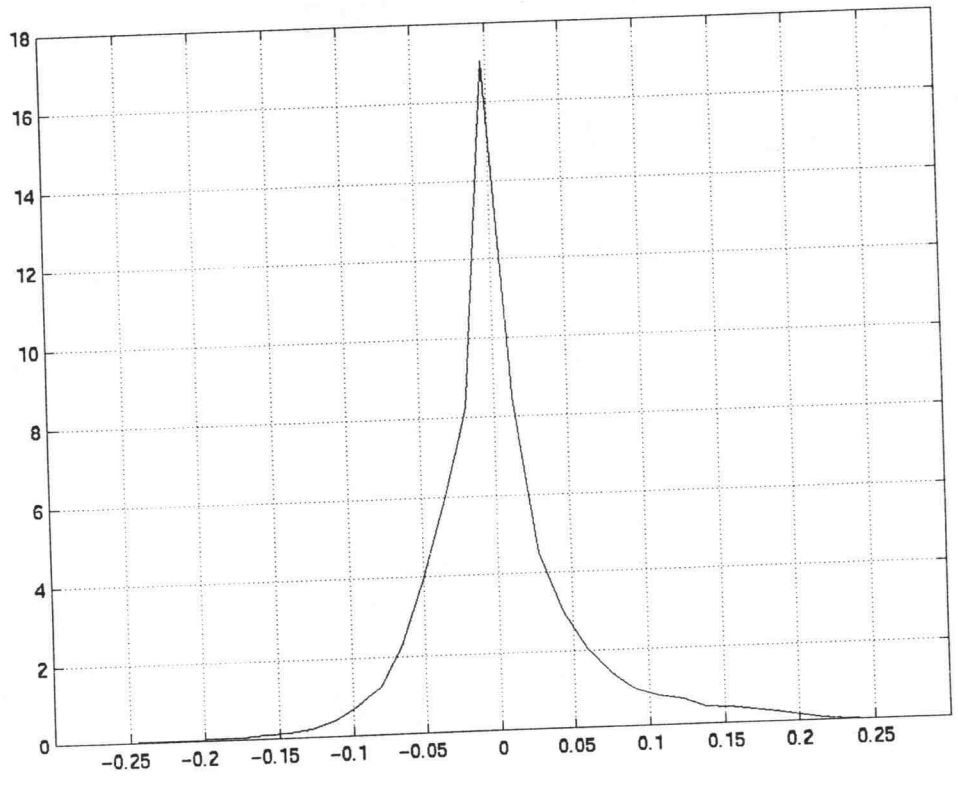


Figure 2.28: Histogram Plot for Input Speech Signal

Note that the difference signal has a dynamic range of -0.0837 to 0.0746 with a standard deviation of 0.0116. The original speech signal has a dynamic range from -0.2427 to 0.2651 with a standard deviation of 0.0508. The smaller dynamic range and

variance relative to the speech signal allows fewer bits to be used to achieve approximately the same difference between quantizer levels.

As with the SBC simulation, the DPCM codec is tested using the speech segment as shown in Figures 2.30 and 2.31. Figures 2.32 and 2.33 show the magnitude of the frequency spectrum and the time domain plots of the reconstructed speech respectively.

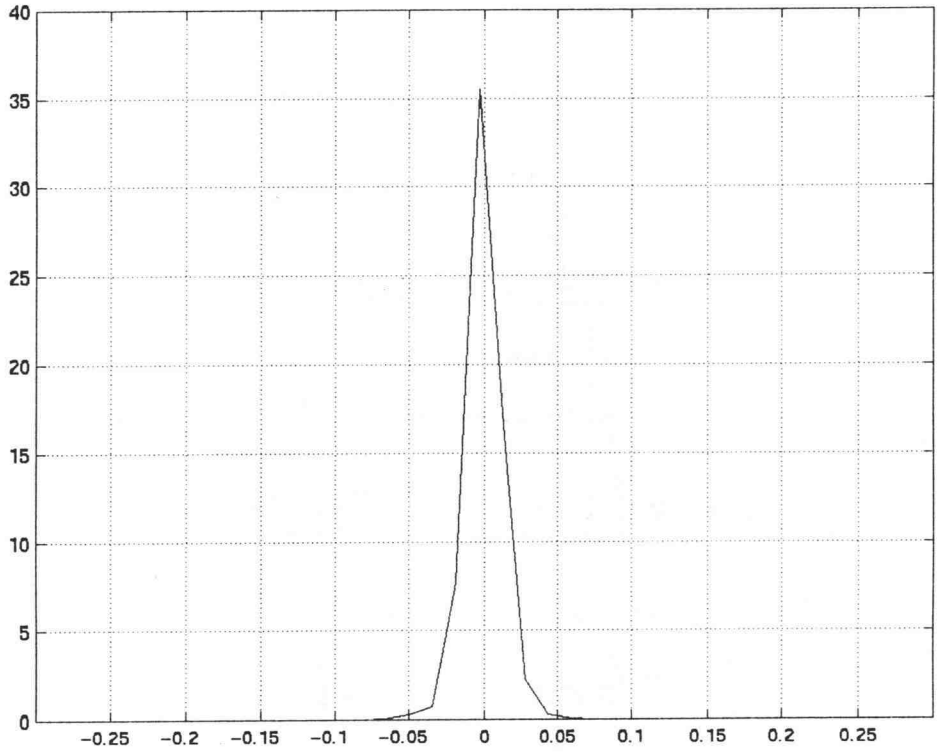


Figure 2.29: Histogram Plot of Difference Signal

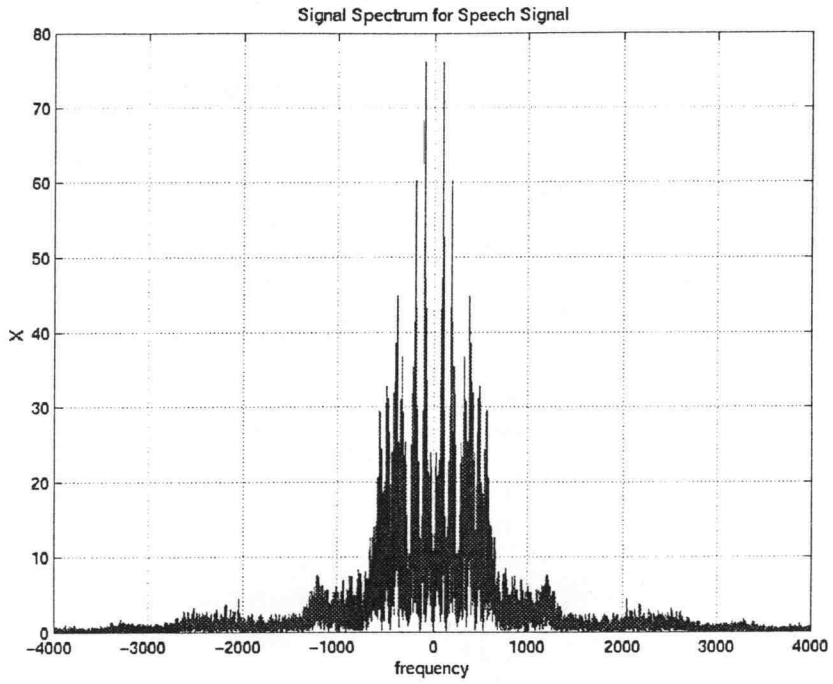


Figure 2.30: FFT of Input Speech Signal

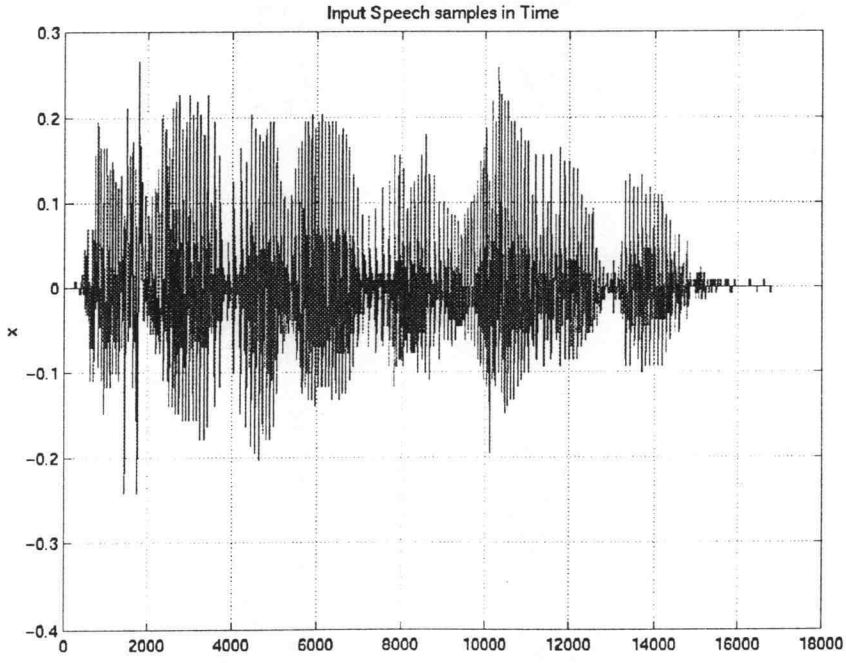


Figure 2.31: Time Domain Plot Of Input Speech Signal

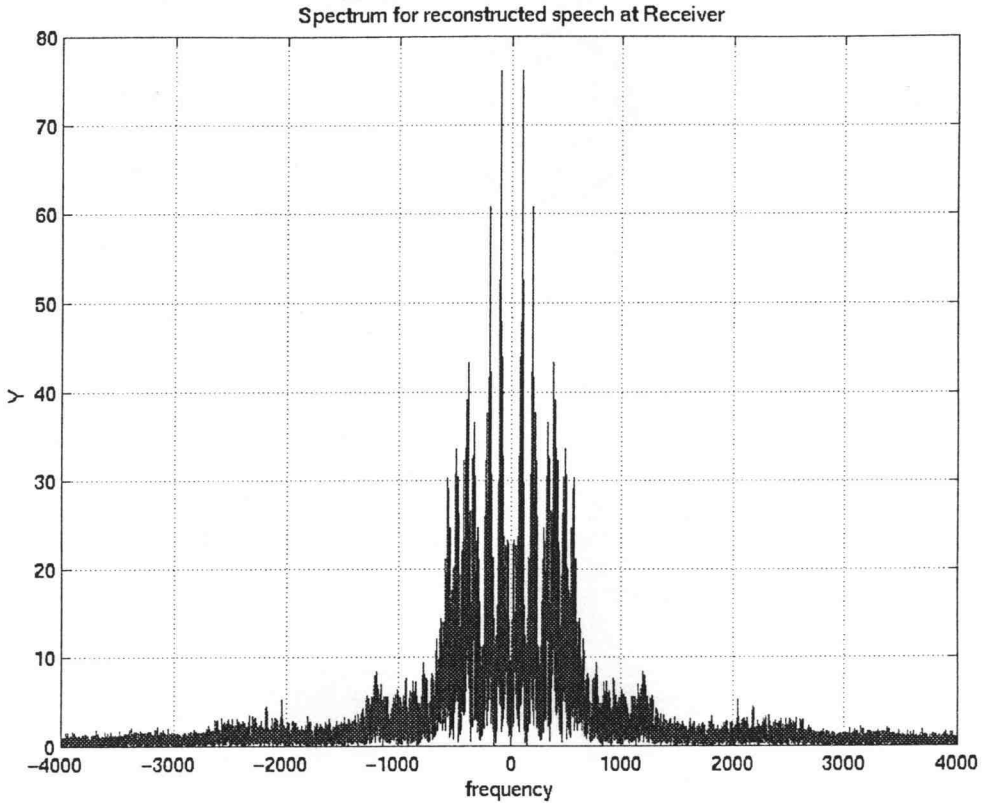


Figure 2.32: FFT of Reconstructed Speech Signal

From a brief visual inspection, the signal as in Figure 2.30 and its spectrum in Figure 2.32 look virtually identical to the original. From informal subjective evaluation the 24 kbps DPCM signal sounds as good as the 64 kbps original PCM signal, and very similar to the 24 kbps SBC speech. The measured SQR is 18.7066 dB, which compares well to the theoretically predicted value of 18.7066 dB. The method for computing these values is discussed next.

