A SPEECH CODER DESIGN FOR LAND MOBILE RADIO COMMUNICATIONS

Thesis submitted in accordance with the requirements of the University of Liverpool for the degree of Doctor of Philosophy

by

Wing-Tak Kenneth WONG

Department of Electrical Engineering and Electronics,
University of Liverpool

June, 1989.

ABSTRACT

The aim of this thesis is to investigate the performance of an APC-AB speech coding scheme when used over land mobile radio (LMR) channels. A comprehensive study of mobile radio channel modelling and simulation, speech coding algorithms, and the effects of channel coding schemes are presented.

Firstly, details of two LMR channel simulation methods based on a Rayleigh fading model for non-coherent frequency shift keying and differential phase shift keying modulation will be described. These simulation algorithms were developed specifically for testing the influence of fading channel errors on low-bit-rate speech coding scheme as used in narrow band single channel per carrier systems. The channel simulation produces a sequence of error pattern due to either slow fading or fast fading (which causes a random FM effect in the demodulated signal). The error sequence is 'modulo-2' added to the binary code sequence produced by a speech coder, thus allowing its effect to be evaluated both subjectively and objectively.

Adaptive predictive coding with adaptive bit allocation (APC-AB) proposed by Itakura[31] for use in LMR systems was simulated. The basic technique of APC-AB involves pitch synchronous time domain segmentation of speech with a dynamic allocation of bits which varies according to the short term energy of each segment. This has been used in conjunction with a sub-band APC-AB coding scheme to achive high speech quality at bit rates of between 12 and 16kb/s. A detailed description of the sub-band APC-AB coder will be presented in the thesis.

The computational complexity of the APC-AB coding scheme as proposed by Itakura can be reduced by the use of an adaptive Line Spectrum Pair(LSP) filter which derives LSP spectral parameter on a sample by sample basis, using a type of LMS algorithm. Theoretical derivation of this new algorithm and results from this reduced complexity APC-AB system will be described.

The simulated APC-AB coder was evaluated by the simulated LMR channel. Without error correction coding, the speech quality of a 16kb/s APC-AB coder is noticeably but slightly affected at average bit error rates of 0.1%. The quality deteriorates rapidly as the average error rate is increased from this value and becomes unintelligible at error rate of 1%. Several channel distortion measurements were performed to find out which APC-AB parameters are most sensitive to channel errors. It was then decided how error correction bits can best be allocated to individual APC-AB parameters. Adopting error correction schemes based on Reed-Solomon codes and extended Hamming codes, it was found that improved speech quality could be achieved with high channel errors by reducing APC-AB bit rates to 13.9 kb/s and using the remaining 2.1 kb/s to protect the most error sensitive parameters. Zero Redundancy error control methods that exploit the transmitting parameter characteristics will also be considered for the LSP coefficients, and a robust scheme is then derived for the transmission of LSP coefficients.

ACKNOWLEDGEMENTS

The author wishes to express his sincere appreciation and gratitude to his supervisor, Dr. B.M.G. Cheetham for his continuous guidance, encouragement and constant friendship throughout the course of this work.

Thanks are also due to Dr. C.C. Goodyear for many stimulating discussions and expert advice; to Professor W. Eccelston for providing research facilities in the Department of Electrical Engineering and Electronics; to the Overseas Research Studentship Committee, the University of Liverpool and British Telecom Research Laboratories for their financial support.

The author is especially indebted to Dr. F. Adachi, Staff Engineer in Nippon Telephone and Telegraph, for his valuable suggestion and discussion on the work of a mobile radio channel simulation system.

Finally, a debt of gratitude is owed to my beloved parents, sister and brother for their support which was often undeserved.

To my

parents,

sister and brother

with

deepest affection and gratitude.

CONTENTS

CHAPTER 1 INTRODUCTION TO DIGITAL LAND MOBILE RADIO SYSTEM

1.1 INTRODUCTION

Since the first land mobile radio (LMR) system was used by the Detroit Police Department in 1921 for police car dispatch [1], there has been an ever increasing interest in such flexible and useful radio systems. Technology in this field has been developing very rapidly in the past two decades. In recent years several systems have appeared on the market, such as TACS (Total Access Communication Systems) in UK and AMPS (Advanced Mobile Phone Services) in America. These systems have met the general needs, but the heavy demand from the increasing number of users means that larger capacity systems with more reliable performance will be required.

At present, LMR telephone systems use a combination of analogue transmission for speech and digital techniques to set up a telephone call. A successful spectrum saving concept called the "cellular" system is also used in areas of heavy traffic requirement. Analogue speech transmission as used in these systems is, however, susceptible to performance degradation associated with multipath propagation and other interference problems in the channel.

Recent advances in the field of VLSI technology and digital signal processing have opened up a wide field of efficient methods to combat these transmission problems. Some examples are low bit rate speech coding, advanced digital modulation, adaptive channel equalisation and error control coding. A higher transmission quality can be obtained through the use of these techniques. The future demands in the LMR systems will also be made on the flexibility to achieve a high degree of voice privacy by

using low cost data encryption method, and the possibility to simplify the methods of ISDN (Integrated Services Digital Network) compatibility.

Consequently, there is considerable interest in developing "all-digital" technology for the next generation of LMR systems. This is especially true in Europe where a Pan-European Public Mobile Communication System is under development and will be put into operation in 1990s [2]. In view of these developments, the aim of the research work described in this thesis is to investigate the requirements of a digital LMR system and to study the effectiveness of a low bit rate speech coding scheme, known as APC-AB, for such a system. The effect of channel errors, as typically encountered in mobile LMR systems, is to be investigated by simulation techniques and the use of error control coding methods specifically related to the use of APC-AB coding is also to be studied.

1.2 SPECTRUM EFFICIENCY FOR LMR SYSTEMS

As the radio spectrum is a non-renewable natural resource, it is necessary to accommodate as many users as possible within available channel allocations. Thus, spectrum efficiency is becoming an important parameter for LMR system design. This becomes a particularly severe problem where systems are to be used in densely populated urban areas.

A typical example of existing LMR systems is the TACS in UK, which can offer 1000 duplex radio phone channels using a frequency division multiple access (FDMA) technique within a channel bandwidth of 890 MHz to 960 MHz. Transmission from mobile to base stations uses channel frequencies between 890-915 MHz, whereas from base stations to the mobiles, frequencies between 935-960 MHz are used. Each band is then sub-divided

into 25 kHz bandwidth channels, and communication is effected using a channel-pair per voice channel.

As the total number of radio channels required for a nation wide service is far beyond the capacity of the available bandwidth, the available radio channels must therefore be re-used several times at different locations within the overall area. This is the cellular system concept [3]. The use of this concept depends on the mechanisms of radio propagation and the use of modulation methods.

Figure 1.2 illustrates a basic seven-cell cluster of frequencies as used in UK. The area to be covered is divided into several hexagonal cells and each cell is served by a radio base station. The base station covers the area up to and into that of the adjacent cell. This leads to the re-use of the same radio frequencies simultaneously in non-adjacent cells on an interference limited basis. When the mobile unit crosses the cell boundary, the received signal may become too low. A Mobile Switching Centre (MSC) checks the adjacent base stations to determine which can provide the best reception. If an idle channel-pair is found from a base station, the call is re-routed and the connection is controlled by the new base station. If no channel is available in the adjacent cells, the call is permitted to degrade until a channel is available. The process is known as "hand-off" and it has to be fast enough to be undetectable by the user.

As there is a limit to the number of channels available per cell, the optimum cell size is determined by the anticipated traffic. Spectrum efficiency can generally be improved by decreasing the cell size in congested areas. However, the cell size cannot be too small as propagation effects could cause unacceptable co-channel interference and there could

be an excessive number of 'hand-off' activities. Typically, cell sizes can be as small as 2 km in diameter in urban areas and as large as 30 km in diameter in suburban areas.

1.3 MOBILE RADIO CHANNEL CHARACTERISTICS

Mobile radio channels are time varying. During transmission, signals suffer from distortion and interference. Much research has been done to identify channel characteristics and to represent these by models controlled by parameters estimated from measurements [6,19,20]. In general, LMR channel behaviour can be characterised as follows,

- Propagation path losses,
- Co-channel and adjacent channel interference,
- Doppler effect due to vehicle motion,
- Ignition and other vehicular noise.
- Short term fading (Rayleigh),
- Long term fading (lognormal),

Propagation path losses and the channel interferences are important parameters for cellular system design. Propagation path loss is primarily due to terrain effects. Measurements in different cities show a general agreement on path loss [4], which is proportional to d^M, where d is the distance between transmitter and receiver and M is typically around 3 and 4. The rate of change of attenuation is of particular interest in estimating interference levels. Rainfall is an attenuation factor in addition to the other causes of path loss.

Co-channel and adjacent channel interference occur in most radio systems. In cellular schemes, this problem can be minimised by careful channel assignment and good control of the transmitting power. For in-

stance, adjacent channels cannot be used in the same cell and the same channel cannot be used in adjacent cells.

As the signal is normally received in a moving vehicle, it is subject to changes in frequency due to Doppler shifts. The Doppler frequency shift can be written as $f_d=v.\cos(\alpha)/\lambda$, where v is the mobile speed, λ is the signal wavelength and α is the incident angle to the mobile with respect to the mobile moving direction. Signals arriving from ahead of the vehicle have a positive Doppler shift (increase in frequency) and those from behind the vehicle have a negative shift (decrease in frequency). The Doppler shifts are proportional to transmission frequency and vehicle speed and may cause very disturbing effects on the received speech signal particularly at higher transmission frequencies and with faster moving vehicles.

Ignition interference from the vehicle itself and from surrounding vehicles can degrade the performance of an LMR installation very severely unless it has been carefully suppressed. Ignition interference is most troublesome in the VHF bands and decreases as the frequency increases.

1.3.1 MULTIPATH PROPAGATION EFFECT

Mobile telephone units are designed to function in a dynamically changing environment, but there is seldom a line-of-sight path between transmitter and receiver in urban areas because the vehicle aerial height is well below that of surrounding buildings. The received signal at the mobile is the resultant of many signals via multiple paths, because of scattering, reflection and diffraction by buildings and other obstructions. This is known as multipath propagation and it results in a

fading effect which is produced by partial reinforcement or cancellation among the signals that reach the receiver by different paths.

A graph showing the power of a typical received signal recorded around the Liverpool City centre is shown in Figure 1.1 to illustrate the fading that occurs due to multipath propagation. This graph shows very rapid fluctuations in power around a mean signal level superimposed on relatively slow variations of the mean level. In general this effect can be resolved into the fast fading component and the slow fading component, and they can be separated statistically [8,10].

The fast fading component is largely caused by reflection from various types of signal scatterers, both stationary and moving. It has been found that the envelope of the received signal is Rayleigh distributed when measured over distances of a few tens of wavelengths, where the mean signal is sensibly constant. The phase of the received signal is uniformly distributed form 0 to 2π. Sometimes, this is referred to as Rayleigh fading or short term fading.

Over longer distance, it can be observed that the local mean value of the Rayleigh fading envelope slowly varies. This effect is mainly caused by the shadowing of the radio signal by buildings and hills. The variation is typically 4 to 8 dB in the LMR environment and this gradual change in the mean level can be statistically characterised by a log-normal distribution. This slow fading effect is also referred to as log-normal fading or long term fading.

As the Rayleigh fading envelope commonly exhibits a much faster and larger variation, typically 35 dB or more below the local mean level, this is the most dominant problem to overcome in LMR system designs.

1.3.2 DELAY SPREAD AND COHERENCE BANDWIDTH

"Delay spread" and "coherence bandwidth" are the important parameters which characterise channels used for high data rate transmission. Their definitions often differ from one reference to another [5,6,7,8,9,10]. These parameters may be best understood by using impulse response analysis on a multipath fading channel.

Figure 1.3 shows the effect characterised by delay spread. If an impulse were to be sent from the base station, the received signal at the mobile would be a train of discrete impulses caused by multiple reflections from the scatterers with different time delays. If the number of scatterers is large, the received impulses become a continuous pulse with a pulse length, T_m, commonly called the delay spread. In data transmission, the data width should be much larger than T_m, which means that the data rate should be much less than $1/T_m$. Otherwise, the receiver may have to make decisions based on overlapping information symbols. This overlapping of bits causes bit errors due to intersymbol interference. Even a drastic increase of transmitting power will not reduce this kind of bit error.

The existence of different time delays in the various frequency components that make up a received signal leads to the important statistical property [8] that two transmitted sinusoids can become essentially amplitude uncorrelated if the frequency separation is large enough. The maximum frequency separation allowed before this effect becomes significant is often called the coherence bandwidth, B_c. Some researchers define B_c as the maximum frequency separation between two sinusoid signals that ensures that the envelope correlation coefficient between the two sinusoids is greater than 0.5. The significant of the coherence bandwidth

can be physically explained as follows. Under multipath propagation conditions, if the transmission signal bandwidth is less than B_c, the entire signal frequency content of the signal will undergo the approximately same attenuation and phase shift through the channel. This is known as frequency non-selective fading. Otherwise, if the signal bandwidth exceeds B_c, the signal is severely distorted by the channel as its frequency components are subject to different gains and phase shifts. This is known as frequency selective fading.

A detailed description of delay spread and coherence bandwidth is beyond the scope of this thesis. Different approaches to explain these parameters can be found in [5,6,7,8,9,10]. However, a simple approximation to the coherence bandwidth is given by [9];

$$B_c \approx \frac{1}{T_m} \tag{1.1}$$

According to measurements performed by Cox [6] in New York city and measurements of Bajwa and Parsons [7] in the city of Birmingham, the delay spread is typically about 3.5 μsec in urban areas, and delay spreads exceeding 8 μsec are very rare. This is equivalent to a coherence bandwidth of about 125 to 300 kHz. To avoid frequency selective fading, the transmission bandwidth of a modulated data should lie within the coherence bandwidth. As the modulated data bandwidth is directly proportional to the data rate [9], the maximum transmission data rate is therefore restricted by the coherence bandwidth. Generally, the coherence bandwidth is smaller in urban areas than it is in suburban areas.

1.4 RECENT DIGITAL LMR COMMUNICATION SYSTEM RESEARCH

Figure 1.4 shows the primary building blocks of a digital LMR system. As the mobile radio channel is very noisy and a high degree of spectrum efficiency is required, sophisticated design is necessary for each block. This is possible by using the latest advanced digital signal processing technology. Recent research areas relevant to the design and performance evaluation of LMR systems can be grouped into the following areas,

- Digital speech processing and channel coding,
- Modulation and multiple access techniques,
- Radio network layout.

An aim of the whole system is the optimization of the interface between each individual block. However, because of the complexity of interdependencies between these functional blocks, an optimal digital LMR system has not yet clearly been defined. The usual approaches tend to parameterize these blocks independently.

1.4.1 SPEECH PROCESSING AND CHANNEL CODING

One reason for the development of digital LMR is the advance of speech coding technology and the possibilities for implementing sophisticated speech coding algorithms in versatile signal processors. This has prompted an examination of various different types of digital coding as possible candidates for digital LMR systems. Though toll quality speech can be achieved by a simple 64 kb/s PCM coder, the spectrum efficiency criterion does not allow such a high data rate to be used in LMR systems.

Recent LMR research shows that the gross bit rate for speech transmission should not exceed 16 kb/s. Possible methods, producing high

quality speech at this rate, include adaptive predictive coding (APC), adaptive transform coding (ATC), residual excited linear predictive coding (RELP), sub-band coding (SBC), multi-pulse linear predictive coding (MPLPC), vector predictive coding (VPC) etc [11,12,13]. Most of these coding methods use knowledge of human speech production and perception characteristics. They are relatively complicated coding schemes, and thereby increase the projected cost of a hardware realisation. For the multi-user LMR system, not only should the cost and its complexity be considered, but also the codec delay, sensitivity against radio transmission errors, size and influence of background noise should also be taken into account.

A distinct advantage in digital transmission is its error correction capability compared with analogue speech transmission. In analogue speech transmission, distorted demodulated speech is not correctable and will often lose intelligibility during fast fading. The structure of channel errors generated in the digital mobile radio transmission channel are random and burst [14,15,16]. Within a fade, the bit error rate may be as bad as 1 erroneous bit in 2 bits. It is a fundamental principle that as the signal coding rate is reduced and the significance of each bit of information is thereby increased [35], the system must be more vulnerable to errors in the data. Thus, the digitised speech must be protected by error correction bits to improve the communication performance and to achieve maximum speech quality at high error rates. Reed Solomon codes, BCH codes or convolutional codes may be used for error correction coding. The application of these codes increases the channel bandwidth when the speech data rate is fixed. With the 16 kb/s gross bit rate limitation of the channel, we need to trade between the bits used

for better speech quality and the bits used for error control. The quantitative trade off between them is still not fully understood in the literature, and is therefore a subject of this project.

1.4.2 MULTIPLE ACCESS TECHNIQUES AND MODULATION METHODS

Multiple access techniques and modulation techniques play a key role in digital LMR systems. Digital LMR systems can be divided into two categories: narrow band and wide band systems. If the transmission bandwidth is near to or wider than the coherence bandwidth (normally about several hundred kHz for 900 MHz LMR systems), it is classified as a wide band system. Otherwise, it is narrow band. Further, because of the Rayleigh fading nature of the channel, constant envelope modulation schemes, such as angle modulation, are normally more advantageous than amplitude modulation schemes.

Current research on narrow band schemes is tending to be focused on single channel per carrier systems with frequency division multiple access (SCPC/FDMA), and narrow band time division multiple access (TDMA) possibly with slow frequency hopping. SCPC/FDMA is very similar to the current TACS system, but the speech is digitally encoded. To keep the interference as low as possible, bandwidth efficient modulation schemes like tamed frequency modulation (TFM) or Gaussian minimum shift keying (GMSK) have been proposed. Space diversity may be used to combat the fading problem. The disadvantage of this type of system is its complexity at the base stations, because it has to carry many channels and high stability is required for the oscillator frequencies. To achieve satisfactory reception, the total frequency error between transmit and receive

frequencies should not more than 3ppm (parts per million). However, such a system could very well coexist with an analogue system for an interim system and take its place in the long term.

Narrow band TDMA systems multiplex in time up to 10 channels, leading to transmission bit rates from 100 to 200 kb/s. The channel bandwidth is therefore around 200 kHz under the assumption of unity bit duration and bandwidth product [9]. This system requires less complicated equipment at the base station and is more tolerant to frequency variations. However, it requires good synchronisation between transmitter and receiver, and TDMA frames must include preambles and guard times to avoid overlapping of transmission, which reduce spectrum efficiency. At bit rates around 200 kb/s a small degree of ISI appears and sophisticated detection methods, such as the maximum likelihood Viterbi algorithm, may be necessary to achieve the required quality. Slow frequency hopping is often used in an attempt to improve the pure narrowband TDMA scheme. In this case, the carrier frequency can be forced to hop cyclically over a given pattern at rates much lower than the symbol rate of the modulating signal, typically every 2 to 4 msec. A distinct hop pattern is assigned to each user. The advantages of this method is the frequency diversity property introduced by the hops; when the separation between two successive hopping frequencies is larger than the coherence bandwidth, the fading at each hop will be independent of the successive one.

Wideband systems include the following :

- Wideband TDMA,
- Direct sequence CDMA (DS/CDMA) and
- Fast frequency hopping CDMA (FFH/CDMA).

In wideband TDMA systems, typically 30 to 60 channels are multiplexed in time on one TDMA frame, resulting in a transmission bit rate

from 600 to 1200 kb/s. Such systems are very flexible for multi-service and multi-bit rate operation and leads to very simple base stations. Since the systems will suffer from frequency selective fading, it has been proposed that a spread spectrum technique should be used at the receiver. The increase in bandwidth occupancy due to the adopted coding scheme is compensated by a higher degree of frequency re-use allowed by the higher robustness to interference.

In DS/CDMA systems long binary codewords are assigned to each user, a carrier is modulated by the codewords at a high transmission rate (2 to 5 Mb/s), and the resulting modulated waveform is used to convey the user's information. This method is very simple in access procedure and substantially reduces the oscillator stability requirements. However, this technique requires complicated hardware in the receivers, and power controls are essential for the individual users.

In FFH/CDMA the hopping rate of the carrier frequency is higher than that of the information digital stream. This technique requires a very precise synchronisation.

Some prospective LMR systems have been described, and the information given will be important when considering the design of the coding schemes for these systems. As different techniques are affected in different ways by propagation effects during transmission, sometimes the channel coding techniques may be optimised taking into account the modulation and access techniques to be used. The design of speech coders can also be improved with a better understanding of channel coding capability.

1.4.3 RADIO NETWORK LAYOUT

Because of the maturity of the theoretical cellular concept, and because it seems to work quite successfully with a high degree of spectrum efficiency in present LMR systems, this concept will probably be used for digital LMR systems in the future. A general measure of spectrum efficiency is specified in terms of 'subscribers served with specified grade of service per bandwidth and area' for spectrum efficiency and is usually abstracted into 'Erlang/MHz/km^2'.

Hexagonal equally-sized tesselations are widely used to cellularise the plane in theoretical models. Often propagation and interference conditions and frequency re-use distances are calculated for these models on the basis of the M-th power law ($3<M<4$). System cost is usually not taken into account in these calculations. Network planners have different opinions on how to use these theoretical results in deciding how to optimise their practical systems. They have to lay out and dimension their networks in real terrain with hills and buildings for inhomogeneous traffic with fluctuations, etc., at minimum cost. Current research is trying to bridge the gap between theoretical predictions and practical measurements in order to improve radio network design methodology.

1.5 POSSIBLE SYSTEM ALTERNATIVES

A brief summary of current research and development in digital LMR systems has been given. Research efforts are generally concerned with improving spectrum efficiency, reducing equipment cost, improving transmission quality and increasing the range of facilities offered by the

system. The evolution of fully digital systems is likely to lead to many new developments because of the power and flexibility of digital signal processing technology.

To initiate the development from analogue to digital, existing narrow band SCPC/FDMA system could be adapted to accommodate digital as well as analogue transmission, assuming that 16 kb/s digitally coded speech can be transmitted in the 25 kHz channels. The implementation of narrow-band TDMA systems would offer many advantages including greater flexibility in allocating transmission capacity, but would offer a low degree of compatibility with existing LMR systems.

DS/CDMA, FFH/CDMA and wideband TDMA systems have some problems from the point of view of maturity in system implementation and frequency sharing compatibility with existing systems.

However, it may be too soon to specify which system should be used, as new techniques are still under research and development. We should keep our mind open to these techniques and further investigation is necessary into the disciplines involved and their inter-relations. Hence, future LMR systems may be quite different from existing radio-phones systems.

This thesis will present a study of a speech coder design for LMR systems. Speech coding techniques, efficient digital signal processing algorithms and possible channel error control coding schemes will be considered. A new channel simulation algorithm was developed to allow testing with the digital speech codec for the SCPC/FDMA mode. The SCPC format is chosen because it is well developed and is currently widely used in many countries (for example the TACS system in England).

1.6 THESIS ORGANISATION

The organisation of this thesis will be as follows : -

Chapter 2: A theoretical model is derived to describe the Rayleigh fading effect and this model is used to simulate Rayleigh fading errors. Results of the computer simulation are verified and it is shown that the simulator can be used to test speech coders at any bit rate under frequency non-selective fading conditions.

Chapter 3: An advanced 'APC-AB' sub-band coding algorithm proposed by Itakura [31] is described. This coder is based on many advanced speech coding techniques such as LSP coefficient coding and adaptive bit allocation in the time domain. Each technique is described in detail and its computer simulation is also discussed. The full complexity of a typical low data rate speech coding scheme is discussed and the functions of the transmitted speech parameters are analysed.

Chapter 4: A new adaptive digital filter called a LSP (Line Spectral Pair) filter is introduced. This filter is adapted using a form of sequential least-mean-square (LMS) algorithm. The rate of convergence of this filter is evaluated by computer simulation. This technique is applied to the APC-AB sub-band coder in place of the LPC autocorrelation method [31] to test its practicability.

Chapter 5: The performance of the simulated sub-band coder is evaluated under conditions that would be encountered in an SCPC system using the simulated Rayleigh fading channel. An error protection method for the speech coder is developed, based on bit-by-bit channel distortion measurements. From the simulation results, it can be concluded that forward error correction coding and interleaving can improve the digitised speech transmission under the Rayleigh fading condition. A 'zero-redundancy'

error correction scheme for some of the transmitted speech parameters is described, demonstrating that with better understanding of the properties of these parameters, an error detection method can be developed which does not require the introduction of addition redundancy in the form of error protection bits.

Chapter 6: Conclusions and recommendations for future work are presented.

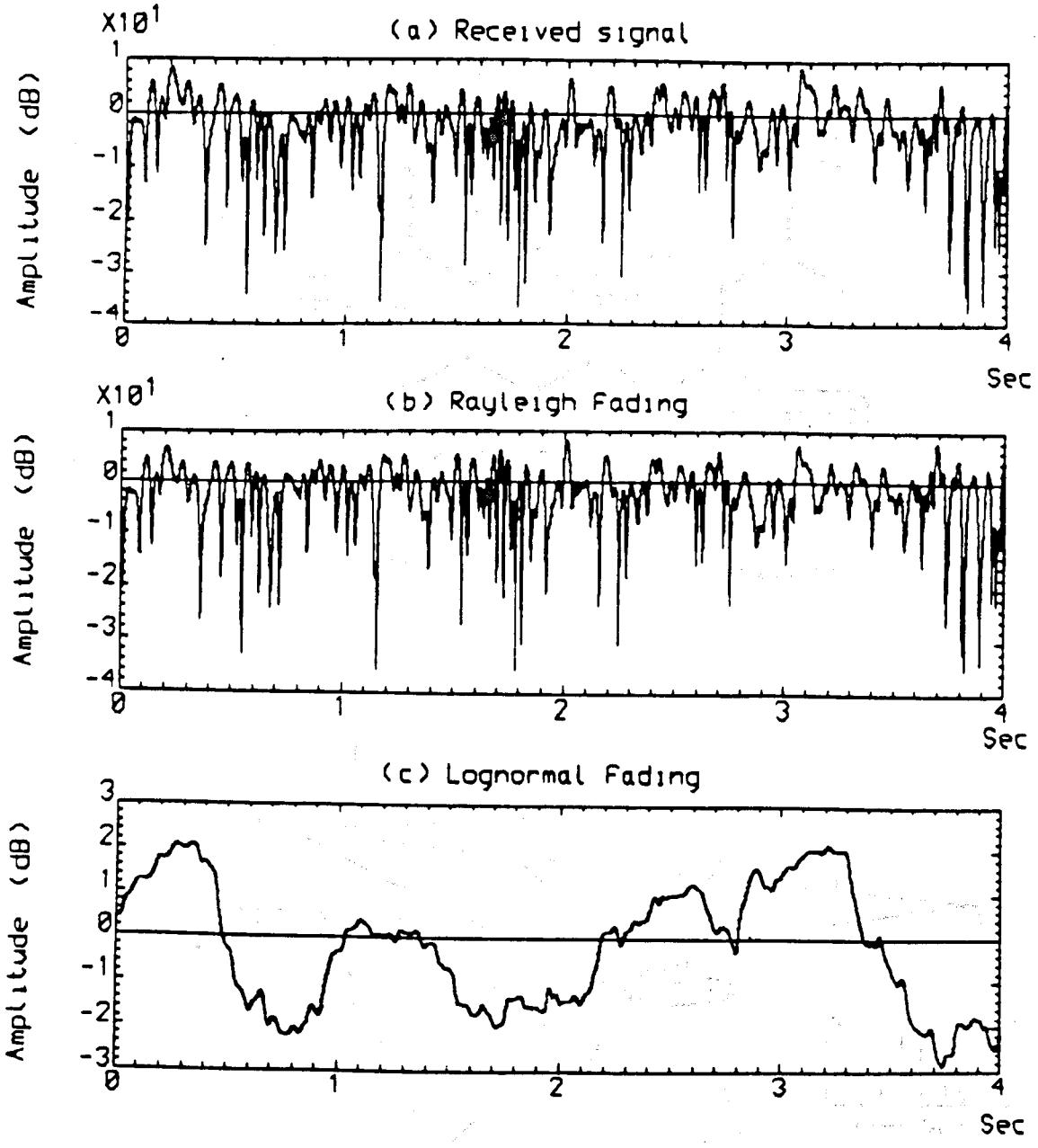


Figure 1.1 Typical received signal variations at 900 MHz measured at a mobile speed of 36 km/hr. Record taken around the Liverpool City Centre.

a) The original received signal.
b) The extracted Rayleigh fading characteristic.
c) The extracted Lognormal fading characteristic.

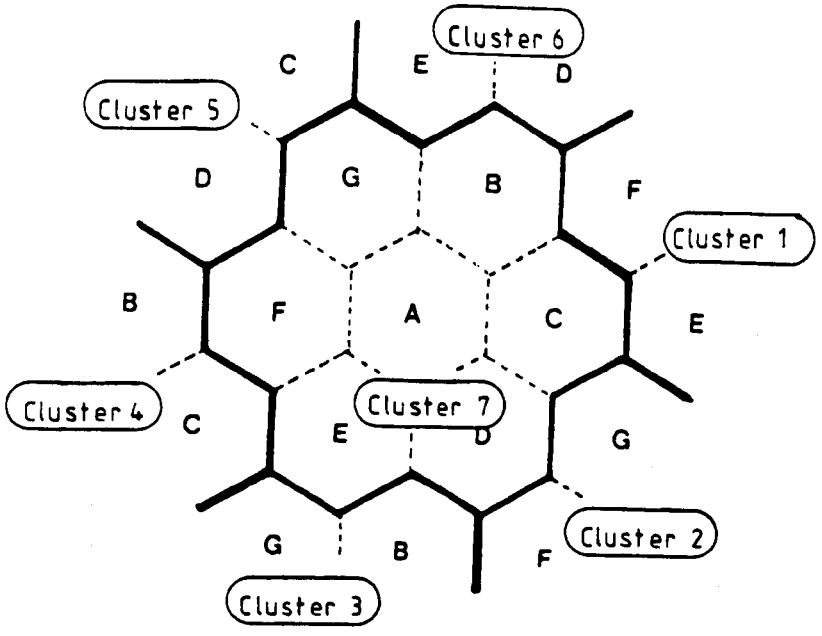


Figure 1.2 Seven-cell radio cluster.

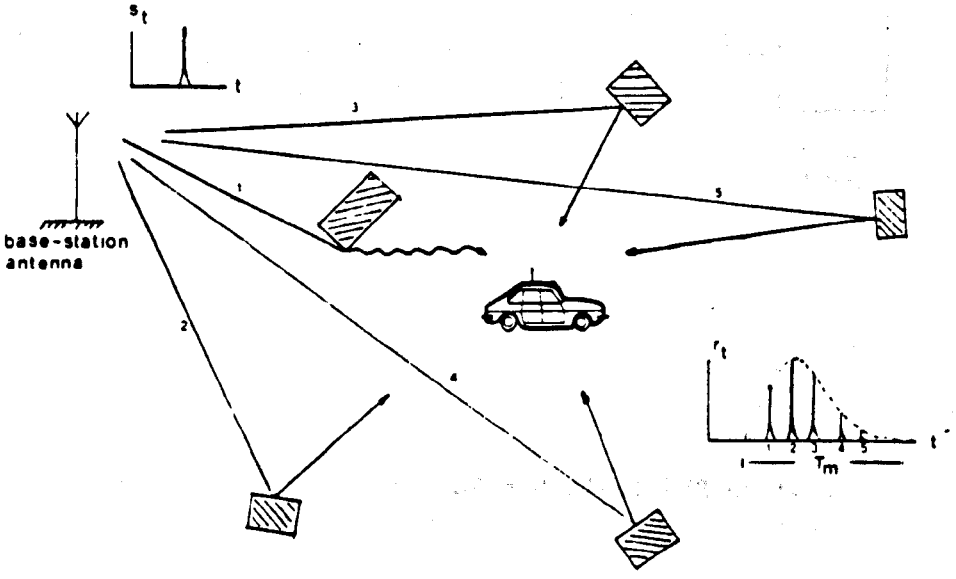


Figure 1.3 Delay spread phenomenon of an Impulse.

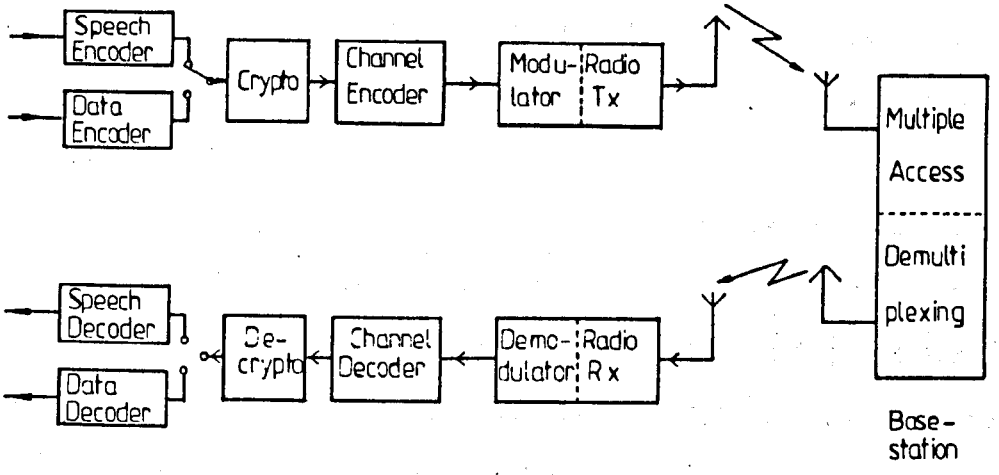


Figure 1.4 Block diagram of a digital LMR system.

CHAPTER 2 MOBILE RADIO CHANNEL SIMULATION

2.1 INTRODUCTION

When developing or assessing mobile radio communication techniques, many field tests may be needed. Such testing can be costly and time consuming and is often inconclusive, because of uncertainty in the statistics of signal variations actually encountered. An alternative approach is to use laboratory testing with signals that simulate the assumed statistical properties of the signal likely to be encountered in the field. Sometimes, it is sufficient to simulate and perhaps emphasis some particular effects for specific investigations concerned with the overall system design.

This chapter describes the design and performance of two digital mobile radio channel simulation algorithms developed by the author. The simulations were based on accepted theoretical work [8] and Arredondo's hardware Rayleigh fading envelope simulator [17,18]. The simulators are capable of generating the effect of Rayleigh fading for different vehicle speeds and carrier frequencies. Algorithms have been developed for generating bit patterns indicating the presence or absence of errors in transmitted bit streams for the cases of non-coherent frequency shift keying (NCFSK) and differential phase shift keying (DPSK). In the DPSK simulation, bit errors introduced by the random frequency modulation (random FM) effect are also faithfully simulated. Space diversity was a promising scheme for LMR systems and this is also incorporated in the channel simulation to make up a sophisticated LMR channel simulation facility.

The output from the simulation algorithm is a sequence of 'error indication' bits which is appropriate to the modulation scheme and space diversity technique under investigation with a chosen mean carrier to noise ratio. This error sequence can be modulo-2 added to a coding algorithm output to simulate the effect of the Rayleigh fading channel on the coded data.

2.2 RAYLEIGH FADING MODEL

Two well-known mathematical approaches to characterising a fading radio channel are the impulse response analysis method [19,20] and the ray path analysis method [8]. The impulse response method develops a time varying system function to characterise the channel and has the advantage of providing convenient access to frequency response measurements. This method is generally used for wide band channel analysis. However, for a narrow band channel where the transmission bandwidth is significantly less than the coherence bandwidth (see Chapter 1), the ray path model is found to be appropriate and simpler [8]. This model is based on an assumed knowledge of the multipath signal statistics, and aims to simulate the effect of statistical properties on the amplitude and phase of the fading signal. As this thesis is mainly concerned with SCPC/FDMA narrow band systems, the ray path model will be used. It is described in the following section and is analysed for Rayleigh fading conditions.

2.2.1 THEORETICAL MODEL

A simple ray path model [8,10] is shown in Figure 2.1. To analyse this model, the following assumptions are made :

a) The signal strength at the receiving antenna is determined by the effect of the superposition of many partial waves which are generated by scattering in the terrain by buildings and by reflection and diffraction,

b) There exists no significant directly transmitted component between transmitter and receiver,

c) The mobile antenna is omnidirectional in azimuth plane, and

d) The signal received at the mobile antenna can have components from all the azimuthal directions, each component being equally likely. Due to reciprocity, the same conditions will also apply to the base station.

If the mobile is travelling in the x-direction with velocity, v, and an unmodulated carrier $A.\cos(\omega_c t)$ is sent to the mobile, the vehicle motion introduces a Doppler shift of ω_i rad/sec in every partial wave. Assuming the signal seen at the mobile consists of N partial waves, then

$$r(t) = \mathrm{Re}\{ \sum_{i=1}^{N} A\, a_i \exp j[(\omega_c + \omega_i)t + \phi_i)] \} \qquad (2.1)$$

where $$\omega_i = \beta\, v \cos \alpha_i \qquad (2.2)$$

$$\beta = 2\pi / \lambda \qquad (2.3)$$

The symbol λ is the wavelength of the carrier ω_c, and ϕ_i for i=1,2,...,N are the random phase angles of the partial waves which are assumed to be random variables uniformly distributed from 0 to 2π. A is a constant proportional to the amplitude of the transmitted signal, and a_i is the i-th path attenuation factor for i=1,2,...,N, which is real. Hence, $A.a_i$ is the real amplitude of the i-th partial wave and the a_i coefficients may be considered normalised so that the ensemble average

$$< \sum_{i=1}^{N} a_i^2 > = 1 \quad . \tag{2.4}$$

Hence,

$$r(t) = \sum_{i=1}^{N} A\, a_i \cos (\omega_c t + \beta v (\cos \alpha_i) t + \phi_i) \quad . \tag{2.5}$$

Put $\psi_i(t) = \beta\, v \cos \alpha_i t + \phi_i$. (2.6)

Then

$$r(t) = A \sum_{i=1}^{N} a_i \cos (\omega_c t + \psi_i(t)) \tag{2.7}$$

$$= x(t) \cos \omega_c t - y(t) \sin \omega_c t \tag{2.8}$$

where

$$x(t) = A \sum_{i=1}^{N} a_i \cos \psi_i(t) \qquad \text{and} \tag{2.9}$$

$$y(t) = A \sum_{i=1}^{N} a_i \sin \psi_i(t) \quad . \tag{2.10}$$

It is sometimes convenient to re-express (2.8) in the following form as

$$r(t) = R(t) \cos (\omega_c t + \theta(t)) \tag{2.11}$$

where

$$R(t) = \sqrt{(x^2(t) + y^2(t))} = \text{the fading envelope} \tag{2.12a}$$

$$\theta(t) = \tan^{-1} [y(t)/x(t)] = \text{the random phase} \tag{2.12b}$$

Signals x(t) and y(t) are known as the quadrature components of the multipath signal. The random multipath process defined by (2.8) will form the basis for much of the theoretical analysis in succeeding sections of this chapter.

By the central limit theorem [21], x(t), y(t) and r(t) may be considered as independent Gaussian random variables. Assuming that they have zero mean and equal variance, the latter implies

$$\langle |r(t)|^2 \rangle = \langle x^2(t) \rangle = \langle y^2(t) \rangle = A^2/2 = \sigma^2 \qquad (2.13)$$

where $\langle \cdot \rangle$ is the ensemble average. Their probability density functions (pdfs) have the form

$$p(z) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(-\frac{z^2}{2\sigma^2}\right) \qquad (2.14)$$

where z can be x(t), y(t) or r(t). The pdf of the envelope R(t) and the pdf of the random phase $\theta(t)$ can be shown to be Rayleigh distributed and uniformly distributed respectively [22]. Their pdfs are

$$p(R) = \frac{R}{\sigma^2} \exp\left(-\frac{R^2}{2\sigma^2}\right) \qquad R \geq 0 \qquad (2.15a)$$

$$p(\theta) = \frac{1}{2\pi} \qquad 0 \leq \theta < 2\pi \qquad (2.15b)$$

The cumulative distribution function of the envelope below any specified value r_e is given by

$$P(R \leq r_e) = \int_0^{r_e} p(R)\, dR = 1 - \exp\left(-\frac{r_e^2}{2\sigma^2}\right) \qquad (2.16)$$

In practice, the mean power σ^2 varies typically by between 4 and 8 dB due to shadowing of the radio signal by buildings and hills levels to a gradual change in log-normal distribution. If there is a wave of significant magnitude arriving directly from the transmitter, the statistics of the

signal will be changed and tend to become Rician distributed. For a Rician distribution, the phase has zero mean and the envelope has a mean value equal to the amplitude of the direct wave. The shape of the distribution is close to that of a Gaussian distribution but slightly different. As Rayleigh fading is the major adverse factor affecting the LMR transmission quality, this Chapter will particularly concentrate on the analysis of this fading effect.

The above analysis of the Rayleigh fading effect is very similar to the analysis presented in Reference [9,22] of narrow band additive white Gaussian noise (AWGN). The output of a narrow band filter excited by Gaussian noise may be represented in the form of r(t) as defined by Equation (2.8) and such signals are encountered in many applications of electronics. For instance the thermal noise output at the i.f. (intermediate frequency) stage of a radio receiver is generally considered as a narrowband Gaussian noise signal [9,22].

2.2.2 RF POWER SPECTRUM

By Equation (2.7), the frequency of the received multipath carrier signal is bounded by the values $\pm\beta v$ ($=\pm 2\pi f_d$), where f_d is the maximum Doppler shift which is very much less than the carrier frequency, f_c. The signal may thus be described as a narrow band random process, with its amplitude and phase randomly varying with an effective bandwidth $2f_d$.

Referring to the assumption (d) in Section 2.2.1, if a sinusoid with frequency f_c is transmitted and arrives from the direction α as shown in Figure 2.1, the fraction of the total incoming power within α to

$\alpha+d\alpha$, where $d\alpha$ is small, is $p(\alpha)d\alpha$. $p(\alpha)$ is the probability distribution of power with arrival angle α, which is assumed to be uniform, i.e.,

$$p(\alpha) = \frac{1}{2\pi} \qquad -\pi \leq \alpha \leq \pi \tag{2.17}$$

If $G(\alpha)$ is the receiving antenna gain, the power contributed to the received signal by plane waves arriving within the angle $d\alpha$ is

$$P_r(\alpha) = \sigma^2 \, G(\alpha) \, p(\alpha) \, d\alpha \tag{2.18}$$

where σ^2 is the mean power that would be received by the receiving antenna. Because of the Doppler shift, the relationship between frequency and angle is

$$f(\alpha) = f_d \cos \alpha + f_c \tag{2.19a}$$

where f_d is equal to v/λ. Whence,

$$d\alpha = - \frac{df}{\sqrt{f_d^2 - (f-f_c)^2}} \tag{2.19b}$$

The power spectral density is

$$S(f) = P_r(\alpha) \left| \frac{d\alpha}{df} \right| \tag{2.20}$$

Since $+\alpha$ and $-\alpha$ give the same Doppler shift, $f(\alpha) = f(-\alpha)$, then

$$S(f) = (P_r(\alpha) + P_r(-\alpha)) \left| \frac{d\alpha}{df} \right| \tag{2.21a}$$

$$= \frac{\sigma^2}{\sqrt{f_d^2 - (f-f_c)^2}} \, (P_r(\alpha) + P_r(-\alpha)) \tag{2.21b}$$

If the antenna is a vertical monopole and therefore

$$G(\alpha) = G(-\alpha) = 3/2$$

then, the power spectrum of the received signal is

$$S(f) = \frac{3\sigma^2}{2\pi f_d}\left[1 - \frac{(f-f_c)^2}{f_d^2}\right]^{-0.5} \qquad |f-f_c| \le f_d \qquad (2.22)$$

$$= 0 \qquad \text{otherwise}$$

This spectrum is plotted in Figure 2.2.

It can now be concluded that for a given transmitted frequency, f_c, the received multipath signal consists of the sum of a large number of sinusoids whose frequencies are bounded within $\pm f_d$ and whose amplitudes and phases are randomly varying within this bandwidth.

2.2.3 LEVEL CROSSING RATE AND DURATION OF FADES

Under Rayleigh fading conditions, the signal envelope experiences very deep fades occasionally. The shallower the fade the more frequently it is likely to occur. Sometimes, the fluctuating received signal may fall 30 to 40 dB below the mean signal level. It would be interesting to know how often the received signal level will fall below a specific level and what is the likely duration of these fades.

The level crossing rate, N_R, is the expected rate at which the envelope crosses a specified level R in the positive direction. If the signal is received from an omnidirectional antenna, N_R is given by [8] :

$$N_R = \sqrt{2\pi}\, f_d\, \rho \exp(-\rho^2) \qquad (2.23)$$

where

$$\rho = \frac{R}{\sqrt{<|r|^2>}} = \frac{R}{R_{rms}} \qquad (2.24)$$

and N_R has a unit of $\sec^{-1}$ (Hz), i.e., the number of crossings per second.

The mean duration of fades, τ, below the level R is defined as

$$\tau = \frac{P(r \leq R)}{N_R} \tag{2.25}$$

Substituting (2.16) into (2.25), the mean fade duration can be written as

$$\tau = \frac{[\exp(\rho^2)-1]}{\sqrt{2\pi} f_d \rho} \tag{2.26}$$

Equation (2.23) and (2.26) shows that N_R and τ primarily depend on the Doppler shift. In data transmission, they can be used to determine the probable number of signaling bits that will be lost during fades.

2.3 RAYLEIGH FADING SIMULATION

In previous sections, it has been shown that the Rayleigh fading carrier is represented by (2.8). The quadrature components are Gaussian and their frequency content is spread within $\pm f_d$. Equation (2.8) suggests a technique for generating a multipath fading signal. The fading spectrum S(f) as given in (2.22) can be simulated by shaping the spectrum of the quadrature components. The Arredondo hardware simulation [17] illustrated in Figure 2.3 is based on this technique.

2.3.1 SPECTRAL SHAPING FILTER DESIGN

In simulation, the quadrature components x(t) and y(t) are obtained by lowpass filtering two independent Gaussian sources. The filter's frequency response should have the same shape as the Doppler fading

spectrum. The lowpass filter's frequency response can be deduced from (2.22), i.e.,

$$S_{lp}(f) = \frac{3\sigma^2}{2\pi f_d}\left[1 - \left(\frac{f}{f_d}\right)^2\right]^{-0.5} \qquad f \le f_d \qquad (2.27)$$

$$= 0 \qquad \text{otherwise .}$$

With reference to the original work by Arredondo [17], the shaping filter is simulated by a low-pass filter with a 6 dB peak at the Doppler frequency and 18 dB/octave roll-off at frequencies above the Doppler frequency as depicted in Figure 2.4. This was found to be an adequate approximation to $S_{lp}(f)$. The required frequency response can be obtained, approximately, using a serial cascade of a 2-pole filter with the poles arranged to produce a 9 dB peak and a 1-pole filter with 3 dB loss at the Doppler frequency. The transfer function of this cascade is :

$$H(s) = \frac{\omega_p^2}{s^2 + 2\xi\omega_p s + \omega_p^2} \times \frac{\omega_p}{s + \omega_p} \qquad (2.28)$$

$$H(s) = \frac{\omega_p^3}{s^3 + s^2\omega_p(1+2\xi) + s\omega_p^2(1+2\xi) + \omega_p^3} \qquad (2.29)$$

where $\omega_p = 2\pi f_p$ is the simulated Doppler fading frequency, and ξ is the damping factor for the 2nd order filter.

Let M_p at f_p be the peak magnitude of the simulated lowpass filter spectrum. It is given as follows [23] :

$$M_p(\text{in dB.}) = -20 \log_{10}(2\xi) - 3 \qquad (2.30)$$

When M_p is 6 dB, ξ is equal to 0.177406694. The simulated Doppler fading frequency, f_p, is calculated by equating the moments of the theoretical and simulated spectrum. Let b_n denote the n-th moment of the theoretical low pass spectrum $S_{lp}(f)$. Then

$$b_n = (2\pi)^n \int_0^{\infty} S_{lp}(f)\, f^n\, df \qquad (2.31)$$

The theoretical maximum Doppler shift has been derived in terms of moments as [4],

$$f_d = (1/2\pi)\ \sqrt{[2b_2/b_0]} \quad . \qquad (2.32)$$

From (2.32), the simulated maximum Doppler shift can be written as

$$f_p = (1/2\pi)\ \sqrt{[2b_2'/b_0']} \quad . \qquad (2.33)$$

b_2' and b_0' are computed using the simulated fading spectrum $H(j\omega)$. The ratio of f_d to f_p was calculated to be about 1.2 [18], hence the simulated maximum Doppler frequency is written as,

$$f_p = f_d/1.2 \quad . \qquad (2.34)$$

2.3.2 DIGITAL LOW PASS FILTER IMPLEMENTATION

A third order IIR filter is implemented to simulate the spectral shaping effect. The filter coefficients are changed according to the vehicle speed and the carrier frequency, and these changes affect the filter bandwidth and hence the simulated Doppler spread. The z-transform system function of the digital filter's impulse response is derived from the Laplace transform of the impulse response a 3rd order analogue filter using the bi-linear transformation. Rewrite equation (2.29) as

$$H(s) = \frac{A}{s^3 + a_2^2 + a_1 s + a_0} \qquad (2.35)$$

where $a_0 = A = \omega_p^3$

$a_1 = (1+2\xi)\ \omega_p^2$

$a_2 = (1+2\xi)\ \omega_p$.

Using the bilinear transformation, substitute

$$s = \frac{2}{T}\frac{(z-1)}{(z+1)} \tag{2.36}$$

into H(s), where T is the sampling period of the filter. Then

$$H(z) = \frac{n_0 z^3 + n_1 z^2 + n_2 z + n_3}{d_0 z^3 + d_1 z^2 + d_2 z + d_3} \tag{2.37}$$

where $n_0 = n_3 = A\,T^3$

$n_1 = n_2 = 3\,A\,T^3$

$d_0 = 8 + 4a_2T + 2a_1T^2 + a_0T^3$

$d_1 = -24 - 4a_2T + 2a_1T^2 + 3a_0T^3$

$d_2 = 24 - 4a_2T - 2a_1T^2 + 3a_0T^3$

$d_3 = -8 + 4a_2T - 2a_1T^2 + a_0T^3$

As $H(z) = \dfrac{V_o(z)}{V_i(z)}$, then

$$v_0(n) = (1/d_0) \sum_{i=0}^{3} n_i\, v_i(n-i) - (1/d_0) \sum_{i=1}^{3} d_i\, v_o(n-i) \tag{2.38}$$

A direct implementation of Equation (2.38) is shown in a signal flow graph in Figure 2.5. The simulated and the theoretical spectra are shown in Figure 2.6 for comparison. Note that for a given value of T, the filter response is determined by ω_p, which is the lowpass filter cut-off frequency. When the low pass filter bandwidth is wider, the fading rate is correspondingly higher as may be deduced from Equation (2.23).

2.3.3 RAYLEIGH FADING ENVELOPE SIMULATION

The validity of the simulator can be established by generating the Rayleigh fading envelope and observing its statistics. A flow chart for the required test program is given in Figure 2.7. This test program generates a sequence of Gaussian distributed random samples which are spec-

trally shaped by two separate spectral shaping filters 'filta1' and 'filta2', which simulate the fading quadrature components x(t) and y(t). The samples are equally spaced in time T seconds apart. From (2.12a), the fading envelope is given by

$$\text{Envlop}(t) = \sqrt{[\, x^2(t) + y^2(t) \,]} \qquad (2.39)$$

Two segments of the simulated fading envelope are shown in Figure 2.8. The amplitude values are normalised to have a zero decibel mean, i.e.,

$$< \text{Envlop}^2(t) > = 1 \qquad (2.40)$$

The two envelopes shown in Figure 2.8 can be considered as representing a 900 MHz received fading signal when a mobile is moving at 30 and 50 km/h respectively. It shows that the simulated signal envelope falls about 25 to 30 dB below the mean signal level, and the fading rate is increasing at higher mobile speed. This simulated envelope can be compared with the real Rayleigh fading envelope as shown in Figure 1.1b. In the time domain comparison, they look very similar. In the following section, statistical evaluation methods are described for the simulated envelope.

2.3.4 SIMULATION PERFORMANCE

The fading simulation performance was evaluated by comparing the theoretical cumulative probability distribution, level crossing rate and fade duration distribution with the corresponding values obtained from the program. The cumulative distribution of the simulated signal envelope is plotted in Figure 2.9 and compared with the theoretical Rayleigh function. The coordinates in the graph are scaled so that the Rayleigh Cumulative distribution appears as a straight line.

The level crossing rate (LCR), N_R, is determined by counting the average number of times, n_s, within a time period P, that the envelope crosses a particular level R with a positive slope. N_R is given by:

$$N_R = \lim_{P\to\infty} \left. \frac{n_s(P)}{P} \right|_R \tag{2.41}$$

Then, (2.41) may be approximated by taking a sufficiently large time period P and dividing $n_s(P)$ by P. The normalised LCR, which is N_R divided by the simulated maximum Doppler frequency, f_p, is plotted in Figure 2.10 and compared with the theoretical LCR as given in (2.23).

The mean fade duration as discussed in Section 2.2.3 is determined by:

$$\tau_R = \lim_{P\to\infty} \left. \frac{n_c(P)}{N_R} \right|_R \tag{2.42}$$

where $n_c(P)$ is the average number of envelope samples below R. Again (2.42) may be approximated by taking a sufficiently large value of P and dividing $n_c(P)$ by N_R. A normalised value of τ_R, which is τ_R multiplied by the f_p, is plotted in Figure 2.11 and compared with the theoretical value as given by (2.26).

Agreement between the envelope statistics of the simulated fading waveforms and the theoretical fading waveforms are very good as may be seen in Figure 2.9 to 2.11. Thus, the simulation algorithm appears reliable, and it will be used for assessing the performance of a speech coder later in this thesis.

2.4 APPLICATION OF THE ENVELOPE SIMULATION

In this research programme, we will look at the effect of fading channel error in digital speech transmission. This will be done by using the fading channel simulator as outlined above. In a hardware simulator as illustrated in Figure 2.3, the fader is linear and transparent to modulation method so that digital code from a source is modulated and RF is fed through the fader. The RF is attenuated and fed directly into the radio's front end. A sequence of erroneous codes may then be obtained after demodulation.

In computer simulation, the direct application of the RF signal is impracticable. Instead, different fading channel algorithms are programmed for different types of modulation schemes to generate fading channel error sequences in accordance with their modulation and demodulation characteristics. For the hardware simulation process, a well designed modulation system must be built and used with the fader to generate the channel error sequences. This hardware will be costly. However, the software method may be cheaper, more adaptable and easier to modify if necessary. In this section, non-coherent frequency shift keying (NCFSK) and differential phase shift keying (DPSK) modulation simulations will be described, and these will be combined with the channel fading effect described in the previous section. Fading error sequences can be generated and the performance of the simulation will be compared with the theoretical predictions.

2.4.1 NCFSK SYSTEM

2.4.1.1 NCFSK Analysis

A simple NCFSK detection scheme is shown in Figure 2.12. If the modulated signal is transmitted through a multipath fading channel, the received signal, $s_r(t)$, is represented by (2.8) and (2.11), i.e.,

$$s_r(t) = x(t)\cos\omega_c t - y(t)\sin\omega_c t \quad (2.43a)$$

$$= R(t)\cos(\omega_c t + \phi(t)) \quad (2.43b)$$

where $\omega_c = \omega_o + \Delta\omega = \omega_1$ when a mark is sent (2.44)

$= \omega_o - \Delta\omega = \omega_0$ when a space is sent.

The narrow band additive white Gaussian noise (AWGN) [9] can be expressed as,

$$n(t) = x_n(t)\cos(\omega_c t) - y_n(t)\sin(\omega_c t) \quad (2.45)$$

$x_n(t)$ and $y_n(t)$ are independent Gaussian random variables. If a mark is sent, the output of the bandpass filter shown in Figure 2.12 would be

$$g_1(t) = R(t)\cos(\omega_1 t+\phi(t)) + n_1(t) \quad (2.46a)$$

$$= R(t)\cos(\omega_1 t+\phi(t)) + x_{n1}(t)\cos(\omega_1 t+\phi(t)) - y_{n1}(t)\sin(\omega_1 t+\phi(t))$$

and

$$g_0(t) = n_0(t) \quad (2.46b)$$

$$= x_{n0}(t)\cos(\omega_0 t) - y_{n0}(t)\sin(\omega_0 t)$$

In the above equations, $n_1(t)$ and $n_0(t)$ are the AWGN in the upper and lower arms of the detector respectively, and $x_{n1}(t)$, $y_{n1}(t)$, $x_{n0}(t)$ and

$y_{n0}(t)$ are independent zero mean Gaussian random variables. The outputs of the envelope detectors are then written as

$$z_1(t) = \sqrt{[\ (R(t) + x_{n1}(t))^2 + y_{n1}^2(t)\ ]} \tag{2.47a}$$

$$z_0(t) = \sqrt{[\ x_{n0}^2(t)) + y_{n0}^2(t)\ ]} \tag{2.47b}$$

If it is assumed that a mark is sent, an error will occur if

$$z_1(t) - z_0(t) \leq 0 \tag{2.48}$$

which means that

$$(R(t) + x_{n1}(t))^2 + y_{n1}^2(t) \leq x_{n0}^2(t)) + y_{n0}^2(t) \tag{2.49}$$

Similarly, if a space is sent

$$z_0(t) = \sqrt{[\ (R(t) + x_{n0}(t))^2 + y_{n0}^2(t)\ ]} \tag{2.50a}$$

$$z_1(t) = \sqrt{[\ x_{n1}^2(t)) + y_{n1}^2(t)\ ]} \tag{2.50b}$$

An error will be made if

$$z_1(t) - z_0(t) \geq 0 \tag{2.51}$$

i.e.,

$$(R(t) + x_{n0}(t))^2 + y_{n0}^2(t) \leq x_{n1}^2(t)) + y_{n1}^2(t) \tag{2.52}$$

Equations (2.49) and (2.52) are the error decision conditions. They are virtually the same and these equations are used in the NCFSK simulation algorithm.

Without fading, if the received carrier signal is $u \cos(\omega_c t)$ and the envelope is nearly constant, the bit error rate of the NCFSK system in the AWGN channel is given by

$$P_e = \tfrac{1}{2} e^{-\gamma/2} \tag{2.53}$$

where $\gamma = u^2/2N$ = the received signal to noise ratio (2.54)
u = the carrier peak amplitude
N = the mean noise power.

When the signal is subject to fading, the instantaneous CNR, γ, will vary with time because of the changes in carrier amplitude. It is interesting to know how γ varies under Rayleigh fading conditions. Now, define the long term average of γ to be Γ, that is

$$\Gamma = \text{mean carrier to noise ratio (CNR)}$$

$$= \frac{\text{mean carrier power}}{\text{mean noise power}} = \frac{\langle R^2(t)\rangle}{2N} = \frac{\sigma^2}{N} \qquad (2.55)$$

$$= \langle\gamma(t)\rangle$$

where $2\sigma^2$ is the mean square value of $R(t)$. As γ is a function of R and R is Rayleigh distributed, the pdf of γ, $p(\gamma)$, is given by [24]

$$p(\gamma) = p(R)\left|\frac{dR}{d\gamma}\right| . \qquad (2.56)$$

Using the relationships given by (2.54), we have

$$\frac{d\gamma}{dR} = \frac{R}{N}, \quad \text{and} \quad \frac{R^2}{2\sigma^2} = \frac{\gamma}{\Gamma} . \qquad (2.57)$$

Then

$$p(\gamma) = (1/\Gamma)\exp(-\gamma/\Gamma) \qquad \gamma \geq 0 . \qquad (2.58)$$

Hence, in a Rayleigh fading situation, γ is exponentially distributed. The probability that γ is less than some specified values γ_s is

$$p(\gamma \leq \gamma_s) = (1/\Gamma)\int_0^{\gamma_s} \exp(-\gamma/\Gamma) \qquad (2.59)$$

$$= 1 - \exp(-\gamma_s/\Gamma) . \qquad (2.60)$$

To obtain the overall probability of error over the fading channel, P_e must be averaged over the pdf of γ. Hence, the average error rate is given as follows :

$$\langle P_e \rangle = \int_0^\infty P_e \, p(\gamma) \, d\gamma \tag{2.61}$$

For a NCFSK system,

$$\langle P_e \rangle = \int_0^\infty \tfrac{1}{2} \exp(-\gamma/2).(1/\Gamma) \exp(-\gamma/\Gamma) \, d\gamma \tag{2.62}$$

$$= \frac{1}{2 + \Gamma} \tag{2.63}$$

2.4.1.2 NCFSK Fading Simulation

The NCFSK fading simulation is based on the conditions given by Equations (2.49) and (2.52). In simulation, it is assumed that a sequence of data from the source is modulated and sent through the fading channel. For each bit of the received sequence the amplitude of the simulated received signal is computed and compared with the noise level at the detector (refer to Equation 2.52) to test if the bit will be received in error. The sampling frequency of the fading envelope is equal to the transmission data rate.

To simulate the error conditions, six independent Gaussian random number sources are required. Four of them for the AWGN of the NCFSK detector, and two of them are used to generate the Rayleigh fading envelope, R, which is written as

$$R = F \sqrt{[\, x_r^2 + y_r^2 \,]} \tag{2.64}$$

where x_r and y_r are the low pass filtered Gaussian random numbers and F is introduced to control the effect of different mean CNR. For clarity, the error testing algorithm is described step-wise as follows, where 'SIGE' and 'NOISE' variables are accumulators to compute the carrier

signal energy and noise energy respectively. GAUs' are the different Gaussian sources.

Step 0 : Initialization,

F = a constant,
SIGE = 0, NOISE = 0, IERR = 0, NPTS = 0,

Step 1 : Compute Rayleigh fading envelope,

$$R = (x_r^2 + y_r^2)^{\frac{1}{2}}$$
$$SIGE = SIGE + R^2$$

Step 2 : Error testing for a transmitting data,

IF $((GAU_3+R)^2 + GAU_4^2) \leq (GAU_5^2+GAU_6^2)$ then
(Invert data symbol) and
IERR = IERR + 1

Step 3 : NOISE = NOISE + 1,
NPTS = NPTS + 1,
IF NPTS = 1000 then EXIT (for 1000 data),
GOTO Step 1.

On EXIT, the r.m.s. value of the carrier signal and the noise power can be computed as :

$$\langle\sigma_s^2\rangle = SIGE/NPTS \; ; \quad \text{and} \quad \langle\sigma_n^2\rangle = NOISE/NPTS \; . \tag{2.65}$$

Then, the simulated mean carrier to noise ratio, CNR, is

$$\Gamma = \langle\sigma_s^2\rangle \; / \; \langle\sigma_n^2\rangle \; . \tag{2.66}$$

The simulated mean error rate, P_e', is

$$\langle P_e'\rangle = IERR/NPTS \; . \tag{2.67}$$

A flow chart of this error testing algorithm is given in Figure 2.13.

If a binary symbol is found to be in error, it is converted from 1 to 0 or vice versa. In fact, if a fading error sequence is also represented in binary form, '1' for an error and '0' for non-error, the

fading effect on the speech code sequence can be simulated by adding the error sequence to the speech code sequence using modulo-2 arithmetic, thus allowing its effect to be evaluated.

The simulation performance was evaluated by plotting the simulated bit error rate performance at different mean CNR as shown in Figure 2.14, and its theoretical bit error rate curve as given in (2.63) has also been plotted in the same Figure for comparison. The theoretical curve and the simulation curve are very close together, and this is evidence of the validity of the simulation. Figure 2.15 also illustrates the LMR channel error patterns at several mean CNR values, which shows the burst error structure generated from the simulation algorithm.

2.4.2 DPSK SYSTEM

2.4.2.1 DPSK Analysis

A functional block diagram of a DPSK system is shown in Figure 2.16. The baseband signal, b_k, is recovered by detecting phase changes for each sampling interval, and its phase reference is derived by comparing the current phase with that at the previous sampling interval. In a Rayleigh fading environment, the received DPSK signal not only suffers from rapid envelope variations, but also the time derivative of the phase varies with time in a random manner, depending on the mobile speed and the carrier frequency. In other words the signal is randomly frequency modulated, and this is known as the random FM effect. In this case, the random FM effect can vary the phase reference by as much as one bit time interval, which results in errors. As these errors cannot be decreased

even when the transmitting power is increased, this is often referred to as irreducible error [8].

As DPSK systems suffer different fading effects as compared with NCFSK systems, and because phase modulation schemes may be used in LMR systems, it is interesting to see how DPSK systems would behave in mobile radio environments.

Consider the differential encoding process. In the generation of a differential encoding bit, d_k, the present data, b_k, and the previous differential encoded data, d_{k-1}, are compared. Assuming bipolar signalling, this means that b_k is either 1 or -1. If there is no difference between b_k and d_{k-1}, then $d_k = 1$, otherwise, $d_k = -1$. This is expressed as :

$$d_k = b_k \times d_{k-1} \tag{2.68}$$

An example of an encoding sequence is shown in Figure 2.17. An arbitrary reference bit is taken as the initial bit of the d_k sequence. The DPSK demodulator is the combination of a mixer and a low pass filter as shown in Figure 2.16. The demodulator performs the inverse function to that of the encoder by comparing the phase angles of the received signal (which may be corrupted by noise) and its one bit delayed version. An example of the decoding process is also given in Figure 2.17. A detailed description of the DPSK process on the fading carrier is now presented.

Assume that the transmitted signal of the DPSK system is

$$S_{DPSK}(t) = A \cos (\omega_c t + \phi_s(t)) \tag{2.69}$$

where A = a constant signal amplitude,
$\phi_s(t)$ = a information phase of the DPSK signal,
$\omega_c(t)$ = the carrier frequency.

The signal at the receiver input is

$$r(t) = S_r(t) + n_r(t) \quad . \tag{2.70}$$

$S_r(t)$ and $n_r(t)$ are the received fading signal and the receiver noise respectively, and they can be written as :

$$S_r(t) = x(t)\cos[\omega_c(t)+\phi_s(t)] - y(t)\sin[\omega_c(t)+\phi_s(t)] \tag{2.71}$$

$$n_r(t) = x_n(t)\cos[\omega_c(t)+\phi_s(t)] - y_n(t)\sin[\omega_c(t)+\phi_s(t)] \tag{2.72}$$

where $x(t)$, $y(t)$, $x_n(t)$ and $y_n(t)$ are independent zero mean Gaussian variables, but $x(t)$ and $y(t)$ are the quadrature components of the fading signal. Their variances are given as follows :

$$\langle x^2(t)\rangle = \langle y^2(t)\rangle = \sigma_s^2 \tag{2.73}$$

$$\langle x_n^2(t)\rangle = \langle y_n^2(t)\rangle = \sigma_n^2 \quad .$$

Thus

$$r(t) = [x(t)+x_n(t)]\cos(\omega_c(t)+\phi_s(t)) - [y(t)+y_n(t)]\sin(\omega_c(t)+\phi_s(t))$$

$$= x_t\cos(\omega_c(t)+\phi_s(t)) - y_t \sin(\omega_c(t)+\phi_s(t)) \tag{2.74}$$

Then x_t and y_t are zero mean Gaussian random variables and have variances:

$$\langle x_t^2\rangle = \langle y_t^2\rangle = \sigma_s^2 + \sigma_n^2 \quad . \tag{2.75}$$

Referring to Figure 2.16, the output of the receiver multiplier is given by:

$$V_1(t) = r(t) \cdot r(t-T) \tag{2.76}$$

$$= [x_t\cos(\omega_c(t)+\phi_s(t)) - y_t\sin(\omega_c(t)+\phi_s(t))] \cdot$$

$$[x_{t-T}\cos(\omega_c(t-T)+\phi_s(t-T)) - y_{t-T}\sin(\omega_c(t-T)+\phi_s(t-T))]$$

where T is the pulse duration.

Put $\phi_{t1} = \phi_s(t)$, and $\phi_{t2} = \phi_s(t-T)$. Hence,

$$V_1(t) = x_t x_{t-T} \cos(\omega_c t+\phi_{t1}) \cos(\omega_c(t-T)+\phi_{t2})$$
$$- y_t x_{t-T} \sin(\omega_c t+\phi_{t1}) \cos(\omega_c(t-T)+\phi_{t2})$$
$$- x_t y_{t-T} \cos(\omega_c t+\phi_{t1}) \sin(\omega_c(t-T)+\phi_{t2})$$
$$+ y_t y_{t-T} \sin(\omega_c t+\phi_{t1}) \sin(\omega_c(t-T)+\phi_{t2})$$

Whence,

$$2.V_1(t) = x_t x_{t-T}[\cos(2\omega_c t-\omega_c T+\phi_{t1}+\phi_{t2})+\cos(\omega_c T+\phi_{t1}-\phi_{t2})]$$

$$- y_t x_{t-T}[\sin(2\omega_c t-\omega_c T+\phi_{t1}+\phi_{t2})+\sin(\omega_c T+\phi_{t1}-\phi_{t2})]$$
$$- x_t y_{t-T}[\sin(2\omega_c t-\omega_c T+\phi_{t1}+\phi_{t2})-\sin(\omega_c T+\phi_{t1}-\phi_{t2})]$$
$$- y_t y_{t-T}[\cos(2\omega_c t-\omega_c T+\phi_{t1}+\phi_{t2})+\cos(\omega_c T+\phi_{t1}-\phi_{t2})] \quad . \quad (2.77)$$

If $V_1(t)$ is low-pass filtered at cut-off ω_c to remove twice carrier terms, and we put $\omega_c T = 2n\pi$ ($n = 1,2,\ldots$), then the LPF output, v_o, is

$$2.V_o = (x_t x_{t-T} + y_t y_{t-T}) \cos \Delta\phi \qquad (2.79)$$

where $\Delta\phi = \phi_{t1}-\phi_{t2}$ which is the received base band signal and it has the value either 0, π or $-\pi$.

In a noise-free system, if there is no phase change, $\Delta\phi = 0$. The LPF output is high as cos $\Delta\phi$ is 1. However, an error will be made if

$$(x_t x_{t-T} + y_t y_{t-T}) \leq 0 \qquad \text{(no phase change case)} \quad . \quad (2.80)$$

If there is a phase change, then $\Delta\phi$ is either π or $-\pi$ and the LPF output should be negative as cos $\Delta\phi$ is -1. Hence, an error will be made if

$$(x_t x_{t-T} + y_t y_{t-T}) \leq 0 \qquad \text{(phase change case)} \quad . \quad (2.81)$$

Equations (2.80) and (2.81) imply that in the fading situation no matter what the phase change is, an error will occur if

$$[x(t)+x_n(t)][x(t-T)+x_n(t-T)]$$
$$+ [y(t)+y_n(t)][y(t-T)+y_n(t-T)] \leq 0 \quad . \qquad (2.82)$$

Without envelope fading and the random FM effect, it can be proved that the bit error rate in an AWGN channel is given by [10]:

$$P_e = \tfrac{1}{2} e^{-\gamma} \qquad (2.83)$$

where γ is the instantaneous signal to noise ratio. In this case, P_e depends on the variation of the signal power relative to noise and a very good phase recovery is assumed.

When the DPSK signal is in the multipath fading situation, the mean bit error rate depends on the rapid changes in the signal envelope and the random FM effect. Due to the fading envelope, the mean bit error rate can be obtained by using (2.61), that is

$$P_1 = \int_0^{\infty} \tfrac{1}{2}.\exp(-\gamma).(1/\Gamma).\exp(-\gamma/\Gamma)\, d\gamma \tag{2.84a}$$

$$= \frac{1}{2(1+\Gamma)} \quad . \tag{2.84b}$$

When the mean receiver CNR, Γ, approaches infinity, P_1 is zero. Then the mean bit error rate due to the random FM effect is given by [8,10]

$$P_2 = \frac{\Gamma(1-\rho)}{2(1+\Gamma)} \tag{2.85}$$

and $$\rho = J_o(2\pi f_d T) \tag{2.86}$$

where ρ is the autocorrelation of the in-phase or the quadrature component of the fading shifted by time T, and $J_o(.)$ is the Bessel function of the first kind of order 0. Then, the total mean error rate is

$$\langle P_e \rangle = P_1 + P_2 = \frac{1+\Gamma(1-\rho)}{2(1+\Gamma)} \quad . \tag{2.87}$$

As Γ tends to infinity, $\langle P_e \rangle$ is $(1-\rho)/2$. The total mean error rate is leveled at this rate, and this is largely due to the random FM effect. This also means that even though the transmitting power is increased, the bit error rate would not diminished. This is the irreducible error because of the random FM effect.

2.4.2.2 DPSK Fading Simulation

Apart from the error decision rule given in (2.82), the DPSK and the NCFSK fading simulation algorithms are nearly the same. In the case of DPSK, four independent Gaussian random number sources are required. Two Gaussian sources are to simulate the receiver noise, and two are used to generate the fading signal. The error testing algorithm is described

step-wise as follows, where SIGE and NOISE are accumulators for the carrier signal energy and the DPSK system noise, respectively. 'F' is a constant to determine the mean carrier to noise ratio (CNR), and this mean CNR would be higher for larger value of 'F'.

Step 0 : Initialization,

F = a constant,
SIGE = 0, NOISE = 0, IERR = 0, NPTS = 0,

Step 1 : Compute Rayleigh fading quadrature components, $x_{r,t}$ and $y_{r,t}$,
$SIGE = SIGE + F^2(x_{r,t}^2 + y_{r,t}^2)$

Step 2 : Error testing for a transmitting data,

$XXT = (F.x_{r,t}+GAU_{1,t})\times(F.x_{r,t-1}+GAU_{1,t-1})$
$YYT = (F.y_{r,t}+GAU_{2,t})\times(F.y_{r,t-1}+GAU_{2,t-1})$
If XXT ≤ -YYT then
(Invert data symbol) and
IERR = IERR + 1 .

Step 3 : $NOISE = NOISE + (GAU_{1,t}^2 + GAU_{2,t}^2)$,
NPTS = NPTS + 1,
IF NPTS = 1000 then EXIT (for 1000 data),
GOTO Step 1.

Assume that 'NPTS' bits of data are transmitted and each bit is compared with the above fading channel algorithm. The simulated mean CNR and the simulated bit error rate may be computed as for the NCFSK system given in (2.65) to (2.67).

The simulation performance was again evaluated by plotting the simulated bit error rate performance at different mean CNR as shown in Figure 2.18, and its theoretical bit error rate curve as defined in (2.87) has also been plotted in the same Figure for comparison. The theoretical curve and the simulation curve are very close together. At maximum Doppler frequency 41.67 Hz as shown in the Figure, the bit error curves become

level beyond 40 dB mean CNR. This is because of the Random FM effect. At a higher Doppler frequency, the bit error rate curves would start to level off at a higher bit error rate so that the transmission quality will become worse.

2.4.3 DIVERSITY SYSTEMS

Diversity systems have been well known in radio communications for many decades [25], and can dramatically improve communication performance in Rayleigh fading channels, even in frequency selective channels. There are many types of diversity scheme, such as space diversity, frequency diversity, polarisation diversity, field component diversity and time diversity etc.

Space diversity seems to be a more suitable technique for use in LMR communications. For space diversity systems, if two receiving antennae are separated sufficiently apart horizontally or vertically, the fast fading fluctuation of the signal received at one antenna tends to be independent of the fluctuation of the signal received on the second antenna. If the system output is obtained by combining the two signals at each instant, this will greatly reduce excessively deep fade effects. Further improvement can be obtained by using more than two antennae, but the system complexity will be correspondingly increased. In space diversity, there are three major combining techniques [22,26]. They are selection diversity, equal gain combining and maximal ratio combining.

A two antennae diversity system is considered to be the most practical, and selection diversity is the simplest compared with the other combining techniques. This will be described in the following sections.

2.4.3.1 Selection Diversity Analysis

For selection diversity, if M antennae are used, the effective received signal is obtained from the antenna bearing the strongest signal, therefore increasing the mean CNR.

It is assumed that the signal on each antenna has Rayleigh statistics and so the CNR on the k-th antenna, γ_k, has exponential statistics as in (2.60). Thus, the probability that CNR on all M antennae are simultaneously less than or equal to γ_s is

$$P_s(\gamma \leq \gamma_s) = [\ 1 - \exp(-\gamma_s/\Gamma)\ ]^M \ . \qquad (2.91)$$

This is the cumulative distribution function of the best signal taken from the M antennae.

The pdf, $p_s(\gamma)$, of the CNR at the output of a diversity system is obtained by differentiating $P_s(\gamma \leq \gamma_s)$, i.e.,

$$p_s(\gamma) = (M/\Gamma).(1 - \exp(-\gamma/\Gamma))^{M-1} \exp(-\gamma/\Gamma) \ . \qquad (2.92)$$

When M is 1, $p_s(\gamma)$ represents the Rayleigh distribution of the CNR in each antenna.

The effectiveness of a diversity system in data communications can be determined by the reduction that is achieved in the bit error rate. If the error rate calculation is based on frequency non-selective fading assumptions, the mean error rate of the diversity system is given by

$$\langle p_d \rangle = \int_0^\infty P_e(\gamma)\ p_s(\gamma)\ d\gamma \ . \qquad (2.93)$$

$P_e(\gamma)$ is the error rate of the digital modulation scheme in the presence of AWGN.

Consider a 2 antenna selection diversity arrangement with NCFSK data. The average error rate is

$$\langle P_{d,2} \rangle = \int_0^\infty \tfrac{1}{2}\exp(-\gamma/2).(2/\Gamma)(1 - \exp(-\gamma/\Gamma)).\exp(-\gamma/\Gamma)\ d\gamma \qquad (2.94)$$

$$= \frac{1}{(2+\Gamma)(4+\Gamma)} \quad .$$

If Γ is very large, $\langle P_{d,2} \rangle$ is approximately equal to $4/\Gamma^2$. When comparing this bit error rate with that obtained without diversity, i.e. with $\langle P_e \rangle$, as given in (2.63)

$$\langle P_{d,2} \rangle \approx 4 \langle P_e \rangle^2 \quad . \qquad (2.95)$$

Hence, if $\langle P_e \rangle$ is 0.01, $\langle P_{d,2} \rangle$ is 4×10^{-4}. The 2-antenna diversity can therefore improve the signal level by about 10 dB, which is known as diversity gain. 4-antenna diversity systems yield 16 dB diversity gain. However, 2-antenna diversity is less complex while still achieving a considerable improvement. The performance of the equal gain combining diversity and maximal ratio combining diversity techniques are generally much better than the selection diversity technique and therefore further improvement can be obtained even with 2-antenna diversity.

For narrowband SCPC systems, diversity is a very useful technique. The following section describes the simulation of a 2-branch selection diversity system.

2.4.3.2 Selection diversity Simulation

In this 2-antenna diversity simulation, it is assumed that the antennae spacing is chosen appropriately so that the receiving signals at the antennae are independent. The received signal from both antennae are fed to a selection circuit, and the stronger signal will then be switched to a demodulator input.

Assuming that a NCFSK detector is used, the simulation of the 2-branch selection diversity system requires 8 independent Gaussian sources. Four of them are used as the quadrature components of two receiving signals and their fading envelopes are given by

$$R_1(t) = \sqrt{[\, x_1(t)^2 + y_1(t)^2 \,]} \qquad (2.96)$$
$$R_2(t) = \sqrt{[\, x_2(t)^2 + y_2(t)^2 \,]}$$

where $x_1(t)$, $y_1(t)$, $x_2(t)$ and $y_2(t)$are Gaussian random numbers which have been spectrally shaped using Doppler spectrum filters. Thus, $R_1(t)$ and $R_2(t)$ are statistically independent envelope samples. As the stronger signal is chosen, the effective received signal CHOICE(t) is therefore

$$\text{CHOICE}(t) = \text{MAX}\,(\, R_1(t),\, R_2(t)\,)\ . \qquad (2.97)$$

Using the NCFSK error decision condition given in (2.52), an error will be made if

$$(\text{CHOICE}(t)+\text{GAU}_5(t))^2 + \text{GAU}_6(t)^2 \;\le\; \text{GAU}_7(t)^2 + \text{GAU}_8(t)^2 \qquad (2.98)$$

where GAUs' are the Gaussian sources for detector noise. The improvement of the 2-branch diversity system can be visualised from Figure 2.19, where two independent simulated fading envelopes are plotted against time in broken lines, and the solid line shows the effective received signal. The diagram clearly shows that the deep fades have been greatly alleviated. The simulated bit error rate and the simulated CNR are determined as in the case of the NCFSK simulation algorithm given in (2.65) to (2.67). The bit error rate performance of the 2-branch diversity simulation is plotted in Figure 2.20 along with the theoretical bit error rate curve. They are very close to each other.

2.5 DISCUSSION AND CONCLUSION

This Chapter has presented a development of LMR channel simulation techniques based on theoretical statistical model of the channel. The simulation algorithms described are applicable to narrow-band single channel per carrier (SCPC) digital LMR systems, where the bandwidth is less than the coherence bandwidth (as described in Chapter 1) and the ratio of bit rate to bandwidth is not greater than about one for each channel. In this case, the effects of frequency selective fading and delay spread are very small and therefore need not be simulated. The simulation is suitable for 900 MHz LMR systems with narrowband channels (25kHz) used for transmitting data rates up to 25 kb/s. As the delays in the LMR channel of over 8 μsec are very rare [4], the highest transmission data rate, before frequency selective fading becomes prominent in such a system, is approximately 1/8μ b/s (= 125 kb/s). The transmission data rate of 25 kb/s is far less than 125 kb/s. Hence, the simulation is valid for the SCPC system in frequency non-selective fading channels.

The basic Rayleigh fading signal as described in (2.8) was simulated. This simulated signal was applied to two commonly used modulation schemes: NCFSK and DPSK, and a selection diversity technique was incorporated. Thus the simulator is able to generate a fading channel error sequence to indicate the occurrence of fading channel errors which would affect data transmission. The performance of the simulator has been evaluated by measuring the fading envelope statistics and bit error rates at different mean carrier to noise ratios, and is consistent with theoretical results [8,10]. The most important results produced by the simulation are the bit error patterns as shown in Figure 2.15, from which the burst and random error characteristics of the LMR channel are deduced.