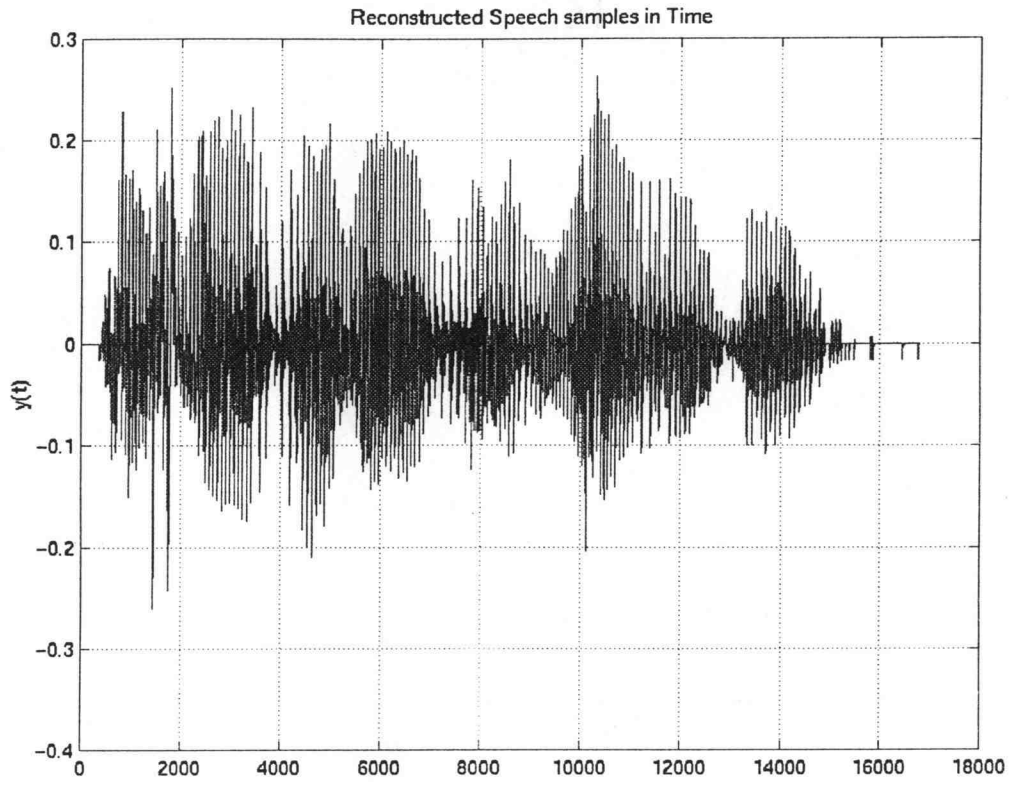


Figure 2.33: Time Domain Plot of Reconstructed Speech Signal

2.2.5 Codec Testing: Comparison of Analytical and Simulated Results

The analytical SQR is computed as derived in equation (2.16). The simulated SQR is computed as the ratio of the signal power to the reconstruction error power, where the reconstruction error is obtained by subtracting the reconstructed signal from the original signal as in equation (2.14). The SQR's computed from analytical results and simulation are presented in table 2.4.

SQR simulated (dB)	SQR PCM (dB)	Gain (dB)	SQR calculated (dB)
18.7066	9.3152	9.3914	18.7066

Table 2.4: Analytical and simulated SQR for DPCM

These values correlate well with existing results of SNR = 18 dB reported in the literature for 24 Kbps DPCM [2]. The theoretical and simulated results match thereby validating both the codec performance and assumptions used in the analysis.

Chapter 3: Channel Model

3.1 Introduction

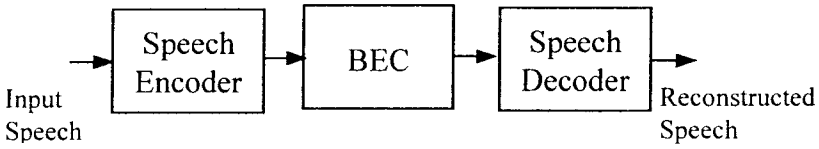
After compression, the encoded speech signal is transmitted over a physical channel to the decoder, which reconstructs the signal. Depending upon the type of the channel, the information reaching the decoder may have been distorted causing degradation in the reconstructed signal. Wireless communication channels contain a myriad of distortions, which can be described as, e.g., fading, AWGN and bursty. Here we model the overall impact of the channel, comprising both the physical channel and the error correction, on the bit errors into the source decoder. We are primarily interested in how the perceptual quality of the reconstructed speech is affected by these errors, i.e. the difference in the speech decoder performance. Recall that the channel here comprises the channel encoder and decoder pair, the modulator/demodulator, and the physical transmission channel, the channel decoder includes error correction.

In an actual wireless system, burst errors at the input of the source decoder can occur for a variety of reasons. For instance, these errors could be from an error correction device that was overloaded. A typical real channel with clustered errors is a mobile radio link. The slow signal fading over such a channel causes bit error patterns in which temporal correlation exists. This thesis tries to explore the effect of these bursts of errors on the two speech coding schemes, DPCM and Subband. Here we assume that the cumulative result of transmitter distortion results in a channel comprising burst errors of

different lengths. This chapter develops and analyzes a model for this type of channel, which we call a Burst Error Channel (BEC).

3.2 Overall Picture

Figure 3.1 illustrates the various blocks in the overall setup for experimentation where the speech encoder is 24 kbps DPCM or subband coder with binary output. The BEC results in burst of bit errors, which in turn cause distortion in the decoded speech. The goal of this thesis is to compare the effect of a burst error channel on the relative performance of the DPCM and subband codecs. For comparison, The bit error rate (BER) is kept constant as the length of the burst errors is increased.



BEC : Burst Error Channel

Figure 3.1: Overall Block Diagram

3.3 The Burst Error Channel

The BEC results in bursts of bit errors, where a " burst " is a string of consecutive errors. The resulting bit error vectors are represented as follows. If binary arithmetic is used, then '1' denotes an error and a '0' denotes no error in the error vector. An error vector e1 might be:

e1 = [0 0 1 0 0 1 0 0 0 0 0 0 1 0 0 0 1 0 0 0 0 1 0 0 0 0 0 0 1 0]

The vector e1 is a vector with single bit errors as is likely to occur in AWGN.

Since the total number of samples is 30 and the number of errors are 6, the BER = 0.2. If errors occur in the same places as e1 but in bursts of length 2 then the error vector becomes, e.g.

e2 = [0 0 1 1 0 1 1 0 0 0 0 0 1 1 0 0 1 1 0 0 0 1 1 0 0 0 0 0 1 1]

Note that now 12 errors occur in the same 30 samples, so the BER is doubled. To keep the BER constant, the number of times that a burst of errors occurs must be halved, e.g.

e2' = [0 0 1 1 0 0 0 0 0 1 1 0 0 0 0 0 0 0 0 0 1 1 0 0 0 0 0 0 0 0]

In this thesis, the BER is fixed and the burst error lengths varied.

3.4 A Filter Model for Burst Errors of Length N

In this analysis, we assume a burst of sample, rather than bit, errors. In contrast, in the simulation the burst errors occur in the bits and then the bits are converted to samples. For example, with 3 bits/sample in DPCM, a burst of 4 bit errors would correspond to 2 sample errors. The bit errors would cause the quantizer to assign a different level to the combinations of bits giving rise to errors.

The goal here is to create an analytical model for bursts of sample errors. Consider the model shown in Figure 3.2 where the output y(n) of the linear system h(n) is a burst of sample errors of length N. As an example, consider a sequence x(n) with single sample errors and the total sample error rate given by SER :

$$x(n) = [\ \Delta\ 0\ 0\ 0\ 0\ -\Delta\ 0\ 0\ -\Delta\ 0\ 0\ \ \Delta]$$

where Δ is the error value at each point, here assumed to be the difference between two consecutive levels in the quantizer. Consider another sequence z(n) which is just x(n) interpolated by 2 (i.e. has a zero added after each sample in x(n)) i.e.

$$z(n) = [\Delta\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ -\Delta\ 0\ 0\ 0\ 0\ 0\ -\Delta\ 0\ 0\ 0\ 0\ 0\ \Delta\ 0]$$

A MATLAB command to generate z(n) could be

```
z[1:2:2*length(x)] = x
```

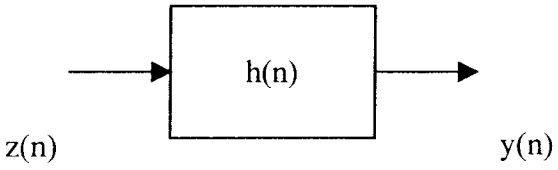


Figure 3.2: Filter Model Parameters

If this sequence is passed through a filter $h(n) = \delta(n) + \delta(n-1)$, the output from that filter is

$$y(n) = [\Delta\ \Delta\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ -\Delta\ -\Delta\ 0\ 0\ 0\ 0\ -\Delta\ -\Delta\ 0\ 0\ 0\ 0\ \ \Delta\ \Delta]$$

then this is equivalent to a burst of length 2 having the same SER as x(n).

In general, let the input z(n) be the vector of single bit errors having a given $\frac{SER}{N}$ where SER is the desired sample error rate and N is the length of the burst in samples. Then if z(n) is generated by white noise with zero mean, the autocorrelation function of z is:

$$R_{zz}(n) = \sigma_z^2\, \delta(n) \qquad (3.1)$$

where σ_z^2 is the Power Spectral Density (PSD) and is a constant for all frequencies i.e.

$$S_{zz}(f) = \sigma_z^2 \tag{3.2}$$

If the length of a burst is much less than the number of bits between bursts, then a LTI filter

$$h(n) = \sum_{i=0}^{N-1} \delta(n-i)$$

will create the output error symbol y(n) having a SER denoted by ser_1 and bursts of length N. Thus

$$S_{zz}(f) = \sigma_z^2 = \text{ser\_1/N} \tag{3.3}$$

The PSD of the output then can be computed as

$$S_{yy}(f) = |H(f)|^2 S_{zz}(f)$$

resulting in the autocorrelation function :

$$R_{yy}(n) = F^{-1}\{ S_{yy}(f) \}$$

$$R_{yy}(n) = [\, h(n) * h(-n) \,] \, (\text{ser\_1/N})$$

$$R_{yy}(n) = [(N - \text{abs}(n))/N] \, \text{ser\_1 for } n \, \varepsilon \, (-N, N) \tag{3.4}$$

Note that $R_{yy}(n)$ is a triangular function as shown in Figure 3.3 for a burst length of length 8 with a power spectral density as shown in Figure 3.4. As length of burst increases, more of the power lies is in the lower frequencies.

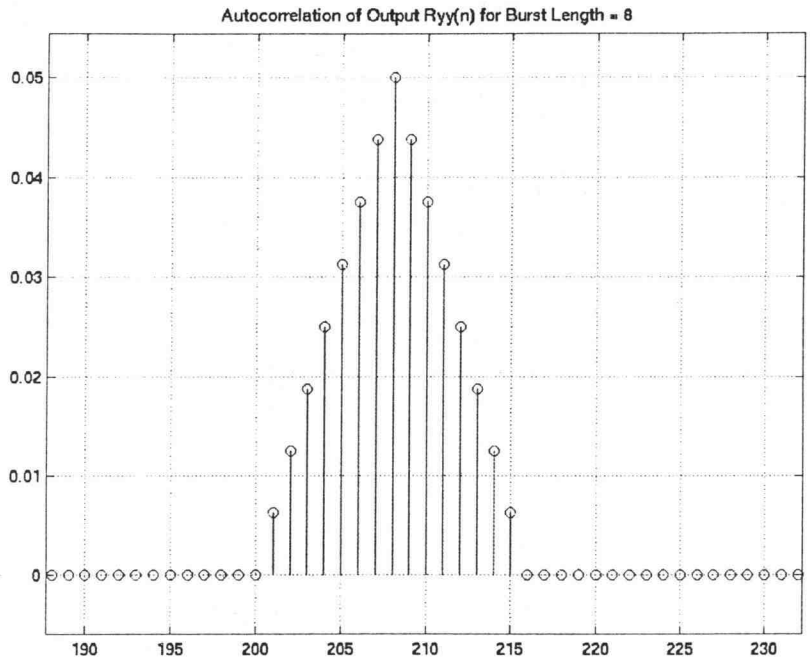


Figure 3.3: Autocorrelation for Burst Length of 8

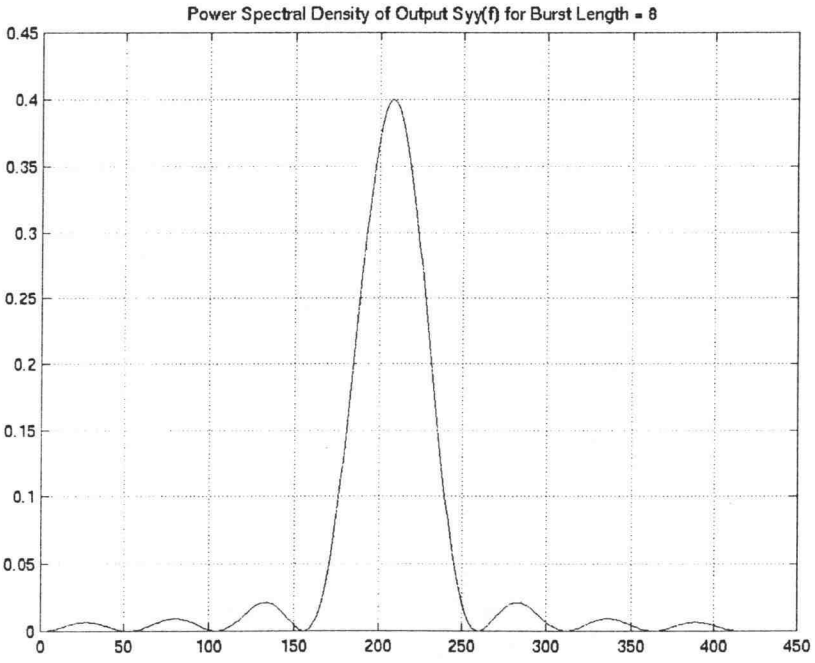


Figure 3.4: Power Spectral Density for Burst Length of 8

Consider a 3-bit quantizer having 8 levels and the difference between any two levels is Δ. Suppose there is an error of only one level, $+\Delta$ or $-\Delta$ at the output of the quantizer. This is an assumption that we make to simplify the problem. If SER is the sample error rate, and assuming the errors of $+\Delta$ and $-\Delta$ are both equi-probable, and N is the burst error length,