This error pattern output can be applied to digitally coded speech, with its error correction code, to assess the effect of fading channel error on the quality of speech received. As the simulation parameters, such as mobile speed and data transmission rate can be changed easily, the burst error characteristics, such as burst error length and frequency can be conveniently measured and analysed.

The simulation techniques described in this Chapter for NCFSK, DPSK and selection diversity also demonstrate a methodology that can be used for simulating the Rayleigh fading effect for other types of system. Furthermore, the simulation is not restricted to 900MHz SCPC radio channels. Many researchers have measured signal statistics in real radio channel over frequency ranges from 50 to 11,200 MHz, which covers from the HF (high frequency) to the SHF (super high frequency) region [8]. They showed that received signals have Rayleigh statistics in mobile radio environment when measured over distances of a few tens of wavelengths when the mean signal is relatively constant. Hence, as long as the radio channel is not frequency selective in transmission, the simulation algorithms used in this Chapter can also be used to evaluate any data transmission for HF, VHF, UHF and SHF channels.

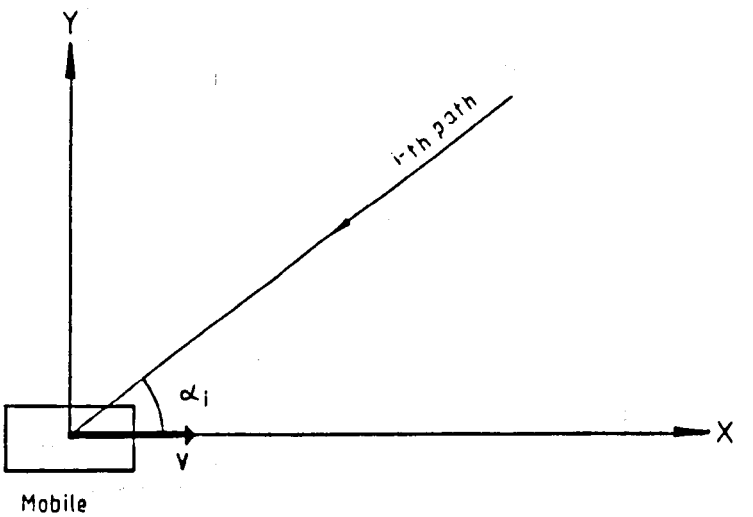


Figure 2.1 The ray-path Rayleigh fading model.

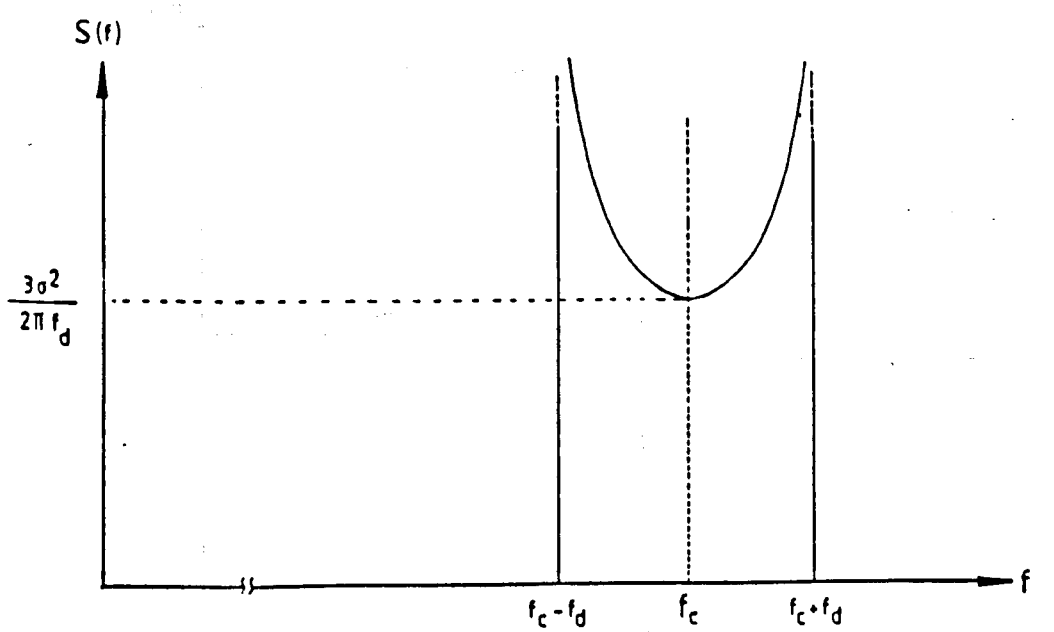


Figure 2.2 Power spectra of the received signal for uniformly distributed arrival angles.

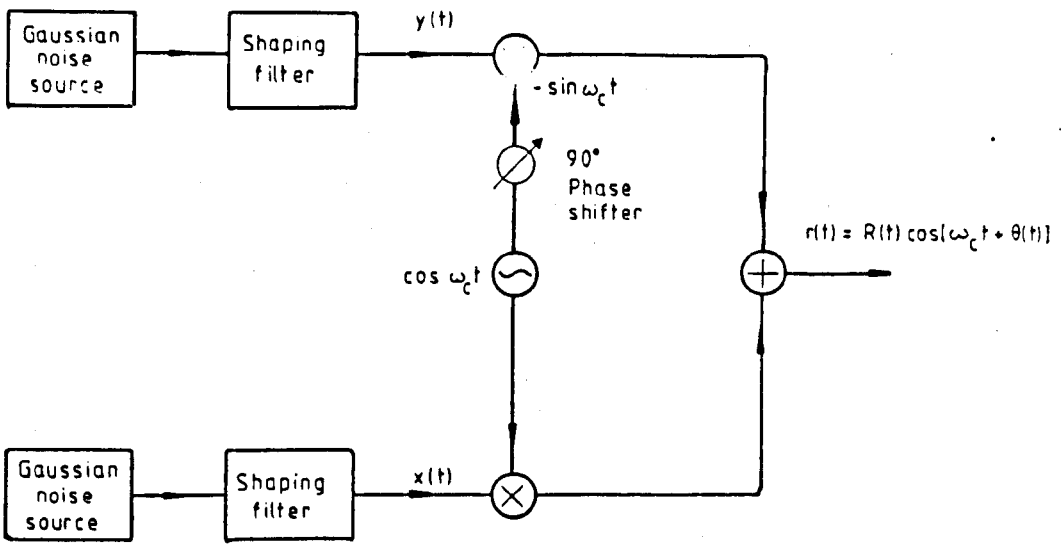


Figure 2.3 Rayleigh fading simulator.

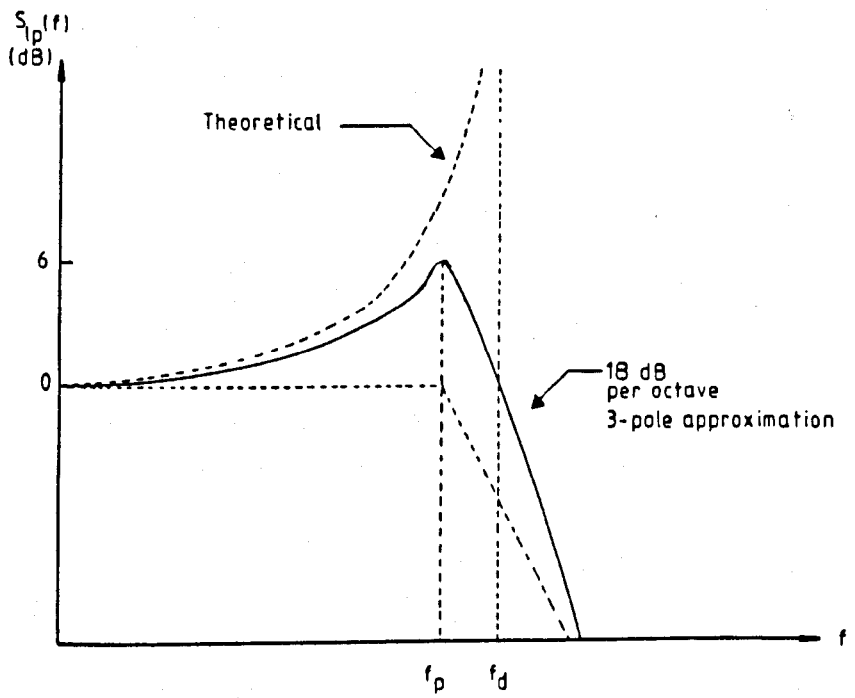


Figure 2.4 Shaping filter frequency response.

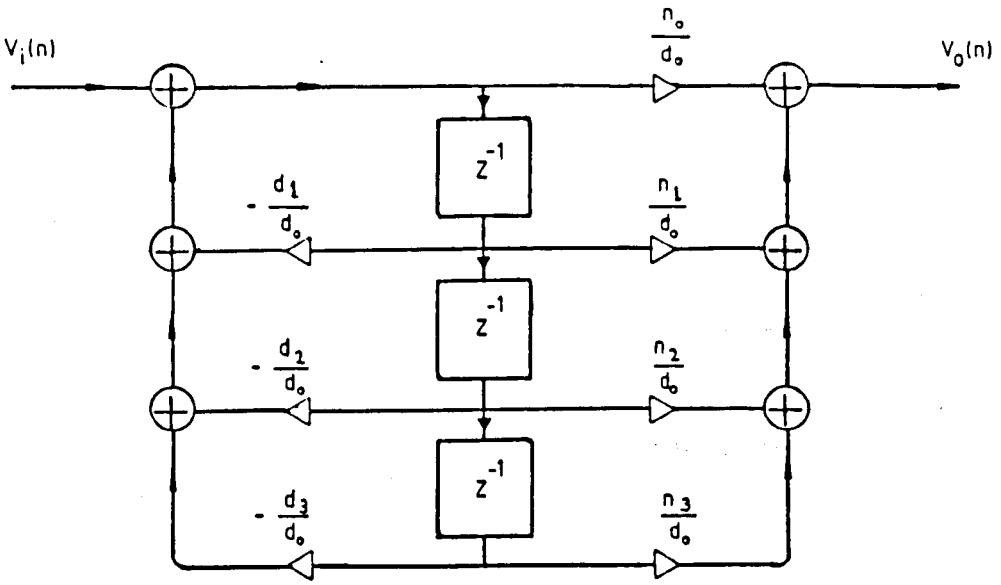


Figure 2.5 The third order IIR filter configuration.

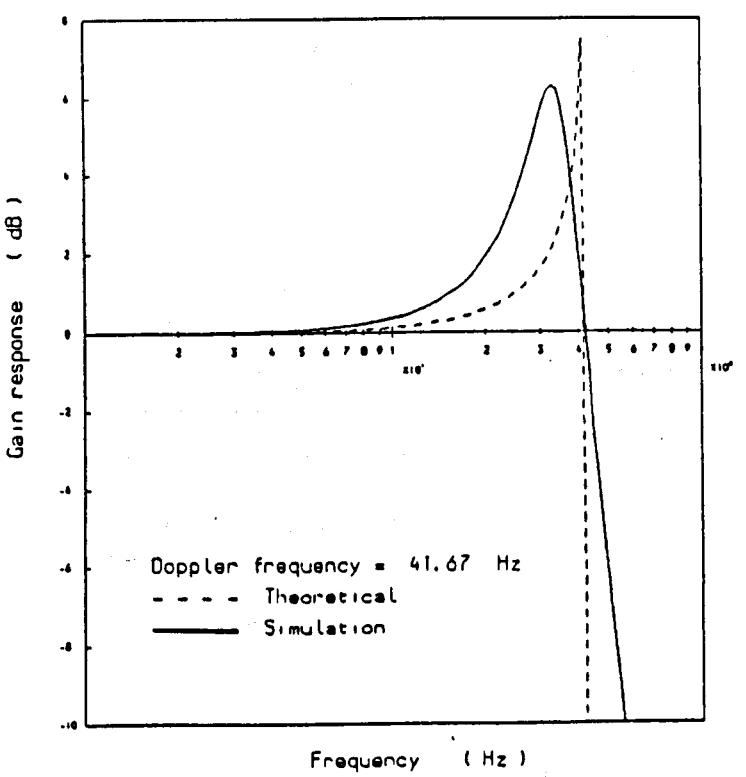


Figure 2.6 Power spectrum for the scattering field model.

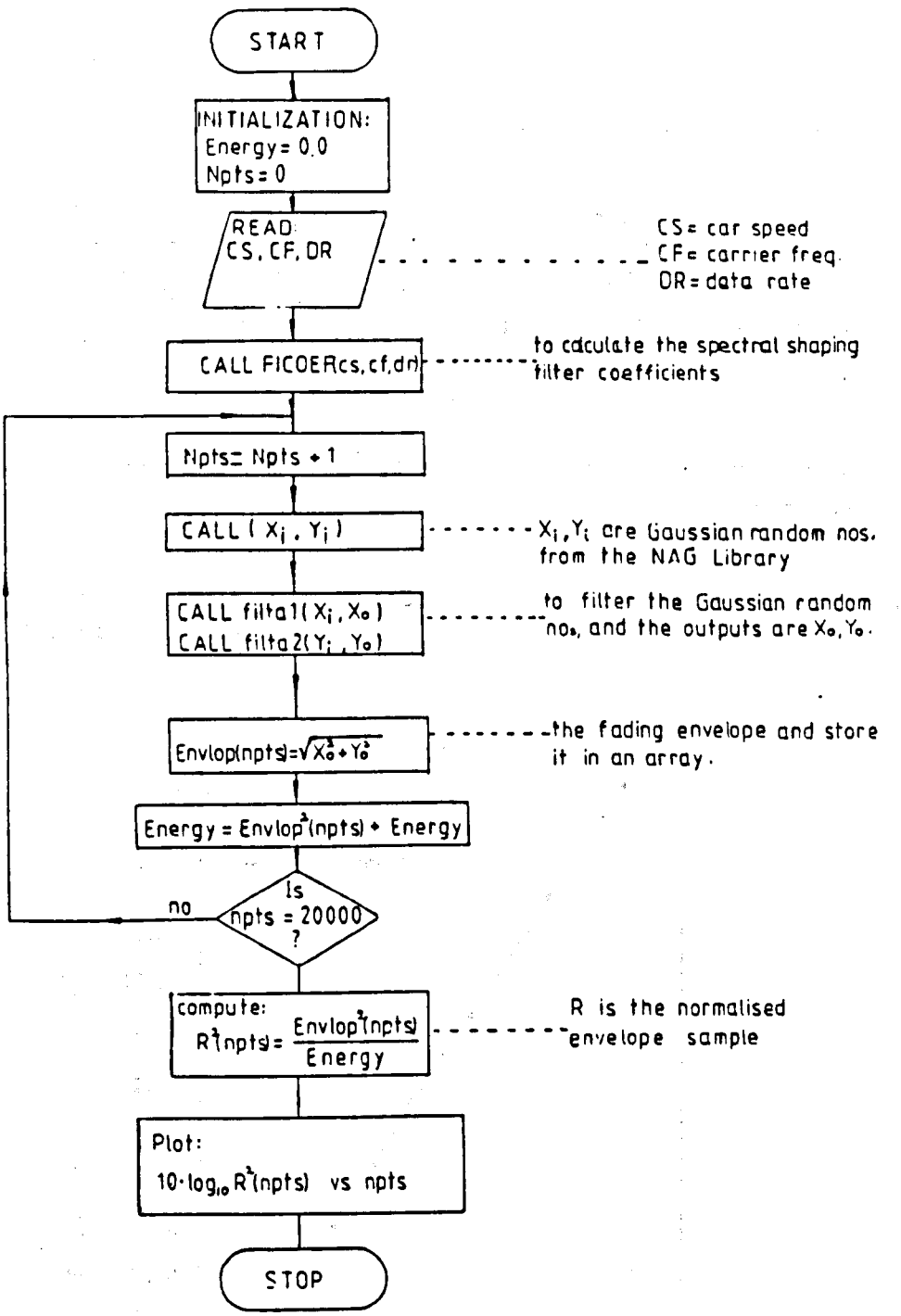


Figure 2.7 Flow chart of the Rayleigh fading envelope simulator.

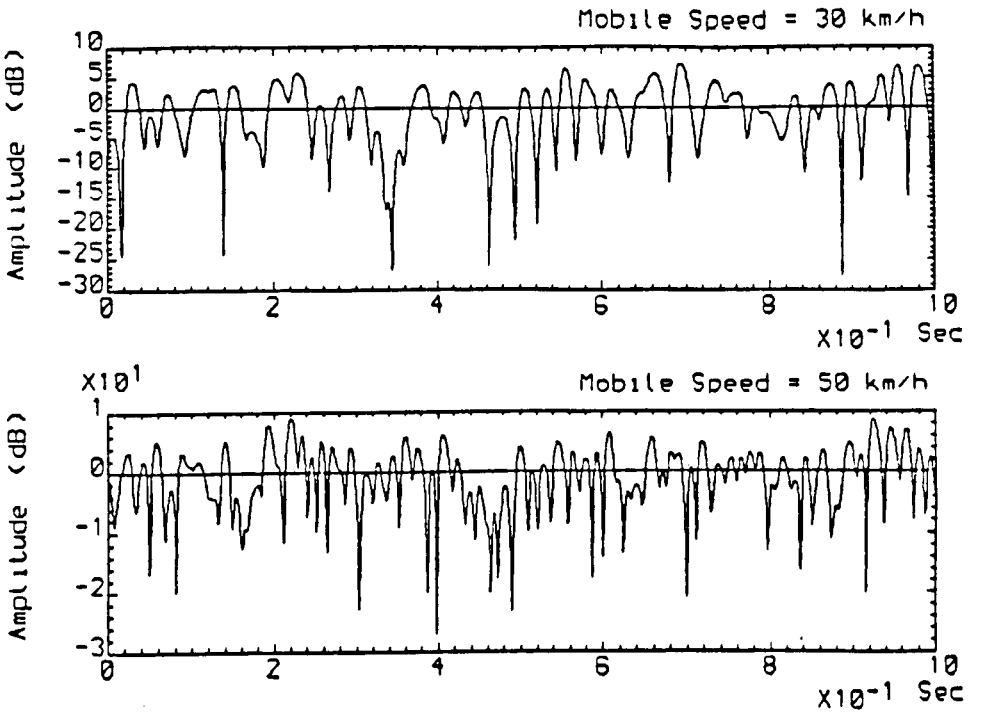


Figure 2.8 Simulated Rayleigh fading envelopes of a 900 MHz carrier for two different mobile speeds.

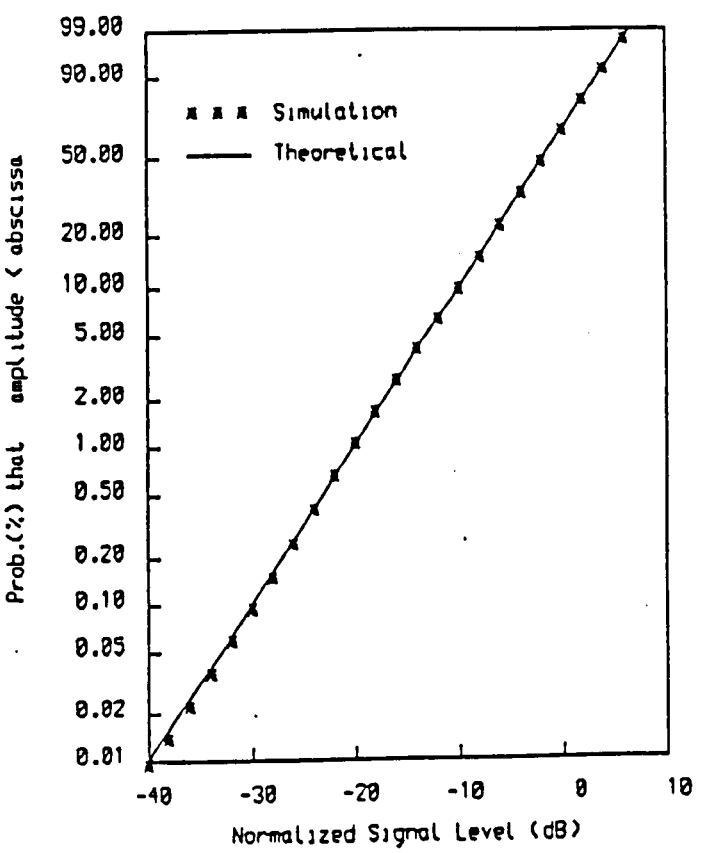


Figure 2.9 Rayleigh plot of the fading envelope.

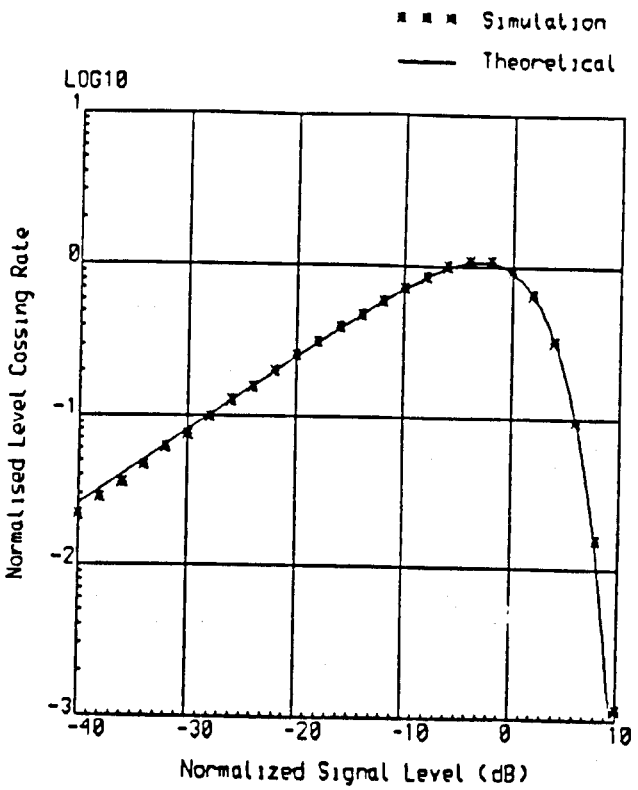


Figure 2.10 Normalised level crossing rates of the envelope.

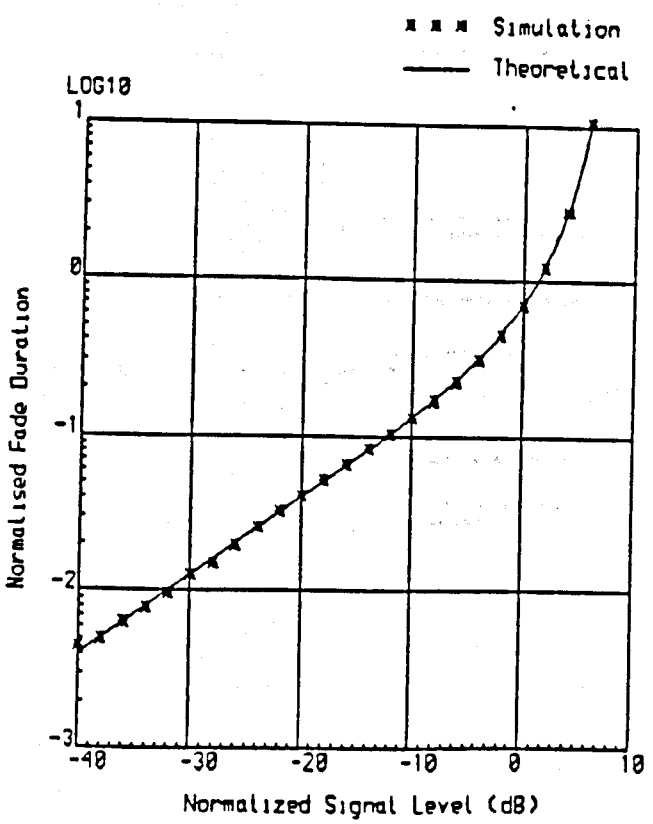


Figure 2.11 Normalised fade durations of the envelope.

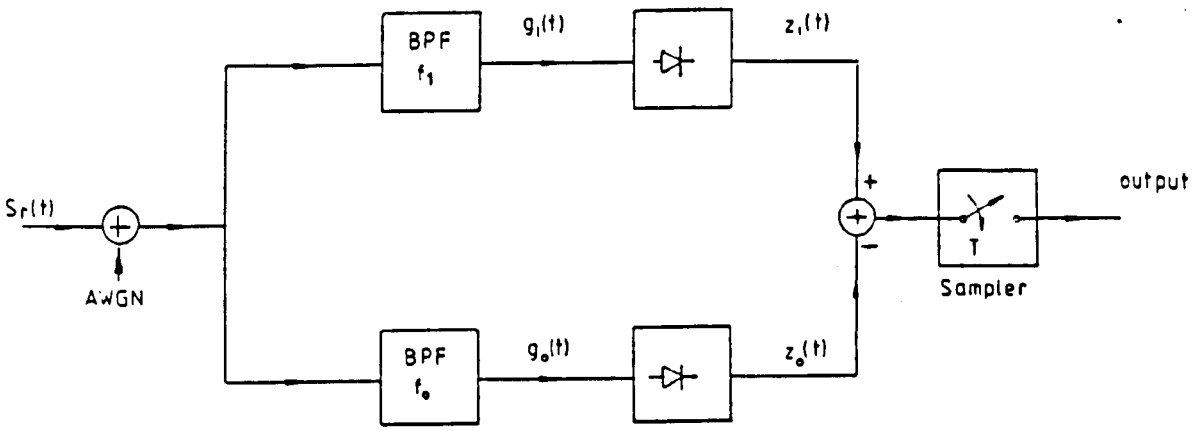


Figure 2.12 Non-coherent frequency shift keying detector.

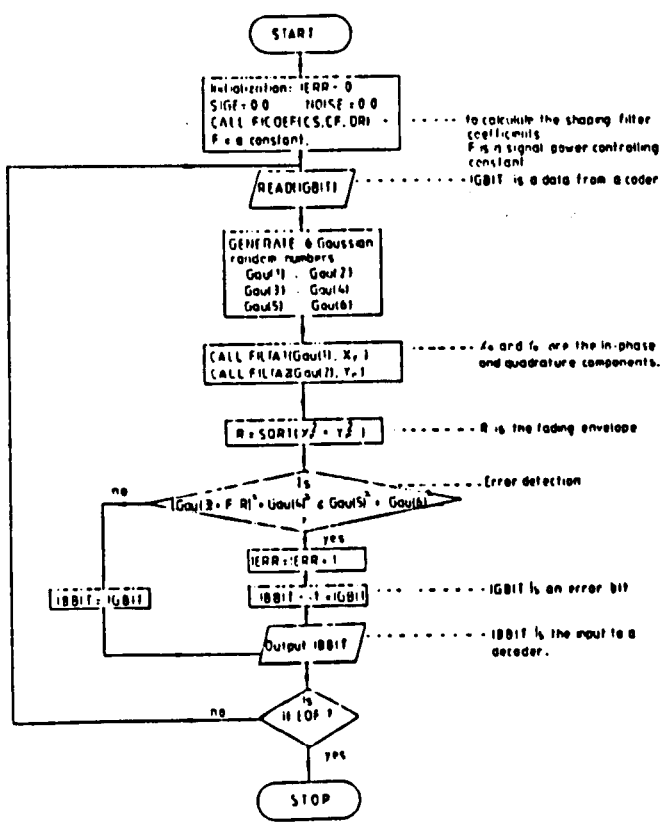


Figure 2.13 Flow diagram for the NCFSK fading simulation.

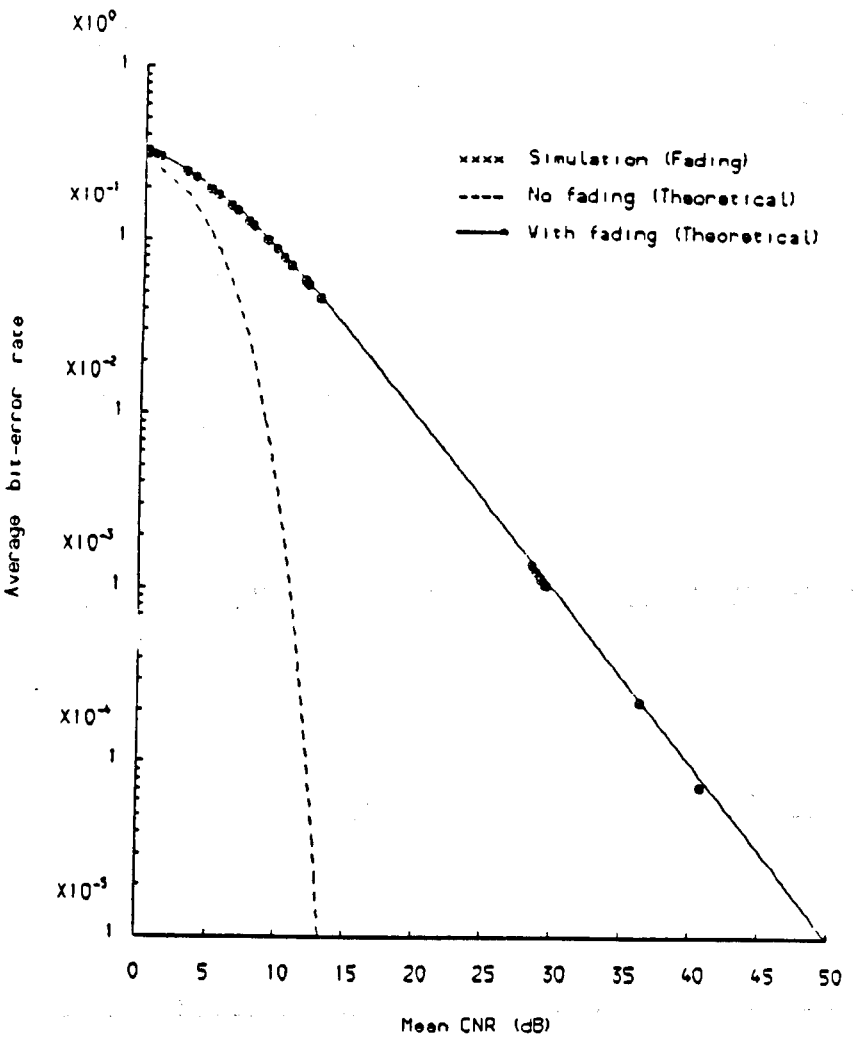


Figure 2.14 Bit-error rate of NCFSK signalling on a Rayleigh fading channel.

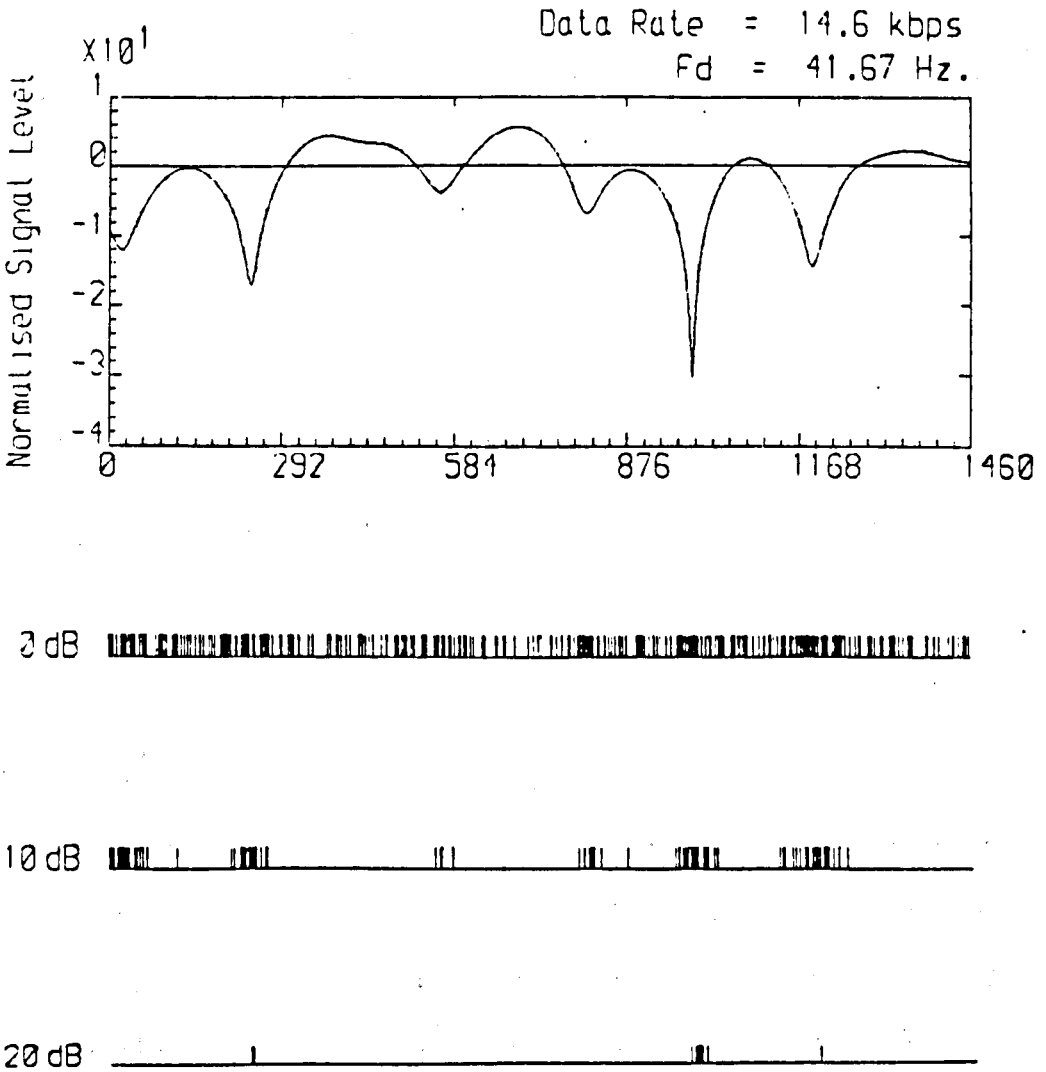


Figure 2.15 NCFSK receiver error pattern at different mean CNR.

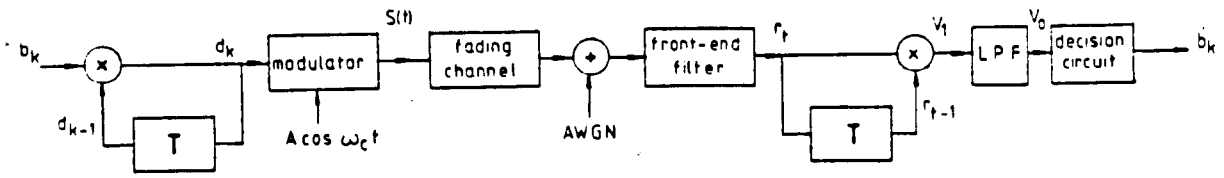


Figure 2.16 Block diagram of a DPSK system.

Message operation		Message coding							
b_k		1	-1	-1	1	1	1	-1	-1
d_k	1	1	-1	1	1	1	1	-1	1
$\phi_s(t)$		0	π	0	0	0	0	π	0
$\Delta\phi$		0	π	$-\pi$	0	0	0	π	$-\pi$
b'_k		1	-1	-1	1	1	1	-1	-1

Figure 2.17 An example of DPSK encoding

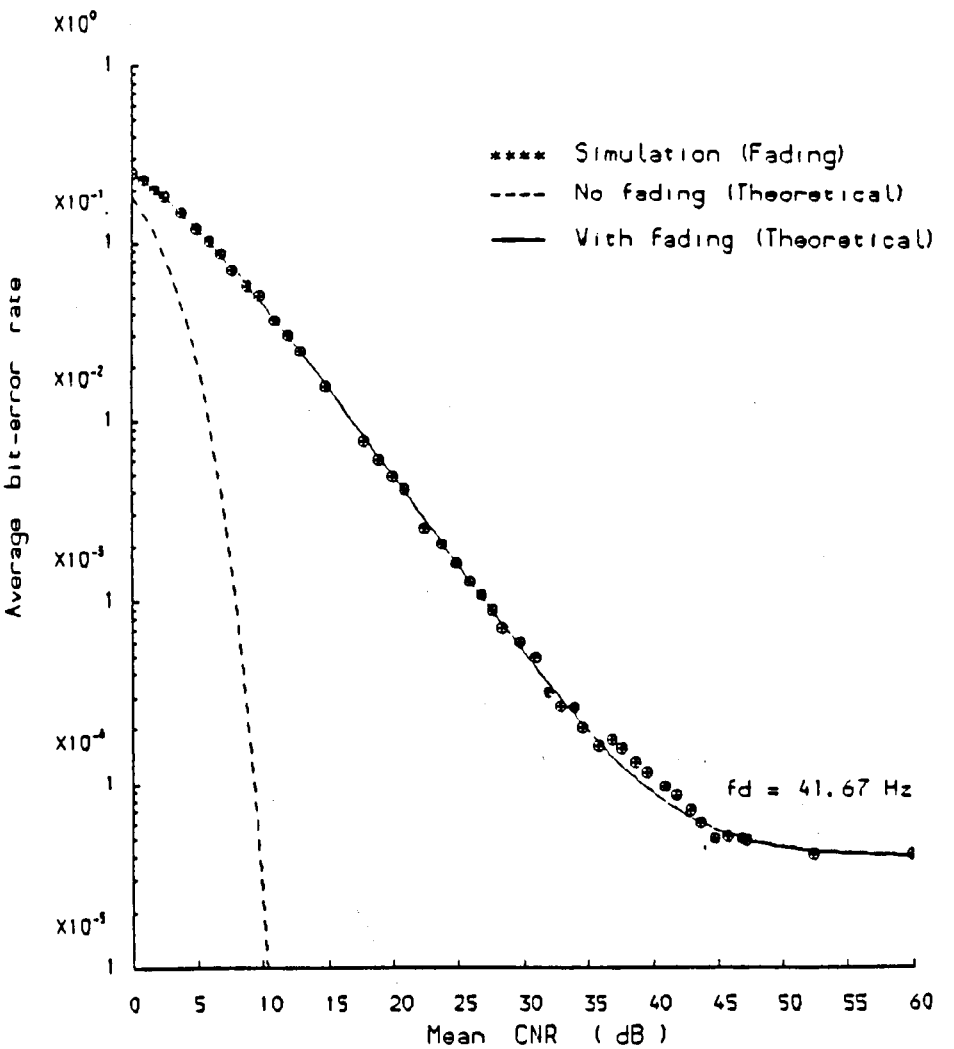


Figure 2.18 Bit-error rate of DPSK signalling on a Rayleigh fading channel.

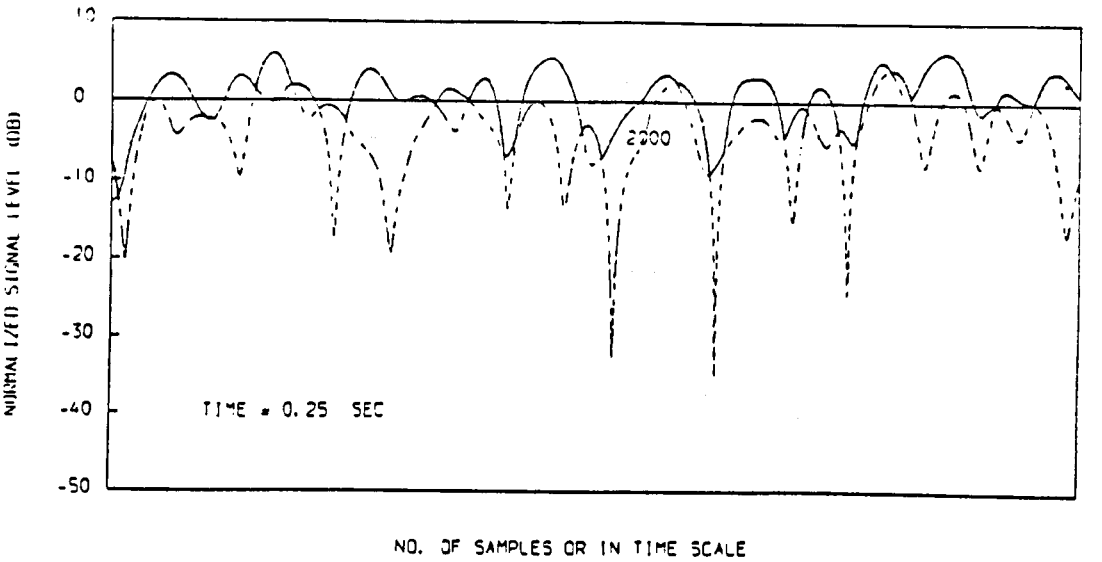


Figure 2.19 Illustration of the function of 2-branch diversity on the fading envelope for f_d = 41.67 Hz.

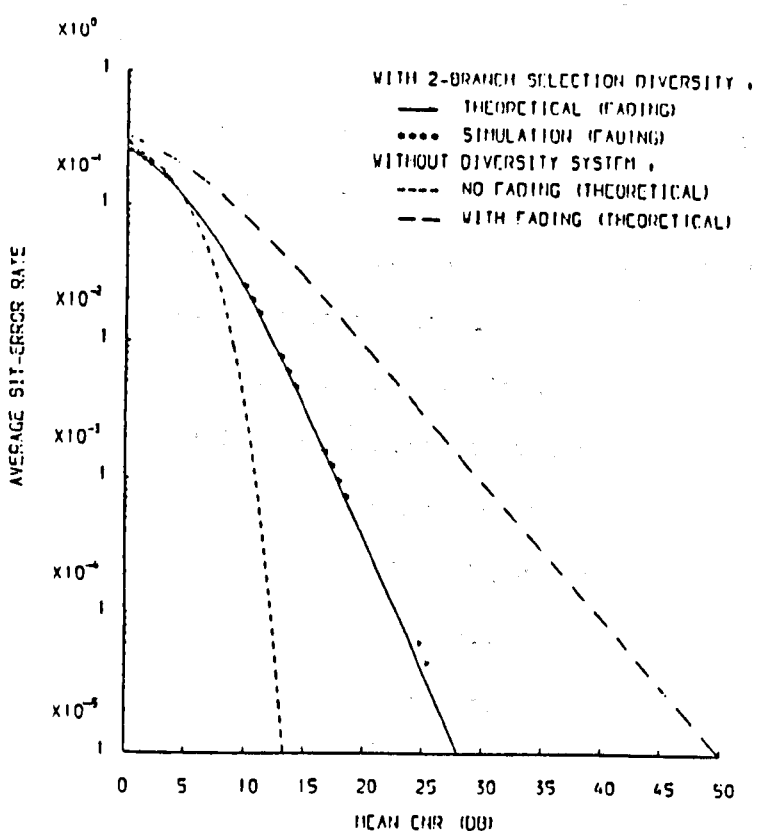


Figure 2.20 Bit-error rate performance of a 2-branch NCFSK system.

CHAPTER 3 ADAPTIVE PREDICTIVE CODING WITH ADAPTIVE BIT ALLOCATION

3.1 INTRODUCTION

To achieve high channel spectrum efficiency for digital LMR systems, it will be necessary to use low-bit rate speech coders which produce speech of a quality comparable with telephone speech. Several types of speech coder have been examined for use in digital LMR systems. Considering the quality of speech coders available at present, it is clear that a 16 kb/s speech coder should be capable of producing good quality speech and that if a degree of error protection can be incorporated into such a coder it could be made to satisfy the spectrum efficiency and quality required for use in LMR systems[2].

One promising possibility is the use of sub-band coding (SBC) methods. At present, there are many different types of SBC [e.g.27,28,29,30], their most important differences being in the methods used to encode the speech parameters. Advanced coding techniques like vector quantization, vector prediction or adaptive bit allocation have been used to code the sub-band signals. A coding technique called 'adaptive predictive coding with dynamic bit allocation' (APC-AB) [31] has been found to produce good quality speech at 16 kb/s or below, and this has been strongly supported [31] as a suitable technique for LMR systems. One of the aims of this research programme is to investigate the influence of fading channel errors on a APC-AB subband speech coding system.

At the cost of being computationally rather expensive as compared with some other low to medium bit rate speech coders, the APC-AB coder achieves high speech quality. The performance of the APC-AB coder is optimized by the use of a combination of several advanced techniques, i.e.

sub-band analysis and synthesis, pitch detection, APC coding, transform coding and adaptive bit allocation. It is believed that before attempting to analyse the performance of this speech coding system in a noisy channel, a good understanding of the APC-AB algorithm must be gained. This Chapter will describe all the techniques used in the speech coder in detail. As the implementation details of the original APC-AB coder proposed by Itakura [31] are not fully available in the literature [31], it has been necessary for the author to devise his own version of the APC-AB coder. The coder was implemented in simulation by running a FORTRAN 77 program on an IBM/760 main frame computer. Its performance will be compared with results published by Honda & Itakura [31].

During the course of this work it was found that the complexity of the original APC-AB coder [31] could be greatly reduced by the use of a new algorithm for extracting line spectrum pair coefficients. This new technique will be described in detail in Chapter 4.

Detailed description of the 16 kb/s APC-AB speech coder simulation is organised as follows. In Section 3.2, the basic coding algorithm is described, and in Section 3.3 the details of the encoding techniques are given. In Section 3.4, some practical aspects of the coder implementation are described. The speech coder was assessed by objective measurements, and informal listening tests. These tests are described in Section 3.5. Finally, the speech coder design is summarised in Section 3.6.

3.2 APC-AB CODER DESCRIPTION

Figure 3.1 shows the basic configuration of the APC-AB encoder and decoder. Figure 3.2 shows a step by step flow chart of the APC-AB algo-

rithm, and each block of the flow chart is described in the following sections.

3.2.1 QMF ANALYSIS AND SYNTHESIS NETWORKS

The input speech is band-limited to 3.2 kHz and sampled at 6.4 kHz by a 12-bit A/D converter. Every 20 msec of speech, which consists of 128 samples, forms a frame which is split into three frequency bands using a two level quadrature mirror filter (QMF) analysis network. This analysis network is implemented by 32 tap FIR filters at both levels. The three sub-band bandwidths are 0 to 0.8 kHz, 0.8 to 1.6 kHz and 1.6 to 3.2 kHz. The QMF network is implemented via the polyphase structure proposed in [32]. In this way, computation and memory requirements are reduced by a factor of four as compared to the conventional structure. The QMF network is described in detail in Section 3.3.1.

3.2.2 ADAPTIVE PREDICTIVE CODING (APC)

After sub-band splitting, the sub-band signals are encoded by an APC scheme, which consists of a long term predictor and a short term predictor for each subband. These predictors are used to remove the correlation among speech samples and thereby produce outputs known as residual signals before quantization. Each residual signal has a smaller power than the corresponding sub-band speech signal resulting in an increased output signal to quantization noise ratio (SNR). The long term predictor coefficients are known as pitch gain, b, and pitch period, M, and the short term predictor coefficients are Line Spectrum Pair (LSP) coeffi-

cients. These coefficients are varied adaptively in time to track the changing properties of the input speech, and such adaptation is based on a forward prediction scheme [33], which transmits the adapted coefficients once every frame.

Unit variance quantizers are used to quantize the sub-band residuals; these unit variance quantizers are optimally designed by minimising the mean square error between the quantizer input and output assuming probability distributions appropriate to the inputs [34]. The quantizer has a gain normalisation stage of gain 1/G followed by the optimum quantizer. Theoretically, G should be the r.m.s. value of the APC residual so that the normalised APC residual has unit variance. G is encoded and transmitted to the receiver along with the encoded residual so that the quantized APC residual can be recovered from the decoded normalised residual by multiplying by the decoded value of G.

As G is computed from the residual and the APC residual becomes available only as the APC loop is in operation, the residual signal should therefore be computed twice: once to calculate G and once to compute the quantized residual signal through the APC loop as shown in Figure 3.1 . In the simulation, each sub-band signal is first filtered by a long term predictor to produce the first residual, then the first residual is filtered by a short term predictor to produce the second residual. From this process, the LSP coefficients are also calculated, and G is computed as the r.m.s. value of the second residual.

Having obtained the predictor coefficients and the quantizer parameter, G, the residual is then calculated and encoded using the corresponding quantized predictor coefficients and quantized G, to correspond to the operation of the receiver; this also yields a smaller mean square

value for the quantization noise as perceived in the received speech than if the unquantized parameters are used.

3.2.3 PITCH SYNCHRONOUS ADAPTIVE BIT ALLOCATION

The residual quantizer produces two kinds of quantization error: round off error and clipping error. Round off error is produced whenever the residual lies within the range of the quantizer, and this causes a degradation in the output speech in the form of broadband background noise. Clipping errors are produced whenever the residual lies outside the range of the quantizer. This may happen for high amplitude pitch signals. Clipping errors cause undesirable degradation in the form of 'pops' or 'clicks', and such effects can be perceived even when the incidence of clipping errors is as low as 0.1 %.

For a 16 kb/s APC-AB speech coder, the average number of available bits per residual sample is 2.5 bits. As a forward prediction scheme is used, some of the bits must be used to transmit side information. Therefore, with less than 2.5 bits/sample available, the method used for residual coding must have provision to deal with regions of high amplitude residual samples (e.g. pitch pulses), while also keeping round-off noise to an acceptable level.

The bit allocation scheme used in the speech coder was proposed by Honda [31], and is known as Pitch Synchronous Energy Concentration Adaptive Bit Allocation. This method divides each pitch period of the second residual into L approximately equal length segments. The segments within each pitch period are indexed in ascending order from 1 to L. Segments for successive pitch periods are combined to form L distinct sets

of samples, each set comprising several samples from each period. The energy per sample, u_{ij}, for each segment is computed, where i and j indicate the i-th sub-band and j-th segment respectively. For voiced speech, pitch pulses are assumed to occur at regular time intervals, and the pitch pulses always contain most of the excitation signal energy. The pitch synchronous strategy arranges the segments so that the pitch pulses occur in segment one. This means that u_{i1} always has the largest power. To achieve this, u_{11} is obtained after an energy search procedure, and a side information parameter known as "segment location" or "sub-interval position", T_{d1}, is derived to specify the distance (in terms of the number of samples or in time) of the very first segment one from the beginning of the first sub-band frame. The segment location for other bands are given by

$$T_{di} = \frac{\omega_i}{\omega_1} T_{d1} \tag{3.1}$$

where ω_i is the ratio of the i-th sub-band bandwidth to the full band bandwidth.

By rate distortion theory [35], the optimal bit allocation, R_{ij}, i.e. the number of bits allocated to subband i in time segment j, may be computed so that the mean square quantization error is minimised. It is found that R_{ij} is given by

$$R_{ij} = R + \frac{1}{2} - \log_2 \left[\frac{\tilde{u}_{ij}}{\prod_{n=1}^{N} \prod_{\ell=1}^{L} (\tilde{u}_{n\ell})^{w_n c_\ell}} \right] \tag{3.2}$$

where R is the average bit rate, c_j is the ratio of the j-th segment length to the frame length and $\tilde{u}_{ij}$ is the residual energy normalised by the bandwidth ratio ($= u_{ij}/\omega_i$). In this way, the quantization bits are dy-

namically allocated to the prediction residuals in accordance with the distribution of their energies in the time segments and in the frequency sub-bands.

Furthermore, unit variance quantizers are applied to each segment and the normalisation gain, G_{ij}, for each segment is given by

$$G_{ij} = \sqrt{u_{ij}} \tag{3.3}$$

The optimal step size for a uniform unit-variance quantizer for a Gaussian distributed input signal was derived by Max [34] as

$$\Delta = g(R) \tag{3.4}$$

where g(R) is a tabulated function of the number of bits R to yield minimum mean square error for a zero mean and unit-variance input signal. Max has computed and presented the optimal step sizes by tabulating g(R) for 1-bit to 32-bit quantizers in [34]. If the residual in subband i and time segment j is assumed to be zero mean and Gaussian, the optimal step size for this residual signal is therefore $\Delta_{ij}G_{ij}$ where $\Delta_{ij} = g(R_{ij})$ and G_{ij} is the normalised gain as defined in (3.3). This method makes the quantizer step size adapt rapidly to local variations in the energy of the residual samples and provides good tracking of the short term amplitude variation of the APC residual.

The pitch structure will not exist in un-voiced speech which tends to be Gaussian noise like. In this case, the sub-interval position, T_{d1}, is set equal to 0 and each sub-band frame is divided into L segments. Hence, even for un-voiced speech, more bits are allocated to segments bearing higher speech energy.

3.2.4 TRANSFORM CODING

The residual energies u_{ij} may be quantized logarithmically and transmitted to the receiver so that the bit allocation for the residual and G_{ij} can be recovered. However, it was found that the logarithmic value of the normalised residual energies, u_{ij}', given by

$$u_{ij}' = \log_e (u_{ij}/U) \tag{3.5a}$$

with

$$U = \frac{1}{N\,L} \sum_{i=1}^{N} \sum_{j=1}^{L} u_{ij} \tag{3.5b}$$

are highly correlated. Thus the encoding of u_{ij}' can be more efficiently achieved by a data compression technique which attempts to remove this correlation. In addition to adaptive predictive coding methods as described in the previous section, there are other efficient coding techniques including the "transform coding" technique [36], which appear to be well suited to this application. A transform coding technique based on the Karhunen Loeve transformation is adopted.

The Karhunen Loeve transformation transforms the set of u_{ij}' parameters into a set of uncorrelated coefficients u_{ij}'' by an orthogonal transformation that diagonalizes the autocovariance matrix of u_{ij}'. The transformed coefficients may then be more efficiently quantized using a linear quantizer with a fixed number of bits assigned according to their statistical variances. That is

$$B_{ij} = B + \frac{1}{2} \log_2 \left[\frac{\lambda_{ij}}{\prod_{n=1}^{N} \prod_{\ell=1}^{L} \lambda_{n\ell}^{\,1/(N.L)}} \right] \tag{3.6}$$

where λ_{ij} and B_{ij} are the variances and the fixed bit allocation for u_{ij}" respectively, and B is the average bit rate for u_{ij}".

The average residual energy U is quantized using its logarithmic value and transmitted to the receiver.

At the receiver, the side information is first decoded, the K-L coefficients are inverse transformed back to the quantized u_{ij}' parameters, and the quantized residual normalisation gain is recovered from the quantized u_{ij}' and the average residual power. Then, the bit allocation for the residual is calculated using the decoded segment position, pitch period and u_{ij}'. Thus, the sub-band signals are recovered by using the decoded LSP coefficients and the bit allocation. Finally, the reconstructed speech is obtained by merging the sub-band signals into a QMF synthesis network.

To summarise, there are six categories of side information parameters, these being the LSP coefficients, the pitch period, pitch gain, the sub-interval position, the KLT coefficients and the average residual energies. These parameters are computed and encoded every frame. The encoded data are multiplexed with the sub-band residual data and transmitted to the receiver. The techniques being used to encode the residual and the side information parameters are in fact advanced data compression techniques. Such techniques require a large amount of computation, and the speech coder complexity is always proportional to the amount of computation needed. A complex coder can turn out to be bulky in size and this is a most undesirable feature for LMR use. However, some methods being used, like quadrature mirror filtering can be considerably simplified. For example, if the filtering process is carefully considered, a scheme can be devised which reduces the computation by a factor of 4 compared

with the direct implementation method. The details of some interesting algorithms will be given in the following sections.

3.3 DESCRIPTION OF CODING TECHNIQUES

3.3.1 QUADRATURE MIRROR FILTERS (QMFS)

Encoding signals in sub-bands offers several advantages: for instance, the available bits can be allocated to each individual sub-band according to perceptual criteria for each sub-band. If the bits are appropriately allocated in different bands, the reconstruction error variance can be separately controlled in each band, to avoid the masking of one frequency range by quantization noise in another frequency range.

In theory, to preserve all the signal information within the transmission bandwidth, the separation between sub-bands should be zero, but in practice the sub-band filters will overlap and cause mutual aliasing of signal energy between sub-band transition region. The amount of aliasing distortion is directly dependent on the degree to which the sub-band filters approximate ideal bandpass filters, and such distortion cannot be removed by ordinary filtering techniques. This problem can be tackled by the use of QMFs. Using such filters, transition region overlap is permissible because in theory [32] aliasing distortion can be cancelled out at the receiver synthesis filters. Practically, this cancellation is obtained down to the level of the quantization noise of the coders.

The basic structure of the QMF analysis and synthesis filter banks is shown in Figure 3.3a. Because of the nature of the QMF structure, the

filter banks can be implemented by a computation saving algorithm which is known as the polyphase QMF algorithm as represented in Figure 3.3b. This section presents a computational structure for the sub-band analysis and synthesis QMF filters.

3.3.1.1 Polyphase Realisation of the QMF Network

To simplify the analysis, a two band splitting method is first considered. The basic two band design is shown in Figure 3.4a. The input signal x(n) is divided into two equally spaced frequency bands by the filters whose impulse responses are $h_1(n)$ and $h_2(n)$ respectively. According to the design requirements for QMF by Esteban [32], these are FIR filters, that is, $h_1(n)$ and $h_2(n)$ are zero for $n < 0$ and $n > N_t-1$, where N_t is the number of taps in filters, and N_t has to be an even number. Both $h_1(n)$ and $h_2(n)$ are obtained from a low-pass prototype design h(n), such that

$$h_1(n) = h(n) \qquad \text{for } n = 0,1,\ldots\ldots,N_t-1 \qquad (3.7)$$

$$h_2(n) = (-1)^n h(n).$$

Thus, the design of one low pass filter h(n) specifies all the coefficient values in the encoder and decoder filters.

If the outputs of the filters are denoted by $w_1(n)$ and $w_2(n)$, then

$$w_1(n) = \sum_{m=0}^{N_t-1} h(m)\, x(n-m) \qquad (3.8)$$

$$w_2(n) = \sum_{m=0}^{N_t-1} (-1)^m h(m)\, x(n-m) \quad .$$

These outputs are then down-sampled by decimating, taking one sample out of every two samples; hence the lower band and the upper band signal, $x_1(n)$ and $x_2(n)$ are expressed as

$$x_1(n) = w_1(2n) = \sum_{m=0}^{N_t-1} h(m)\, x(2n-m) \tag{3.9}$$

$$x_2(n) = w_2(2n) = \sum_{m=0}^{N_t-1} (-1)^m\, h(m)\, x(2n-m)$$

Equation (3.8) indicates that for each speech input, the analysis filters require $2.N_t$ multiplication and addition operations.

Taking advantage of the fact that the $h_1(n)$ and $h_2(n)$ filters are derived from a common low pass filter h(n), the polyphase structure can be derived from (3.8), which can be decomposed to

$$x_1(n) = f(n) + g(n) \tag{3.10}$$

$$x_2(n) = f(n) - g(n)$$

where

$$f(n) = \sum_{m=0}^{\frac{1}{2}N_t-1} h(2m)\, x[2(n-m)] \tag{3.11}$$

$$g(n) = \sum_{m=0}^{\frac{1}{2}N_t-1} h(2m+1)\, x[2(n-m)-1]$$

This analysis polyphase structure is shown in Figure 3.3b.

For the synthesis filter bank, its output is given by [32],

$$x'(n) = 2\,[\, y_1(n) - y_2(n)\,] \tag{3.12}$$

where

$$y_1(n) = \sum_{m=0}^{N_t-1} h(m)\, v_1(n-m) \tag{3.13}$$

$$y_2(n) = \sum_{m=0}^{N_t-1} (-1)^m\, h(m)\, v_2(n-m)$$

v(m) is the output of the sampling rate expander, and the transfer function of the expander can be written as

$$v_1(m) = x_1(m/2) \qquad m = 0, \pm 2, \pm 4, \ldots \tag{3.14}$$

$$= 0 \qquad \text{otherwise .}$$

Then, the polyphase realisation of the synthesis filter bank output is represented by

$$x'(2n) = 2 \sum_{m=0}^{\frac{1}{2}N_t-1} h(2m)\, f'(n-m) \tag{3.15}$$

$$x'(2n+1) = 2 \sum_{m=0}^{\frac{1}{2}N_t-1} h(2m+1)\, g'(n-m)$$

where

$$f'(n) = x_1(n) - x_2(n) \tag{3.16}$$

$$g'(n) = x_1(n) + x_2(n)$$

This 2-band synthesis network is depicted in Figure 3.3b.

It can be proved that when N_t is even, the total time delay, T_d of this 2-band splitting and merging system can be written as

$$T_d = (N_t - 1)\, T_s \tag{3.17}$$

where T_s is the speech sampling period.

3.3.1.2 Control and Data Flow of the 2-Band Polyphase Structure

As Figure 3.3b shows, the input signal to the polyphase structure is separated into two sets by a "commutator". Because the sampling rates of $x_1(n)$ and $x_2(n)$ are one half of the input sampling rate, the commutator is thus required to execute a two-cycle control operation to generate $x_1(n)$ and $x_2(n)$. Cycle zero is associated with even sampling times $n = 0,2,4,\ldots,$ and these samples are filtered by the odd tap coefficients filter, G(z). Then, cycle 1 is associated with odd sample times $n = 1,3,5,\ldots,$ and these samples are filtered by the even tap coefficient filter, F(z). At the end of the odd sample times, the sums and differences of the two filter outputs are taken to produce $x_1(n)$ and $x_2(n)$ as given by (3.9), where $x_1(n)$ and $x_2(n)$ are the lower band and upper band samples respectively. In this way, the computational load is shared equally between cycles. As the even or odd tap filters only have $N_t/2$ filter coefficients, on average each input, x(n) of the analysis filter bank needs $N_t/2$ multiplication and addition operations, which is approximately one fourth the computation required for the conventional structure.

For the synthesis filter bank, the sum and difference of the subband samples are first computed to produce inputs for the G(z) and F(z) filters, respectively. At cycle 0, the difference signal is filtered by even filter coefficients to reconstruct the even sample values of the output x'(0), x'(2), ... At cycle 1, the odd filter coefficients are used to compute odd samples of the output x'(1), x'(3), x'(5)..... . The flow chart of the 2-band QMF polyphase algorithm is given in Figure 3.4.

3.3.1.3 Sub-band Structure of APC-AB

For the APC-AB coder, the full-band speech is split into three sub-bands as shown in Figure 3.5. By extending the 2-band SBC concept, the three band system can easily be realised by cascading the 2-band SBC in a tree structure as shown Figure 3.6. Stage 1 QMF first splits s(n) into lower band and upper band, then stage 2 splits the lower band into band 1 and band 2. The upper band from stage 1 is called band 3, in which delay lines are added to compensate for the delay of the stage 2 QMF filters. From (3.17), stage 2 QMF introduces (N_t-1) samples delay. The delay lines included in the third band encoder and decoder are given by

$$D_1 = N_t/2 \qquad (3.18)$$

$$D_2 = N_t/2 - 1$$

In this case, alias-free reconstruction is achieved without noticeable degradation.

In the codec simulation, 32-tap FIR filters are used in both stages. The filter coefficients were obtained from [33]. As the FIR filter is a linear phase filter, the overall phase response of the entire system is always linear. The entire time delay, T_{dt}, of a multi-band QMF system can be shown as

$$T_{dt} = (N_t-1)\,(2^{\ell}-1)\,T_s \qquad (3.19)$$

where ℓ is number of 2-band QMF stages and it is assumed that N_t tap FIR filters are used for all stages. For the 2-stage 3-sub-band system, the overall system delay due to the QMFs is $93T_s$.

3.3.2 PITCH DETECTION

Pitch detection is used in many types of low bit rate speech coder. Its function is to estimate the pitch period of segments of voiced speech and to calculate the pitch gain from which a voiced/unvoiced decision can be made. For the APC-AB coder, accurate pitch detection is very important as the value obtained will be used by the long term prediction filter to remove the pitch structure from the speech. The estimated pitch period will also be used for the pitch synchronous bit allocation scheme and hence accurate detection is needed to achieve maximum performance from the scheme.

3.3.2.1 Autocorrelation Method

Many methods can be used to estimate the pitch period [37]. A reliable method for pitch extraction in a noisy environment is to examine the signal's short term autocorrelation function (ACF) [38,39]. Generally, the ACF of a block of speech containing k samples is calculated for time shifts of between 2.5 and 20 msec which covers the expected range of speech pitch periods from 50 Hz to 400 Hz, accommodating both low pitch male speech and high pitch female speech. The main peak in the function is detected, which should correspond to the pitch period of the speech segment.

Now, if the speech sampling rate is 6.4 kHz, the estimated pitch period, M, can be written as [40],

$$M = \arg \left[\max_{16<I<128} \rho(I) \right] \qquad (3.20a)$$

where the normalised autocorrelation function is:

$$\rho(I) = \frac{\sum_{j=1}^{k} s(j)\, s(j-I)}{\left[\sum_{j=1}^{k} s(j-I)^2 \sum_{j=1}^{k} s(j)^2 \right]^{\frac{1}{2}}} \qquad (3.20b)$$

M is a value of I for which the normalised ACF, $\rho(I)$, is maximum. k (=128) is the number of samples in a 20 msec speech frame, and no samples from the previous frame are involved in this ACF calculation. The pitch gain is defined as

$$b = \frac{\sum_{j=1}^{k} s(j)\, s(j-M)}{\sum_{j=1}^{k} s(j-M)^2} \qquad (3.21)$$

For voiced speech, b will be close to unity, but for noise-like unvoiced speech which has little correlation M samples apart, b will be near to zero.

Calculating the ACF is an arithmetic-intensive procedure, and may cause overflow problems in fixed point aritnmetic. A major difficulty with this method is that the periodicities of the formants may be confused with the pitch [38]. In this case, substantial peaks of autocorrelation are not clearly seen and this may lead to a wrong pitch detection.

The pitch detection algorithm described above was used in the original APC-AB system proposed by Itakura & Honda [31]. However, in early experiments carried out by the author, it was found to be insufficiently reliable for the high demands of the pitch synchronous bit allocation scheme, owing to the difficulties, referred to above, of confusing first formant and pitch period. A more reliable pitch detection algorithm is

then required. Methods of reducing the effects of formants to make the pitch periodicity more prominent have been proposed [e.g.40] using spectrum flattening [40]. One of these techniques using the ACF is suggested by Sondhi [41,42]. This method involves centre clipping of the input signal against adaptive thresholds to quantize it to values of -1, 0 and +1. ACF calculation on these quantized values allows fixed point arithmetic to be used more accurately and probably simplifies the multiplication operations of the ACF computation [42]. This method can be modified by using decimation/interpolation to further reduce the required computation. This method was used for the APC-AB speech codec and is described in the following section.

3.3.2.2 Modified Autocorrelation Method

A block diagram of the modified Sondhi's algorithm is shown in Figure 3.7. The input speech is first lowpass filtered by a 9-tap FIR filter to reduce the effects of the higher formant structure on the ACF. The filter cut-off should be above the maximum pitch frequency which is 600 Hz in the codec design.

The filtered signal d(n) is then down sampled by a $\frac{1}{2}$-rate decimator. In this case, the ACF computation is reduced by half. The output of the decimator can be expressed as

$$d'(n) = d(2n) \qquad n = 1,2,.....,k/2. \qquad (3.22)$$

Over the long term, speech should be a zero mean process, but with short term measurements, considerable bias can exist. This bias can lead to shape distortion of the desired ACF and this will result in wrong pitch

period estimation. Hence, the mean should be extracted before the ACF calculation, and the mean extracted signal, $d_m(n)$ is written as

$$d_m(n) = d'(n) - m' \qquad n=1,2,.....,k/2. \quad (3.23a)$$

$$m' = \frac{1}{k/2} \sum_{i=1}^{k/2} d'(n) \quad (3.23b)$$

The higher formants are eliminated by the filter h(n), but the first formant often exists between 200 Hz and 1 kHz. The method to remove the effect of the first formant prior to the ACF calculation is called centre-infinite clipping. The clipping level is set at 60 percent of the smaller of the peak absolute sample value as measured in the first one-third and the last one-third portion of the current speech frame, that is

$$C_1 = 0.6 \text{ Min } [\underset{1\le n\le k/6}{\text{Max}}(|d_m(n)|), \underset{k/3\le n\le k/2}{\text{Max}}(|d_m(n)|)] . \quad (3.24)$$

Following the calculation of the clipping level, the clipped signal, c(n),is calculated as

$$c(n) = \begin{cases} 1 & : \text{ if } d(n) > C_1 \\ 0 & : \text{ if } |d(n)| < C_1 \\ -1 & : \text{ if } d(n) < -C_1 \end{cases} \quad (3.25)$$

where $n = 1,2,....k/2$. The nonlinear process preceding the ACF computation is done to approximately flatten the signal spectrum, thereby enhancing the periodicity of the signal. This effect is shown by comparing the power spectrum of the original signal and the clipped signal [43].

Due to the down sampling process, the ACF is then searched for its maximum value within the time delay from 3200/400 (= 8) to 3200/50 (= 64). The search process is presented as

$$M_d' = \arg \{ \underset{9<m<63}{\text{Max}} \rho'(m) \} \quad (3.26a)$$

$$\rho'(m) = \sum_{i=1}^{\frac{1}{2}k-m+1} c(i)\, c(i+m-1) \qquad m = 1,2,.....k/2. \quad (3.26b)$$

In this case, $\rho'(1)$ is the ACF at zero delay. After the search procedure, M_d' is extracted. However, as the time scale is compressed by a factor of two, M_d', is not the required pitch period. The 'decimated' pitch period, M_d' should be re-scaled by multiplying the factor two.

A more accurate value for the pitch period may be obtained by re-scaling the time by a quadratic interpolation technique, which is described in Appendix 1, and the pitch period, M', can be estimated by the following rule :

$$M' = \begin{cases} 2\,M_d' + 1 & \text{if } 2.\Delta d \geq \Delta n \\ 2\,M_d' - 1 & \text{if } 2.\Delta d \leq -\Delta n \\ 2\,M_d' & \text{otherwise} \end{cases} \quad (3.27)$$

where $\Delta d = r(2M_d'-2) - r(2M_d'+2)$

$= \rho(M_d'-1) - \rho(M_d'+1)$

and $\Delta n = r(2M_d'-2) - 2.r(2M_d') + r(2M_d'+2)$

$= \rho(M_d'-1) - 2.\rho(M_d') + \rho(M_d'+1)$.

The estimated pitch period, M', is the distance between two peaks. Due to the conventional definition of the period of a signal, the pitch period is re-defined as :

$$M = M' - 1 \ . \quad (3.28)$$

Then the pitch gain is computed as follows,

$$b = \frac{\sum_{j=1}^{k-M} s(j)\, s(j+M)}{\sum_{j=1}^{k-M} s(j+M)^2} \quad . \tag{3.29}$$

Referring to (3.29), the pitch gain is related to the autocorrelation of the speech signal. Usually, the value of b would be higher for the quasi-periodic voiced signal than the noise-like unvoiced speech signal, so that a voiced and unvoiced decision can be made from this value.

3.3.2.3 Illustrative Measurements of Pitch Detection

Figure 3.8 illustrates an example of pitch detection on a typical frame of voiced speech, and the ACF of this speech segment is shown in Figure 3.8b. In this case, the ACF peaks due to the vocal tract response are bigger than the peak due to the periodicity of the vocal excitation so that the periodicity due to the pitch pulses are not clearly shown. Thus, a simple procedure of picking the largest peak in the ACF would fail to indicate the correct pitch period in such a case. If the original signal is downsampled and centre-infinite peak clipped as drawn in Figure 3.8c, the ACF of this processed signal shown in Figure 3.8d is significantly different from the shape of the ACF as in Figure 3.8b. The multi-peak structure is enhanced on the new ACF. Thus, a more reliable pitch period estimation can be obtained by the peak searching process.

Figure 3.9b and 3.9c show pitch period and pitch gain contours as obtained using the pitch detection algorithm detailed above for the segment of speech illustrated in Figure 3.9a. These diagrams provide a visual impression of the continuity of the pitch period from one frame to the

next for voiced speech. During voiced speech, the pitch period contour is relatively continuous and the pitch gain is often above 0.2. For unvoiced speech, the estimated pitch period is random, and the pitch gain is often below 0.2.

Furthermore, histograms of pitch period and pitch gain for a 30 sec segment of low voice male and a 30 sec segment of high pitched female speech are shown in Figure 3.10. As expected, the pitch period of female speech is concentrated in the small pitch period region, which corresponds to a higher pitch voice. The pitch period for the male speech is lower. From the pitch gain histogram, the values of pitch gain are bounded within -0.5 to 1.5 and two humps appears in the histograms. The right hump is due to the voiced speech and the left hump is due to the unvoiced speech. In calculating the bit allocation for the residual signal, the algorithm should be different for voiced and unvoiced speech as described in Section 3.2.3.. A threshold value of 0.2, which is a value between the two humps, has been chosen as the voicing decision for the bit allocation scheme.

3.3.3 ADAPTIVE PREDICTIVE CODING (APC)

After sub-band splitting, the sub-band signal is digitally encoded for transmission, but this process causes quantization error in the reconstructed signal. For most digital signal processing systems, the signal to quantization noise ratio, SNR, is a useful performance criterion. In general, if the dynamic range of the quantizer is set only to enclose the statistical maximum and minimum values of a signal, the quantization noise power would be larger for a large variance signal than a signal with small variance [50].

Speech that is sampled at the Nyquist rate exhibits very significant correlation between successive samples. One consequence of this correlation is that the variance of the first difference

$$e_{n1} = x_n - x_{n-1} \tag{3.30a}$$

is smaller than the variance of the speech, x_n, itself. With basic statistical notation

$$\begin{aligned} \langle e_{n1}^2 \rangle &= \langle (x_n - x_{n-1})^2 \rangle \\ &= \langle x_n^2 \rangle + \langle x_{n-1}^2 \rangle - 2\langle x_n x_{n-1} \rangle \\ &= \langle x_n^2 \rangle \, [2(1-C_1)] \end{aligned} \tag{3.30b}$$

where C_1 is the correlation coefficient between the successive samples, and is defined as

$$C_1 = \frac{\langle x_n x_{n-1} \rangle}{\langle x_n^2 \rangle} \tag{3.30c}$$

If C_1 is greater than 0.5 as for voiced speech, e_{n1} has smaller variance than x_n. As a result, it is advantageous to quantize e_{n1} instead of x_n, and this permits a smaller quantization error variance. Thus, a better SNR is achieved. At the receiving end, the speech can be reconstructed by adding up the received signal differences.

APC is an efficient speech coding method which uses linear prediction analysis based on a speech prediction model [44]. This method operates by predicting each input sample from the past history of the waveform. The difference between each input sample and its predicted value can be very efficiently encoded and sent to the receiver. The difference signal is sometimes referred to as the residual signal. To illustrate the concept of linear prediction take simple case where the prediction of x_n is derived from the previous sample, x_{n-1}, that is the prediction is:

$$x_n' = a\, x_{n-1} . \tag{3.31}$$

x_n' can be considered as a predictor output and a is the predictor coefficient. The variance of the residual signal would become

$$\begin{aligned} \langle e_n^2 \rangle &= \langle (x_n - a.x_{n-1})^2 \rangle \\ &= \langle x_n^2 \rangle (1+a^2-2aC_1) \quad . \end{aligned} \tag{3.32}$$

where C_1 is as defined in (3.30c). The minimum value of $\langle e_n^2 \rangle$ would be obtained when $a = C_1$, giving

$$\langle e_n^2 \rangle_{min} = \langle x_n^2 \rangle (1-C_1^2) \quad . \tag{3.33}$$

For values of C_1 greater than 0.5, $\langle e_n^2 \rangle_{min}$ is always smaller than $\langle e_{n1}^2 \rangle$ as given by (3.30b). In order to use APC coding effectively, a higher order predictor must be used (as will be outlined in a subsequent section) and the predictor coefficients must be changed periodically and sent to the receiver for reconstructing the original speech. In this way, the encoder can adaptively keep track of the changing properties of the speech, and make the residual encoder more efficient.

In fact, this can be considered as a redundancy removal process. Before quantization, spectral features are removed and this results in a residual signal which has a smaller dynamic range than the input speech. Thus, the system performance can be improved either by finer quantization of the residual to improve the SNR of the speech, or by using fewer bits/sample to encode the residual to achieve the same SNR of the speech.

For the APC-AB coder, the speech redundancy is removed by a two-level process which is carried out by two predictors as shown in Figure 3.11a; the first residual is produced by a long term predictor, C(z), which removes redundancy due to the vocal tract excitation, and the second residual is produced by a short term predictor, A(z), which removes redundancy contributed by the vocal tract shape. The predictor coefficients are updated regularly every 20 msec to improve the system performance.

3.3.3.1 Long Term Prediction

The first level of prediction is based on the redundancy due to the quasi-periodic nature of the vocal tract excitation signal. Such redundancy can be removed using a long term predictor which produces a coarsely predicted signal

$$x_n' = \beta\, x_{n-M} \tag{3.34}$$

where M is a variable delay , and β is the gain of the predictor. Subtracting x_n' from the current speech sample x_n generates the first residual whose n-th sample is given by

$$e_{1n} = x_n - x_n' \,. \tag{3.35}$$

The goal is to select M and β so that the time averaged prediction error, E, over a block of N samples is minimised, where E is given by

$$E = (1/N) \sum_{n=1}^{N} (x_n - \beta.x_{n-M})^2 = \langle (x_n - b.x_{n-M})^2 \rangle \,. \tag{3.36}$$

By partially differentiating E with respect to β and setting the result to zero:

$$\frac{\partial E}{\partial \beta} = -2\langle (x_n - \beta.x_{n-M})\, x_{n-M} \rangle = 0 \,. \tag{3.37}$$

Hence, $$\beta = \frac{\langle x_n\, x_{n-M} \rangle}{\langle x^2_{n-M} \rangle} \,. \tag{3.38}$$

By substituting β into (3.36), M should be selected to minimise the error E in the given time frame. The optimal value of M is usually equal to the pitch period for voiced speech, and the corresponding value of β is the pitch gain as given in (3.29).

When long term prediction is applied to the APC-AB coder, it is assumed that the gain changes in each sub-band are the same, and that the

sub-band pitch period ,M_i, is proportional to the full band pitch period M as :

$$M_i = M \omega_i \tag{3.39}$$

where ω_i is the i-th sub-band to full band bandwidth ratio. As the values of M and β are obtained from the pitch detection algorithm as described in Section 3.3.2.2, the long term error minimisation calculations do not need to be carried out in each sub-band. Then, the sub-band long term predictor transfer function in the z domain can be written as :

$$C_i(z) = 1 - \beta z^{-M_i} . \tag{3.40}$$

In order to ensure the stability of the pitch predictor, the value of the pitch gain is set within the range of 0 to 1.

In general, M_i may be a non-integer value. The digital sub-band must be therefore interpolated, using a technique known as digital phase shifting [45], to obtain samples at non integer sample times if prediction is to be performed. This technique is based on the idea of first interpolating the signal to a high sampling rate, then using an integer delay, and finally decimating the signal back to the original sampling rate. Figure 3.12b illustrates the basic operation of the phase shifter. The input to output transfer function can be derived from the following relationships :

$$V(z) = X(z^D) \tag{3.41}$$

$$W(z) = H(z)\, z^{-(M-I)} \tag{3.42}$$

$$Y(z) = \sum_{\ell=0}^{D-1} W(e^{-j2\pi\ell/D}\, z^{1/D}) \tag{3.43}$$

where D is the ratio of the high sampling rate to the input signal sampling rate. H(z) is an linear phase FIR band pass filter and it has an

overall delay of I samples. If H(z) has a unity pass-band gain and removes all but the ℓ-th band, only the ℓ-th term in (3.43) is significant. Hence,

$$H(e^{j\omega}) \simeq \begin{cases} e^{-j\omega I} & : \omega \text{ in } \ell\text{-th band} \\ 0 & : \text{otherwise} \end{cases},$$

and

$$Y(e^{j\omega}) = W(e^{j(\omega+2\pi\ell)/D}) \quad . \tag{3.45}$$

Whence,

$$\frac{Y(e^{j\omega})}{X(e^{j\omega})} \simeq e^{-j\omega M/D}\, e^{-j2\pi iM/D} = e^{j(\omega\theta+\phi)} \tag{3.46}$$

The term $\theta = -M/D$ denotes the phase slope and the term $\phi = -2\pi\ell M/D$ denotes the phase offset due to integer band sampling.

For the three sub-band APC-AB coder, D is given by

$$D_i = 1 / \omega_i \tag{3.47}$$

Figures A.2.1 (a), (b) and (c) in Appendix A show the gain responses of the linear phase FIR filters of the interpolation/decimation process used in the phase shifter network. These filters have a large number of taps and make the coder rather complex. A more efficient way of predicting a non-integer number of samples ahead is to first remove correlation for integer delay at a low sampling rate and then to implement the fractional delay predictor with a phase shifter filter. The simplified structure is shown in Figure 3.12c, and this can be related by

$$y_{in} = \sum_{m=0}^{Q_i} h_i(mI_i - m_{0i})\, x_i(n-m-m_{1i}) \tag{3.48}$$

where $Q_i = (N_{ti}-1)/D_i$

$m_{0i} = \text{Mod}(M, D_i)$

$m_{1i} = \text{Int}(M/D_i) - D_i.(N_{ti}-1)/2$

N_{ti} = number of taps for the i-th sub-band phase

I_i = group delay of the i-th sub-band phase shifter filter.

m_{1_i} is the integer delay at the low sampling rate. As m_{1_i} has to exceed zero, an important constraint that must be met is that

$$N_{ti} \leq 2.M_{min} + 1 \tag{3.49}$$

where M_{min} is the minimum allowed pitch delay. If $h_i'(k) = h_i(kI_i - m_{0_i})$, Equation (3.40) becomes

$$C_i(z) = 1 - \sum_{k=0}^{Q_i} b_i(k)\, z^{-k-m_{1_i}} \tag{3.50}$$

where $b_i(k) = b\, h_i'(k)$.

In this way, if $N_{ti} = 32$ and $D_i = 4$, only 8 multiplications per sample at the sub-band sampling rate are required to implement the pitch predictor.