$$\text{probability of no error is } p(0) = 1 - \frac{SER}{N}$$

$$\text{Probability of an error of } +\Delta \text{ is } p(+\Delta) = \frac{SER}{2N}$$

$$\text{Probability of an error of } -\Delta \text{ is } p(-\Delta) = \frac{SER}{2N}$$

$$\text{And thus the total error power } k_1 = \Delta^2 * \frac{SER}{2N} + \Delta^2 * \frac{SER}{2N}$$

$$k_1 = \Delta^2 * (\frac{SER}{N})$$

If for a 3 bit quantizer with a dynamic range of 1 and Δ (= 1/4) is only one level, we can compute the value for the above constant as

$$k_1 = (0.25)^2 * (\frac{SER}{N})$$

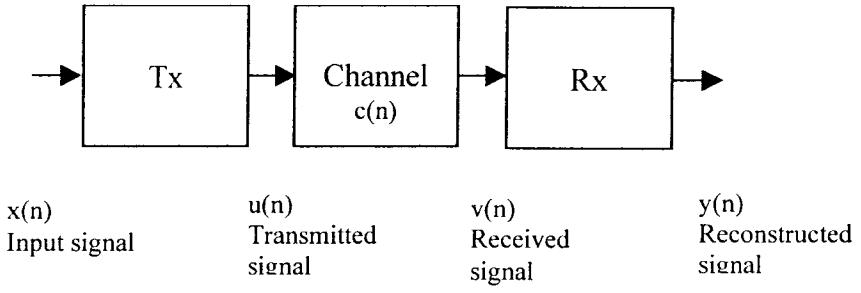
$$R_{yy}(n) = k_1 [\, h(n) * h(-n) \,]$$

$$R_{yy}(n) = [(N - abs(n))/N] \, (0.25)^2 * SER \qquad \text{for } n \in (-N, N) \qquad (3.5)$$

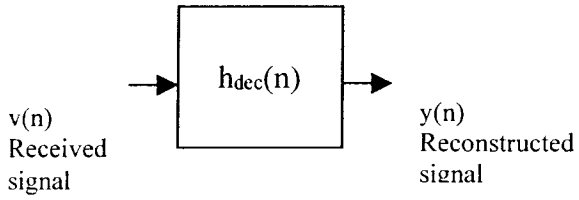
Note that since these errors are random, there exist a number of combinations of error sequences at the output of the quantizer which have not been discussed here. If all

those cases are considered, the error power would be greater than what is obtained in the above expression (3.5). The simplifications are used here to facilitate a preliminary analysis.

3.5 Analysis of Burst Errors Impact on DPCM



In the Receiver :



Tx : Transmitter
Rx : Receiver
$h_{dec}(n)$: Impulse Response of Decoder

Figure 3.5: Analysis Variables for DPCM

Figure 3.5 shows the setup used here for the analysis of Differential Pulse Code Modulation (DPCM) for burst errors. The burst errors occur in the channel and corrupt the transmitted signal. If there are no errors, then c(n) = 0, and thus the received signal is identical to the transmitted signal, i.e. v(n) = u(n). If errors occur then $c(n) \neq 0$, and thus v(n) = u(n) - c(n). When errors occur,

$$y(n) = v(n) * h_{dec}(n) = [u(n) - c(n)] * h_{dec}(n) = u(n) * h_{dec}(n) - c(n) * h_{dec}(n)$$

In the above equation if we analyze each term, we obtain:

$$u(n) * h_{dec}(n) = \text{output without channel errors} = x(n) - q(n)$$

$$c(n) * h_{dec}(n) = \text{errors introduced due to the channel.}$$

The total reconstruction error is just the sum of the reconstruction and channel errors, or

$$r(n) = x(n) - y(n) = q(n) + c(n) * h_{dec}(n)$$

Total reconstruction error variance (assuming zero mean) can therefore be written as

$$\sigma_r^2 = E[\, r^2(n)\,] = E[\, (q(n) + c(n) * h_{dec}(n))^2\,]$$

$$= E[\, q^2(n)\,] + E[\, (c(n) * h_{dec}(n))^2\,] + 2E[\, q(n)\,[c(n) * h_{dec}(n)\,]\,]$$

$$= \sigma_q^2 + 2 \sum_{k=0}^{\infty} h_{dec}(n)\, E[\, q(n)c(n-k)\,] +$$

$$\sum_{k=0}^{\infty} \sum_{l=0}^{\infty} h_{dec}(k)\, h_{dec}(l) E[\, c(n-k)c(n-l)] \qquad (3.6)$$

for LTI systems.

Since the quantization noise is due to the inherent quantizer resolution, we can reasonably assume that there is no correlation between the quantization noise and the channel errors and thus,

$$E[\, q(n)c(n-k)\,] = 0 \qquad (3.7)$$

Substituting equation (3.7) in (3.6) yields

$$\sigma_r^2 = \sigma_q^2 + \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} h_{dec}(k)\, h_{dec}(l) E[\, c(n-k)c(n-l)]$$

Now,

$$E[\, c(n-k)c(n-l)\,] = R_{cc}(k-l) \qquad (3.8)$$

Thus we can write, combining all the above conditions,

$$\sigma_r^2 = \sigma_q^2 + \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} h_{dec}(k)\, h_{dec}(l)\, R_{cc}(k-l)$$

For N = 2,

$$R_{cc}(m) = R_{yy}(m) = \Delta^2 \left(\frac{SER}{N}\right) [\, 2\delta(m) + \delta(m-1) + \delta(m+1)]$$

Now if:

$$R_{dec}(n) = \sum_{k=0}^{\infty} h_{dec}(k+n)\, h_{dec}(n)$$

Then $$\sigma^2_r = \sigma^2_q + \Delta^2 \left(\frac{SER}{N}\right) [2R_{dec}(0) + R_{dec}(1) + R_{dec}(-1)]$$

The general expression for the total reconstruction error power for a burst of length N is:

$$\sigma_r^2 = \sigma_q^2 + \Delta^2 \frac{SER}{N} \sum_{n=-(N-1)}^{(N-1)} [N - |n|] R_{dec}(n) \qquad (3.9)$$

3.6 Implementation and Comparison for DPCM

To use the above expression for computing the reconstruction error power for DPCM, the h_{dec} for each block was used to compute the error power from equation (3.8) for that block and then the computed SNR obtained. Since h_{dec} is an IIR filter, the autocorrelation of the impulse response was computed by taking the inverse FFT of the power spectral density of the filter when input to the filter is white noise. The simulated SNR was as obtained via MATLAB simulations and is described in detail in Chapter 4. As in section 3.4, a simplified case where the sample errors caused by bit errors correspond to an error of 1 quantizer level. Also, assume that the number of bit errors

caused by a burst of length N bits in a total of T bits translates to (N/3) sample errors in a total of (T/3) samples where 3 bits/sample are used In this case, the sample error rate(SER) can be assumed to be same as the bit error rate. Note that this is the best case, i.e., the fewest number of sample errors given the number of bit errors. For DPCM a 3 bit quantizer has been used and so $\Delta = 0.25$.

For this simplified model, Figure 3.6 shows a plot of the analytical and simulated SNR values for DPCM for a SER of 0.001

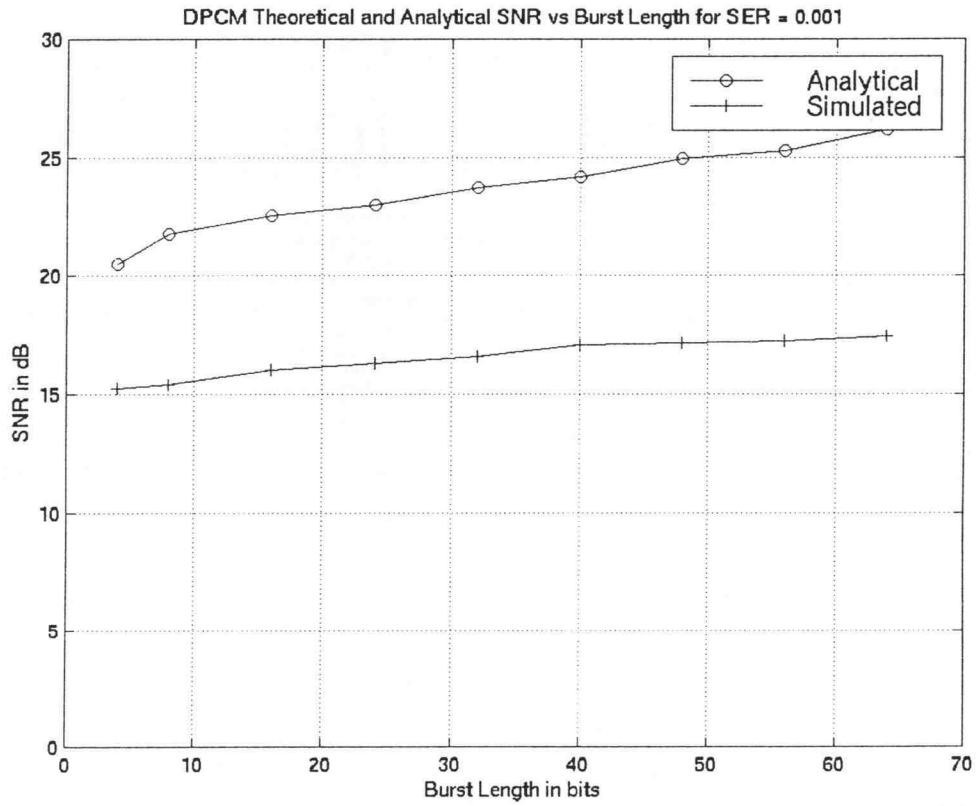


Figure 3.6: DPCM Analytical and Simulated SNR for SER = 0.001

It is important to note here that the analytically computed values are higher than the simulated values. Recall that for analysis, the simplifying assumption used was that

all the errors are such that the output of the quantizer changes only by one level. This is the best case which gives the lowest error power. For all other combinations of errors at the quantizer output causing the actual samples to change by more than one level, the error power will be much larger giving a lower value for SNR as is obtained in the simulated values. Both the analytical and simulated SNR values increase with increasing burst lengths in bits. The reason for this increase is argued in Chapter 4.

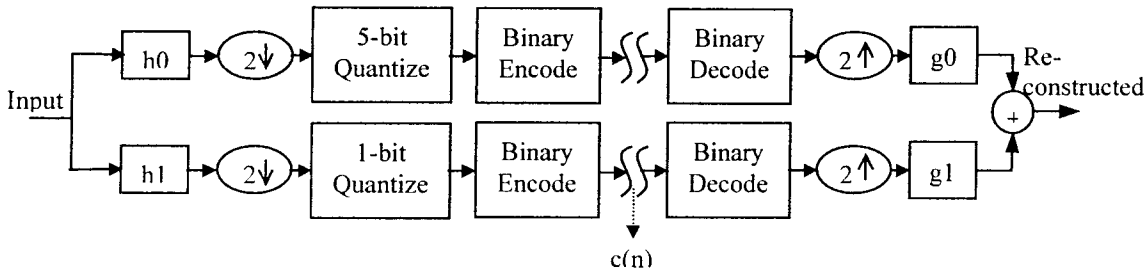


Figure 3.7: Analysis Variables for Subband

3.7 Analysis of Burst Errors Impact on SBC

The goal in this section is to analyze the impact of burst errors on SBC, using the same procedure as described in the previous section. Consider a 2 channel Subband system as shown in Figure 3.7. As with DPCM the total errors is the sum of the quantization noise q(n) and the out of decoder r(n) channel errors c(n) out of the decoder , i.e.,

$$r(n) = q(n) + c(n) * g_0(n) + c(n) * g_1(n)$$

where c(n) are the channel errors and $g_0(n)$, $g_1(n)$ are the impulse responses of the filters in the decoder. Thus the total error power, σ_r^2 , can be written as(assuming zero mean, without loss of generality),

$$\sigma_r^2 = E[\, r^2(n)]$$

$$= E[\, (q(n) + c(n) * g_0(n) + c(n) * g_1(n))^2\,]$$

$$= E[q^2(n)] + E[(c(n) * g_0(n))^2] + E[(c(n) * g_1(n))^2] + 2E[q(n)(c(n) * g_0(n))] \quad +$$

$$2E[q(n)(c(n) * g_1(n))] + 2E[(c(n) * g_0(n))(c(n) * g_1(n))] \qquad (3.10)$$

To simplify this expression in equation (3.10) we make the following assumptions:

- The correlation between the inherent quantization noise and the channel errors is zero, i.e., E[q(n)c(n)] = 0 (also used for DPCM).
- The correlation between the filters $g_0(n)$ and $g_1(n)$ is negligible since they occupy distinct frequency bands. As a result the cross terms are zero in equation (3.10). While this simplifying assumption is not true for QMF filters, but is reasonable given that transition bands are narrow relative to filter bandwidth.