3.3.3.2 Short term Prediction

In addition to long term prediction, short term prediction is also used to further reduce the speech redundancy due to the vocal tract shape. The short term predictor is derived from an all-pole model of the vocal tract transfer function. It is assumed that the 1st residual sample e_{1n} can be approximately predicted as a linear combination of a number of immediately preceding samples. To simplify the mathematical notation, let e_{1n} be denoted by x_n. Hence, the predicted first residual sample can be expressed as:

$$x_n' = - \sum_{k=1}^{p} a_k\, x_{n-k} \tag{3.51}$$

where a_k, $1 \leq k \leq p$, are real constants known as linear predictive (LP) coefficients, and p is the order of the predictor which should be ideally equal to at least twice the number of speech formants within the speech

transmission bandwidth. The second residual is the error between the actual and the predicted first residual given by

$$e_{2n} = x_n - x_n' \quad . \tag{3.52}$$

e_{2n} is obtained by passing x_n through a filter A(z), where A(z) is given by

$$A(z) = 1 + \sum_{k=1}^{p} a_k z^{-k} \quad . \tag{3.53}$$

Now, defining the vectors

$$\underline{x}^T = (x_{n-1} \; x_{n-2} \; \ldots\ldots\ldots \; x_{n-p})$$

$$\underline{a}^T = (a_1 \; a_2 \; \ldots\ldots\ldots \; a_p) \quad .$$

Then, Equation (3.52) becomes

$$e_{2n} = x_n + \underline{a}^T \underline{x} \tag{3.54}$$

The linear predictive (LP) coefficients are obtained by minimising the mean square value of the residual signal with respect to $\underline{a}$ over a block of speech. The squared 2nd residual error (m.s.e.) is

$$e_{2n}^2 = x_n^2 + 2 \, \underline{a}^T \underline{x} \, x_n + \underline{a}^T \underline{x} \quad \underline{x}^T \underline{a} \quad . \tag{3.55}$$

Taking expectations, the m.s.e. may be written as

$$\langle e_{2n}^2 \rangle = r_0 + 2 \, \underline{a}^T \underline{p} + \underline{a}^T R \, \underline{a} \tag{3.56}$$

where $r_i = \langle x_n \, x_{n-i} \rangle \qquad i = 0,1,\ldots,p$

$$\underline{p}^T = (r_1 \; r_2 \; \ldots\ldots\ldots \; r_p)$$

where

$$R = \begin{bmatrix} r_0 & r_1 & \ldots\ldots\ldots\ldots & r_{p-1} \\ r_1 & r_0 & \ldots\ldots\ldots\ldots & r_{p-2} \\ . & . & . \; . \; . \; . \; . \; . \; . \; . & . \\ . & . & . \; . \; . \; . \; . \; . \; . \; . & . \\ . & . & . \; . \; . \; . \; . \; . \; . \; . & . \\ r_{p-1} & r_{p-2} & \ldots\ldots\ldots\ldots & r_0 \end{bmatrix}$$

To find the minimum m.s.e., consider the effect of changing $\underline{a}$ to $\underline{a} + \partial\,\underline{a}$ in (3.56), then

$$\langle \partial e_{2n}^{\,2} \rangle = 2\,\partial\,\underline{a}^T\underline{p} + \partial\,\underline{a}^T R\,\underline{a} + \underline{a}^T R\,\partial\,\underline{a} \,. \tag{3.57}$$

As R is symmetric,

$$\langle \partial e_{2n}^{\,2} \rangle = 2\,\partial\,\underline{a}^T(\,\underline{p} + R\,\underline{a}\,) \tag{3.58}$$

and it follows that $e_{2n}^{\,2}$ is minimum when

$$\underline{a} = -R^{-1}\underline{p} = \underline{a}_{opt} \tag{3.59}$$

A fixed filter with these coefficient will give the minimum m.s.e. given by

$$\langle e_{2n}^{\,2} \rangle = r_0 - \underline{p}^T R^{-1}\underline{p} \tag{3.60}$$

The well known autocorrelation method [44] may be used to compute the predictor coefficients. In this method, each complete speech frame is used for analysis, and the signal is Hamming windowed so that the signal becomes zero outside the window. Thus,

$$x_n \equiv \begin{cases} x_n\,\omega_n & 0 \le n \le N-1 \\ 0 & \text{otherwise} \end{cases} \tag{3.61}$$

where N is the number of samples in a frame and ω_n is the Hamming window function. The autocorrelation function is then given by

$$r_i = \sum_{n=0}^{N-i-1} x_n\,x_{n+i} \,. \tag{3.62}$$

Substituting the auto-correlation values in (3.59), the LP coefficients can be calculated by inverting R.

The calculated LPC coefficients could be encoded and sent to the receiver. At the receiver, the decoded coefficients would then be used for by an all pole synthesis filter

$$H(z) = 1 \,/\, A(z) \tag{3.63}$$

to reconstruct the first residual signal. However, these LPC coefficients do not make good transmission parameters because quantization or channel

errors can cause the poles of the synthesis filter, H(z), to move outside the unit circle in the z-plane, causing unstable waveforms. An alternative approach is to use PARCOR coefficients, which can be calculated from the LP coefficients. These PARCOR coefficients can be efficiently encoded using an inverse-sine transformation [46,47,48], and the predictor stability is assured if the magnitudes of the PARCOR coefficients are less than unity. Recently, a new set of parameters known as Line Spectrum Pair (LSP) coefficients have been proposed by Itakura [57]. LSP coefficients are an alternative way of representing the predictor filter, which are less susceptible to quantization error than PARCOR coefficients and hence more efficient for encoding.

LSP coefficients are closely related to the zeros of A(z) but are much easier to calculate. For a p-th order predictor, the LSP coefficient are labelled θ_1, ω_1, θ_2, ω_2, ..., $\theta_{p/2}$, $\omega_{p/2}$ assuming p is even. They can be computed from the LP coefficients and, by the "LSP theorem" [57], the LSP coefficients for a stable predictor must obey,

$$0 < \theta_1 < \omega_1 < \theta_2 < \omega_2 \ldots\ldots\ldots\ldots < \theta_{p/2} < \omega_{p/2} < \pi. \qquad (3.64)$$

Further details of this theorem will be discussed in Chapter 4.

For the APC-AB coder, a 4th order short term predictor is applied to each of the three sub-bands. The LP coefficients are first derived for each sub-band frame containing N_{si} samples, where

$$N_{si} = N_T \omega_i \qquad (3.65)$$

N_T is the number of full band speech samples and ω_i is the sub-band ratio. The LP coefficients are then transformed to LSP coefficients using a method that will be described in Section 4.2.4. When the coder is used at 16 kb/s, the LSP coefficients are encoded by 3 bits per coefficient.

Although the LSP coefficients lie in the finite interval $(0,\pi)$, it is inefficient to have the quantizer dynamic range across the whole of this interval for each coefficient. To optimally design the quantizer for each coefficient, statistics and histograms of individual LSP coefficients are computed and plotted for determining reasonable ranges for each coefficient. The results are illustrated in Figure 3.14, and the maxima and minima of the coefficient distributions are given in Table 3.1 (see Figure 3.13) along with their means and the standard deviations. All the LSP coefficients appear to be approximately truncated Gaussian in distribution, and their variances are almost the same. Before quantization, a zero mean and unit variance process is applied to each coefficient in which each coefficient is first subtracted from the corresponding mean and then divided by the corresponding standard deviation so that each coefficient should have zero mean and unit variance. The LSP coefficients are then uniformly quantized by the Max [34] unit variance quantizer, which has been specifically designed to minimise the quantization noise for a unit variance Gaussian input signal.

It was found that a straight 3-bit quantization per LSP coefficient would sometimes make the quantized LSP set fall into unstable ordering not satisfying the condition given in (3.64). The reason can be seen from Figure 3.14 which shows that the distributions of LSP coefficients partly overlap each other. Therefore the dynamic ranges of the quantizers will also overlap each other. In cases, where the difference between two successive LSP coefficients is very small, the quantized LSPs may overlap to produce an unstable order. When this happens various remedies are available for keeping the synthesis filter stable. For example, the receiver could refer to and perhaps re-use the previous stable LSP coeffi-

cient set if the current quantized LSP coefficient set is unreliable. Using this approach, the system performance will be degraded, especially when an unstable condition happens at a transition region of voiced to unvoiced speech or the onset of a period of silence. It is desirable, therefore, to find another more efficient quantization scheme for LSP coefficients.

An improved LSP quantization scheme proposed by Soong and Juang [49] is now investigated. This method uses differential encoding to encode the LSP differences so that the decoded LSP coefficients can not fall into unstable ordering. For the same speech data base as was used to produce Figure 3.14, the distribution of the differences between adjacent LSP coefficients are plotted in Figure 3.15. The maxima, the minima, the mean and the standard deviation of the LSP coefficient differences are given in Table 3.2 (see Figure 3.13). A simple encoding strategy as shown in Figure 3.16 is used to encode the LSP coefficient differences. The process referred to earlier for producing zero mean and unit variance is applied to each LSP coefficient difference which may then be quantized by the Max uniform quantizer. During quantization, the quantization error is effectively fed back to the input. To avoid the possibility of a negative difference signal, which would cause instability in the reconstructed LSP coefficients, the absolute values of the LSP differences are quantized. This guarantees that the decoded LSP coefficients are always in the correct order for stability.

In simulation, each LSP coefficient difference was uniformly quantized by 3 bits, and no instability condition was observed. This scheme was therefore adopted to encode the LSP coefficients on the APC-AB coder.

3.3.4 ADAPTIVE RESIDUAL QUANTIZATION

This section describes the residual quantization scheme for the APC-AB speech coder which synchronously allocates more bits to the high energy samples in the second residual than to the lower energy samples. As the bit allocation scheme is derived from the second residual in each sub-band, the second residuals (unquantized versions) must first be computed by passing the sub-band signals through the predictors as shown in Figure 3.11a. The bit allocation is then carried out in two stages: the first stage to compute the segmental residual energies, and the second stage to calculate the bit allocation scheme for the residual in each segment.

3.3.4.1 Segmental Energy Calculation

The second residual of each sub-band is split into a small number, L, of time-domain segments. The way this is done depends on whether the speech is voiced or unvoiced. The voicing decision is derived from the pitch gain: if the pitch gain is less than 0.2, the speech frame is considered as unvoiced speech otherwise it is voiced. When the speech is considered unvoiced, each sub-band frame is divided into L equal segments according to a fixed strategy. For voiced speech, the second residual of each sub-band frame is split into segments whose definitions are determined by the pitch period and the position of the highest energy parts of the signal. The strategy is to divide each sub-band pitch period, M_i, into L approximately equal length portions and to categorise each portion as belonging to a particular segment depending on its position within the pitch period, i.e. all the first portions form segment one,

the second portions form segment two, etc. L=4 was chosen for the 16 kb/s APC-AB coder and an example of how a particular block of voiced speech is partitioned is illustrated in Figure 3.17.

As M_i/L may not be an integer, the length of 1st to (L-1)th segment is set to

$$I_s = \left\lfloor \frac{Mi}{L} \right\rfloor + \left\lfloor \frac{Mod(M_i,L)}{L-1} \right\rfloor \tag{3.66}$$

The square bracket $[\bullet]$ implies the integer value of $\bullet$ The length of the L-th segment is

$$I_\ell = M_i - (L-1).I_s. \tag{3.67}$$

The segments at the beginning and at the end of the frame may have fewer than I_s or I_ℓ samples.

Each segment is indexed with a number from 1 to L, and the residual energy in each segment, u_{ij}, is calculated, where i=1,2,....,N, j=1,2,...,L, and N is the number of sub-bands. For convenience, u_{i1} is chosen in such a way that it always has the highest energy. A delay measurement known as the segment location or the sub-interval position, T_{di}, is used to indicate the location of the very first segment from the beginning of the frame. This segmentation arrangement is better understood using set representation. Let $N_j(M_i,T_{di})$ be the set contain the j-th segment residual in the i-th sub-band within $N_j(M_i,T_{di})$. T_{di} was given in (3.1) and T_{d1} is obtained via a searching process given by

$$T_{d1} = \arg\left[\max_{0\le m<M_1} u_{11(m)} \right] \tag{3.68}$$

and

$$u_{11}(m) = \frac{1}{|N_1(M_1,m)|} \sum_{n \in N_j(M_i,T_{di})} e_1^2(n) \qquad (3.69)$$

T_{d1} should be within the range from 0 to M_1, and $e_1(n)$ is the first band residual. Having obtained the segment positions, the residual energy per sample of the other segments can be computed as

$$u_{ij} = \frac{1}{|N_j(M_i,T_{di})|} \sum_{n \in N_j(M_i,T_{di})} e_i^2(n) \qquad (3.70)$$

T_{d1} is transmitted to the receiver to recover the segmentation. As the maximum value of M_1 is 32, this can be noiselessly encoded by 5 bits. In the following, an optimal bit allocation scheme is developed to dynamically allocate the quantization bits to the residual according to the energy distribution among u_{ij}.

3.3.4.2 Rate Distortion Function

The bit allocation in each frame is computed to minimise the speech coder distortion. Such distortion can be measured as the mean square error between the encoder input, s_n, and the decoder output s_n', which is given by

$$D = \langle (s_n - s_n')^2 \rangle . \qquad (3.71)$$

With the assumption that the sub-band splitting and merging does not cause any distortion the coder distortion may be expressed in terms of the sub-band signals. Equation (3.71) can then be written as

$$D = \sum_{i=1}^{N} \langle (x_{in} - x_{in}')^2 \rangle . \qquad (3.72)$$

where x_{in} and x_{in}' are the outputs from the QMF analysis filter bank and the decoded sub-band signals respectively. As the sub-band signal is en-

coded by the segmentation quantization scheme, the error in the decoded sub-band signal is equal to the sum of the quantization error in each segment. Let D_{ij} be the distortion per sample in segment j for subband i, then the coder distortion can be written as

$$D = \sum_{i=1}^{N} \sum_{j=1}^{L} c_j D_{ij} \tag{3.73}$$

where $c_j = N_j(M_i, T_{di}) / N_{si}$ (3.74)

and N_{si} is the sub-band frame length.

It can be shown that [33] the distortion, σ_q^2, of a quantizer, of bit rate R can be expressed as

$$\sigma_q^2 = k.2^{-2R} \sigma_x^2 \tag{3.75}$$

where σ_x^2 is the mean square value of the quantizer input, and k is a term which is dependent on the pdf of the quantizer input. Using this relationship, the coder distortion is

$$D = \sum_{i=1}^{N} \sum_{j=1}^{L} k.\, 2^{-2R_{ij}}\, c_j\, u_{ij} \tag{3.76}$$

where R_{ij} is the number of bits allocated to each segment j in subband i. The bit allocation of the speech coder is computed to minimise the distortion, D, with the constraint of a given average bit rate given by

$$R = \sum_{i=1}^{N} \sum_{j=1}^{L} \omega_i\, c_j\, R_{ij} \,. \tag{3.77}$$

Using the Lagrange multiplier optimisation method [68], a Lagrange function, F, is defined as the sum of the distortion, D, and the product of a constant λ with a 'constraint function', which becomes zero when (3.77) is satisfied:

$$F = \sum_{i=1}^{N} \sum_{j=1}^{L} k.\, 2^{-2R_{ij}}\, c_j\, u_{ij} + \lambda\left(R - \sum_{i=1}^{N} \sum_{j=1}^{L} \omega_i\, c_j\, R_{ij}\right) \,. \tag{3.78}$$

The constant λ is known as the Lagrange multiplier. The necessary conditions for a constrained minimum of D can be obtained by differentiating F with respect to R_{ij} and setting the result to zero, i.e.,

$$\frac{\partial F}{\partial R_{ij}} = 0 \tag{3.79}$$

so that

$$R_{ij} = (1/2)\log_2(\tilde{u}_{ij}/\lambda) + (1/2)\log_2(-2.K.\ln 2) \tag{3.80}$$

where $\tilde{u}_{ij}$ are the normalised segmental residual energies ($=u_{ij}/\omega_i$). Substituting (3.80) into (3.77) and using the fact that ω_i and c_j are normalised to 1,

$$a^{\left(\sum_{i=1}^{N}\sum_{j=1}^{L}\omega_i c_j\right)} = \prod_{i=1}^{N}\prod_{j=1}^{L} a^{\omega_i c_j} = a .$$

Whence

$$R = \frac{1}{2}\left[\log_2 \prod_{i=1}^{N}\prod_{j=1}^{L} \tilde{u}_{ij}^{\omega_i c_j} + \log_2(1/\lambda) + \log_2(-2k.\ln 2)\right] . \tag{3.81}$$

Combining (3.80) and (3.81), the optimum bit allocation can be obtained as

$$R_{ij} = R + \frac{1}{2}\log_2\left[\frac{\tilde{u}_{ij}}{\prod_{n=1}^{N}\prod_{\ell=1}^{L}\tilde{u}_{n\ell}^{\omega_n c_\ell}}\right] . \tag{3.82}$$

3.3.4.3 Bit Re-assignment and Quantization

The bit allocation scheme [51] computed from (3.82) must be rounded to non-negative integers. To do this, a simple iterative algorithm is used to determine each value of IR_{ij}, i.e., the number of bits allocated to segment j in band i. The algorithm proceeds by setting IR_{ij} equal to $[R_{ij}+r]$, where $[\bullet]$ is the integer of $\bullet$, and r is the incremental bit rate which is adaptively changed for each iteration until all the available bits are allocated. This bit re-assignment algorithm must satisfy the following constraints :

$$\begin{aligned} &\text{(i)} && IR_{ij} \geq R_{min,ij} \geq 0 \\ &\text{(ii)} && IR_{ij} \leq R_{max} \\ &\text{(iii)} && \sum_{i,j} w_i c_j IR_{ij} \leq R \end{aligned}$$

where $R_{min,ij}$ is the minimum bit rate for segment j in i-th band, and R_{max} is the maximum allowable bit rate.

Condition (i) ensures that IR_{ij} is an non-negative integer. In the iterative procedure, the bit allocation for segment-1 in each sub-band, IR_{i1}, is forced to be greater than or equal to one so that $R_{min,i1} = 1$. This constraint is perceptually important in the coded speech, as it effectively prevents a reduction in the bandwidth for the coded speech, and accordingly reduces spectral envelope distortion. For the case of $R_{ij} = 0$, all residual samples in the segment are quantized to zero.

To prevent the over design of a quantizer, condition (ii) sets the maximum allowable value of IR_{ij} equal to R_{max}(R_{max}=7 was used), on the assumption that if the bit rate is greater than R_{max} for a segment, the

enhanced quality of coding will be perceptually irrelevant. The bit assignment algorithm directs irrelevant bits to segments that can benefit from finer quantization.

Finally, condition (iii) guarantees that the transmission bit rate will not over-step the total number of bits per frame.

The iterative scheme aims to find a small value r, which may be added to all values of R_{ij} to ensure that when they are integerised, the mean bit rate is exactly equal to R. The routine starts by choosing r to be an arbitrary value 0.5. This value of r is added to R_{ij} for each time segment in each subband, the resulting values are integerised, and the mean bit rate is calculated using the formula (3.77). If this new bit rate is larger than R, r is decreased by a small amount Δr. If the new bit rate is smaller than R, r is increased by Δr. If the bit rate is equal to R, the routine terminates. At each iteration, Δr is halved so that at each iteration r will become closer to the required solution. A step by step description of the algorithm is given as follows:

Step 1. Initialization:-
Set r to 0.5, Δ to 0.5 and $R_{max} = 7$

Step 2. Calculation of new mean bit rate 'IR':-
Set IR = 0,
Repeat the following 4 steps for all values of i in the range 1 to N and for all values of j in the range 1 to L:-

(i) Set IR_{ij} equal to the integer of $(R_{ij}+r)$,

(ii) If $IR_{ij} < R_{min,ij}$ then $IR_{ij} = R_{min,ij}$.

(iii) If $IR_{ij} > R_{max}$ then $IR_{ij} = R_{max}$.

(iv) Add $\omega_{ij} c_j IR_{ij}$ to IR,

Step 3. If IR < R then add Δr to r,
divide Δr by 2 and go to Step 2,

Step 4. If IR > R then subtract Δr from r,
divide Δr by 2 and go to Step 2,

Step 5. Finish. Value of r is correct.

This iterative algorithm requires only a small number of iterations, typically about 4.

Having computed the bit allocation for a given frame, the residuals in each segment are uniformly quantized by Max quantizers. The step sizes for 1 to 8-bit quantizers are given in Figure 3.18 and a graph showing the quantization levels for a 3-bit quantizer is shown in Figure 3.19. In the current implementation, the residual e_n is quantized to a level L_n as follows:

$$\begin{aligned} L_n &= \text{Min}(\text{sign}(1,e_n),0) && e_n/\Delta < 1 \\ &= \text{Max}(\text{Min}([e_n/\Delta],\ell_n-1),-\ell_n) && e_n/\Delta \geq 1 \end{aligned}$$

where $[\bullet]$ stands for "integer part of $\bullet$", $\ell_n = 2^{R-1}$ is the number of levels above the d.c. level. L_n will converted to a binary representation before transmission. Then, the quantized value is

$$e_n' = \Delta.(L_n+0.5) \quad . \tag{3.83}$$

Figure 3.20 shows the bit allocation scheme adopted for a frame of voiced speech of duration 20 msec. It can be seen that the quantization bits are distributed non-uniformly over different speech intervals within each pitch period, as well as over the sub-bands. It is believed [31] that this bit allocation method will provide a better signal to quantization noise ratio as compared with what could be achieved at the same bit rate, using a less sophisticated quantization method.

3.3.5 TRANSFORM CODING

The segmental residual power, u_{ij}, must be sent to the receiver to re-compute the bit allocation scheme and the unit variance quantizer parameters for each segment in each sub-band. It was found that the logarithmic values of the normalised residual powers, u_{ij}', given by (3.5a) are highly correlated between segments. Histograms of u_{ij}' are plotted in Figure 3.21. Apart from u_{11}', the distribution curves of u_{ij}' lie in layers, which may imply some correlation between u_{ij}'s. The correlation matrix of u_{ij}', which indicates cross-correlation between u_{ij}'s, is given in Figure 3.22. It shows that in each sub-band the correlation among u_{ij}' is fairly high (>0.7). Thus, u_{ij}' are quantized by using a transform coding scheme to improve bit reduction efficiency. Transform coding requires a sequence of two operations. The first is a linear transformation that transforms a set of statistically dependent parameters into a set of "more independent" coefficients. The second operation is an individual quantising and coding of each of the coefficients. The number of bits required to code the coefficients is based on a distortion rate function which minimises the mean square quantization error over all the coefficients under a constraint of fixed bit rate. The following section describes the basic theory of transform coding and the bit allocation method for the transformed coefficients is also given.

3.3.5.1 Theory of One Dimensional Transformation

Let $\underline{x}$ be a vector containing m elements which are statistically dependent parameters:

$$\underline{x}^T = (x_1 \; x_2 \; \ldots\ldots\ldots\ldots \; x_m) \qquad (3.84)$$

Assume that each element has a variance σ_{xi}^2. Using transform coding, this vector $\underline{x}$ may be linearly transformed by an mxm matrix A to produce a vector $\underline{y}$:

$$\underline{y} = A\,\underline{x} \tag{3.85}$$

The elements of $\underline{y}$ are referred to as the transformed coefficients, each of these elements having a variance of σ_{yi}^2. Suppose that the transformed coefficients are quantized, thus leading to a vector $\underline{y'}$, transmitted to the receiver, and re-transformed using the inverse matrix A^{-1}, i.e.,

$$\underline{x'} = A^{-1}\underline{y'} \tag{3.86}$$

The vector $\underline{x'}$ is a reconstruction of vector $\underline{x}$, and it is desired to choose A to make $\underline{x'}$ as close as possible to $\underline{x}$.

In order to achieve ultimate coding efficiency, the variances σ_{yi}^2 are required to differ as much as possible [52]. This requirement is met by the use of the Karhunen-Loeve Transformation (KLT), because of its orthogonalizing properties. A matrix, A, is said to be orthogonal if and only if its rows form an orthonormal set, that is

$$\underline{A}_i^T \underline{A}_j = \begin{cases} 1 & \text{for } i=j \\ 0 & \text{for } i\neq j \end{cases} \tag{3.87}$$

where $\underline{A}_i$ denotes the i-th row of A, expressed as a vector for i=1,2,..,m. It follows that if A is orthogonal,

$$A^{-1} = A^T \tag{3.88a}$$

and this implies that

$$A^T A = A A^T = I \tag{3.88b}$$

The KLT matrix is computed from the covariance of $\underline{x}$ given by

$$R_x = \langle \underline{x}\,\underline{x}^T \rangle \tag{3.89}$$

Denoting the eigenvectors of R_x by $\underline{A}_i$, and its eigenvalues, λ_i, it follows that

$$R_x \underline{A}_i = \lambda_i \underline{A}_i \qquad i = 1,2,\ldots\ldots,m \tag{3.90}$$

As R_x is a real symmetric matrix, its eigenvalues are real, and there are exactly m eigenvectors which are orthogonal and can be chosen to be orthonormal. From this fact, the KLT matrix A is defined by the requirement that its rows be the normalized eigenvectors of R_x, and that the eigenvectors must be ordered into a sequence such that a monotonic decrease of the corresponding eigenvalues is achieved :

$$\lambda_1 \geq \lambda_2 \ldots\ldots\ldots \geq \lambda_m \tag{3.91}$$

From (3.85), the covariance of the transform coefficients can be written as :

$$R_y = \langle \underline{y}\,\underline{y}^T \rangle = \langle A\,\underline{x}\,\underline{x}^T A^T \rangle \tag{3.92a}$$

$$= A\,R_x\,A^T$$

$$= \begin{bmatrix} \lambda_1 & 0 & \ldots\ldots\ldots\ldots & 0 \\ 0 & \lambda_2 & \ldots\ldots\ldots\ldots & 0 \\ \cdot & \cdot & \cdot\cdot\cdot\cdot\cdot\cdot\cdot\cdot & \cdot \\ \cdot & \cdot & \cdot\cdot\cdot\cdot\cdot\cdot\cdot\cdot & \cdot \\ \cdot & \cdot & \cdot\cdot\cdot\cdot\cdot\cdot\cdot\cdot & \cdot \\ 0 & 0 & \ldots\ldots\ldots\ldots & \lambda_m \end{bmatrix}$$

$$= \lambda\,. \tag{3.92b}$$

It turns out that the KLT diagonalizes R_y, where the transform coefficients are uncorrelated. Hence, the components of $\underline{x}$ are transformed into components that are of differing importance. The variances of the coefficients are

$$\sigma_{yi}^2 = \lambda_i \qquad i = 1,2,\ldots m \tag{3.93}$$

$$= \text{eigenvalues of } R_x$$

and, as has been mentioned before, are as different from one another as possible.

In the simulation, each 20 msec speech frame requires 12 power parameters, u_{ij}', as i=1,2,3 and j=1,2,3,4. These form a vector of 12 elements arranged as follows:

$$\underline{x}^T = (u_{11}'\ u_{12}'\ \ldots\ldots\ u_{21}'\ u_{22}'\ \ldots\ldots\ u_{31}'\ \ldots\ldots\ldots\ u_{34}')$$

The covariance matrix R_x was computed from representative speech data consisting of a 30 second segment of male speech and a 30 second segment of female speech. The KLT matrix was derived from the eigenvectors of R_x, and the eigenvectors were arranged in the order required to satisfy (3.91). Figure 3.22 shows the KLT matrix that was obtained and the corresponding eigenvalues. It can be seen that the first eigenvalues which determine the variances of the first of the transformed coefficients are much larger than the later ones. The KLT matrix obtained from this representative data is stored in the encoder and also in the receiver so that a KLT transform can be performed before transmission and an inverse KLT can be performed at the receiver.

3.3.5.2 Bit Allocation for the Transformed Coefficients

The favourable properties of the KLT are largely due to the fact that the transformed coefficients are linearly uncorrelated. It is therefore reasonable to quantize the coefficients independently. The coefficients bearing larger variances can be encoded by more bits, and coefficients which are of relatively low variances can be discarded entirely or coded with a small number of bits.

The m.s.e. (mean square error) measure adopted when using transform coding is the sum of the variances of the errors increased in the quantization of individual transformed coefficients. Using (3.75), the m.s.e. is thus given by

$$D = \frac{1}{N_c} \sum_{i=1}^{Nc} k_i \, 2^{-2R_i} \lambda_i \qquad (3.94)$$

where N_c is the total number of coefficients. R_i and λ_i are the number of quantization bits and the variance of the i-th coefficients respectively for i=1,2,..,N_c. The m.s.e. is minimized with the constraint of a given average bit rate, B_0, defined as:

$$B_0 = \frac{1}{N_c} \sum_{i=1}^{Nc} R_i \qquad (3.95)$$

Using the Lagrange multiplier method as in Section 3.3.4.2; the optimum bit allocation scheme is to allocate B_i bits to transformed coefficient u_i'', where :

$$B_i = \frac{1}{2} \log_2 \frac{\lambda_i}{\left[\prod_{n=1}^{N_c} \lambda_i \right]^{1/N_c}} \qquad \text{for } i=1,2,..,N_c . \qquad (3.96)$$

The B_i values depend on the distribution of coefficient variances and arrange themselves in such a way that coefficients with larger variance can be quantized by more bits so as to minimizes the total quantization error.

In the coder simulation, an average bit rate of 2 bits per coefficient is used. The bit allocation for each coefficient is computed from the coefficient variances obtained by analysing representative speech segments, i.e. 30 seconds of male speech and 30 seconds of female speech. The calculated bit allocations are rounded to the nearest integers under the constraint that the total number of bits per frame be equal to $N_c.B_0$. Although the integer bit assignment may result in a slight reduction in performance as compared with the optimal, this degradation cannot be avoided. The mean, standard deviation and the bit allocation strategy adopted for the transform coefficients are given in Figure 3.22.

Histograms and theoretical Gaussian distribution curves for the transformed coefficients are also plotted in Figure 3.23 for comparison. This Figure illustrates that all but two of the transformed coefficients, u_k'' for k=1,2,..,12, are approximately Gaussian distributed the exceptions being u_1'' and u_4''. With the assumption of a Gaussian distribution for the coefficients, optimal Gaussian uniform Max quantizers are used to achieve minimum m.s.e. performance. The quantization process is the same as is used for LSP coefficient quantization. Each transform coefficient is first changed to a zero mean and unit variance parameter by subtracting its mean and dividing by its standard deviation, then the coefficient is uniformly quantized using the step sizes calculated by Max.

3.4 OBJECTIVE MEASUREMENTS FOR APC-AB CODER

The practically of a speech coder depends upon subjective quality which must be assessed by listening tests involving a large group of trained and untrained listeners [50]. Such tests are very time consuming and laborious, and it has not been practical to carry out formal subjective tests during the course of this work.

Objective measurements made by computing the discrepancy between the original speech and the decoded speech are useful for many types of coder, although their relationship to subjective results are often questioned. Various 'more reliable' objective measurements, for example signal to noise ratio, spectral distortion measure, cepstral distortion measures etc, have been summarised in [53]. It is believed that for certain classes of coder, especially waveform coders such as the APC-AB codec, the signal to noise ratio measure is related to the perceived

quality. Thus a coder with a higher signal to noise ratio will generally have a better perceived quality. In this dissertation, two commonly used time domain measures: long term signal to noise ratio, SNR_{long}, and segmental signal to noise ratio, SNR_{seg}, have been used to provide an indication of comparative performance.

The two signal to noise ratio measures are defined as:

$$SNR_{long} = 10 \log_{10} \frac{\Sigma\, s^2(k)}{\Sigma\, (s(k)-r(k))^2} \quad dB \qquad (3.97)$$

and

$$SNR_{seg} = (1/M) \sum_{m=0}^{M-1} 10 \log_{10} \frac{\Sigma_{j=1,J}\, s^2(j+mJ)}{\Sigma_{j=1,J}\, [s(j+mJ)-r(j+mJ)]^2} \quad dB \qquad (3.98)$$

where k ranges over all the samples of the testing speech utterance, and s(n) and r(n) are the original speech sample and the reconstructed speech sample respectively. M is the number of segments and J is the length of a speech segment. For the 16 kb/s APC-AB coder, J=128, corresponding to a segment length of 20 msec. The measured signal to noise ratio using a 2 second block of male speech for the 16 kb/s APC-AB coder is given as follows:

$$SNR_{long} = 19.31 \text{ dB}, \quad \text{and} \quad SNR_{seg} = 16.89 \text{ dB}.$$

Figure 3.24 illustrates the time variation of the segmental SNR for the 2 second block of male speech, which shows that for high energy segments, e.g. for vowel sounds, the SNR value often exceeds 20 dB.

From informal listening tests, the 16 kb/s speech coder was able to produce toll ~ quality speech. Mean opinion scores have been quoted by Itakura and Honda [30].

3.5 CONCLUSION

A 16 kb/s APC-AB coder simulation has been described in detail. As only brief descriptions of the APC-AB coder algorithms were found in the published literature [31,51], some techniques being used in the coder simulation which are described in this thesis may differ from the original APC-AB coder described by Itakura and Honda [31].

During the development of the coder simulation, some computation saving methods such as the use of a polyphase quadrature mirror filter structure, a simplified phase-shifter circuit and a faster scheme for pitch detection have been investigated. A new method for reducing the computational complexity of deriving LSP coefficients from LPC coefficients was also developed and investigated during the course of this project [64,71]. This new scheme will be described in detail in the next Chapter.

The description of the APC-AB coding technique used in this thesis not only illustrates the complexity of a typical medium bit rate speech coding system for LMR applications, but also allows us to understand the importance of each transmission parameter to the decoded speech quality. This knowledge will be useful in understanding the influence of channel error on the APC-AB coding system. Analysing the effect of channel errors on a speech coding system is one of the aims of this thesis which will be discussed in Chapter 5.

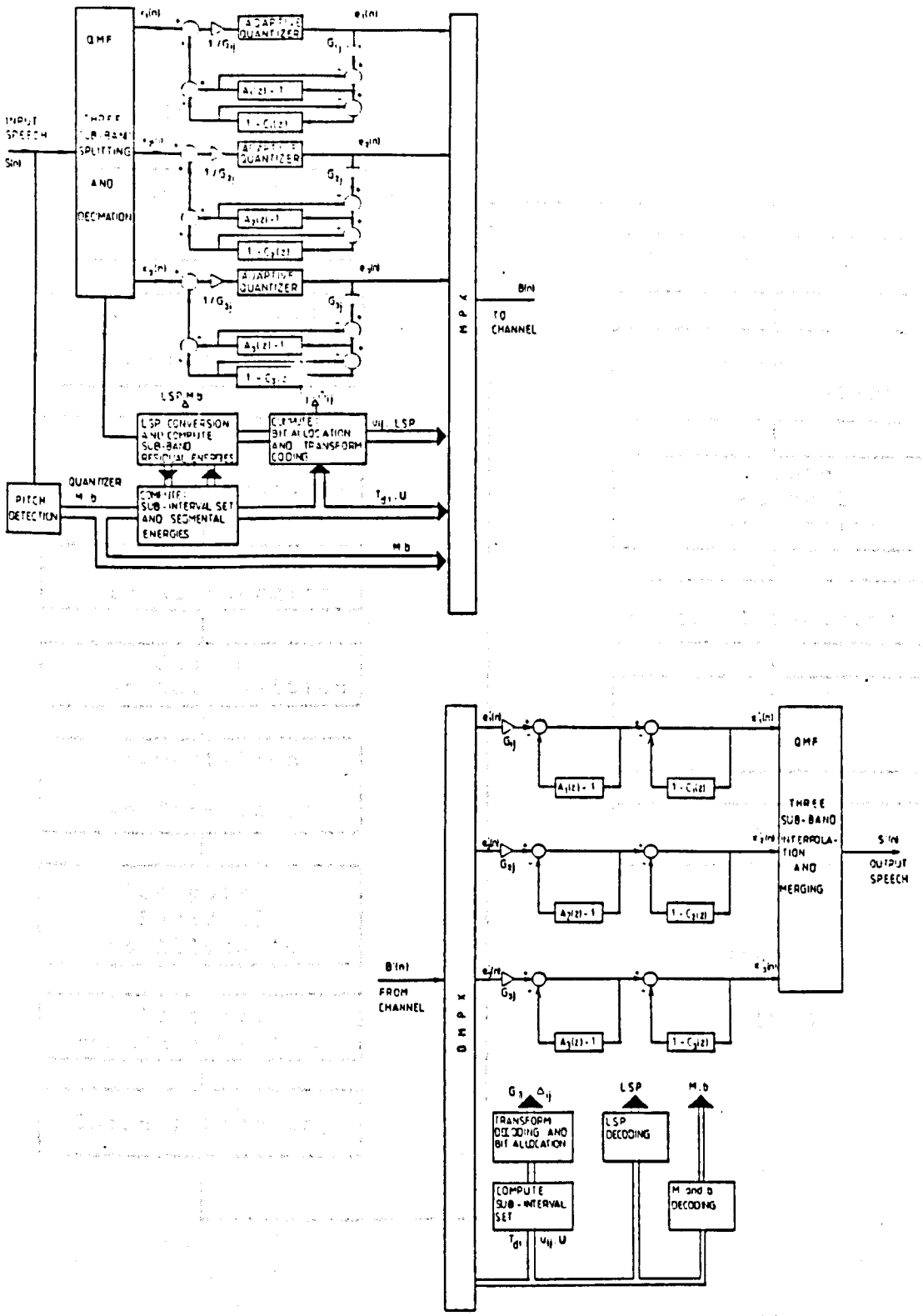


Figure 3.1 Block diagram of APC-AB encoder and decoder.

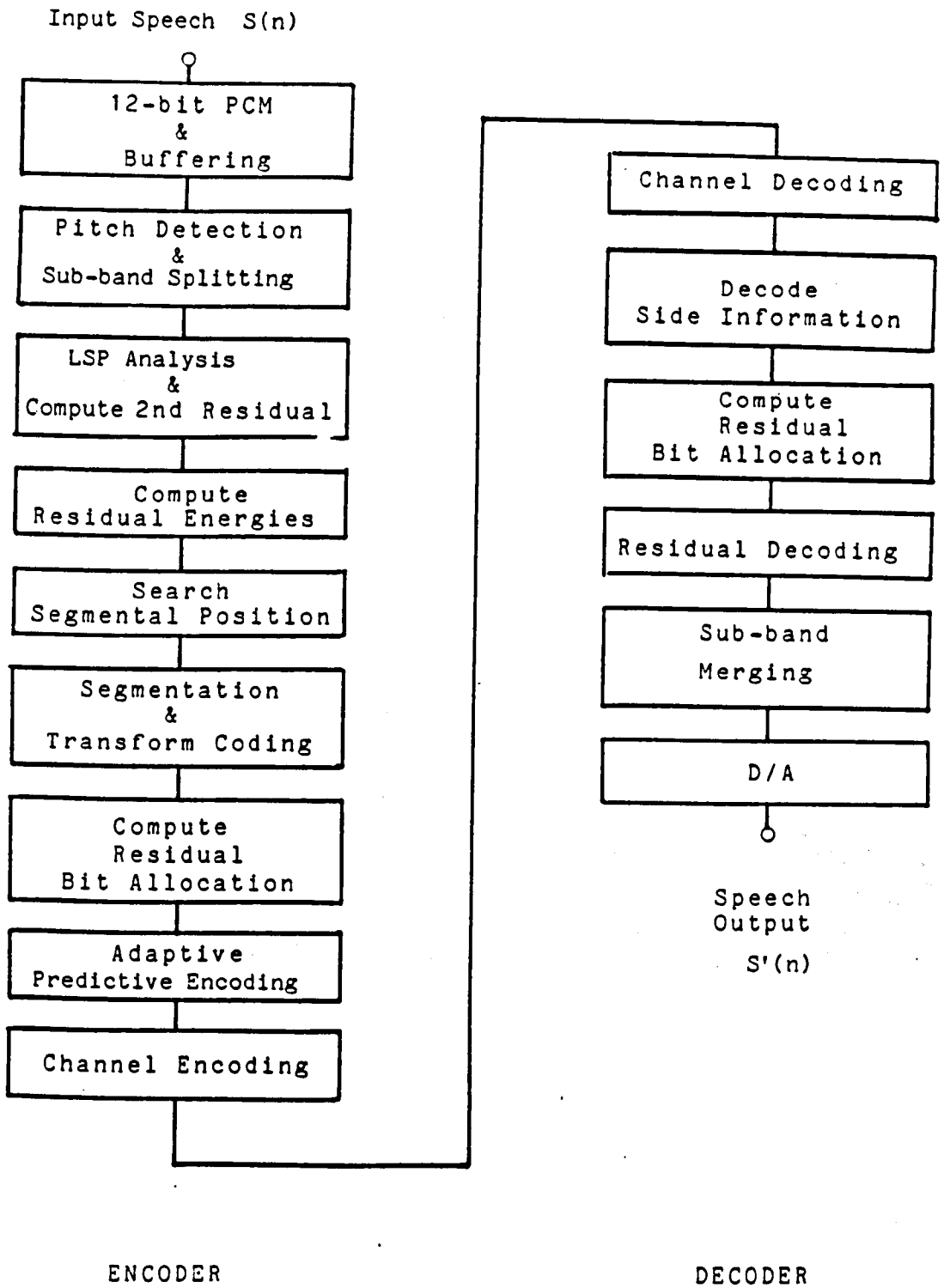
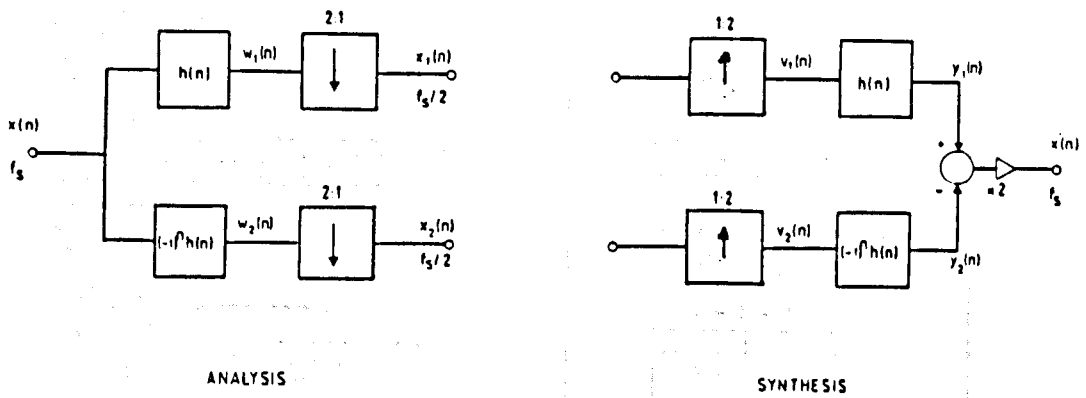
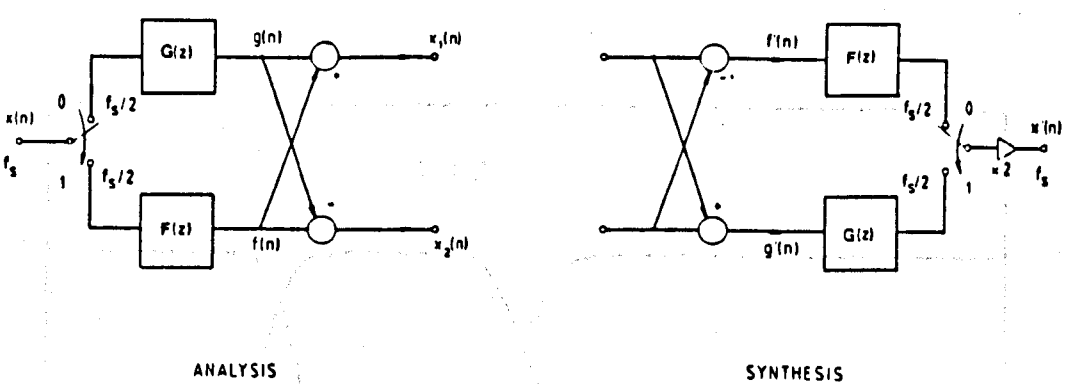


Figure 3.2 Flow chart of APC-AB encoder and decoder algorithms.



(a)



(b)

Figure 3.3 Two-band QMF structure
(a) Basic configuration, (b) Equivalent polyphase structure.

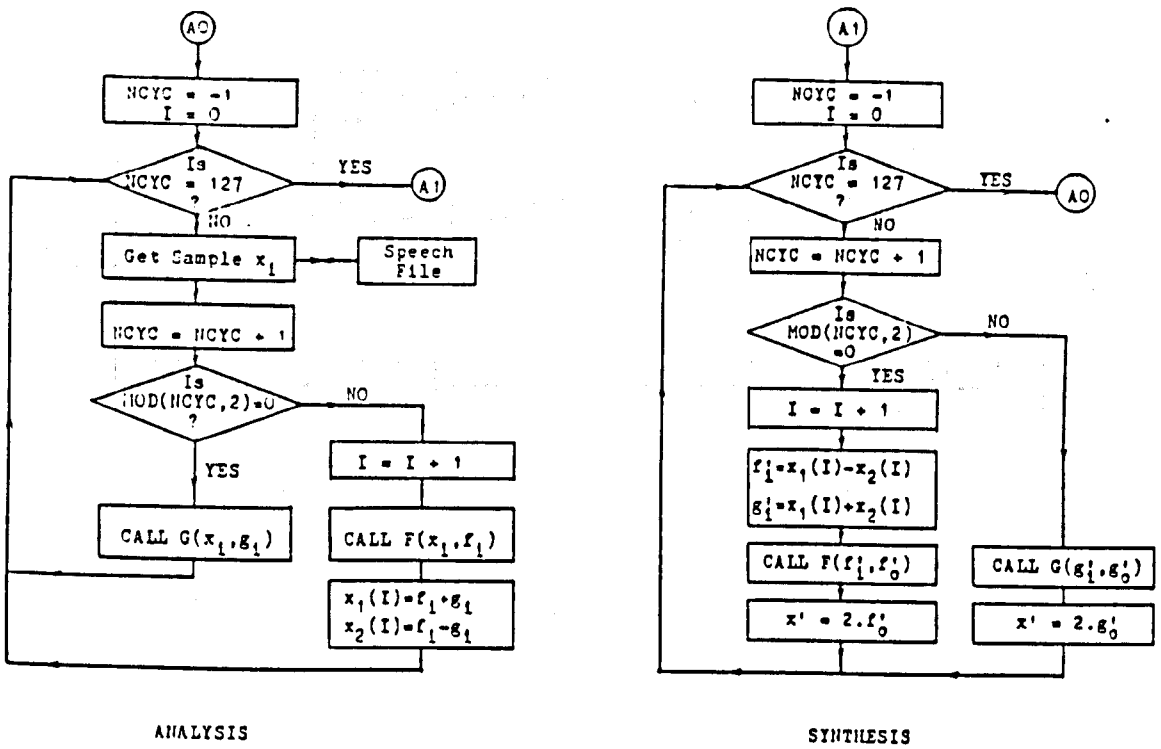


Figure 3.4 Flow chart of the 2-band QMF polyphase algorithm.

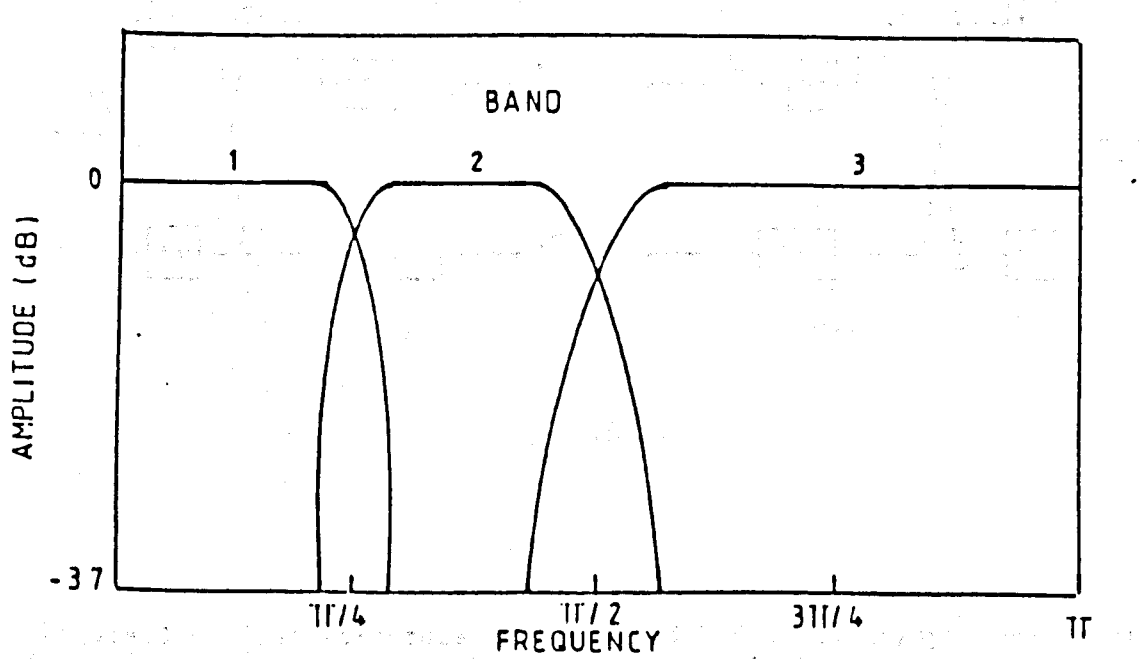
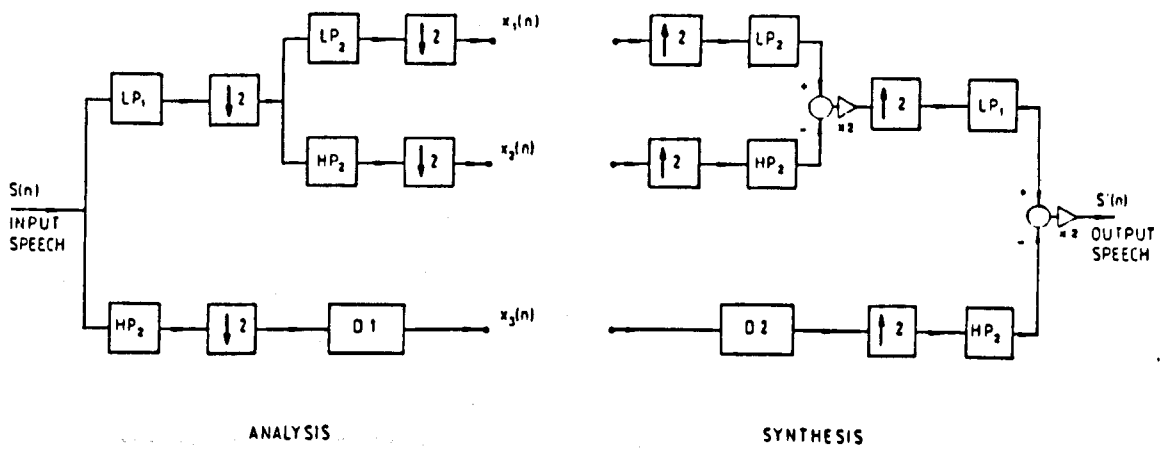
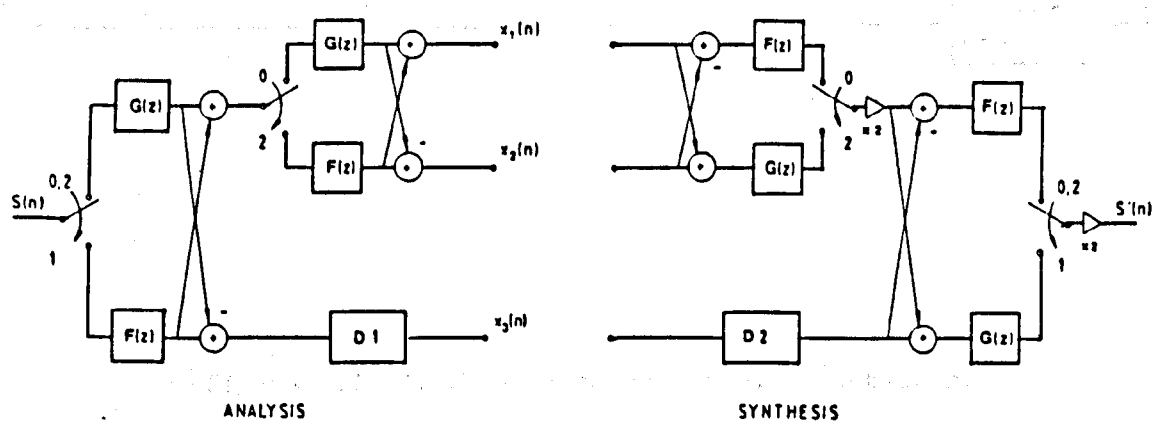


Figure 3.5 Frequency response for 32-tap FIR QMF filter for three subbands APC-AB coder.



(a)



(b)

Figure 3.6 Tree structure of 3-subband QMF analysis/synthesis system. (a) Basic configuration, (b) Equivalent polyphase structure.

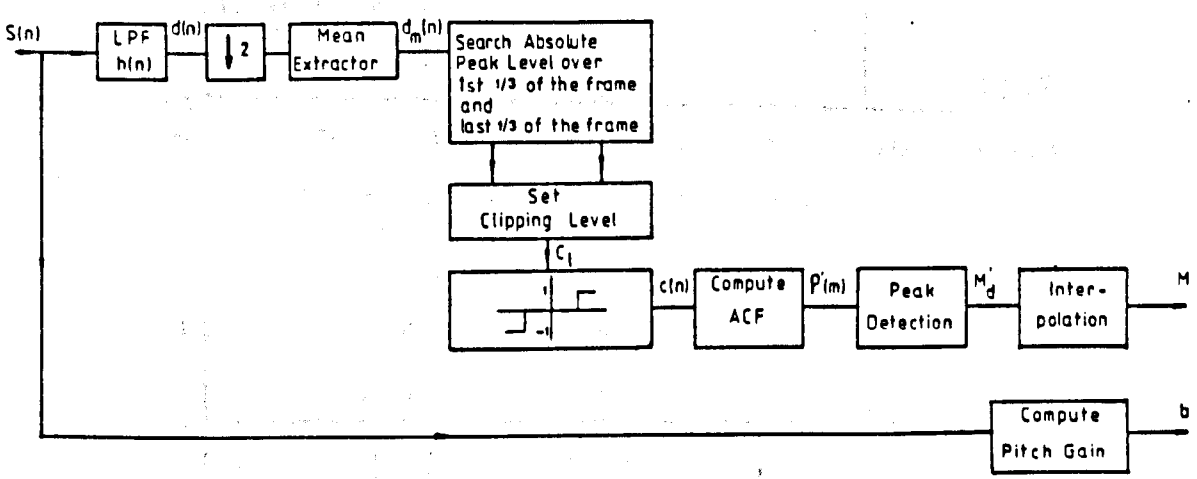
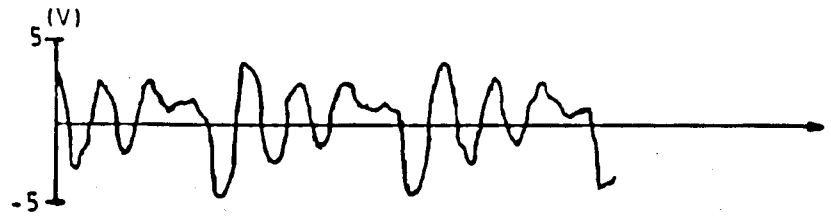
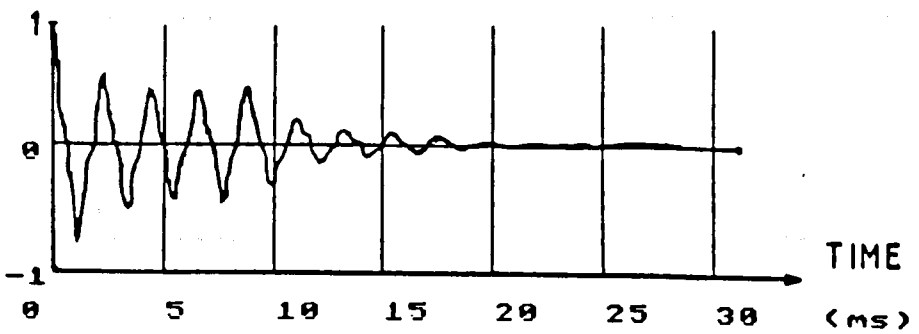


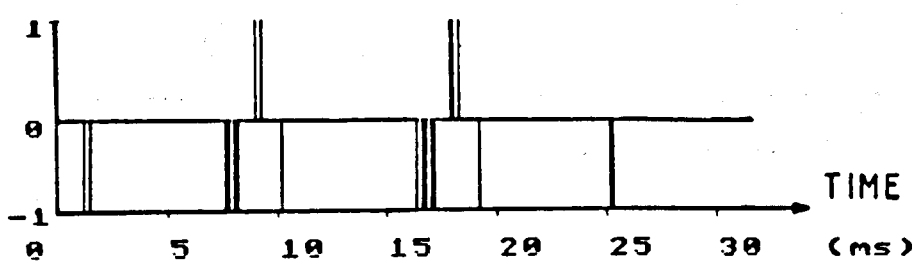
Figure 3.7 Block diagram of the pitch detector.



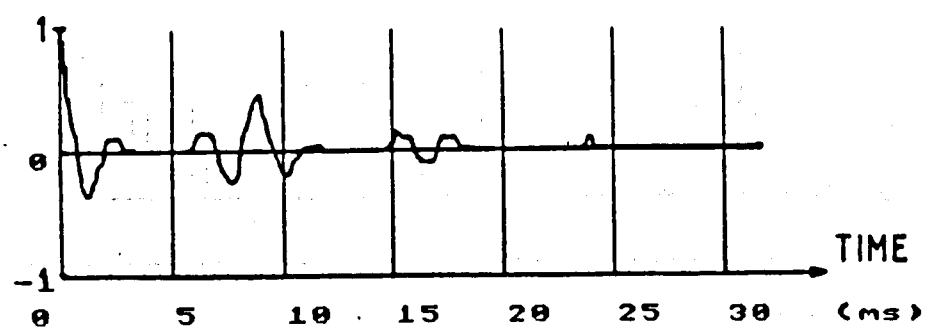
a) original speech



b) normalised ACF of original speech



c) centre-infinite peak clipped signal



d) normalised ACF of clipped signal

Figure 3.8 Pitch detection.

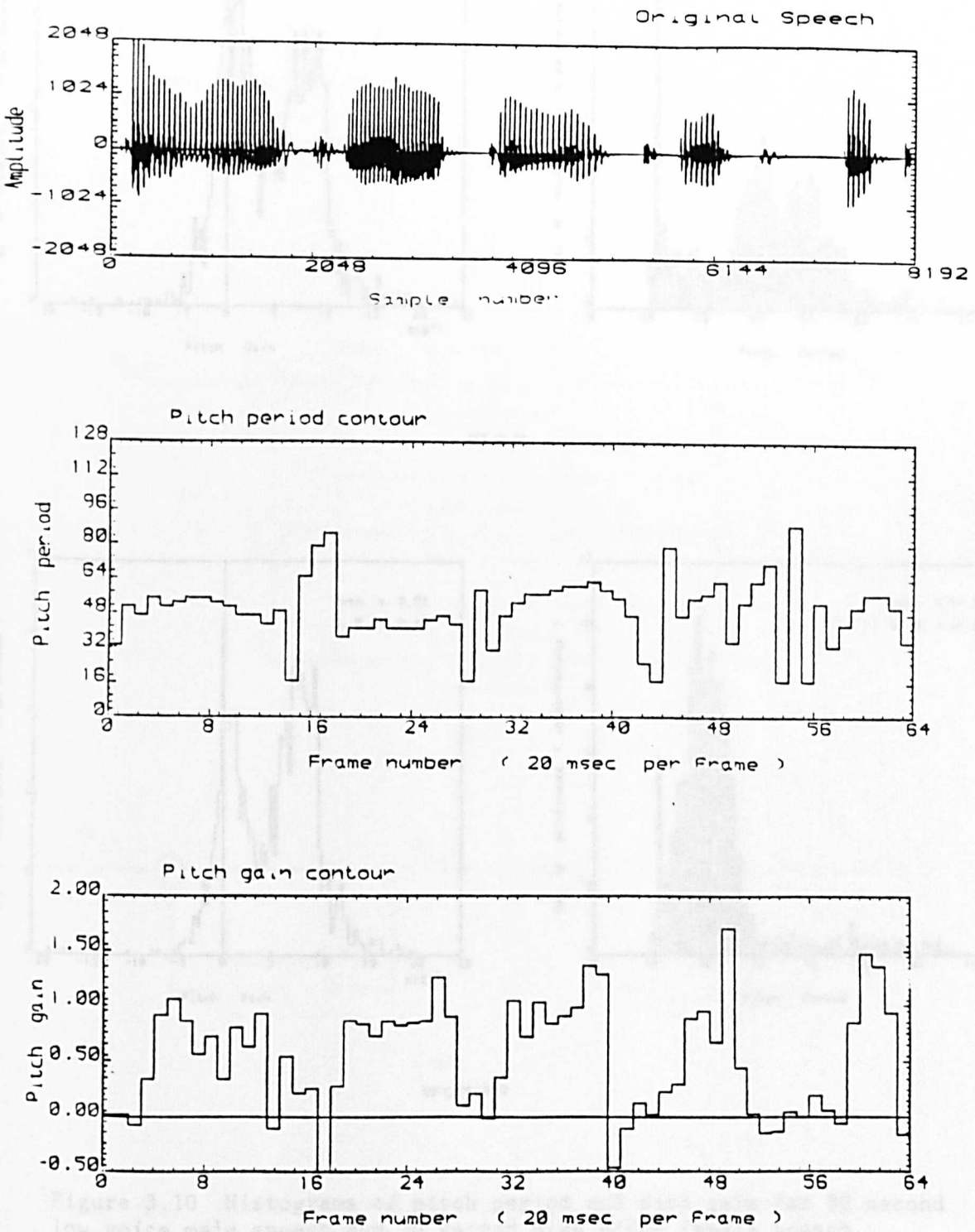


Figure 3.9 Illustration of time variations of pitch period and pitch gain of 1.28 second speech.

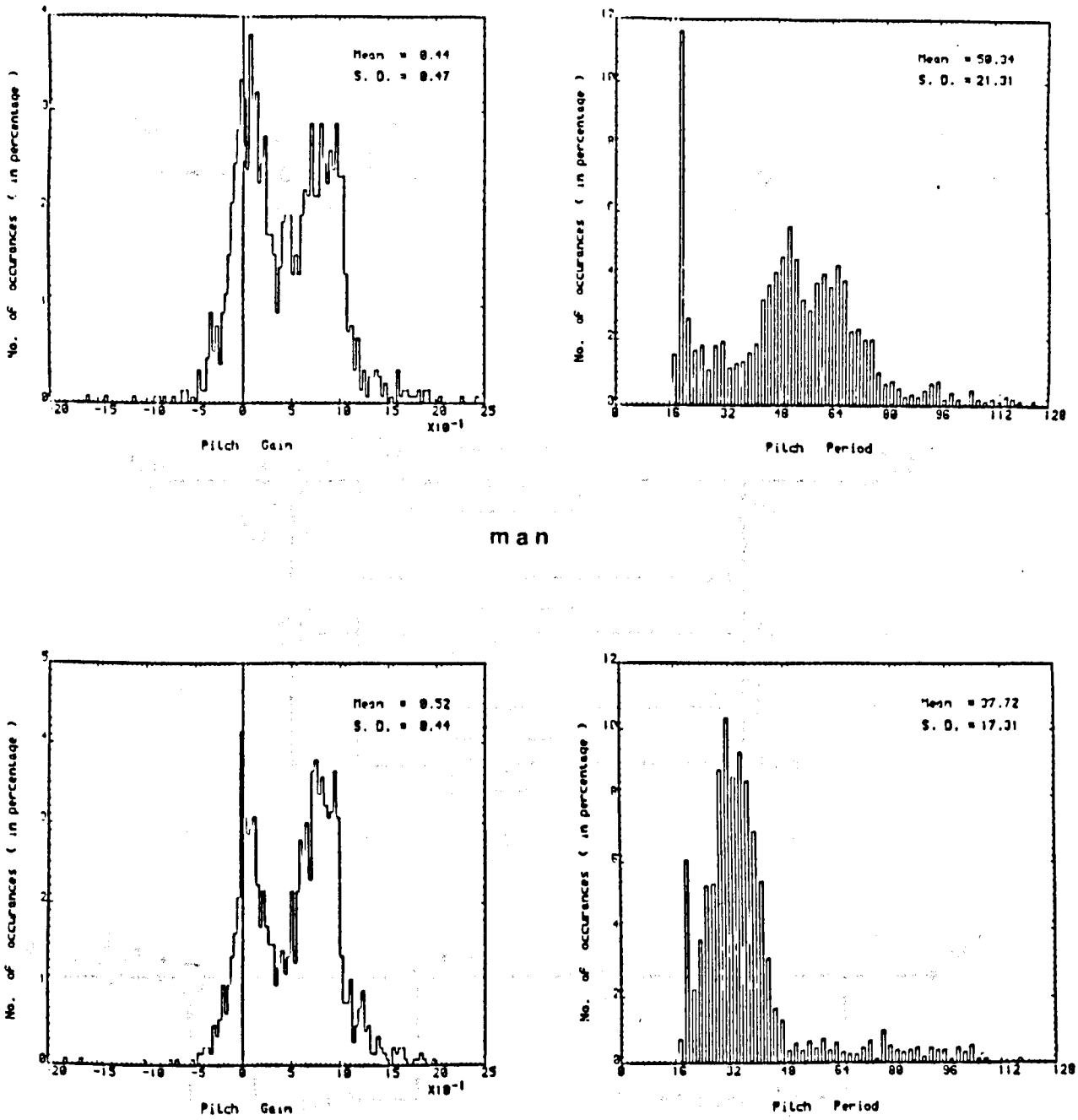


Figure 3.10 Histograms of pitch period and pith gain for 30 second low voice male speech and 30 second high pitch female speech.

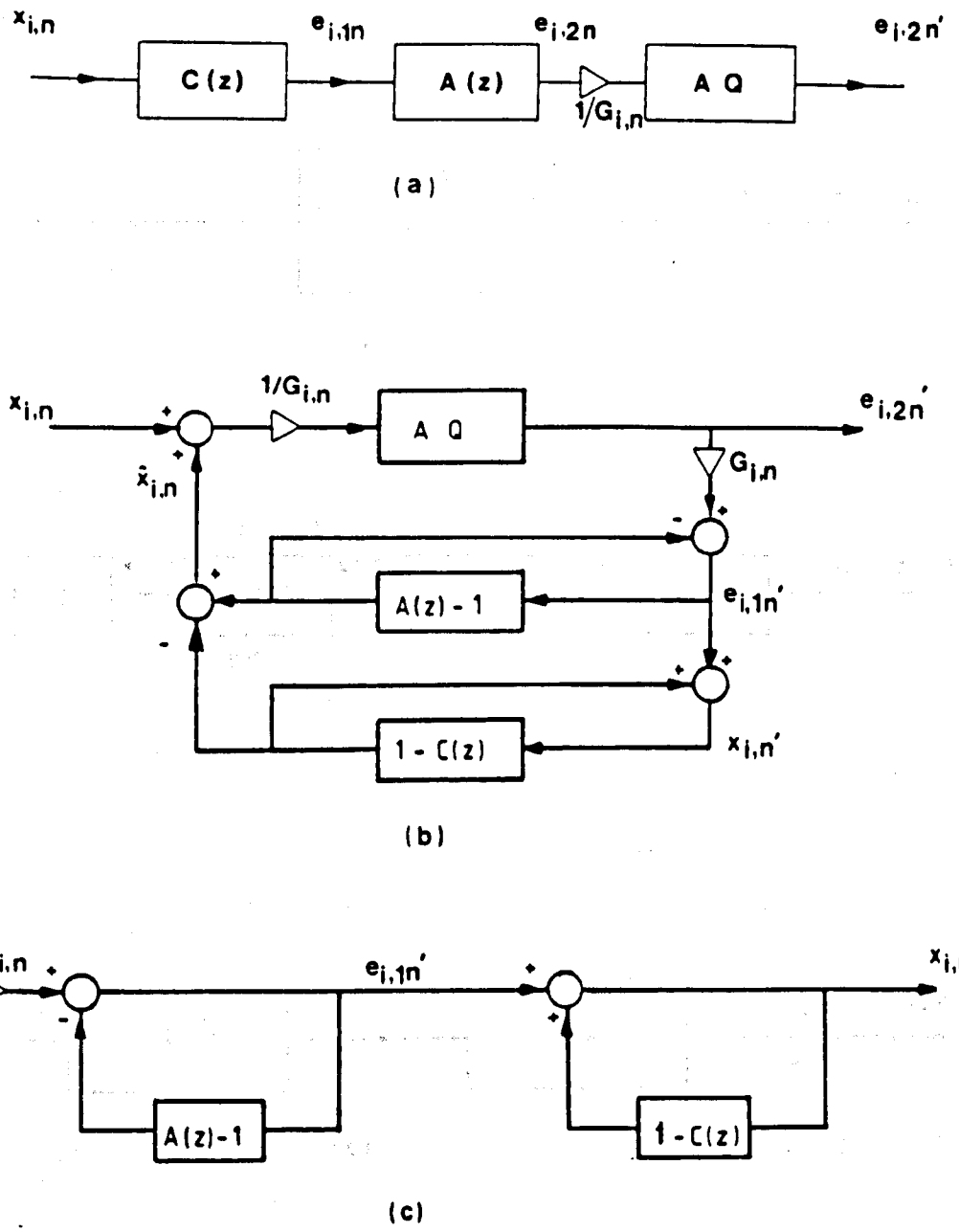


Figure 3.11 Adaptive predictive coding (APC) for sub-band speech.
a) APC encoder,
b) equivalent structure for APC encoder,
c) APC decoder.

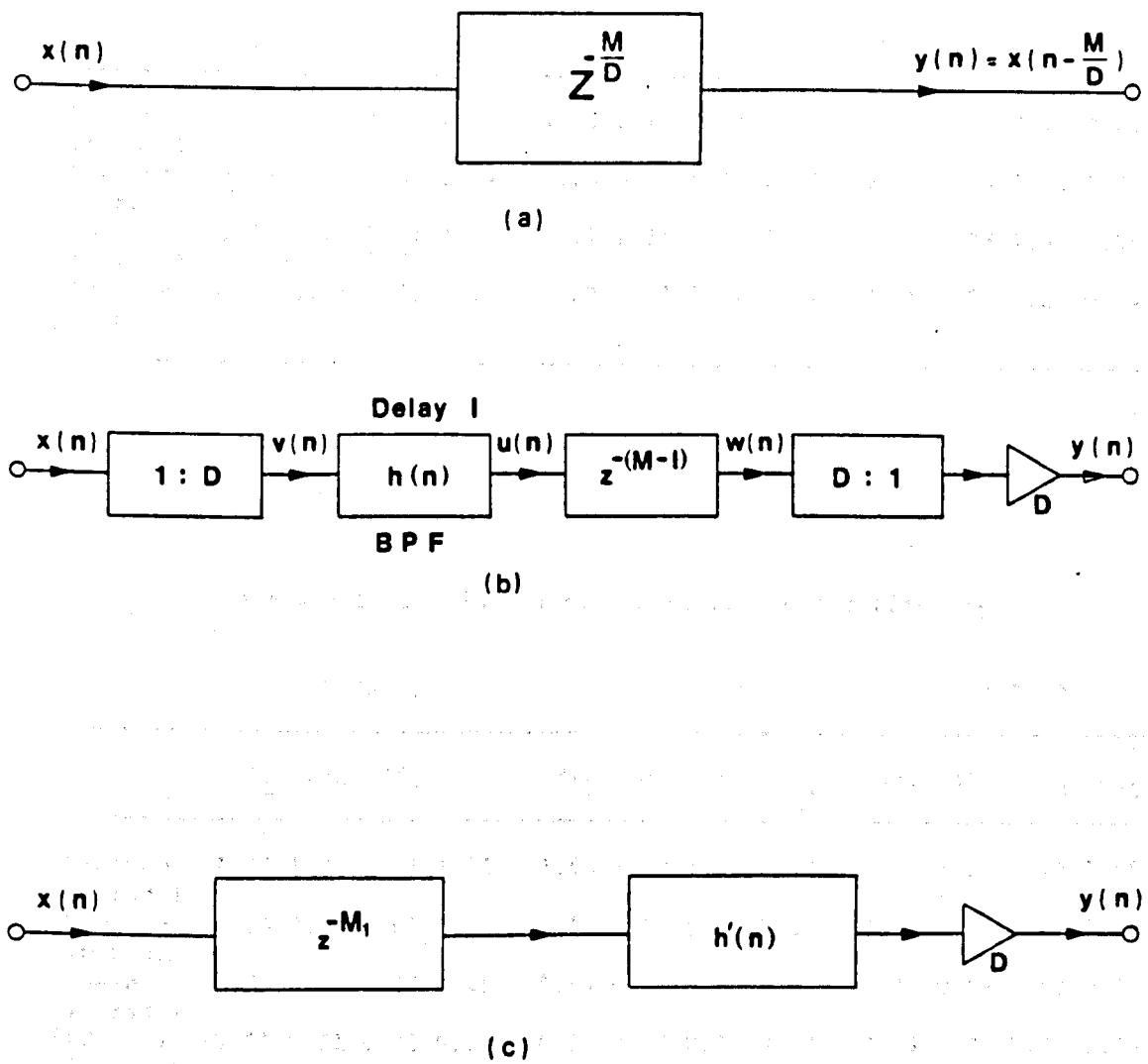


Figure 3.12 Non-integer delay circuit and its equivalent representations.

Table 3.1 Statistics of LSP coefficients

	BAND 1				BAND 2				BAND 3			
	θ_1	ω_1	θ_2	ω_2	θ_1	ω_1	θ_2	ω_2	θ_1	ω_1	θ_2	ω_2
Maximum (in rad.)	1.30	1.82	2.46	2.82	1.63	2.19	2.68	3.02	1.50	2.05	2.74	3.07
Minimum (in rad.)	0.26	0.54	1.06	1.44	0.14	0.49	1.07	1.80	0.31	0.64	1.01	1.59
Mean (in rad.)	0.78	1.19	1.75	2.20	0.64	1.21	1.90	2.50	0.82	1.28	1.96	2.59
Std. dev. (in rad.)	0.21	0.19	0.23	0.27	0.22	0.27	0.24	0.20	0.19	0.20	0.26	0.24

(a)

Table 3.2 Statistics of LSP coefficient differences

	BAND 1				BAND 2				BAND 3			
	Δf_1	Δf_2	Δf_3	Δf_4	Δf_1	Δf_2	Δf_3	Δf_4	Δf_1	Δf_2	Δf_3	Δf_4
Maximum (in rad.)	1.29	1.26	1.41	1.31	1.64	1.63	1.85	1.53	1.49	1.20	1.75	1.58
Minimum (in rad.)	0.26	0.03	0.07	0.04	0.14	0.09	0.06	0.07	0.31	0.05	0.07	0.05
Mean (in rad.)	0.78	0.41	0.57	0.45	0.64	0.57	0.69	0.61	0.83	0.46	0.68	0.62
Std. dev. (in rad.)	0.21	0.21	0.22	0.23	0.22	0.22	0.25	0.22	0.19	0.21	0.26	0.26

(b)

Figure 3.13

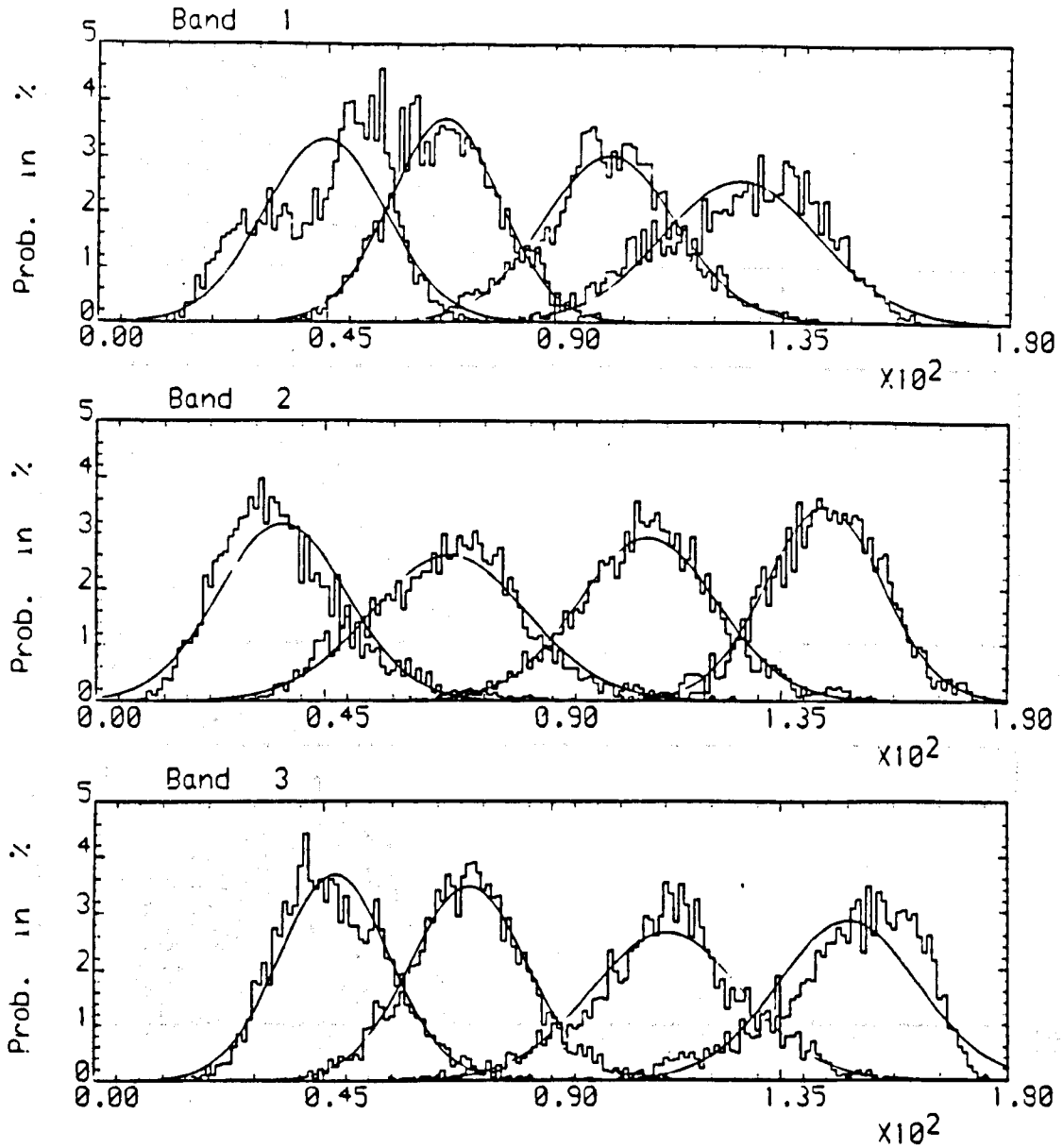


Figure 3.14 Histograms of the LSP coefficients.

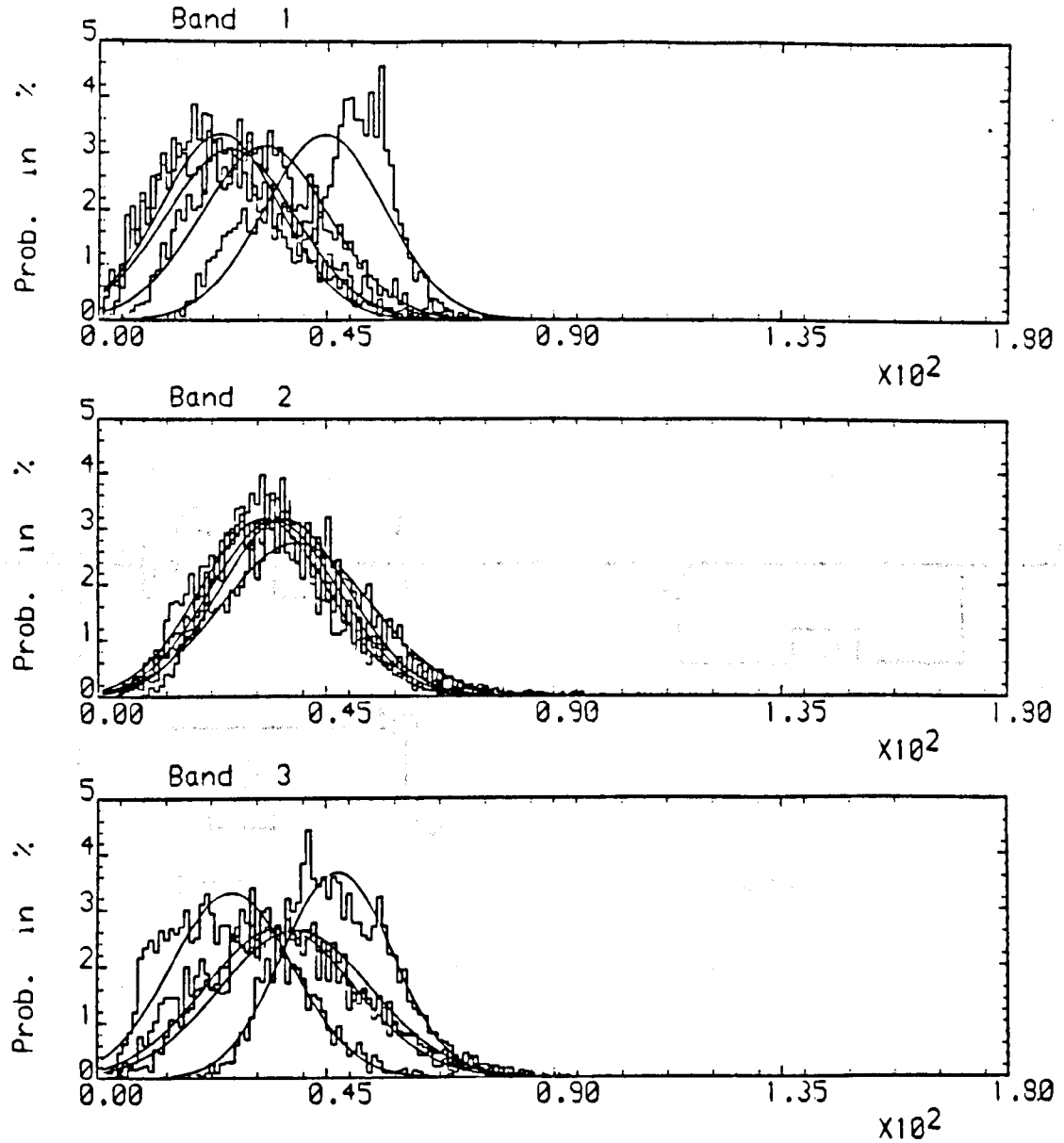


Figure 3.15 Histograms of the LSP coefficient differences.