With these assumptions, equation (3.10) is simplified as follows:

$$\sigma_r^2 = E[q^2(n)] + E[(c(n) * g_0(n))^2\,] + E[(c(n) * g_1(n)\;\;)^2]$$

Since c(n) is random and the receiver filters are fixed, then,

$$\sigma_r^2 = \sigma_q^2 + \sum_{k=0}^{\infty}\sum_{l=0}^{\infty} g_0(k)\, g_0(l) E[c(n-k)c(n-l)] +$$

$$\sum_{k=0}^{\infty}\sum_{l=0}^{\infty} g_1(\,k)\, g_1(l) E[c(n-k)c(n-l)] \qquad (3.11)$$

Now substituting for Rcc(n) as in equation (3.8), equation (3.11) becomes:

$$\sigma_r^2 = \sigma_q^2 + \sum_{k=0}^{\infty}\sum_{l=0}^{\infty} g_0(k)\, g_0(l) R_{cc}(k-l) + \sum_{k=0}^{\infty}\sum_{l=0}^{\infty} g_1(k)\, g_1(l)\, R_{cc}(k-l)$$

Now if :

$$Rg_0(n) = \sum_{k=0}^{N} g_0(k+n) g_0(n) \quad \text{and}$$

$$Rg_1(n) = \sum_{k=0}^{N} g_1(k+n) g_1(n) \quad \text{then}$$

$$\sigma_r^2 = \sigma_q^2 + \Delta^2\left(\frac{SER}{N}\right)\{ [N - abs(n)]\, Rg_0(n) + [N - abs(n)]\, Rg_1(n)\} \text{ for } n \in (-N, N) \quad (3.11)$$

Comparing this the expression of total reconstruction error of subband to DPCM, we find that they both involve deterministic ' correlation ' functions but in the case of subband there are two filters instead of just one. Again assume that in the best case the sample error rate is same as the bit error rate, and the quantizer error corresponds to only a single level quantizer. With these assumptions, the values for the computed SNR have been obtained. It is important to note here that in SBC, multiple (two here) quantizers are used, each of which encodes the sample levels using different number of levels. The bits are multiplexed using TDM from each channel. For the 2 channels, 24 kbps system used in this analysis, one frame comprises of four samples from channel 1 and four samples from channel 2. The resulting frame can be designated as [1 1 1 1 2 2 2 2] as one frame (where the numbers indicate the samples from the given channel numbers). Frames are transmitted at a rate of 4000 per second, resulting in a 24 kbps rate.

Given this frame arrangement and the number of bits per sample, the probability of an error in channel 1 is:

The probability of an error in the first four samples = 20/24 = 5/6 (3.12)

Similarly,

The probability of an error in the next two samples = 4/24 = 1/6 (3.13)

Using equation (3.5) and the above results, the value of the total error power at the output of the quantizer can be computed as:

$$k_1 = (\frac{SER}{N}) [\frac{5}{6}\Delta^2_1 + \frac{1}{6}\Delta^2_2] \quad \text{where } \Delta_1 = 0.0625;\ \Delta_2 = 0.5;$$

3.8 Implementation and Comparison for Subband

As with DPCM, the SNR is computed analytically for different burst lengths assuming that the quantizer error for SBC is only a single level. Also sample error rate is assumed to be the same as the bit error rate, 0.001 in this case. Figure 3.8 shows analytical and simulated SNR values for SBC in the presence of burst errors of different lengths.

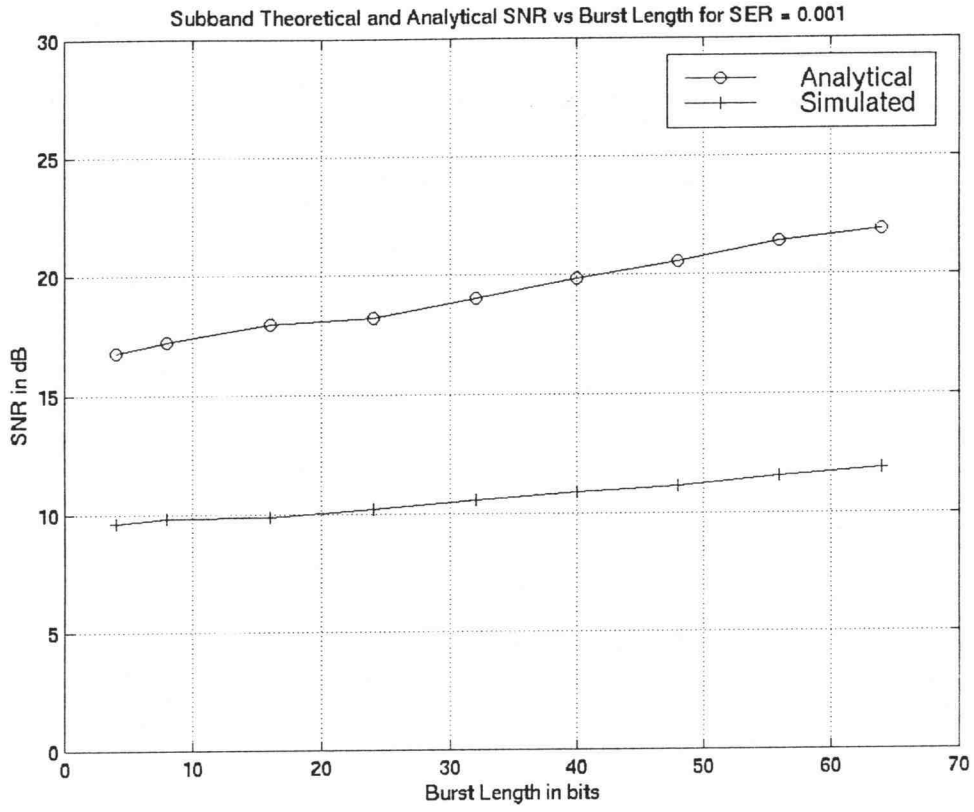


Figure 3.8: Subband Analytical and Simulated SNR for SER = 0.001

As with DPCM, calculated SNR is greater than the simulated values since the sample errors may be greater than the single-quantizer level error assumed here. Note that, as with DPCM, the SNR increases with increases in burst length. The reasons for this increasing trend are discussed in Chapter 4.

Chapter 4: Simulation Performance Comparison

4.1 Performance Measures

4.1.1 Introduction

Evaluating performance requires the use of measures. For speech, it is well understood that while SNR is widely used as a quantitative measure of relative performance, it is not a good measure of perceived quality. The purpose of this chapter is to summarize three performance measures of the quality of the reconstructed speech as used in this thesis. These measures are then used to compare the two 24kbps speech codecs. Signal to Noise Ratio (SNR), Segmental Signal to Noise Ratio (SNRSEG), and the Mean Opinion Score (MOS) are the measures of coder quality used in this thesis. DPCM and SBC performance are compared using these different measures.

4.1.2 Signal to Noise Ratio (SNR)

The reconstruction error r(n) in digital coding is defined as the difference between encoder input x(n) and decoder output y(n) i.e. r(n) = x(n) – y(n). The signal to noise ratio, SNR, is defined as the ratio of input signal variance σ_x^2 to reconstruction error variance σ_r^2; i.e.,

$$SNR = \frac{\sigma_x^{\ 2}}{\sigma_r^{\ 2}} \tag{4.1}$$

which can be expressed in dB as

$$SNR_{dB} = 10\log_{10} SNR \tag{4.2}$$

The reconstruction error variance σ_r^2 is the mean square error (MSE). Codec designs that minimize this quantity are called minimum mean square error (MMSE) designs. SNR is easy to compute both analytically and through simulations and so is a widely used measurement technique. The segmental SNR, discussed next, provides better correlation with perceived quality by accounting for the time varying nature of SNR during speech.

4.1.3 Segmental Signal to Noise Ratio (SEGSNR)

In the ongoing quest for a subjectively meaningful objective measure of coder quality, several refinements of the conventional SNR have been proposed and used in speech and image work. An important class of SNR refinements, used widely in speech coding, are those that recognize the fact that the speech signal is non-stationary and that the same amount of noise has may have different perceptual values depending on the signal to which it is added. The segmental SNR (SEGSNR) measure is based on a log weighting that converts component SNR values to dB values and then averages those values.

The segmental SNR is defined as

$$\begin{aligned} SEGSNR(dB) &= E[SNR(m)(dB)] \\ &= \frac{1}{M}\sum_{m=1}^{M} SNR_{dB}(m) \end{aligned} \tag{4.3}$$

where SNR(m) is the conventional SNR for segment m, and the expectation is a time average over all segments of interest in an input sequence. An appropriate segment length

in speech, which is typically used, would be of the order of 16-20 msec, the same as used for Linear Predictive analysis. The following flow chart shows the steps required to compute the segmental SNR.

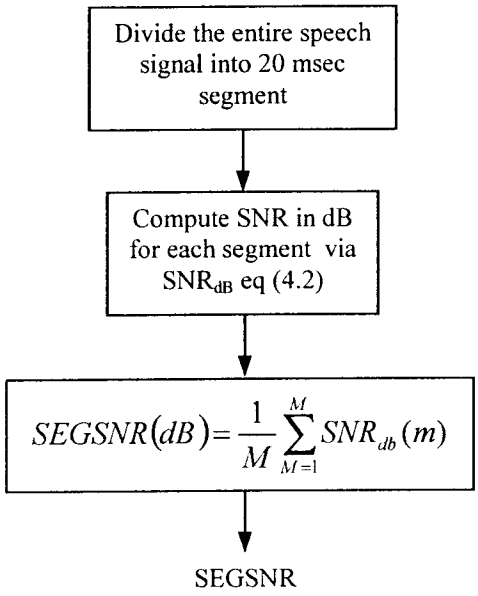


Figure 4.1: Steps to Compute Segmental SNR.

4.1.4 Limitations of SNR and SEGSNR

While SEGSNR takes into account the time varying nature of speech it still is not a very good measure of perceived speech and quality. The insufficiency of the SNR measurements has to do with the fact that the quantization error sequence has signal dependent or signal correlated components. This signal dependent noise does not have the same annoyance value as independent additive noise of equal variance [5, 15]. Consequently, the perceptual quality provided by a quantizer with signal correlated errors cannot be completely described by the ratio of signal power σ_x^2 to reconstruction error

power σ_r^2. Additionally, different frequencies have different perceptual importance. This can be illustrated by a simple example of a high frequency tone in the reconstructed speech signal. In this case, σ_r^2 can be quite small and the SNR can be very high, but the perceived speech quality can be extremely poor [2]. Thus it is also important to look at a subjective measure for complete experimentation and analysis.

There exist different types of objective measures, which represent accurately the subjective quality of speech. The Bark Spectral Distance [7,14] (BSD), Mel Scale Distance [28, 29] (MSD), Auditory Frequency Weighted Spectral Distance [31] (AUSD) are some psycho-acoustically motivated measures of perceptual quality of speech coders. The use of these measures may be a thesis in themselves. In this thesis, we have used the simplest yet most subjective measure, the mean opinion score test as described in the next section.

4.1.5 Mean Opinion Score (MOS)

In order to perform a subjective quality assessment a commonly used measure is the mean opinion score (MOS). In this thesis, MOS has been used to perform an informal subjective quality assessment of the reconstructed speech. MOS scores require subjective testing, but are accepted as a norm for comparative rating of different systems.

MOS is a quality test involving the recruitment of an ensemble of subjects, each of whom classifies a coder output on an N-point quality scale; for example, on a 5-point scale for signal quality. The rating scale employed in MOS testing is illustrated in the following table together with a general description of the levels of distortion typically associated with each numerical score. A MOS score is a mapping of perceived levels of

distortion into either the descriptive terms " excellent, good, fair, poor, unsatisfactory " or into equivalent numerical ratings in the range 5-1. The numerical mapping permits the ranking of coders and comparisons with other objective measures. However, MOS lumps different kinds of distortions together, providing very little insight into the causes of distortion. The final results from these tests is a pooled average judgement called the Mean Opinion Score (MOS) for the ensemble of listeners.

The MOS test in this thesis was done by averaging the judgement of 13 people To minimize the effect of external noise, headphones were used instead of speakers. Also, the tests were conducted in my office in the evenings over a period of three days. The people helping to do the MOS test were told to compare the sound quality as if they were hearing the speech over the phone.

Rating	Speech Quality	Level of Distortion
5	Excellent	Imperceptible
4	Good	Just perceptible but not annoying
3	Fair	Perceptible but slightly annoying
2	Poor	Annoying but not objectionable
1	Unsatisfactory	Very annoying and objectionable

Table 4.1: Descriptions in the Mean Opinion Score (MOS)

4.2 Simulations

4.2.1 Introduction

In this section, the DPCM and SBC codecs described in Chapter 2 are evaluated with respect to their performance given the BEC described in Chapter 3. The results will show the burst error performance of the two codec types for fixed bit error rates of 0.001 and 0.05 for two kinds of speech segments, a voiced segment " We were away a year ago" and an unvoiced segment " She sells sea shells sea shore ". Voiced sounds like "ahh" are produced by the resonant cavity of the throat, while unvoiced sounds like "shh" are more noise like and produced in our mouths. As explained above, the performance measures are SNR, Segmental SNR and MOS.

4.2.2 Performance for Burst Errors

For these simulations, the BER has been fixed to a given value (0.001, 0.05) and the two codec performances simulated for burst errors of increasing lengths. Since the errors are being added to the bits, different number of speech samples will be in error depending on the number of quantizer levels and the location of the error. For example, for a 3 bit quantizer, a burst of 8 bit errors would correspond to either 3 or 4 sample errors. For a 7 bit quantizer, a burst of 8 bit errors would transform to only 2 sample errors. The two values of BER were chosen to exhibit two cases, one in which there is a single error in every 1000 bits and one in which there is an average of 50 errors in every 1000 bits. The bit error rates have been chosen so that they are significant enough to

cause perceptual degradation in the final output. The values graphed were obtained by averaging values obtained by repeating the simulation three times.

4.2.2.1 SNR Performance Comparison for a Voiced Speech Segment

Figures 4.2 and 4.3 show the SNR performance of the two codecs for different burst lengths for a BER of 0.001 and 0.05, respectively. For both BERs, both DPCM and Subband show an increase in SNR with increasing burst lengths in bits. Recall that this trend of increasing SNR with increase in burst length in bits was also predicted analytically in chapter 3. For a BER of 0.05, the SNR again increases with increase in burst length for both Subband and DPCM.

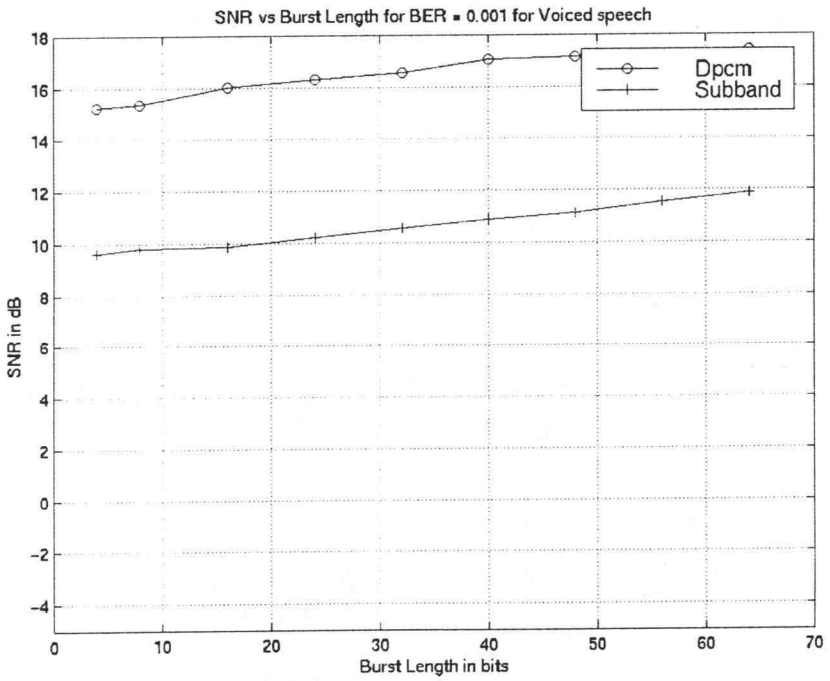


Figure 4.2: SNR vs. Burst Length for BER = 0.001 for Voiced Speech

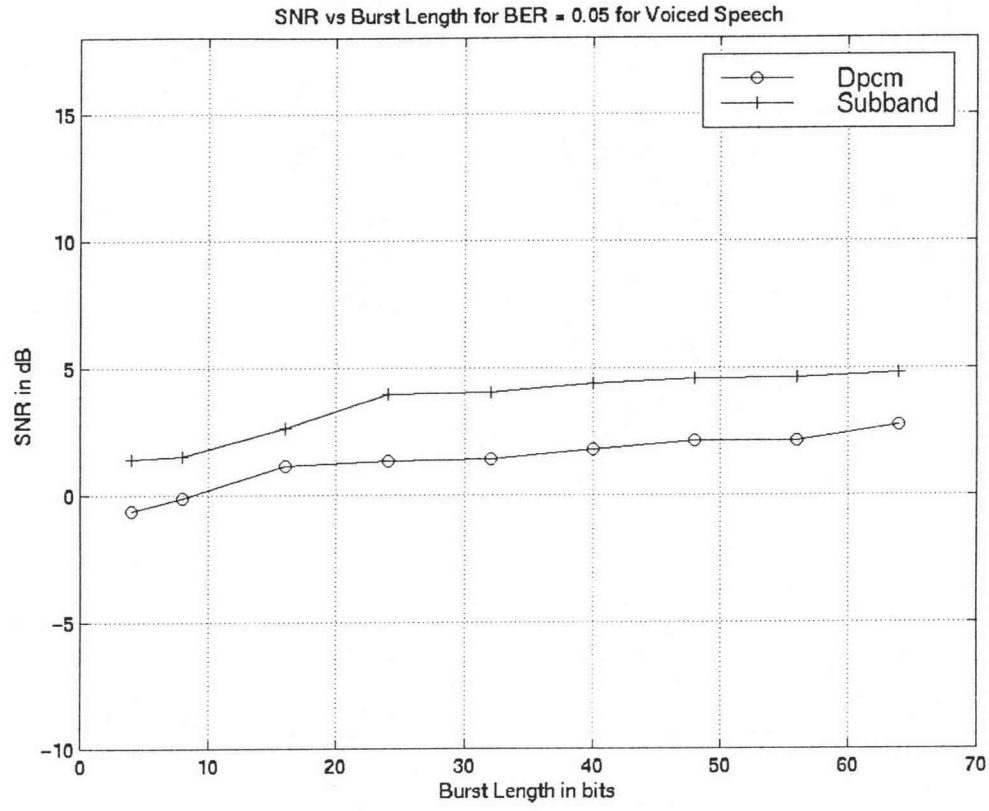


Figure 4.3: SNR vs. Burst Length for BER = 0.05 for Voiced Speech

While increasing SNR with burst length is not intuitive, this effect can best be understood by examining the errors at the input and output of the binary decoder and the source decoder. The goal is to determine whether the increase in SNR is an artifact of the bits to samples conversion, or a result of decreased error propagation in the source decoder. Here we show that the primary effect is the later effect.

Figure 4.4 shows the noise power at the *input of* source decoder (DPCM/Subband) vs the burst length in bits for a BER of 0.05. While the BER is fixed, the effective input noise power initially reduces, with increase in burst length in bits. Note that the input noise power is high for single bit errors (or burst length = 1), and

decreases rapidly until a burst length of 8 bits, after which it levels out and becomes nearly constant. Both DPCM and SBC binary decoders exhibit the same trend of reducing decoder input noise power with increase in burst length in bits. Since SBC uses three different binary decoders, namely, 1-bit, 3-bit and 7-bit the effect of error power reduction is more evident here. DPCM uses a 3-bits/sample decoder.

Recall that the burst errors are being added to the bits and not directly to the sample values. For DPCM, for example, three bits is equivalent to one sample. Thus, when a burst error occurs in bits, it transforms to a lower number of sample errors. When single bit errors occur randomly, each bit error transforms into a sample error. When the errors occur in bursts, several bit errors may cause only one sample error. The error magnitude depends on the type of binary coding used. For e.g. if Gray Coding was used as the coding technique in the binary decoder, the error magnitude would be different.. For burst errors therefore, the total number of sample errors out of the binary decoder reduce, since the samples can only take certain specific values given by the levels in the decoder. The effective noise power out of this binary decoder (or into the source decoder or the input noise power) reduces. But this reduction is only significant up to a burst length = 16 bits, after which it becomes constant.

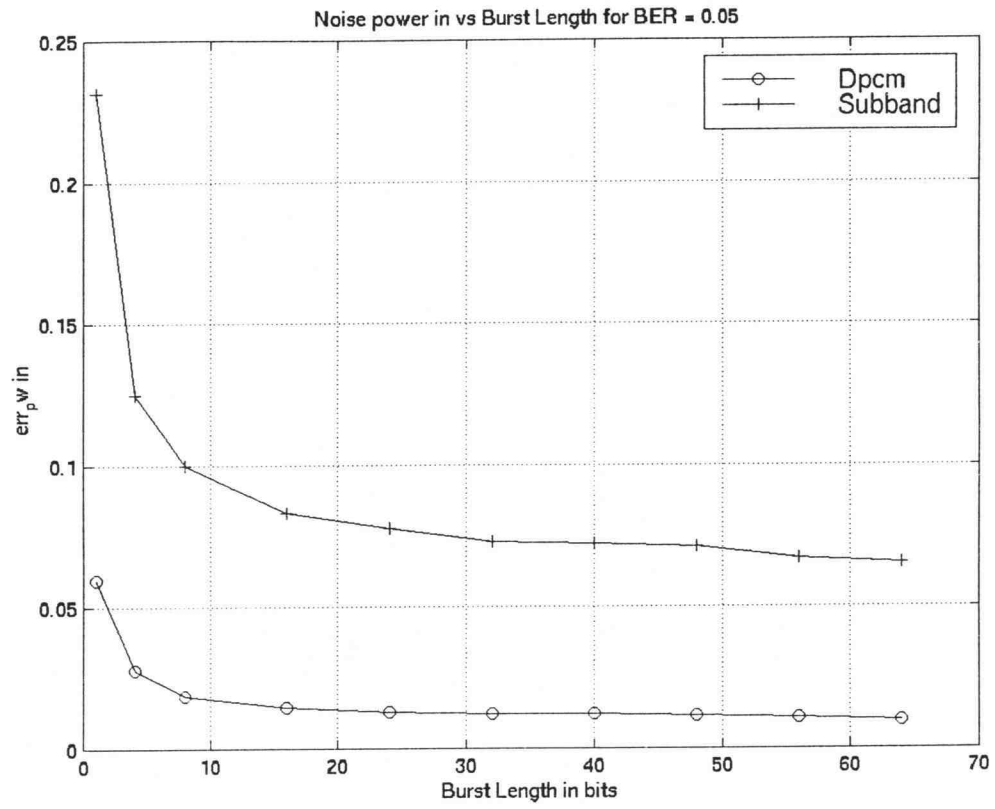


Figure 4.4: Noise Power into Source Decoder vs. Burst Length in Bits

Consider now the effect of the source decoder on the error power. Note the term decoder now refers to the source decoder (DPCM/SBC) unless mentioned otherwise. Figure 4.5 shows the percentage of samples in error *out* of the decoder vs the percentage samples in error in to the decoder for both DPCM and SBC. The burst length in bits increases right to left on the x axis. Note that the percentage samples in error out of the decoder reduces with the increase in burst length. At the output of the decoder, the samples are requantized to 8000 samples/sec and 8 bits per sample to give the 64 kbps toll quality speech. Thus the samples can again take specific values at the output of the

source decoder. The *number* of samples in error out reduces with the increasing burst length in bits and the SNR increases.