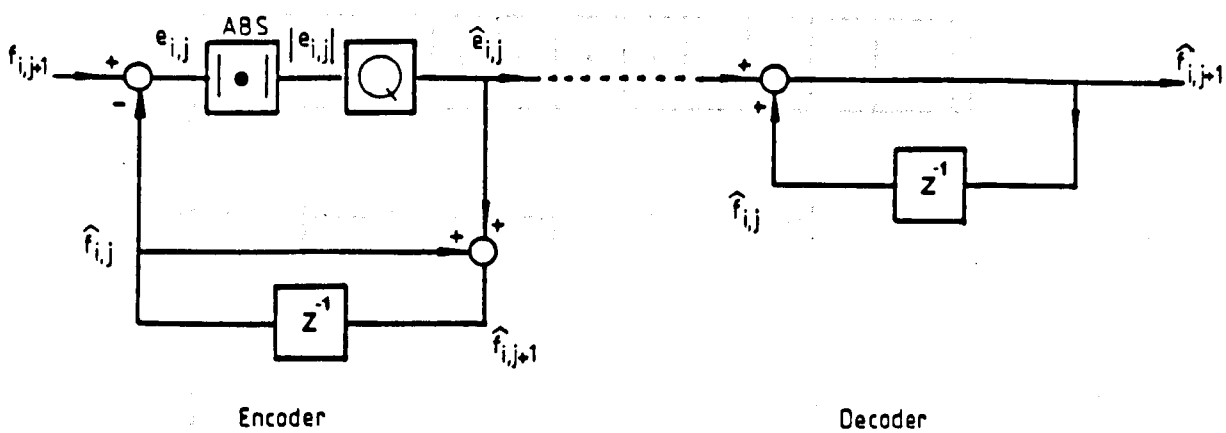


Figure 3.16 Encoder and decoder for the quantization of the LSP coefficient differences.

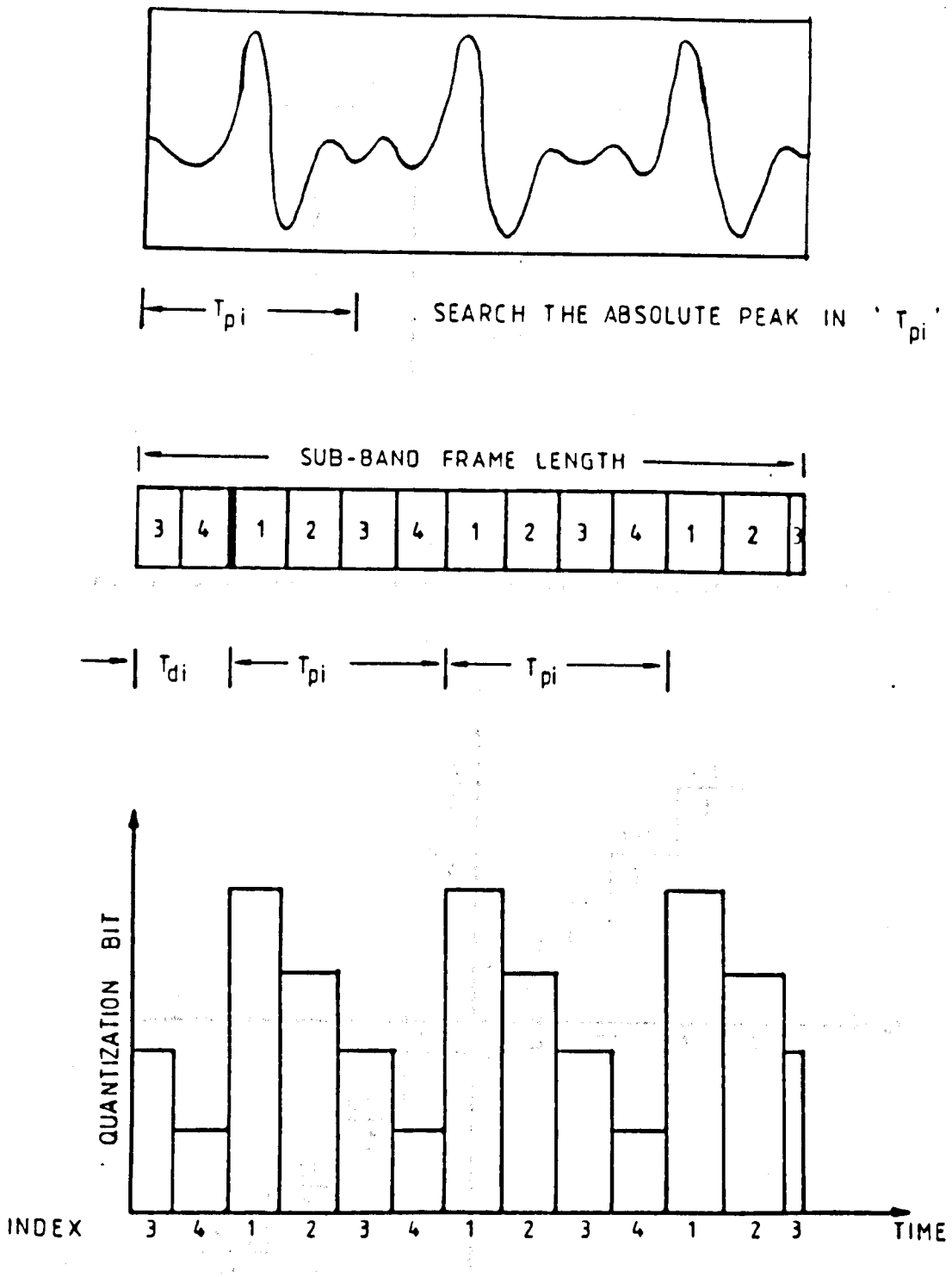


Figure 3.17 Method for segmenting pitch period and the pitch synchronous bit allocation.

Bit number	Step size
0	—
1	1.5956
2	0.9957
3	0.5860
4	0.3352
5	0.1881
6	0.1041
7	0.0569
8	0.0308

Figure 3.18 Optimum step size for Max uniform quantization.

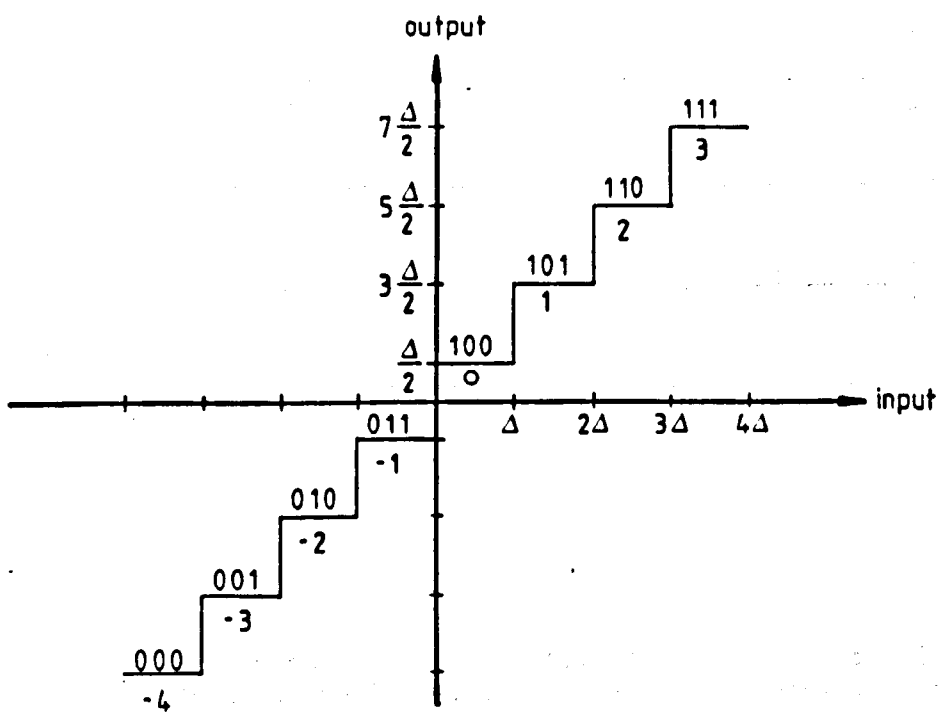


Figure 3.19 Uniform quantization using a 3-bit quantizer

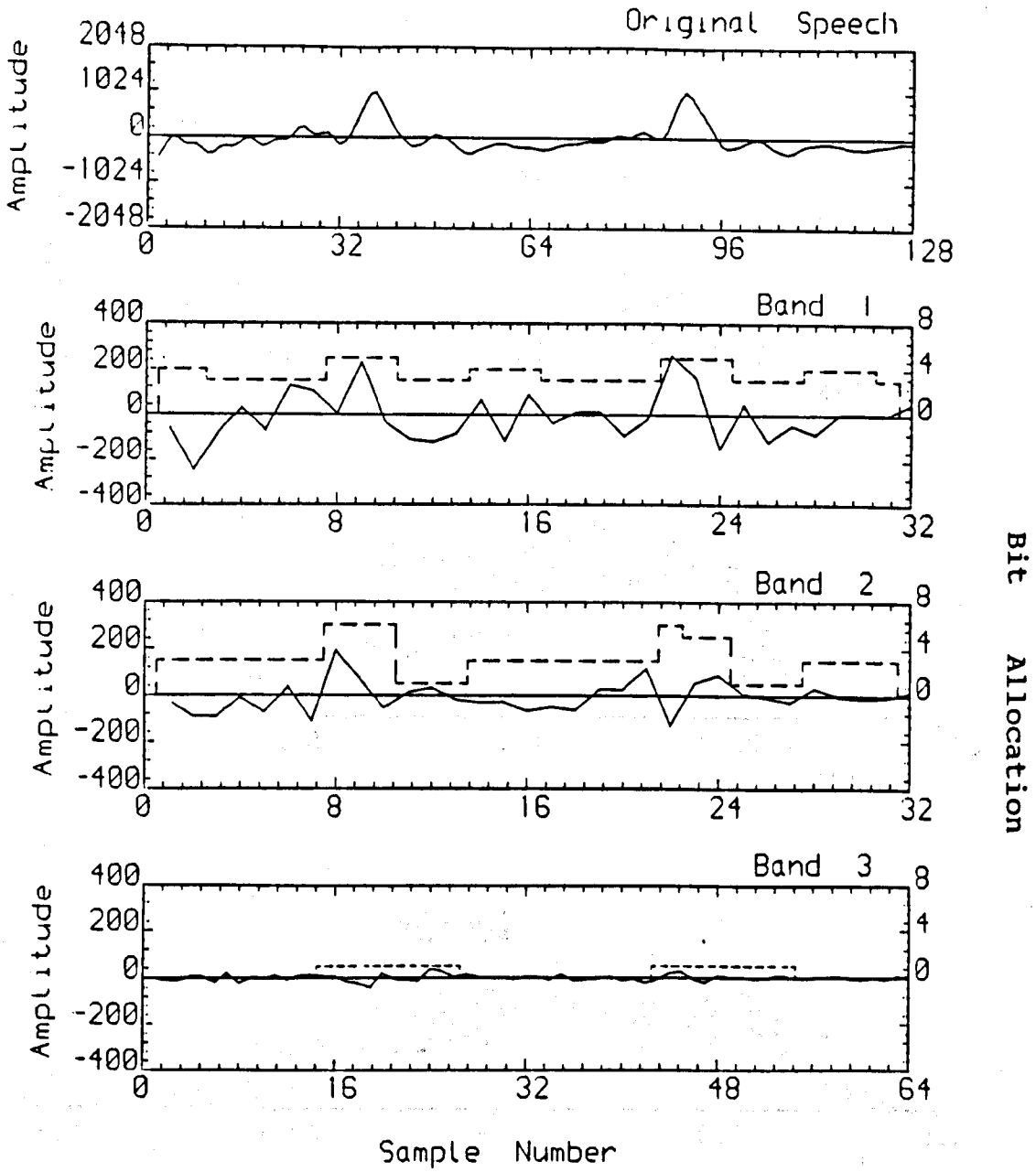


Figure 3.20 Graphs showing how the bit allocation (dotted line) in frequency bands 1, 2 and 3 varies with residual siganl amplitude (continuous line).

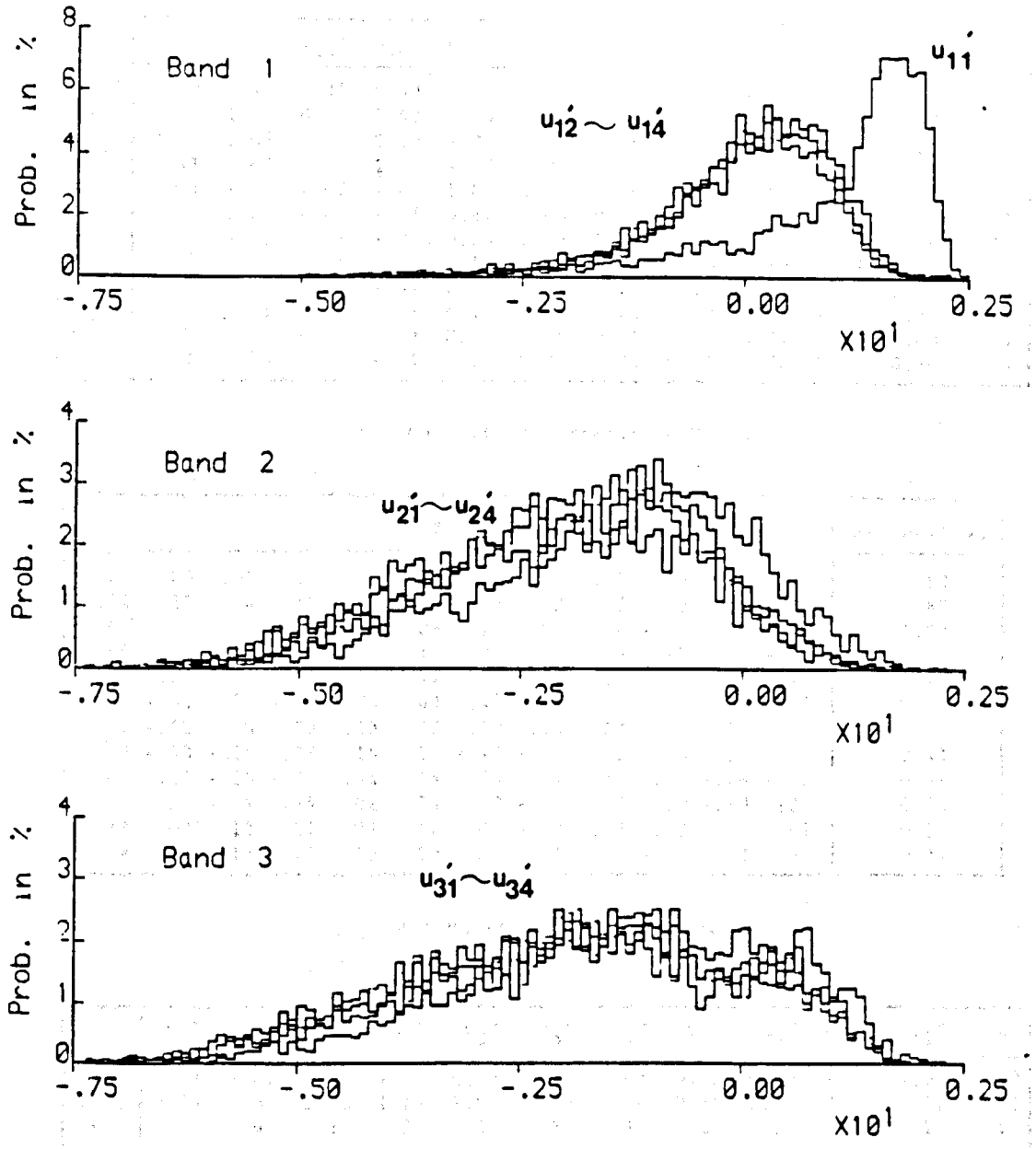


Figure 3.21 Frequency distributions of the normalised logarithmic residual energies.

	u_{11}'	u_{12}'	u_{13}'	u_{14}'	u_{21}'	u_{22}'	u_{23}'	u_{24}'	u_{31}'	u_{32}'	u_{33}'	u_{34}'
u_{11}'	1.00	0.42	0.36	0.43	-.04	-.32	-.34	-.25	-.45	-.60	-.63	-.61
u_{12}'	0.42	1.00	0.46	0.42	-.04	0.02	-.06	-.08	-.36	-.28	-.32	-.33
u_{13}'	0.36	0.46	1.00	0.50	-.07	-.01	0.08	-.03	-.34	-.28	-.21	-.27
u_{14}'	0.43	0.42	0.50	1.00	-.06	-.08	-.02	0.09	-.37	-.36	-.33	-.25
u_{21}'	-.04	-.04	-.07	-.06	1.00	0.64	0.54	0.59	0.19	0.10	0.02	-.02
u_{22}'	-.32	0.02	-.01	-.08	0.64	1.00	0.74	0.64	0.14	0.28	0.26	0.02
u_{23}'	-.34	-.06	0.08	-.02	0.54	0.74	1.00	0.67	0.12	0.24	0.36	0.30
u_{24}'	-.25	-.08	-.03	0.09	0.59	0.64	0.67	1.00	0.19	0.19	0.22	0.32
u_{31}'	-.45	-.36	-.34	-.37	0.19	0.14	0.12	0.19	1.00	0.81	0.71	0.70
u_{32}'	-.60	-.28	-.28	-.36	0.10	0.28	0.24	0.19	0.81	1.00	0.84	0.77
u_{33}'	-.63	-.32	-.21	-.33	0.02	0.26	0.36	0.22	0.71	0.84	1.00	0.84
u_{34}'	-.61	-.33	-.27	-.25	-.02	0.20	0.30	0.32	0.70	0.77	0.84	1.00

(a) Normalized Correlation Matrix

K L T Matrix												Eigenvalues
.20	.03	.01	.02	-.25	-.36	-.39	-.33	-.27	-.26	-.40	-.38	39.53
.11	.13	.14	.15	.38	.38	.36	.32	-.32	-.33	-.32	-.31	6.80
-.01	.27	.41	.36	-.53	-.06	.23	.03	-.38	-.07	.25	.27	1.97
.45	.40	.36	.37	.26	-.10	-.24	.02	.43	.22	-.05	-.03	1.42
-.06	.26	.16	-.23	-.02	.41	.09	-.59	-.11	.32	.20	-.41	1.13
.21	-.35	.28	-.09	.24	-.43	.49	-.27	.13	-.25	.31	-.12	0.69
-.42	-.50	.44	.29	-.13	.02	-.10	.15	.13	.26	-.16	-.37	0.49
.30	-.43	-.28	.56	.15	.27	-.15	-.28	-.28	.08	.20	.11	0.43
.47	-.18	-.15	-.04	-.56	.26	.35	.04	.37	.06	-.26	-.10	0.41
-.29	.28	-.49	.44	.02	-.35	.40	-.11	.08	.20	-.10	-.24	0.38
-.32	.02	.13	.17	.07	.24	.07	-.49	.33	-.37	-.36	.41	0.38
-.14	.12	-.15	.17	-.15	.20	-.20	.13	.34	-.54	.52	-.35	0.31

(b) KLT Matrix and its eigenvalues

	u_1''	u_2''	u_3''	u_4''	u_5''	u_6''	u_7''	u_8''	u_9''	u_{10}''	u_{11}''	u_{12}''
mean	5.31	-.42	-.13	-.05	0.02	0.00	-.01	0.02	0.02	-.01	-.01	0.00
S.d.	3.36	2.57	1.40	1.19	1.06	0.83	0.70	0.65	0.64	0.62	0.61	0.55
B_i	4	4	3	2	2	2	1	1	1	1	1	1

(c) Statistics of transformed coefficients and its bit allocations

Figure 3.22

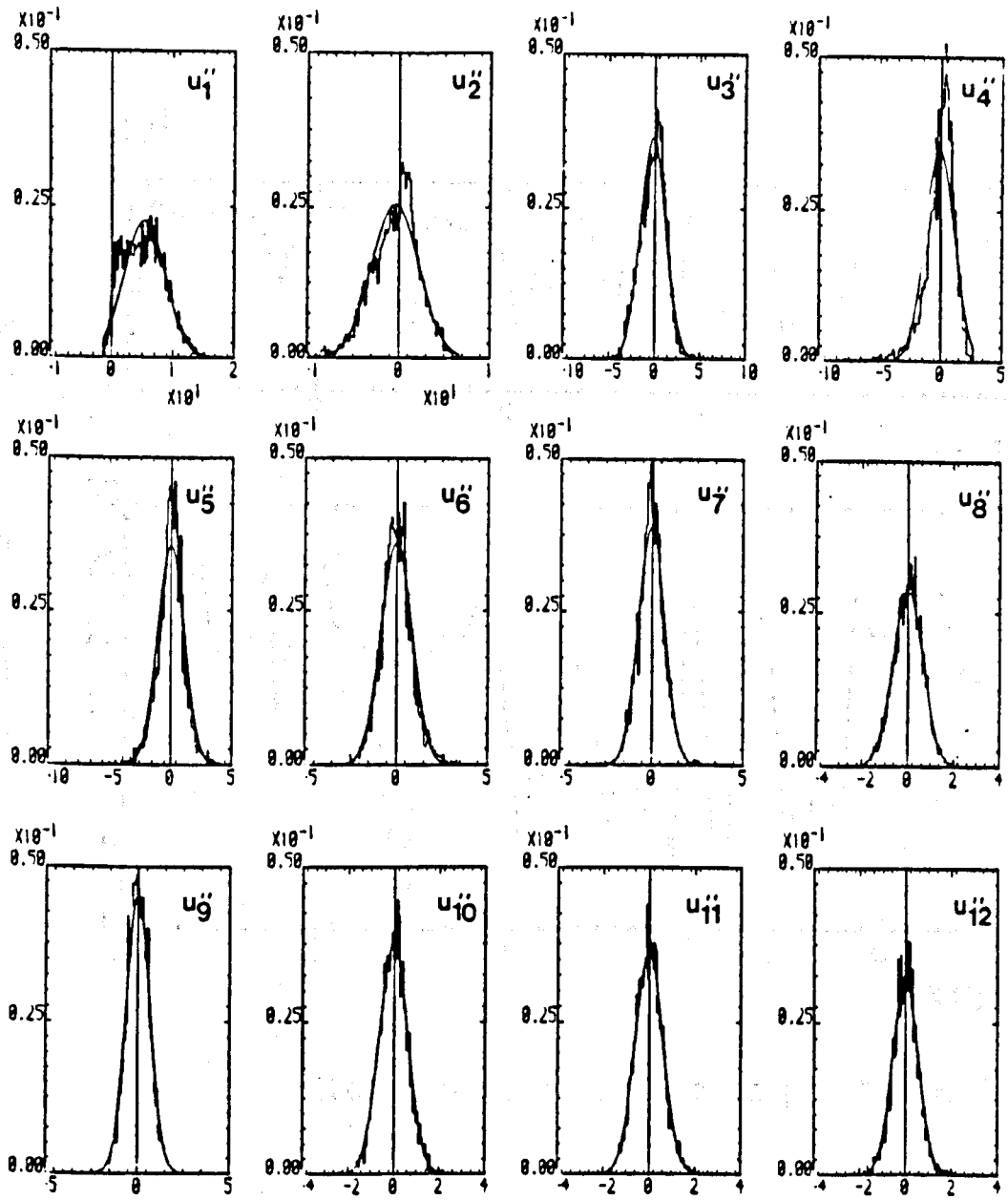


Figure 3.23 Histograms and theoretical Gaussian distribution function for the transform coefficients.

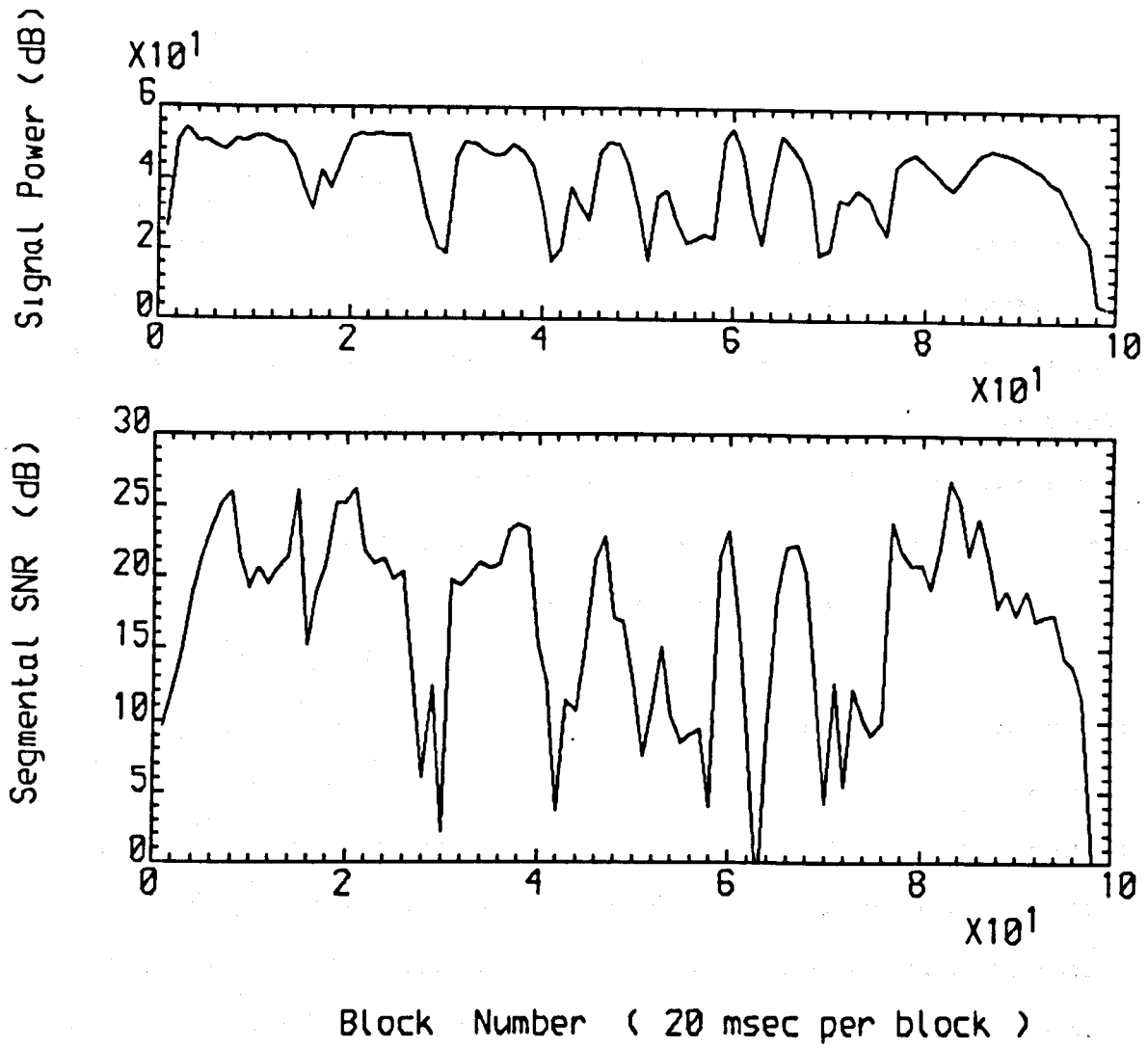


Figure 3.24 SNR_{seg} performance of 16 kbs^{-1} APC-AB coder.

CHAPTER 4 LINE SPECTRUM PAIRS (LSP)

4.1 INTRODUCTION

Line Spectrum Pair coefficients are linear prediction parameters, which may be used to model the human vocal tract. Speech is produced by a process in which the generation of the vocal tract excitation (voiced or unvoiced) and the articulation by resonances and anti-resonances in the vocal tract may be considered independently. Using linear predictive (LP) analysis, a time sequence of speech samples may be transformed into a set of filter coefficients for an FIR vocal tract analysis filter $A_p(z)$, whose impulse response samples are known as LP coefficients. The articulation due to the vocal tract shape can be modelled by an IIR synthesis filter, as illustrated in Figure 4.1a, whose transfer function $H_p(z)$ is the reciprocal of $A_p(z)$.

Linear prediction may be used to remove the redundancy in speech that is due to vocal tract resonances. It is thus possible to achieve better coding efficiency than is possible by straight forward waveform coding by transmitting quantised LP coefficients and an excitation signal. However, the LP coefficients are highly inter-dependent, and coarse quantization of these coefficients can produce large spectral distortion and can make the inverse filter become unstable. It was found [55] that the LPC coefficients must be quantized with no less than about 10 or 11 bits for each coefficients to assure stable operation. Such high bit rate for the coefficients would be difficult to accommodate in a low bit rate system.

PARCOR coefficients, often referred to as reflection coefficients or k coefficients, have been found to overcome some of the quantization

disadvantages of LP coefficients. The k coefficients can be made less sensitive to the effects of quantization noise which will therefore cause less spectral distortion [47,48]. Figure 4.1b shows a non-recursive 'lattice' digital analysis filter whose multipliers are the k coefficients. The synthesis filter is a recursive lattice structure which can be made equivalent to the filter in Figure 4.1a by a suitable choice of k coefficients. It may be shown that the recursive lattice filter will not be unstable as long as the magnitudes of all the k coefficients are smaller than unity. Hence, for a stable synthesis filter, all k coefficients will be less than one in modulus, and this makes quantization of k coefficients more convenient than it is for LP coefficients. A weakness of reflection coefficients is that, as for LP coefficients, a change in one coefficient causes spectral changes in the entire frequency band of a speech signal. Hence, it is difficult to relate the effect of quantising individual k coefficients to perceptual criteria.

Recently, an alternative representation of the LPC synthesis filter, based on Line Spectrum Pair (LSP) coefficients [56,57] has been reported to provide an even better coding efficiency than is possible with PARCOR coefficients. LSP coefficients are much more closely related to the frequencies and bandwidth of speech formants than k or LP coefficients. Each LSP coefficient is effectively a frequency. An error in one LSP coefficient tends to affect the synthesised spectrum only locally, near to the frequency of the LSP coefficient. This means that using LSP coefficients allows quantization effects to be distributed in the frequency domain in accordance with properties of auditory perception to save bits. For example, coarser quantization may be used for the higher fre-

quency formant positions since errors in these may be less perceptually objectionable.

The method originally proposed [57] for calculating LSP coefficients was to obtain LP coefficients via a standard autocorrelation analysis technique based on a stored block of speech samples and then to carry out a transformation from LP to LSP coefficients using root finding computations. This technique is mathematically quite complicated. A new method of calculating LSP coefficients, based on the least mean square (LMS) algorithm of Widrow [58], has been developed to simplify the computation. The new method will be referred to as the LMS-LSP method. This Chapter describes the fundamental concept of LSP coefficients, and the new technique for deriving them.

The organisation of this Chapter is as follows. Section 4.2 describes the analysis and synthesis LSP filters, conventional methods for deriving LSP coefficients, and the basic properties of these coefficients. In Section 4.3, the LMS-LSP algorithm for deriving LSP coefficients is presented, and a performance measure for this technique is defined. In Section 4.4, the LMS-LSP method is extended by introducing a second order optimisation technique to enhance the rate of adaptation. This modified method and results obtained using it are also presented. Section 4.5 describes the application of the LMS-LSP scheme to the APC-AB speech coder to assess the practicability of this new algorithm.

4.2 FUNDAMENTAL CONCEPTS OF LSP COEFFICIENTS

As previously mentioned, the human speech production system can be represented by an LP model as in Figure 4.1a, and the transfer function of the vocal tract model can be written as

$$H_p(z) = \frac{1}{A_p(z)} \qquad (4.1)$$

$A_p(z)$ is the inverse vocal tract transfer function given by

$$A_p(z) = 1 + \sum_{i=1}^{p} a_i z^{-i} = \prod_{i=1}^{p} (1 - z_i z^{-1}) \qquad (4.2)$$

where p is the order of the LP model, a_i is the i-th LP coefficient and z_i is the i-th zero of $A_p(z)$. For a stable system, $A_p(z)$ has to be minimum phase which means that all the poles of $H_p(z)$ must be inside the unit circle in the Argand z-plane.

In general, the a_i coefficients can be calculated by autocorrelation analysis as described in Section 3.3.3.2. These coefficients may be used with an analysis filter $A_p(z)$ to transform speech samples into prediction residue samples. The original speech can be reconstructed from the residue by using the synthesis filter, $H_p(z)$ as described in Section 3.3.3.2.

Alternatively, the LP model can be realised in the form of a recursive lattice structure as in Figure 4.1b. A p-th order lattice filter has the relations:

$$A_n(z) = A_{n-1}(z) + k_n z^{-1} B_{n-1}(z), \qquad A_0(z) = 1 \qquad (4.3a)$$

$$B_n(z) = k_n A_{n-1}(z) + z^{-1} B_{n-1}(z), \qquad B_0(z) = 1 \qquad (4.3b)$$

$$B_n(z) = z^{-n} A_n(z^{-1}) \qquad (4.3c)$$

where $A_n(z)$ and $B_n(z)$ are defined as the n-stage forward and backward going signals respectively and the k coefficients are k_n for n=1,2,..p. This lattice structure can be thought of as describing the vocal tract as a series of short, connected tubes with uniform cross-sectional area. Each k coefficient determines what proportion of the sound pressure travelling wave energy is reflected at each boundary between sections [44].

4.2.1 LSP ANALYSIS FILTER

Consider a p-th order lattice filter augmented by an extra stage as depicted in Figure 4.2 with $k_{p+1} = -1$ or $+1$. The system function thus produced is a (p+1)-th order polynomial, $P_p(z)$, when $k_{p+1}=-1$ and a different (p+1)-th order polynomial, $Q_p(z)$, when $k_{p+1}=1$. $P_p(z)$ can be considered to represent an entirely opened acoustic tube consisting of p+1 sections [56]. $Q_p(z)$ can be considered as representing an entirely closed (p+1)-th section acoustic tube. The new polynomials are

$$P_p(z) = A_p(z) - z^{-1}B_p(z), \qquad k_{p+1} = -1 \tag{4.4a}$$

$$Q_p(z) = A_p(z) + z^{-1}B_p(z), \qquad k_{p+1} = 1\ . \tag{4.4b}$$

It may be shown that if $A_p(z)$ is minimum phase, the zeros of $P_p(z)$ and $Q_p(z)$ are all on the unit circle and that they are interlaced with each other, i.e. before each pair of zeros of $P_p(z)$ there will be a zero of $Q_p(z)$ and vice-versa. This will be shown below in Section 4.2.3. The zeros of $P_p(z)$ and $Q_p(z)$ can therefore be expressed in terms of angular frequencies and these frequencies are called the Line Spectrum Pairs coefficients.

Combining (4.4a) and (4.4b), the LP analysis filter system function is

$$A_p(z) = \tfrac{1}{2}\,[P_p(z) + Q_p(z)] \ . \qquad (4.5)$$

A realisation of $A_p(z)$ is drawn in Figure 4.3 and will be referred to as an 'LSP filter'.

To show that all the zeros of $P_p(z)$ are on the unit circle in the z-plane, $A_p(z)$ expressed by (4.2) is substituted into (4.4a). Thus,

$$P_p(z) = 1 + \sum_{i=1}^{p} a_i\, z^{-i} - z^{-(p+1)} \left(1 + \sum_{i=1}^{p} a_i\, z^{i}\right) \qquad (4.6a)$$

$$= \sum_{i=0}^{p+1} f_{p,i}\, z^{-i} \qquad (4.6b)$$

where $f_{p,0} = 1$

$f_{p,i} = a_i - a_{p+1-i}, \qquad i = 1,2,.....,p$

$f_{p,i} = - f_{p,p-i+1}$ (anti-symmetric property).

When p is even, z = 1 is a zero of P(z). Hence,

$$P_p(z) = (1 - z^{-1})[1 + b_1 z^{-1} \ ..... + b_1 z^{-p+1} + z^{-p}] \qquad (4.7a)$$

$$= (1 - z^{-1})\, C_{p-1}(z) \ . \qquad (4.7b)$$

When p is odd, z = ±1 are zeros of P(z). Hence,

$$P_p(z) = (1 - z^{-2})[1 + b_1 z^{-1} \ ..... + b_1 z^{-p+2} + z^{-p+1}] \qquad (4.8a)$$

$$= (1 - z^{-2})\, C_{p-2}(z) \ . \qquad (4.8b)$$

$C_{p-1}(z)$ is a symmetric polynomial and the first and last coefficients are unity. As the coefficient of the last term of $C_{p-1}(z)$ is unity, this implies that the product of the radii of all the zeros is unity. This can only be true when all the zeros of $P_p(z)$ are on the unit circle. Further, the coefficients of $C_{p-1}(z)$ are real so that the zeros of $C_{p-1}(z)$ occur in complex conjugate pairs. Thus, when p is even, $P_p(z)$ may be expressed as

$$P_p(z) = (1 - z^{-1}) \prod_{i=1}^{p/2} (1 - \exp(j\omega_i) z^{-1})\, (1 - \exp(-j\omega_i) z^{-1}) \qquad (4.9a)$$

$$= (1 - z^{-1}) \prod_{i=1}^{p/2} (1 + c_i z^{-1} + z^{-2}) \qquad (4.9b)$$

When p is odd,

$$P_p(z) = (1 - z^{-2}) \prod_{i=1}^{\frac{1}{2}(p-1)} (1 + c_i z^{-1} + z^{-2}) \qquad (4.10)$$

where

$$c_i = -2 \cos \omega_i \qquad \text{for} \quad 0 < \omega_i < \pi,$$

ω_i is an angle of the i-th zero of $P_p(z)$ the unit circle.

Similarly, to show that all the zeros of $Q_p(z)$ are all on the unit circle in the z-plane, $A_p(z)$ of (4.2) is substituted in (4.4b). Thus

$$Q_p(z) = 1 + \sum_{i=1}^{p} a_i z^{-i} + z^{-(p+1)} \left(1 + \sum_{i=1}^{p} a_i z^{i}\right) \qquad (4.11a)$$

$$= \sum_{i=0}^{p+1} g_{p,i} z^{-i} \qquad (4.11b)$$

where $g_{p,0} = 1$

$g_{p,i} = a_i + a_{p+1-i}, \qquad i = 1,2,.....,p$

$g_{p,i} = g_{p,p-i+1}$ (symmetric property).

When p is even, z = -1 is a zero of $Q_p(z)$. Hence,

$$Q_p(z) = (1 + z^{-1}) D_{p-1}(z) . \qquad (4.12)$$

When p is odd, $Q_p(z)$ has no zero on the real axis in the z-plane. Then, $D_{p-1}(z)$ has similar form as $P_p(z)$ as in (4.7) and (4.8). Thus, when p is even, Q_p may be expressed as

$$Q_p(z) = (1 + z^{-1}) \prod_{i=1}^{p/2} (1 - \exp(j\theta_i) z^{-1}) (1 - \exp(-j\theta_i) z^{-1}) \qquad (4.13a)$$

$$= (1 + z^{-1}) \prod_{i=1}^{p/2} (1 + d_i z^{-1} + z^{-2}) . \qquad (4.13b)$$

When p is odd,

$$Q_p(z) = \prod_{i=1}^{\frac{1}{2}(p+1)} (1 + d_i z^{-1} + z^{-2}) \qquad (4.14)$$

where

$$d_i = -2 \cos \theta_i \qquad \text{for} \quad 0 < \theta_i < \pi.$$

θ_i is the angle of the i-th zero on the unit circle in the z-plane.

The parameters ω_i and θ_i are the Line Spectrum Pair coefficients, or alternatively as the Line Spectrum Frequencies.

From (4.5), the LP analysis filter can be realised as a parallel combination of the $P_p(z)$ and $Q_p(z)$ filters. Based upon (4.9b) and (4.13b), a 4-th order LSP analysis filter example is shown in Figure 4.4a.