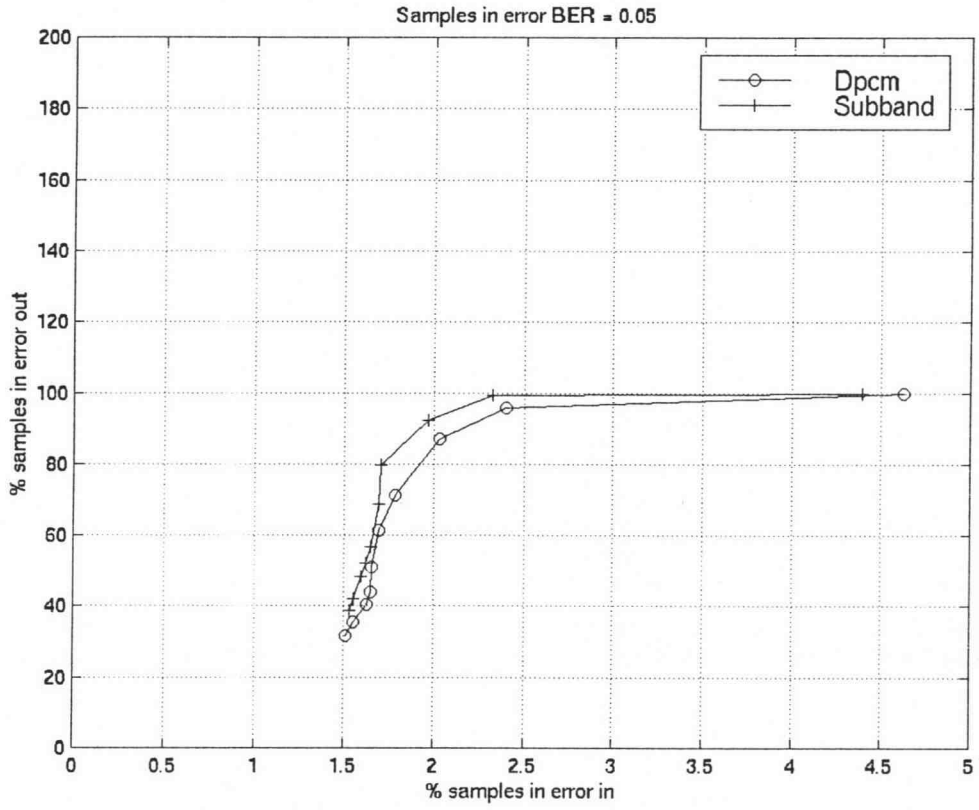


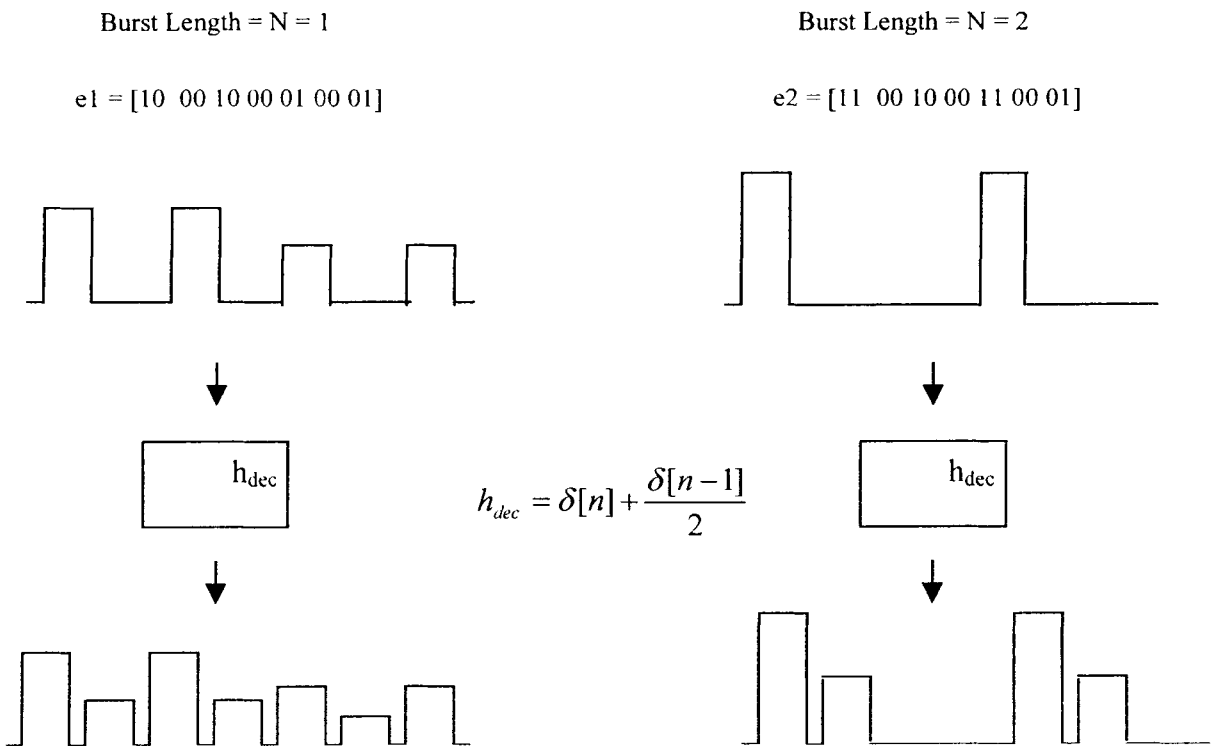
Figure 4.5: % Samples in Error out vs. % Samples in Error in Source Decoder

The effect of source decoder reducing the number of sample errors with increasing length of bursts can be explained by the phenomenon of error propagation in this decoder, and is illustrated in Figure 4.6. Suppose the BER is fixed at 2/7 and the quantizer in the binary decoder has a resolution of 2 bits/sample. In a simple case, assume that the source decoder $h_{dec}(n)$ can be modeled by an impulse response:

$$h_{dec}(n) = \delta(n) + \frac{\delta(n-1)}{2} \tag{4.4}$$

which is a reasonable assumption as all filters after convolution with the input produce artifacts. The second term in the above equation models the error propagation. Note from the Figure 4.6, for a burst length = 1 or single random errors, the error propagation due to the decoder transfer function is more as compared to when the burst length = 2. For a higher burst length, the error propagation is less, resulting in reduction of the error power at the output of the source decoder. The SNR therefore increases.

In DPCM, the signal passes through a single stage of filtering. This filter is an all-pole Infinite Impulse Response (IIR) filter. Ideally an IIR filter has an infinite roll off and its impulse response continues on forever, but practically the length of the filter is limited by its time constant. Beyond a given range, the impulse response is negligible. In particular, it is negligible when its effect is < 1 quantization level. Here the effective length of the filter is within 30-35 for each block. For SBC, the filters in each stage are Finite Impulse Response (FIR) filters each of length 30.



h_{dec}: Impulse Response of source Decoder

Figure 4.6: Error Propagation in the Source Decoder

4.2.2.2 SNR Performance Comparison for the unvoiced speech segment

The performance of both the codecs deteriorates for the unvoiced speech sample. DPCM shows a lower signal to noise ratio and so does SBC. The overall trend though of the SNR increasing with increasing burst length into the source decoder remains the same.

The Linear Prediction model used for DPCM is for voiced sounds produced by the resonating cavity in our throats. Therefore, for an unvoiced speech signal, the prediction error is greater and the performance of DPCM goes down. Recall that

unvoiced sounds are more noise like and so have a higher amount of energy in the high frequency sections as compared to the voiced case. Thus when a low-resolution quantizer is used to encode the high frequency content, more information is lost here than in the voiced case. So SBC performance also deteriorates for unvoiced sounds. The deterioration in DPCM is greater as LPC assumed a voiced model, yielding a higher prediction error for unvoiced speech. Figures 4.7 and 4.8 compare the performance of DPCM and subband for the speech sample " She sells sea shells sea shore " for BER of 0.001 and 0.05.

The SNR for both DPCM and SBC increases with increasing burst length at BER of 0.001 and 0.05. The absolute SNR values are lower than the corresponding voiced case due to reasons mentioned above.

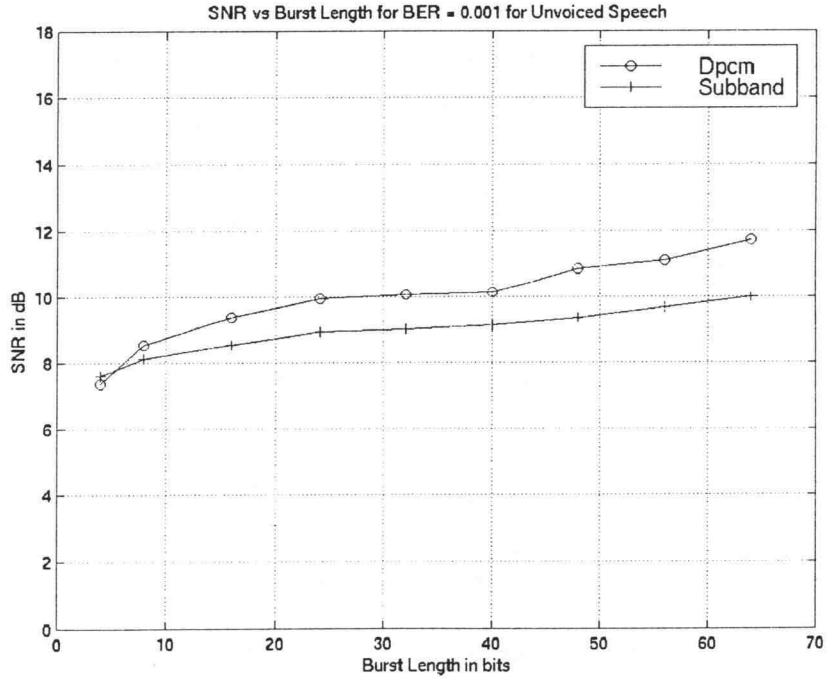


Figure 4.7: SNR vs. Burst Length for BER = 0.001 for Unvoiced Speech

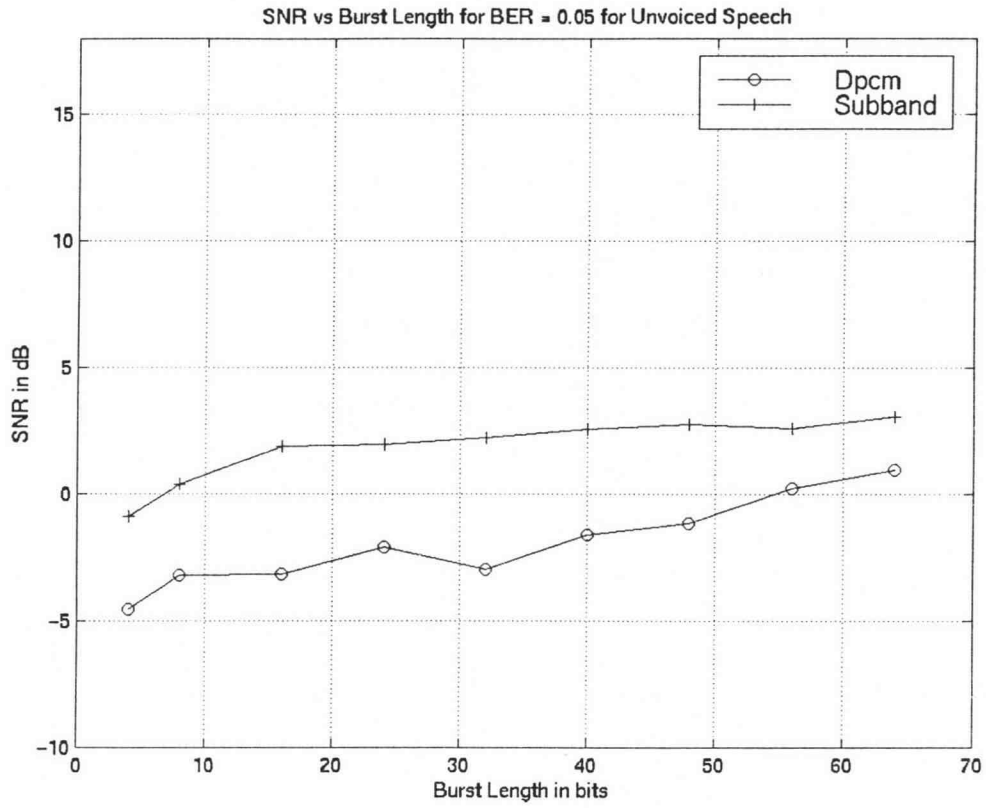


Figure 4.8: SNR vs. Burst Length for BER = 0.05 for Unvoiced Speech

4.2.2.3 Segmental SNR Performance Comparison for Voiced speech

SEGSNR obtains the SNR(dB) for each block (of 160 samples) in both DPCM and SBC and then obtains the average SNR(dB). Figure 4.9 and Figure 4.10 show how the Segmental SNR changes with increasing burst lengths for BER's of 0.001 and 0.05 respectively. Both use the voiced signal input.

For BER = 0.001 and 0.05, the Segmental SNR for DPCM and SBC increases again with increasing burst length. However, the rate of increase is less than that exhibited for SNR, particularly for SBC, where the curves are relatively flat.

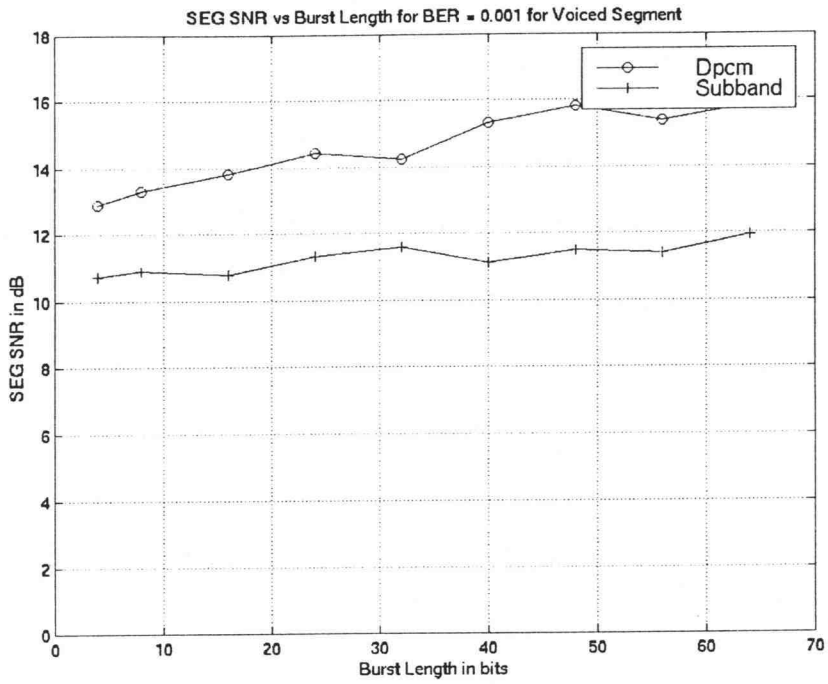


Figure 4.9: SEG SNR vs. Burst Length for BER = 0.001 for Voiced Speech

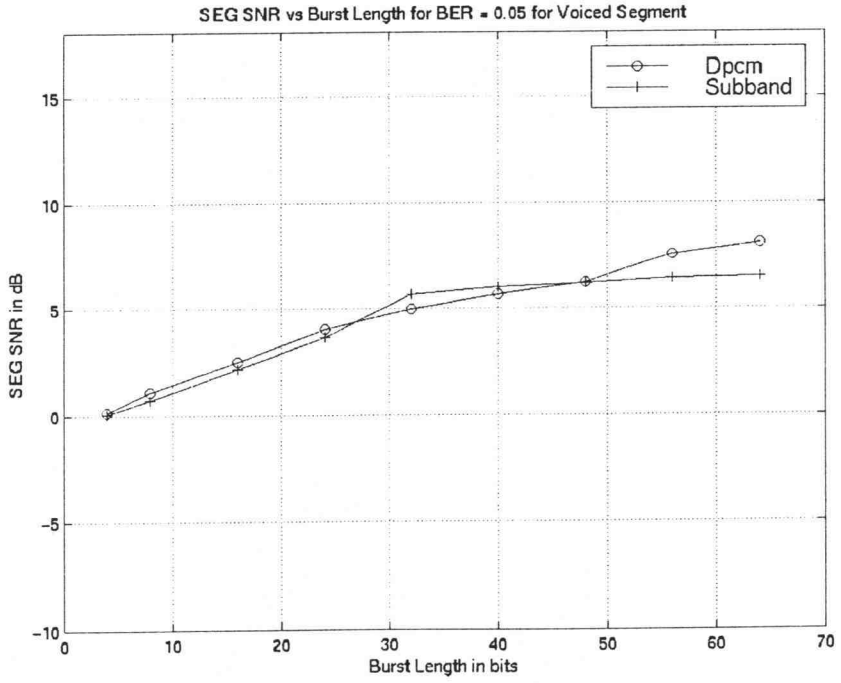


Figure 4.10: SEG SNR vs. Burst Length for BER = 0.05 for Voiced Speech

Recall that segmental SNR is the geometric mean of the SNR values per block. When a higher burst length error occurs, the number of occurrences of this string is over a fewer blocks as compared to the shorter bursts for a given fixed error rate. Now, for a longer burst length, there are a greater number of blocks where no errors occur. The SNR (dB) for an errorless block is very high (it is equal to the SQR). On computing the SEG SNR, a higher number is therefore obtained.

4.2.3.4 Segmental SNR performance for unvoiced speech segment

The trend of increasing segmental SNR with burst length remains for unvoiced speech as well. Figure 4.11 shows the segmental SNR performance for an unvoiced speech segment given a BER = 0.05.

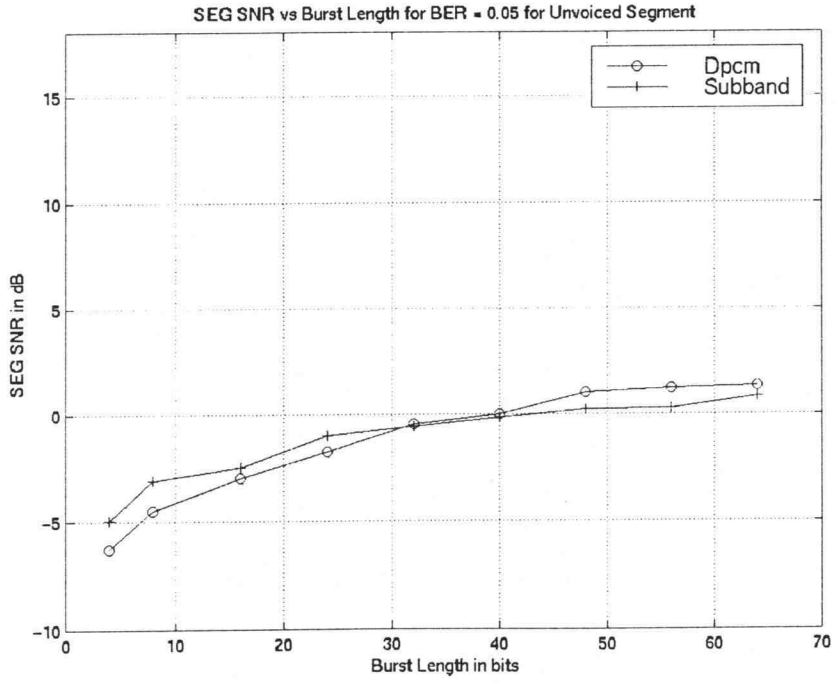


Figure 4.11: SEG SNR vs. Burst Length for BER = 0.05 for Unvoiced Speech

All results show the same trend: increasing quality for increasing burst lengths(with a fixed BER). The quality here is measured by SNR and SEG SNR. The consistency of the trends substantiates the observations. Since it is well known that SNR does not correlate well with perceptual quality, performance of DPCM and SBC for increasing burst errors is evaluated using subjective MOS next.

4.2.3.5 Mean Opinion Score Performance Comparison for voiced speech

The MOS scores were obtained using 13 subjects as described in section 4.1.5. Recall that the MOS scale has 5 levels and a maximum score of 5 corresponds to "Excellent" while a score of 1 corresponds to unsatisfactory. The listeners compared the various decoded sound files with the original toll quality speech. The sound files for different burst lengths were ordered randomly. Data obtained from one subject was not shown to another person to preclude any bias in judgement.

The MOS results for Voiced speech are shown in Figures 4.12 and 4.13, for BER of 0.05 and 0.001 respectively. Figure 4.12 reflects the same trend of increasing perceived quality with increasing burst length. For SBC at a BER of 0.05, the sound quality is close to the "Poor" section (MOS 2.4) for bursts of lengths 4 and 8 and then improves for higher burst lengths and is then rated as " Fair " (MOS 3.57). There is a slight improvement when we go from 24 bits to 32 bits of burst error lengths. The improvement in MOS scores with increasing burst length in bits correlates well with the SNR and segmental SNR trends.

Overall for DPCM the speech quality is "Good " (MOS 4.35) for burst lengths of 16 and higher. It deteriorates for lengths of 4 and 8 burst bit errors (MOS 3.7). For DPCM the quality seems to improve when we go from 16 to 24 bit burst errors.

When comparing the original and decoded speech in both cases, the listeners found that the SBC speech segment lost its individuality and seemed to be more synthetic. To the listeners, this effect was more annoying than the background-like noise more evident in the DPCM – coded speech.

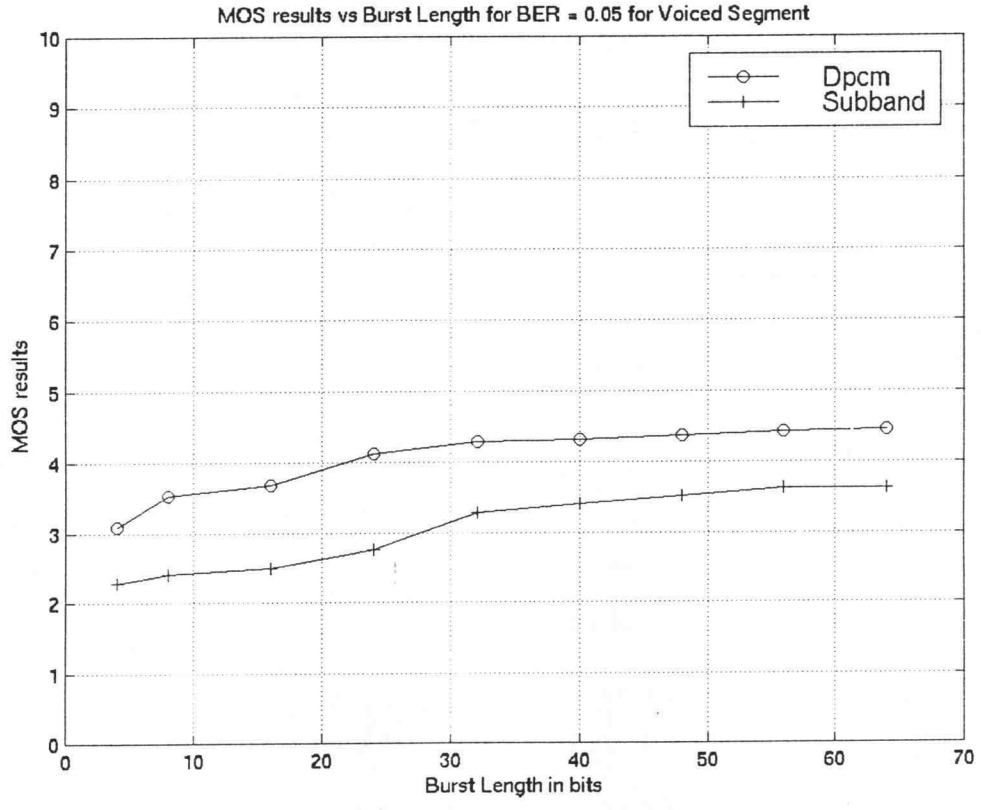


Figure 4.12: MOS Results for Voiced Speech at a BER of 0.05

Note that in SBC the position of the burst of errors is crucial. If the errors occurred in regions, which had low frequency information, then the distorting effects

would be more pronounced than the case when the errors occur in the higher frequency regions. In standard subband decoders there is more error protection on the crucial bits containing low frequency information. Also the propagation of the errors is more due to the different stages of filtering, so greater number of regions are effected.

At a BER of 0.001, it was difficult for the listeners to find any noticeable difference in the speech quality for various levels of burst errors. Both subband and DPCM files seemed to be the same and it gave a flat MOS score of 4 as shown in Figure 4.13. Note that MOS score for a standard 24 Kbps Subband codec with fixed bit allocation is 3.9 which corresponds to the "Good" rating [2]. In standard texts, subjective MOS ratings for DPCM communication quality speech at bit rates of 24 to 32 kbps are on the order of 3.5 or higher [2]. For the informal listening tests in this thesis, the MOS for the very low BER was four, which correlates well with the standard results.

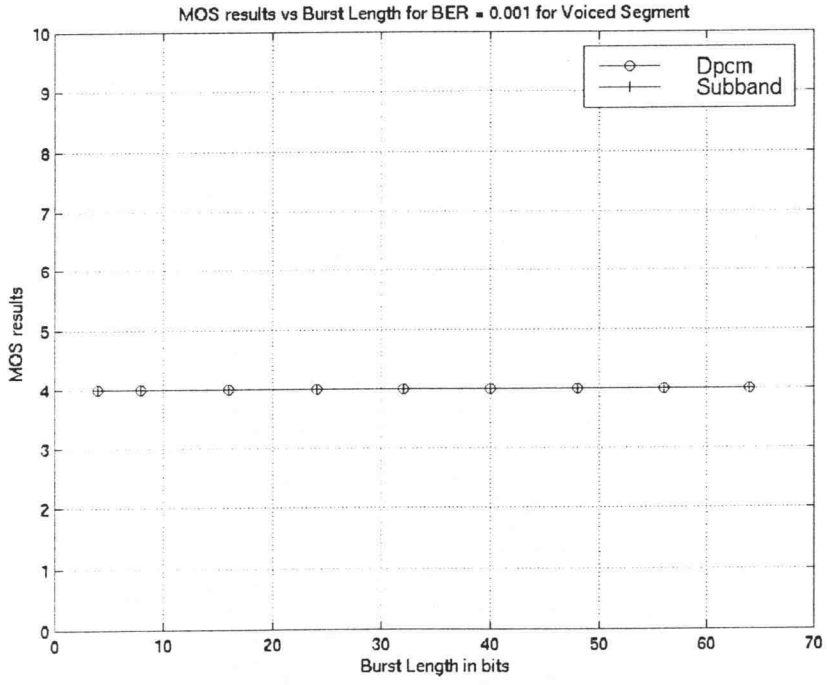


Figure 4.13 : MOS results for Voiced Speech at a BER of 0.001

4.2.3.6 Mean Opinion Score Performance For Unvoiced Speech

Results for MOS for unvoiced speech are shown in Figures 4.14 and 4.15 for BER of 0.05 and 0.001 respectively. Recall that both DPCM and Subband generally have relatively poor performance for unvoiced speech. The deterioration is even more evident in the presence of transmission errors. For a BER of 0.05, the MOS quality exhibits the common trend of improving quality with increasing burst length. For DPCM, the quality is " Poor" (MOS 2.8) for lower burst length of errors and improves to " Fair " (MOS 3.57) for higher burst lengths (Figure 4.14). SBC performance deteriorates too and is close to " Unsatisfactory " (MOS 1.5) for low burst lengths and improves to " Poor " (MOS 2.3) for higher burst lengths.

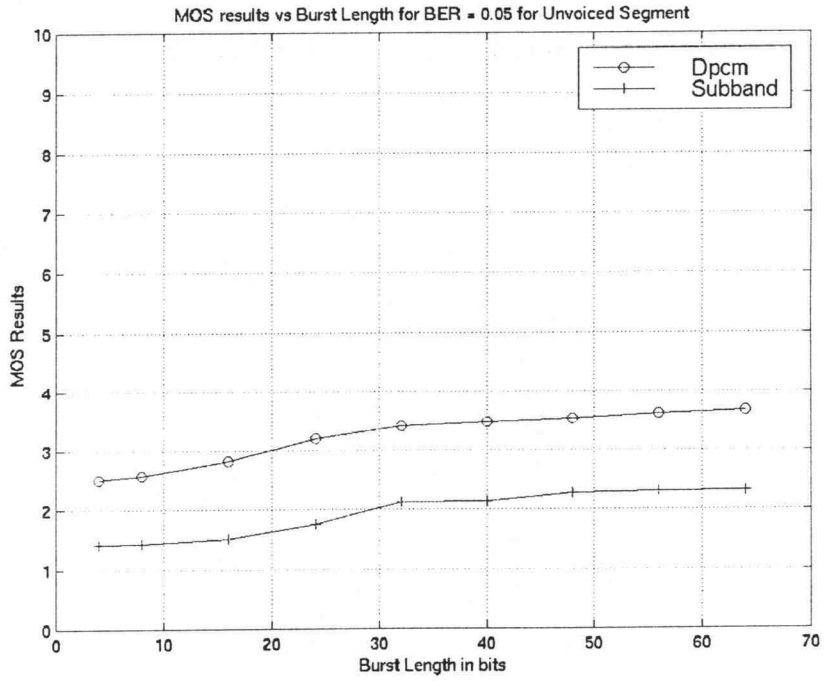


Figure 4.14: MOS Results for BER of 0.05 for Unvoiced Speech

For a BER of 0.001 (Figure 4.15), the MOS curve shows a slight initial increase in MOS rating (from 3.61 to 4 for SBC and 3.93 to 4 for DPCM) with increasing burst length, which levels off starting at a 24 Kbps bit burst length. The listeners found it to be very difficult to discern any changes with increasing burst length.

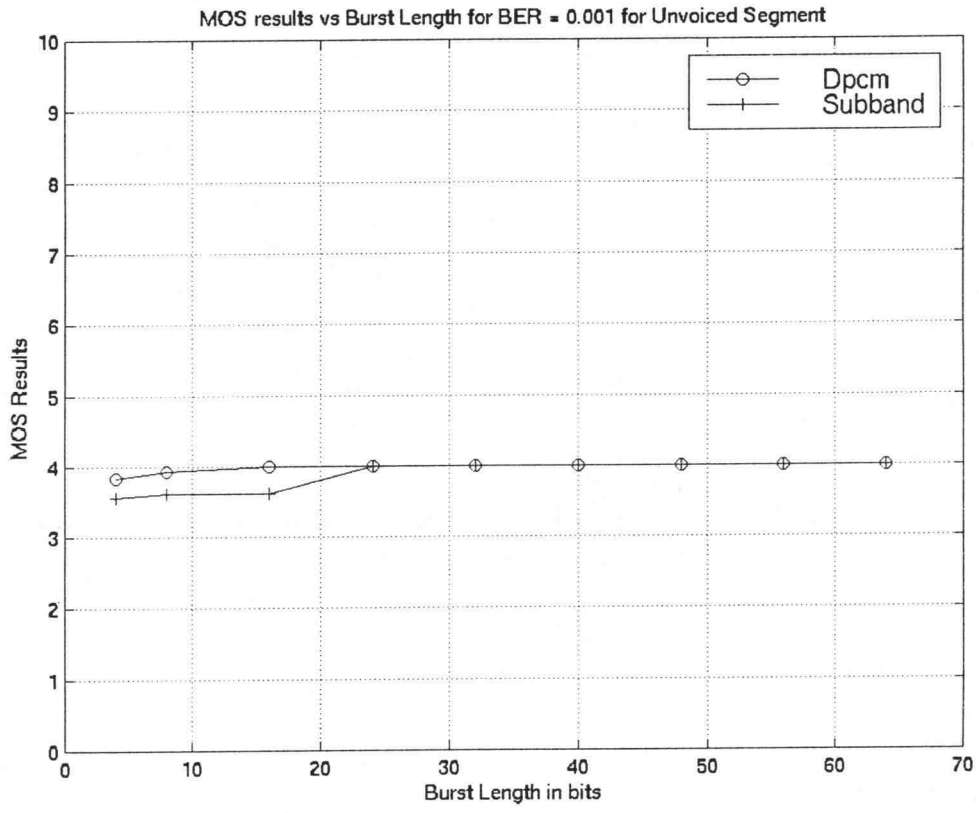


Figure 4.15: MOS results for a BER of 0.001 for Unvoiced Speech

4.2.4 Discussion of results

With BER fixed, increasing the burst length in bits improved the perceptual quality of the decoded speech. For both DPCM and SBC, the number of samples in error out of the source decoder reduced with increasing burst length. As predicted by theory, the SNR also increased with increasing bit bursts. The increasing segmental SNR and the

rising MOS scores all correlate well together. For DPCM encoded speech, there seemed to be less perceptual distortion of the speech itself, but has more background-like noise. SBC-coded speech, on the other hand, sounded more synthetic, which was rated as more perceptually annoying in the informal listening tests.

Chapter 5: Conclusion

5.1 Summary Of Results

The goal in this thesis is to do an evaluation of the relative performance of the major speech compression techniques, in the presence of uncorrectable errors caused by distortion arising during the transmission process. We started with first developing the 24 Kbps DPCM and Subband Codecs for achieving compression. These algorithms reflect the two fundamental techniques used in speech compression standards. DPCM looked into eliminating the redundancies in time whereas Subband uses fewer resources for the perceptually unimportant information by looking in the frequency domain to achieve compression. Note that this is a preliminary study into the performance of the above two codecs in the presence of burst errors.

Performance of these codecs in the burst error channel was tested for bit error rates of 0.001 and 0.05 to account for the cumulative effect of errors occurring in an overloaded channel codec. As predicted by theory, both source codecs showed an increase in the SNR values with increasing burst length for fixed BERs of 0.05 and 0.001. With increase in burst length in bits, the increasing segmental SNR and the rising MOS scores all correlate well together. For the range tested, the quality of the decompressed sound files improved with increasing burst lengths.

From the MOS test, distortion in DPCM sound files seemed to be less degrading perceptually in the informal subjective tests. In particular, for DPCM encoded speech, the distortion sounded more like random background noise. SBC-coded speech, on the other hand, sounded more synthetic, which was rated as more perceptually annoying in the

informal listening tests. Most of the high frequency information here was encoded using only 1 bit. Thus if this information was lost, the quality of sound would be altered but there would be no random background noise.

Based on the results obtained above, it could be hypothesized that for fixed BER, the codec performance would continue to get better for longer bursts, however, only up to a certain point, after which it would deteriorate rapidly. At this point, whole segments of data would be lost, and the speech would become non-comprehensible. In order to substantiate this hypothesis, more simulations are necessary.

It is important to note that this is a preliminary study into the effects of burst errors on the performance of the two codecs. While evaluating the results, consideration of the limitations of these experiments is important as described in the next section.

5.2 Limitations of Experiments

The following points list the limitations of the simulations that have been carried out in this thesis.

1. The results obtained here are based on the use of two short and specific speech samples.
2. MOS results are not very reliable, as informal listening does not bring out accurately whether one speech sample is better or worse than a slightly different one. Moreover the subjects here were not trained listeners for distortion measurement.
3. The binary encoder/decoder used in this thesis do not use Gray Coding.
4. In the quantizer implementation as done in this thesis, the data is scaled by the maximum value in the entire data vector and then quantized. If there is only one large

data value, then the rest of the data points get scaled by this one value. This is not an efficient way of implementing a quantizer.

5. Lower bit rates are more consistent with compression required for wireless. However, these generally require a combination of predictive and transform coding techniques. Here we worked to evaluate these separately.

5.3 Future Work

The degradation and distortion introduced by the burst errors is diverse in nature and this study is a first step in understanding them. Thus these results provide only estimates of the performance of the 24 kbps DPCM and SBC codecs in the presence of burst errors and there is a large scope for more work. For evaluating the speech quality, for example, it would be worthwhile to use a better perceptually motivated measure other than MOS. Also, to do a comprehensive study, more experimentation at lower bit rates, higher BER channels and different types of speech samples is required. This would enable us to evaluate the performance of the codecs better and draw definite conclusions. Gray Coding could be incorporated in the binary encoder/decoder and the effect of burst vs. single errors evaluated. Also the quantizer could be implemented more efficiently by using data points lying within a specified range which then get scaled accordingly.

BIBLIOGRAPHY

1. Rabiner, L.R./ Schafer, R.W. 1978. *Digital Processing of Speech Signals*. Prentice-Hall, Inc., Englewood Cliffs, New Jersey 07632.

2. Jayant, N.S. / Noll, Peter. 1984. *Digital Coding of Waveforms: Principles and applications to speech and video*. Prentice-Hall, Inc, Englewood Cliffs, New Jersey 07632.

3. Proakis, J. G./ Manolakis, D.G. 1996. *Digital Signal Processing: Principles, Algorithms, and Applications*. Prentice-Hall, Upper Saddle River, New Jersey 07458.

4. Steele, Raymond. 1992. *Mobile Radio Communications*, Pentech Press, Publishers-London.

5. Jayant, N.S. 1974. *Digital Coding of Speech waveforms: PCM, DPCM, and DM Quantizers,* Proceedings of the IEEE, May 1974.

6. Wolf, J.K. 1966. *Effects of channel errors on delta modulation*, IEEE transactions on Communicaiion Technology, Vol. COM-14, pp. 2-7, Feb 1966.

7. Wang, Shihua / Sekey, Andrew / Gersho, Allen. 1992. *An objective measure for predicting subjective quality of speech coders*. IEEE Journal on selected areas in communication, vol 10, number 5, pp. 819 - 829, June 1992.

8. Atal B. S./ Cuperman, V./ Gersho, A. 1991. *Advances in Speech Coding*. Kluwer Academic Publishers.

9. Akansu, A. N./ Smith, M.J.T. 1996. *Subband and Wavelet Transforms : Designs and applications*. Kluwer Academic Publishers.

10. Jayant, N. 1992. *Signal Compression: Technology Targets and Research Directions*. IEEE Journal on selected areas in Communications. vol 10, Number 5, pp 796-818, June 1992.

11. Haykin, S. 1995. *Communication Systems*. John Wiley and Sons.

12. Oppenheim, A. V./ Schafer, R. W. *Discrete Time Signal Processing*. 1997. Prentice-Hall

13. Gibson, J.D. *The Mobile Communications HandBook*. 1996. CRC Press. IEEE Press

14. Horstein, Michael. 1994. *Efficient Communication through Burst Error Channels by means of Error Detection.* IEEE Transactions on Information Theory. Vol. 40, pp 186-192, 1994.

15. Noll, P. *The performance of PCM and DPCM Speech Coders in the presence of Independent and Correlated Errors*. Conference Record, 1975 IEEE International Conference on Communications, June 1975.

16. Lippmann, R. *A technique for Channel Error Correction in Differential PCM Picture Transmission*. Conference Record, 1973 IEEE International Conference on Communications, June 1973.

17. Candy, J.C. 1974. *Limiting the propagation of Errors in One-Bit Differential Codec's*. Bell System Technical Journal, vol. 27, pp. 379-423, October 1974.

18. Khansari, M. 1998. *Performance Of Predictive Coders Over Noisy Channels with Feedback*. Hewlett Packard Laboratories, Palo Alto, California. COMM 9.5, ICASSP '98.

19. Out, H. H. / Sayood, K. 1998. *A Joint Source/Channel Coder with Block Constraints*. Department of Electrical Engineering, University of Nebraska, Lincoln. COMM 9.2, ICASSP '98.

20. Kafedziski, V. / Morrell, D. 1998. *Joint Source Channel Coding Over Channels With Intersymbol Interference Using Vector Channels and Discrete Multitone*. Telecommunication Research Center, Arizona State University, Tempe. COMM 9.1, ICASSP '98.

21. Gersho, A. / Gray, R. M. 1991. *Vector Quantization and Signal Compression*. Kluwer Academic Publishers.

22. Wicker, S. B. 1995. *Error Control Systems for Digital Commuincation and Storage*. Prentice-Hall International, Inc. Englewood Cliffs, New Jersey 07632.

23. O'Neal, J. B. / Stroh, R. W. 1972. *Differential PCM for Speech and Data Signals*. IEEE Transactions on Communications, October 1972.

24. User's Guide, 1995, *The Student Edition of Matlab*, Prentice-Hall, Englewood Cliffs, New Jersey 07632

25. Comstock, D./Gibson, J.D. 1984. *Hamming Coding of DCT Compressed Images over Noisy Channels*. IEEE Transactions on Communication, Vol. COM-32, pp. 856-861, July 1984.

26. Chang, K. Y. / Donaldson, R.W. *Analysis, optimization, and sensitivity study of differential PCM systems operating on noisy communication channels* . IEEE Transactions on Communication. Vol. COM-20, pp. 338-350, June 1972.

27. Farvardin, N. 1990. *A Study of Vector Quantization for Noisy Channels*. IEEE Transactions on Information Theory, Vol. 36, pp 799 – 809, July 1990.

28. Spanias, A. 1994. *Speech Coding: A Tutorial Review.* Proceedings IEEE, pp 1541-1582, October 1994.

29. Kitawaki, N. *Objective Quality Evaluation for Low Bit rate Speech Coding Systems.* IEEE Journal on Selected Areas in Communication, pp 242 – 248, Feb.1988.

30. Lechleider, J. W. *The optimum combination of block codes and receivers for arbitrary channels.* IEEE Transactions on Communications, Vol. 38, pp 615-621, May 1990.

31. Quackenbush, S./ Barnwell, T./ Clements, M. 1988. *Objective Measures of Speech Quality.* Prentice Hall, Englewood Cliffs, NJ.

32. Gilber, E. N. 1960. *Capacity of a Burst Noise Channel.* Bell Systems Technical Journal, Vol. 39, pp. 1253-1265, Sept 1960.

33. Elliott, E.O. 1963. *Estimates of Error Rates for Codes on Burst Noise Channels.* Bell System Technical Journal, Vol. 42, pp. 1977-1997, Sept 1963.

34. Salami, Redwan / Laflamme, Claude / Massaloux, D. 1994. *A Toll Quality 8 kn/sec Speech Coder for the Personal Communications System (PCS).* IEEE Transactions on Vehicular Technology, pp 808-816, August 1994.