4.2.2 LSP SYNTHESIS FILTER

The LP synthesis process is the inverse of the analysis process. The basic configuration of the LSP synthesis filter is drawn in Figure 4.3, and it has the transfer function

$$H_p(z) = \frac{1}{A_p(z)} = \frac{1}{\frac{1}{2}\,[P_p(z) + Q_p(z)]} \qquad (4.15)$$

As the synthesis filter contains a feedback path from the output to the input, to realise the filter in practice, (4.15) is modified to :

$$H_p(z) = \frac{1}{1 + [A_p(z)-1]} = \frac{1}{1 + \frac{1}{2}[(P_p(z)-1)+(Q_p(z)-1)]} \qquad (4.16)$$

The feedback element is the quantity in the square bracket.

When p is even, the transfer function of the feedback path is

$$A_p(z) - 1 = \tfrac{1}{2}[(P_p(z)-1) + (Q_p(z)-1)] \qquad (4.17)$$

$$= \frac{1}{2}\left\{\left[(1-z^{-1})\prod_{i=1}^{p/2}(1+c_i z^{-1}+z^{-2}) - 1\right] + \left[(1+z^{-1})\prod_{i=1}^{p/2}(1+d_i z^{-1}+z^{-2}) - 1\right]\right\}$$

$$= \frac{1}{2}\left\{\left[\prod_{i=1}^{p/2}(1+c_i z^{-1}+z^{-2}) - 1 - z^{-1}\prod_{i=1}^{p/2}(1+c_i z^{-1}+z^{-2})\right] + \left[\prod_{i=1}^{p/2}(1+d_i z^{-1}+z^{-2}) - 1 + z^{-1}\prod_{i=1}^{p/2}(1+d_i z^{-1}+z^{-2})\right]\right\}$$

$$= \frac{z^{-1}}{2}\left\{\left[(c_1+z^{-1}) + \sum_{i=1}^{\frac{1}{2}p-1}(c_{i+1}+z^{-1})\prod_{j=1}^{i}(1+c_j z^{-1}+z^{-2}) - \prod_{i=1}^{p/2}(1+c_i z^{-1}+z^{-2})\right] + \left[(d_1+z^{-1}) + \sum_{i=1}^{\frac{1}{2}p-1}(d_{i+1}+z^{-1})\prod_{j=1}^{i}(1+d_j z^{-1}+z^{-2}) + \prod_{i=1}^{p/2}(1+d_i z^{-1}+z^{-2})\right]\right\}. \quad (4.18)$$

Likewise, when p is odd, the transfer function of the feedback path can be shown as

$$A_p(z) - 1 = \frac{z^{-1}}{2}\left\{\left[(c_1+z^{-1}) + \sum_{i=1}^{\frac{1}{2}(p-3)}(c_{i+1}+z^{-1})\prod_{j=1}^{i}(1+c_j z^{-1}+z^{-2})\right.\right.$$

$$- z^{-1} \prod_{i=1}^{\frac{1}{2}(p-1)} (1+c_i z^{-1}+z^{-2}) \Bigg] \Bigg\}$$

$$+ \left[(d_1+z^{-1}) + \sum_{i=1}^{\frac{1}{2}(p-1)} (d_i+1+z^{-1}) \prod_{j=1}^{i} (1+d_j z^{-1}+z^{-2}) \right] \Bigg\} \quad (4.19)$$

Figure 4.4b shows the structure of a 4-th order LSP synthesis filter, and in general it may be deduced that a p-th order speech synthesis filter requires p multiplications and 3p+1 additions to compute a filter output. The amount of computation is only slightly more than for a transversal filter or lattice filter synthesis system.

4.2.3 LSP PROPERTY

The LSP coefficients are determined from the zeros of $P_p(z)$ and $Q_p(z)$. If $A_p(z)$ is minimum phase, the LSP coefficients of $P_p(z)$ alternate with those of $Q_p(z)$,

$$0 < \theta_1 < \omega_1 < \theta_2 < \omega_2 < .... < \theta_{\frac{1}{2}p} < \omega_{\frac{1}{2}p} < \pi \quad (4.20)$$

This property is useful for determining whether the LSP synthesis filter is stable, and will now be proved.

The transfer function of an LP analysis filter, $A_p(z)$, can be expressed as in (4.2):

$$A_p(z) = \prod_{i=1}^{p} (1 - z_i z^{-1}) \quad (4.21a)$$

Whence, the transfer function of the lattice backward path is

$$B_p(z) = z^{-p} A_p(z^{-1}) \quad (4.21b)$$

$$= \prod_{i=1}^{p} (z^{-1} - z_i) \quad (4.21c)$$

The frequency response of the analysis filter may be expressed as :

$$A_p(e^{j\omega}) = R(\omega)\, e^{j\phi(\omega)} \tag{4.22}$$

where $R(\omega)$ and $\phi(\omega)$ are the magnitude and phase response respectively. If all the coefficients of $A_p(z)$ are real, the complex conjugate function can be written as

$$A_p^*(e^{j\omega}) = A_p(e^{-j\omega}) . \tag{4.23}$$

Hence, by (4.21b) the frequency response of $B_p(z)$ is

$$B_p(e^{j\omega}) = e^{-j\omega p}\, R(\omega)\, e^{-j\phi(\omega)} . \tag{4.24}$$

Using (4.22), (4.24) and (4.4a), the frequency response of the $P_p(z)$ filter is

$$P_p(e^{j\omega}) = R(\omega)\, e^{-j\phi(\omega)} - e^{-j\omega} e^{-j\omega p}\, R(\omega)\, e^{-j\phi(\omega)} \tag{4.25a}$$

$$= j2R(\omega)\, e^{-j(p+1)\omega/2} \sin[\phi + (p+1)\omega/2] . \tag{4.25b}$$

Similarly, the frequency response of the $Q_p(z)$ filter is

$$Q_p(e^{j\omega}) = 2\, R(\omega)\, e^{-j(p+1)\omega/2} \cos[\phi + (p+1)\omega/2] . \tag{4.26}$$

As ω is varied from 0 to 2π, $\phi+(p+1)\omega/2$ increases from ϕ to $(p+1)\pi$. $P_p(z)$ and $Q_p(z)$ each have p+1 zeros on the unit circle. To obtain these zeros, we set $P_p(e^{j\omega})=0$ and $Q_p(e^{j\omega})=0$. As $A_p(z)$ is minimum phase, it cannot have a zero on the unit circle and therefore $R(\omega)$ is non-zero for all ω. Hence, when $P_p(e^{j\omega})=0$, it follows that

$$\sin[\phi + (p+1)\omega/2] = 0, \tag{4.27}$$

so that $\omega = \omega_n = 2(n\pi - \phi)/(p+1)$: $n = 0,1,\ldots$.

Similarly, for the $Q_p(z)$ filter, when $Q(e^{j\omega}) = 0$, it follows that

$$\cos[\phi + (p+1)\omega/2] = 0, \tag{4.28}$$

so that $\omega = \theta_n = [(2n+1)\pi - 2\phi]/(p+1)$: $n = 0,1,\ldots$.

Thus, the $P_p(z)$ and $Q_p(z)$ filters have zeros at ω_n and θ_n respectively. As the expressions are sine and cosine functions of the same argument, the zeros of $P_p(z)$ and $Q_p(z)$ interleave each other.

A fourth order example is shown in Figure 4.5. If there are two complex conjugate pole pairs at z_1 and z_2, where

$$|z_1| = 0.9, \quad \arg(z_1) = \pm 20^\circ \quad ; \qquad |z_2| = 0.7, \quad \arg(z_2) = \pm 70^\circ$$

the LSP coefficients are on the unit circle and alternate. It can be shown that the pair, θ_1 and ω_1, is enclosing the pole z_1, and the pair, θ_2 and ω_2, is enclosing z_2. It can also be shown that if a pair of LSP coefficients are close, this indicates a strong resonance near or between the pair [59]. This interesting property is useful for formant extraction in the field of speech synthesis [81].

4.2.4 CONVERSION OF LP COEFFICIENTS TO LSP COEFFICIENTS

LSP coefficients can be obtained from a set of LP coefficients by means of a transformation. Conversely, given a set of LSP coefficients, it is possible to transform them to LP coefficients. The conversion from LP coefficients to LSP coefficients involves a process which is equivalent to finding zeros of the $P_p(z)$ and $Q_p(z)$ filters. The following section describes the computational steps for the conversion of LP coefficients to LSP coefficients.

Assuming that a set of LP coefficients, $\{a_1, a_2, \ldots, a_p\}$, where p is the order of the LPC model, are computed by LP analysis using one frame of speech samples, the procedure for deriving the LSP coefficients from the LP coefficients is:

a) Calculate the polynomial coefficients of $P_p(z)$. By (4.6b), the coefficients are

$$f_{p,0} = 1; \quad f_{p,p+1} = -1, \tag{4.29}$$
$$f_{p,i} = a_i - a_{p+1-i} \qquad i = 1,2,\ldots,p$$

where $f_{p,i}$ is the coefficient of z^{-i} in the polynomial $P_p(z)$.

b) Likewise, by (4.13) the polynomial coefficients of $Q_p(z)$ are calculated as

$$g_{p,0} = 1; \quad g_{p,p+1} = 1 \tag{4.30}$$
$$g_{p,i} = a_i + a_{p+1-i} \qquad i = 1,2,\ldots,p$$

where $g_{p,i}$ is the coefficient of z^{-i} in the polynomial $Q_p(z)$.

c) When p is even, $z = 1$ is a zero of $P_p(z)$ which can be factored out from $P_p(z)$ to simplify computation. Thus,

$$P_p(z) = (1-z^{-1})\, C_{p-1}(z) \tag{4.31}$$

and the polynomial coefficients of $C_{p-1}(z)$, $f_{p-1,i}$ say, can be computed recursively as:

$$f_{p-1,0} = f_{p,0} = 1 \tag{4.32}$$
$$f_{p-1,i} = f_{p,i} + f_{p-1,i-1} \qquad i = 1,2,\ldots,p\ .$$

Likewise, $(1+z^{-1})$ can be factored out from $Q_p(z)$. Thus,

$$Q_p(z) = (1+z^{-1})\, D_{p-1}(z) \tag{4.33}$$

and the polynomial coefficients of $D_{p-1}(z)$, $g_{p-1,i}$ say, can be computed as:

$$g_{p-1,0} = g_{p,0} = 1 \tag{4.34}$$
$$g_{p-1,i} = g_{p,i} - g_{p-1,i-1} \qquad i = 1,2,\ldots,p\ .$$

When p is odd, $z = \pm 1$ are zeros of $P_p(z)$. Thus,

$$P_p(z) = (1-z^{-2})\, C_{p-2}(z) \tag{4.35}$$

and the polynomial coefficients of $C_{p-2}(z)$, $f_{p-2,i}$ say, are

$$f_{p-2,0} = f_{p,0} = 1, \tag{4.36}$$

$$f_{p-2,1} = f_{p,1}$$

$$f_{p-2,i} = f_{p,i} + f_{p-2,i-2} \qquad i = 2,3,\ldots,p-2.$$

For odd p, a zero of $Q_p(z)$ on the real axis in the z-plane is not necessary present.

d) Use a root finding algorithm, for example the Newton-Raphson method, to determine the zeros of the factorised polynomials. Then, the LSP coefficients are computed as the arguments of the complex zeros.

If p is small (say less than 9) [54], analytical methods may be more efficient than root finding methods. For one such analytical methods, the second order factors of $P_p(z)$ as in (4.9b) or (4.10) are multiplied out. Then, the polynomial coefficients which are expressed in terms of the c_i coefficients are compared with the coefficients of polynomials $C_{p-1}(z)$ (if p is even) or $C_{p-2}(z)$ (if p is odd) respectively. As the coefficients of $C_{p-1}(z)$ and $C_{p-2}(z)$ have symmetrical properties, the LSP coefficients of a p-th order system may be derived by first solving a set of p/2 non-linear equations to obtain the c_i coefficients in (4.10). Similarly, the LSP coefficients of the $Q_p(z)$ filter can also be determined analytically by solving the equations.

Other algorithms for converting LP coefficient into LSP representation are given in references [55,60,61,62].

4.2.5 CONVERSION OF LSP TO LP COEFFICIENTS

LP coefficients can be computed from LSP coefficients by expressing the coefficients of the polynomial $A_p(z)$, in terms of LSP filter coefficients. When p is even,

$$A_p(z) = \frac{1}{2}\left[(1-z^{-1})\prod_{i=1}^{p/2}(1+c_i z^{-1}+z^{-2}) + (1+z^{-1})\prod_{i=1}^{p/2}(1+d_i z^{-1}+z^{-2})\right]. \quad (4.37)$$

When the product terms are multiplied out, the resulting polynomial is in the form

$$A_p(z) = 1 + a_1 z^{-1} + a_2 z^{-2} + \ldots\ldots\ldots + a_p z^{-p} . \quad (4.38)$$

Comparing the equal power terms of (4.37) and (4.38), the equivalent LP coefficients are then computed from the LSP filter coefficients. This process is the same for an odd order analysis filter.

Alternatively, the LP coefficients can be obtained by calculating p+1 impulse response samples of the LSP analysis filter, by passing a unity impulse through the digital LSP analysis filter realised as shown in Figure 4.4a.

4.3 SEQUENTIAL ADAPTIVE LSP METHOD

The computational method for calculating LSP coefficients described in the preceding section involves an autocorrelation function calculation, matrix inversion and an root finding algorithm. This is referred to as a block method, since each calculation is based on a block of speech samples. LPC parameters can also be obtained in a continuous process using a recursive algorithm that sequentially updates the parameters upon the arrival of each new speech sample. Such adaptive techniques, which provide new estimates for the LPC parameters for every speech sample, are often called sequential predictive analysis methods. There are two advantages of sequential methods as compared with block methods. Firstly, they provide a choice of various sets of filter coefficients within a particular speech interval [63]. Having such a choice

might be important in obtaining consistent spectral estimates that are not as affected by the quasi-periodic nature of voiced speech. Secondly, the hardware implementation of these predictive methods can be made simpler.

In this section, a new adaptive predictive scheme [64] is investigated for extracting the LSP coefficients. By this method, LSP coefficients are obtained directly, without intermediate conversion from LP coefficients. The new method is based on Widrow's LMS steepest descent algorithm which is applied to a specially chosen filter structure.

4.3.1 THE RECURSIVE LMS ALGORITHM

In Section 3.3.3.2, it was proved that the mean square error output of an LPC analysis filter can be expressed as given in (3.56):

$$E[e^2(n)] = r_0 + 2\,\underline{a}^T\underline{p} + \underline{a}^T R\,\underline{a} \tag{4.39}$$

where $\underline{a}$ is the filter weight vector defined by (4.2) that could be converted to a set of LSP coefficients. R, r_0 and $\underline{p}$ are autocorrelation functions defined in (3.56). For a stationary input signal, the mean square error is a quadratic function of the filter weights and can be pictured as a concave hyperparabolic surface, a function with a unique minimum. At minimum, the corresponding weight vector, $\underline{a}_{opt}$, is known as the optimal Wiener weight vector, which is evaluated as

$$\underline{a}_{opt} = -R^{-1}\underline{p} \tag{4.40}$$

and the minimum of the error surface is written as

$$E[e^2(n)]_{min} = r_0 - \underline{p}^T R^{-1}\underline{p} \tag{4.41}$$

which is known as Wiener minimum mean square error (m.m.s.e.).

If the steepest descent method is applied to compute $\underline{a}_{opt}$, the weight vector is sequentially adjusted with the objective of seeking the minimum of the error surface, and the changes in the weight vector are proportional to the negative of the gradient of the mean squared error function [58]. Accordingly,

$$\underline{a}_{n+1} = \underline{a}_n + \mu\ (-\nabla_n) \tag{4.42}$$

where μ is a small scalar which controls the stability and rate of convergence of this iterative scheme, and ∇_n is the gradient vector of the mean squared error function at time instant n which is defined as

$$\nabla_n = (\ \xi_1\ \ \xi_2\ \ \ldots\ldots\ldots\ \ \xi_p\ )^T \tag{4.43}$$

where

$$\xi_i = \frac{\partial\ E[e^2(n)]}{\partial\ a_i}.$$

By Widrow's LMS scheme, a stochastic approximation is taken to estimate the gradient vector by considering the square of each instantaneous error sample $e^2(n)$ as an estimate of the mean square error, i.e.,

$$\nabla_n \approx \nabla_n' = (\ \xi_1'\ \ \xi_2'\ \ \ldots\ldots\ldots\ \ \xi_p'\ )^T \tag{4.44}$$

where

$$\xi_i' = \frac{\partial\ e^2(n)}{\partial\ a_i}.$$

The gradient estimate of this LMS algorithm is asymptotically unbiased because if the expected value of the gradient estimate is considered as the average of an arbitrarily large number of samples :

$$E[\ \nabla_n'\ ] = \frac{1}{N} \sum_{k=1}^{N} \nabla_n' \tag{4.45}$$

$$= \frac{1}{N} \sum_{k=1}^{N} \frac{\partial e^2(k)}{\partial \underline{a}_n}$$

$$= \frac{\partial}{\partial \underline{a}_n} \left[\frac{1}{N} \sum_{k=1}^{N} e^2(k) \right]$$

$$\approx \frac{\partial}{\partial \underline{a}_n} E[e^2(n)] = \nabla_n$$

which is the true gradient of the error surface.

By extending the LMS approach to the LSP filter system, the LSP filter coefficient update formula is expressed as two equations:

$$\underline{c}_{n+1} = \underline{c}_n - \mu_c \nabla_{n,c} \tag{4.46a}$$

$$\underline{d}_{n+1} = \underline{d}_n - \mu_d \nabla_{n,d} \tag{4.46b}$$

where $\nabla_{n,c}$ and $\nabla_{n,d}$ are the estimated gradient vectors for the upper arm and lower arm of the LSP filter respectively. The gradient vectors are defined as

$$\nabla_{n,c} = (\omega_{c,1} \quad \omega_{c,2} \quad \cdots\cdots\cdots\cdots \quad \omega_{c,\frac{1}{2}p})^T \tag{4.47}$$

$$\nabla_{n,d} = (\omega_{d,1} \quad \omega_{d,2} \quad \cdots\cdots\cdots\cdots \quad \omega_{d,\frac{1}{2}p})^T$$

where

$$\omega_{c,i} = \frac{\partial e^2(n)}{\partial c_i} \quad \text{and} \quad \omega_{d,i} = \frac{\partial e^2(n)}{\partial d_i} \quad \text{for} \quad i=1,2,..,p/2.$$

Constants μ_c and μ_d are the convergence factors. In order to adapt the LSP filter using the LMS algorithm, the gradient estimate components $\omega_{c,i}$ and $\omega_{d,i}$ are required. In the following section, the derivation of these gradient estimates will be described.

4.3.2 COMPUTATION OF THE GRADIENT VECTOR

For the LSP analysis filter, the output error signal can be expressed as

$$e(n) = \tfrac{1}{2} [\, p(n) + q(n) \,] \tag{4.48}$$

where p(n) and q(n) are the outputs of the upper and lower arm of the LSP filter respectively. As p(n) is independent of $d_i(n)$ for all i and q(n) is independent of $c_i(n)$ for all i, the gradient estimate of the LSP filter can then be written as :

$$\omega_{c,i} = e(n) \frac{\partial\, p(n)}{\partial c_i(n)} \tag{4.49}$$

$$\omega_{d,i} = e(n) \frac{\partial\, q(n)}{\partial d_i(n)} \qquad \text{for } i = 1, 2, \ldots, p/2.$$

To simplify the analysis, the convergence factors are set equal. They are

$$\mu = \mu_c = \mu_d \tag{4.50}$$

Hence, for even p, (4.46a) and (4.46b) may be more conveniently be expressed as

$$\underline{c}_{n+1} = \underline{c}_n - \mu\, e(n)\, F(n) \tag{4.51}$$
$$\underline{d}_{n+1} = \underline{d}_n - \mu\, e(n)\, G(n)$$

where

$$F(n) = (f_1(n)\ f_2(n)\ \ldots\ldots\ldots\ldots\ f_{\frac{1}{2}p}(n))^T \tag{4.52}$$
$$G(n) = (g_1(n)\ g_2(n)\ \ldots\ldots\ldots\ldots\ g_{\frac{1}{2}p}(n))^T .$$

where

$$f_i(n) = \frac{\partial p(n)}{\partial c_i(n)} \quad \text{and} \quad g_i(n) = \frac{\partial q(n)}{\partial d_i(n)} \quad \text{for } i=1,2,..,p/2 .$$

Taking the z-transform, the transformed gradient vectors can be obtained by differentiating (4.9b) and (4.13b) to obtain:

$$\frac{\partial p(z)}{\partial c_i} = z^{-1}(1 - z^{-1}) \prod_{\substack{\ell=1 \\ \ell \neq i}}^{p/2} (1 + c_\ell z^{-1} + z^{-2}) \tag{4.53}$$

and similar expressions for the derivatives of Q(z) may be obtained. If $P_i(z)$ is defined as the transfer function between the input to the first second order filter section and the output of the i-th second order filter section on the upper arm of the LSP filter, (4.53) can be written as

$$\frac{\partial p(z)}{\partial c_i} = z^{-1} P_i(z) \prod_{l=i+1}^{p/2} (1 + c_l z^{-1} + z^{-2}) \tag{4.54}$$

This expression can be considered as the transfer function of p/2 'adjoint' filters, each of which will output a gradient estimate component when fed with the same input as the adaptive filter. An arrangement for a sixth order adaptive LSP filter is drawn, Figure 4.6, which shows that the 'adjoint' filter can be combined with the LSP filter to produce an augmented filter which generates $f_i(n)$ and $g_i(n)$, as well as the output e(n).

Note that (4.54) can be expressed as

$$\frac{\partial p(z)}{\partial c_i} = \frac{z^{-1} P(z)}{1 + c_i z^{-1} + z^{-2}} = z^{-1} P(z) A_i(z) \tag{4.55}$$

where $A_i(z)$ corresponds to the inverse of the i-th second order section of the P(z) filter, it seems that a computational more efficient method for generating the gradient from p(n) can be constructed as displayed in Figure 4.7. However, as $A_i(z)$ is a recursive filter and its poles will all lie on the unit circle in the z-plane. This will result in an unstable filter and cause unstable gradient estimations. Hence, this is not an advisable method and the gradient estimates should be computed from (4.54).

With the gradient estimate components now available, the LSP-LMS algorithm (4.51) may be implemented and the optimal weights can in principle be approached from any arbitrary initial weight vector. For each input sample at time n, the gradient vectors and the analysis filter output e(n) are computed and this information is then used to update the LSP coefficient vectors $\underline{c}_{n+1}$ and $\underline{d}_{n+1}$ for use with the next input sample. With an appropriate value of the adaptation coefficient μ, the mean square value of e(n) is reduced towards its minimum possible value as the coefficient vectors are gradually changed, This process should converge on the stationary point (m.m.s.e) regardless of the choice of the initial coefficient vector.

During adaptation, statistical variation in gradient estimates or a value of μ too large could cause one or more of the zeros of the analysis filter to move outside the unit circle resulting in an unstable synthesis filter. However, if the LSP filter coefficients are constrainted to ensure that

$$-2 < d_1 < c_1 < d_2 < \ldots\ldots\ldots < d_{\frac{1}{2}p} < c_{\frac{1}{2}p} < 2 \qquad (4.56)$$

as described in Section 4.2.3, the zeros of the analysis filter will all remain inside the unit circle of the z-plane. Hence, the above inequality is used at each iteration to ensure synthesis filter stability during adaptation.

4.3.4 PERFORMANCE TEST METHOD OF ADAPTIVE FILTER

The LMS scheme described above would behave differently for non-stationary as opposed to stationary signals. Filtering non-stationary signals is a major area of concern especially when the stochastic prop-

erties of the signals are unknown a priori [65,66]. With non-stationary signals being input to the adaptive filter, the bowl-shaped surface of the mean square function moves around, and the orientation and curvature of the bowl may also be changing. The adaptive process has to track the minimum of this surface fast enough to cope with the time varying nature of the input signal. Speech must in general be considered as an non-stationary signal. However, its statistics vary sufficiently slowly that it may be assumed to be stationary over intervals of up to 30 or 50 msec. This makes the LMS adaptive filter a possible candidate for analysing speech signals if the adaption algorithm is fast enough to adjust to the speech characteristics within such a short period.

The speed of convergence depends on the convergence factor as given in (4.51), which controls the step size between two iterations. Large values of μ produce a faster adaptation towards the minimum, but the LSP coefficient vector, will have a larger oscillation around the optimal coefficient vector which causes greater steady state estimation noise. For smaller values of μ, the adaptation takes smaller steps to approach the m.m.s.e. but will take more iterations to approach to m.m.s.e.. In either case, the adaptation would not settle exactly on the m.m.s.e., because the variance of the fluctuations of the adapting coefficients would never be exactly zero, even after the adaptation algorithm has reached an equilibrium or steady state.

With different values of μ, the performance of the adaptation system will behave differently. It is desired to choose an optimal μ which will be best suited to a particular system and the input signal. This section, details of a method of studying the behaviour of the adaptive process, and of choosing an optimal μ for adaptation on speech.

4.3.4.1 Performance Measure by Learning Curve

A way to assess the performance of an adaptive process is to plot a 'learning curve' [67]. As the basic idea of adaptive prediction is to adjust the filter parameters to minimise the mean square of the error output of the adaptive filter, it is therefore logical to judge the performance by plotting the expected mean square error at each stage of the 'learning' process as a function of the number of adaptive iterations.

Figure 4.8 shows a system which was used to study the response of adaptive filters to a stationary input signal. The stationary input signal, s(n), was obtained by driving a fixed filter, having an all-pole transfer function of the form given in (4.1), with a random Gaussian noise source g(n). Before starting the system, fixed pole positions were chosen for the fixed filter of the 'plant', and all the filters were started from zero states. The initial weights for the adaptive LSP filter were set at values equivalent to $\underline{a}_n = 0$. Then the operation of the system was as follows:

a) Generate a sequence of N coloured noise samples, s(n), by passing Gaussian noise through the fixed filter.

b) Input the coloured noise sequence to the adaptive LSP filter, and produce an LSP coefficient vector for each of the N samples.

c) Copy each LSP coefficient vector to a 'copy' filter, which has the same structure and length as the adaptive LSP filter. For each vector, supply a sequence of coloured noise as input and compute the mean square error, $E[e^2(n)]$.

d) A value of mean square error is therefore available for each of the N adaptation steps as a measure of how well the filter has adapted.

The result of this investigation is a learning curve similar to the broken curve in Figure 4.9. The smoothed learning curve shown by the continuous curve in Figure 4.9 is an ensemble average of 50 learning curves, each starting with the same initial coefficient vector and each having a different input derived from the same stochastic process. This smoothed learning curve reveals the underlying exponential nature associated with convergence of the coefficient vector. The horizontal broken line represents the optimum Wiener m.s.e. level that is equal to the variance of the white Gaussian noise. When the adaptive filter is in steady state or in equilibrium, its behaviour should be very similar to that of the inverse of the plant; the generated error signal e(n) should be very similar to the Gaussian noise signal g(n). Then, the ensemble learning curve should approach the Wiener minimum level.

An optimal value of μ can be estimated by measuring various ensemble averaged learning curves at different values of μ, which would give different adaptation speeds and equilibrium mean squared errors. In general, two criteria are employed in assessing the performance of an adaptive system. The first is known as misadjustment, and indicates the amount by which the output error power $E[e^2(n)]$ exceeds the m.m.s.e., ε_o^2. We now define the misadjustment, M, as this excess error power at equilibrium expressed as a fraction of ε_o^2. Thus,

$$M = \frac{| E[e^2(n)] - \varepsilon_o^2 |}{\varepsilon_o^2} \qquad (4.57)$$

and this is often expressed as a percentage. $E[e^2(n)]$ is the error signal power computed from the copy filter when the adaptive system is in equilibrium. The second performance criterion is the learning time which measures the number of iterations required to approach equilibrium. The

learning time is expressed as the number of iterations required for the output power of the adaptive filter to fall between 90% and 10% levels drawn between the input power level and its equilibrium level.

The computation of the settling level from the learning curves was carried out by computer simulation. It was found that the calculation of the mean squared error at each iteration, as described in step 'C' was computationally very expensive. As when the mean squared error is required to be calculated in equilibrium. In some cases, many thousands of iterations are required before the system reaches equilibrium, which means that the mean squared error at each iteration of the learning curve must be computed as the mean square of this large number of error samples. The computational effort required at step 'C' can be dramatically reduced by using (4.39).

To use (4.39), it is required to calculate the LP coefficient vector, $\underline{a}_n$, at each iteration. The correlation vector $\underline{p}$ and the matrix R for the coloured noise must also be calculated as described in Section 3.3.3.2. For the LSP adaptive system, the equivalent LP coefficient vector $\underline{a}_n$ can easily be obtained from the LSP coefficient vector at each iteration. The autocorrelation function which shows the long term characteristics of the coloured noise, is only calculated once for the whole learning curve. An accurate estimate of the autocorrelation function is required in order to acquire a more accurate mean square error estimation. Two methods can be used to estimate the autocorrelation functions. The simplest is to take a block of N coloured noise samples and compute

$$r_i = \frac{1}{N-i-1} \sum_{n=0}^{N-i-1} s_n s_{n+i} \qquad i = 0,1,2,\ldots,N-1 \quad (4.58)$$

The disadvantage of this method is that N must be very large to obtain an accurate estimate. Alternatively, the autocorrelation function can be computed from the impulse response of the fixed filter. If the Gaussian noise, g(n), has zero mean and unit variance, s(n) is the convolution of g(n) and the impulse response of the fixed filter h(n). This is written as

$$s(n) = \sum_{j=-\infty}^{\infty} g(j)\, h(n-j) \; . \tag{4.59}$$

Substituting (4.59) into (4.58), the autocorrelation function is

$$r_i = \frac{1}{N-i-1} \sum_{n=0}^{N-i-1} \sum_{j=-\infty}^{\infty} \sum_{\ell=-\infty}^{\infty} h(n-j)\, h(n+i-1)\, g(i)\, g(\ell) \tag{4.60}$$

Then

$$E[g(i)\, g(\ell)] = \begin{cases} 0 & j \neq \ell \\ 1 & j = \ell . \end{cases}$$

Hence,

$$r_i = \frac{1}{N-i-1} \sum_{n=0}^{N-i-1} h_n\, h_{n+i} \qquad i = 0,1,2,\ldots,N-1 \tag{4.61}$$

For a stable system, h(n) decays towards zero as n increases and therefore the number of terms needed to provide a good estimate of r_i, using (4.61) is generally not excessive. Typically a few hundred samples are more than sufficient. If the variance of the Gaussian noise is σ^2 rather than unity as was previously assumed, the autocorrelation function r_i' is

$$r_i' = \sigma^2 r_i$$

and the coloured noise power is σr_o. The autocorrelation function $\underline{p}$ and R can therefore be computed. Having found $\underline{p}$ and R, they can be used to compute the mean squared error at each iteration.

The learning curves shown in Figures 4.9 and 4.10 were produced by the method described above. They will be described in detail in the following sections.

So far, a performance test method for an adaptive algorithm has been described in detail. From this test, two system parameters, misadjustment and learning time can be calculated for a particular value of μ. In practice it is desirable to choose an "optimal" or preferred μ for optimal system performance. In the next section an empirical process to choose a preferred μ for a system is presented, and the performance of the adaptive LSP system will also be evaluated.

4.3.4.2 Evaluation of the Adaptive LSP Scheme

A coloured noise signal was produced by passing zero-mean, white Gaussian noise through a 4-th order all pole filter with pole positions fixed at

$$|z_1| = 0.9, \qquad \arg(z_1) = \pm 20^\circ \ ; \qquad \text{and}$$
$$|z_2| = 0.7, \qquad \arg(z_2) = \pm 70^\circ .$$

The variance of the Gaussian noise source was chosen as 0.25 and this produced a coloured noise signal of variance 6.667. The coloured noise signal was applied as input to an adaptive LSP filter as illustrated in Figure 4.8. During the adaptation process, LSP filter stability was ensured by ignoring any coefficient set updates which would put the LSP coefficients in an unstable order. Learning curves of the system for an ensemble average of 50 trails are illustrated in Figure 4.10. It may be seen that the learning time becomes smaller as μ is increased. For μ greater than 0.002, it is seen that the ensemble curves exhibit spikes;

the spikes are increasing in frequency as μ rises. The appearance of these spikes is taken as an indication of instability or the onset of instability in the adaptive system. This is due to the value of μ being too large to guarantee coefficient convergence.

The dependence of misadjustment, M, and the learning time, τ, on different values μ, which were computed from the ensemble average of 50 learning curves, are plotted in Figure 4.11. Figure 4.11b shows how M and τ are related as μ changes along the curve, and this is known as a performance curve which is thought to be a good indication of the performance of an adaptive scheme. To choose a preferred value of μ, a compromise is needed between fast adaption and small misadjustment. The preferred value must not be large enough to incur a risk of instability. From Figure 4.11, it was deduced that a value of μ of about 0.003 should give a reasonable compromise between the conflicting needs for small misadjustment and short learning time.

Using the preferred μ, a learning curve of 250 iterations is plotted in Figure 4.9. The broken curve is the learning curve of a Gaussian noise sequence and the continuous curve is the ensemble average of 50 learning curves. The system did converge towards the Wiener m.m.s.e.. However, it is apparent that the gradient estimates are very noisy in this example. This is because the gradient estimates of the LMS algorithm are obtained by taking the squares of instantaneous error samples as estimates of the mean square error. Significant noise will thus be present in these estimates during adaptation. For speech, this amount of gradient estimation noise may be unacceptable. Reducing μ reduces the noise in the gradient estimates, but also results in a slower adaptation time and possibly larger errors in tracking the changes in speech characteristics.

In the next section, an enhanced adaptive LSP algorithm will be described. This enhanced scheme improves the adaptation performance and dramatically minimises the adaptation noise.

4.4 IMPROVED ADAPTIVE LSP ALGORITHM

The previous section described a simple adaptive strategy which is in fact based on a first order Taylor series approximation for a function to be minimised [68]. This strategy could be improved if the Hessian matrix, H, of the Taylor series could be provided, thus allowing a second order method to be used to give a much faster convergence rate than the first order method. Second order method are, however, computationally expensive, and for real time operation it is necessary to consider some form of approximation. A modified LMS algorithm is now proposed based on such an approximation.

4.4.1 SECOND ORDER STEEPEST DESCENT METHOD

This enhanced method can be considered as an approximation to the generalised Newton-Raphson method [68]. The coefficient update formula for the Newton-Raphson method is :

$$\underline{a}_{n+1} = \underline{a}_n + \mu H^{-1} (- \nabla_n) \qquad (4.62)$$

where $\underline{a}$ is the coefficient vector, μ is a scalar convergence factor, and H is the Hessian matrix:

$$H = \begin{bmatrix} h_{11} & h_{12} & \cdots\cdots\cdots & h_{1p} \\ h_{21} & h_{22} & \cdots\cdots\cdots & h_{2p} \\ \cdot & \cdot & \cdots\cdots\cdots & \cdot \\ \cdot & \cdot & \cdots\cdots\cdots & \cdot \\ \cdot & \cdot & \cdots\cdots\cdots & \cdot \\ h_{p1} & h_{p2} & \cdots\cdots\cdots & h_{pp} \end{bmatrix} \tag{4.63}$$

whose elements are

$$h_{ij} = \frac{\partial^2 E[e^2(n)]}{\partial a_i \, \partial a_j} \tag{4.64}$$

and $E[e^2(n)]$ is the mean square error. The gradient vector ∇_n is defined by (4.43). It may be shown that [68] when evaluated at the m.m.s.e., H is a positive definite symmetric matrix. Direct computation of (4.62) is not practical in a real time implementation, because the Hessian matrix, which involves second derivatives, must be computed and inverted at each iteration. However, it is possible to approximate H, by noting that

$$h_{ij} = \frac{\partial^2 E[e^2(n)]}{\partial a_i \, \partial a_j} = E\left[\frac{\partial^2 e^2(n)}{\partial a_i \, \partial a_j}\right] \tag{4.65}$$

$$= 2\,E\left[\frac{\partial e(n)}{\partial a_i}.\frac{\partial e(n)}{\partial a_j}\right] + 2\,E\left[e(n)\,\frac{\partial^2 e(n)}{\partial a_i \, \partial a_j}\right] .$$

By the principle of orthogonality[44], if the mean square value, $E[e(n)]$, of the residual is minimised, the residual e(n) is orthogonal to the past filter inputs, s_{n-i}, that is

$$E[e_n s_{n-i}] = 0 \qquad \text{for} \quad i = 1,2,\ldots\ldots,p \; . \tag{4.66}$$

As $\partial^2 e(n)/\partial a_i \partial a_j$ is a linear summation of the p delayed elements of the input sequence, the second term on the right hand side of (4.65) can, therefore, be neglected if it is assumed that the adaptive filter is adapting near the optimum. Hence, each element of the Hessian matrix can be approximated by

$$h_{ij} \simeq h_{ij}' = 2\,E\left[\frac{\partial e(n)}{\partial a_i}.\frac{\partial e(n)}{\partial a_j}\right] \tag{4.67}$$

and the approximated Hessian matrix can be written as

$$H' = 2\,E[\,\underline{G}_n\underline{G}_n^T\,] \tag{4.68}$$

where

$$\underline{G}_n = (\,\partial e_n/\partial a_1 \quad \partial e_n/\partial a_2 \quad \cdots\cdots \quad \partial e_n/\partial a_p\,)^T$$

Note that positive definiteness of H is an important factor to ensure the stability of an adaptive system based on (4.62). If the true Hessian matrix is not positive definite, as may be the case when it is evaluated at a point other than the minimum, the adaptive process as given in (4.62) may not converge. This situation is most likely to occur at some distance from the minimum. However, if the approximated Hessian matrix is computed as in (4.68), it can be proved that H' is always positive definite so that convergence can be guaranteed. Hence, from the point of view of stability, it may be argued that using the approximated Hessian matrix is safer than using the true one.

Extending the idea to the LSP filter, with

$$\underline{a}_n = (\,c_{1,n} \;\; c_{2,n} \;\cdots\; c_{\frac{1}{2}p,n} \;\; d_{1,n} \;\; d_{2,n} \;\cdots\; d_{\frac{1}{2}p,n}\,) \tag{4.69a}$$

it follows that

$$\underline{G}_n = (\,f_1(n) \;\; f_2(n) \;\; \cdots\; f_{\frac{1}{2}p}(n) \;\; g_1(n) \;\; g_2(n) \;\; \cdots\; g_{\frac{1}{2}p}(n)\,) \tag{4.69b}$$

where $f_i(n)$ and $g_i(n)$ are the first order gradient estimates of the error function, which have been described in the previous section.

Although the approximate H' is simpler than the true Hessian matrix, the inversion of H' still requires a lot of computation. It has been found that a further simplification may be made by assuming that when the adaptive system is near the optimum, the cross-correlation between two first order gradient estimates will tend to be small.

Thus

$$h_{ij}' = 2\,E[(\partial e(n)/\partial a_i)^2] \qquad \text{for } i = j \tag{4.70}$$

$$= 0 \qquad \text{for } i \neq j \ .$$

With this simplification, the simplified Hessian matrix H' becomes a diagonal matrix, and the positive definiteness property is still maintained. The inverse of H', D, is also a diagonal matrix, and the diagonal elements of D are the reciprocals of the diagonal elements of H' Thus,

$$D = [H']^{-1} \tag{4.71a}$$

and the diagonal elements of D are

$$r_{ij} = 1/h_{ij}' \ . \tag{4.71b}$$

The need to compute the inverse of the Hessian matrix has therefore been eliminated. Equation (4.62) is then written as

$$\underline{a}_{n+1} = \underline{a}_n + \mu R e_n \underline{G}_n \tag{4.72a}$$

or

$$a_{i,n+1} = a_{i,n} + \mu e_n r_{ii} \cdot \frac{\partial e_n}{\partial a_i} \tag{4.72b}$$

for $i = 1,2,.....,p$.

Comparing (4.72) with (4.46), the effective convergence factor of this 2nd order method can be written as

$$\mu' = \mu r_{ii} \tag{4.73}$$

which is adaptively changing according to the "second order" gradient estimates, and the extra computations required for each iteration of this second order method is

p x (2 multiplications + 1 division + 2 addition/subtraction).

4.4.2 EVALUATION OF THE 2ND ORDER ADAPTIVE LSP SCHEME

The performance of the pseudo second order adaptive LSP method was judged from the performance curve in Figure 4.12. In calculating the

performance, the same all pole model was used as in the calculation of the performance curve in Section 4.3.4.2. As r_{ii} is equal to the reciprocal of the expectation of $2(\partial e(n)/\partial a_i)^2$ for i=1,2,..,p, this was obtained by averaging ten consecutive values of $1/[2(\partial e(n)/\partial a_i)^2]$ stored in a 10-sample sliding window. The preferred convergence coefficient as estimated from the performance curve was

$$\mu_{opt} = 0.010 \qquad \text{(2nd order LMS-LSP)},$$

which gave 2% misadjustment at adaptation time n=108 samples. When compared with the performance of the fixed μ LMS/LSP scheme with an estimated optimal convergence coefficient

$$\mu_{opt} = 0.003 \qquad \text{(fixed } \mu \text{ LMS/LSP)},$$

which gave 2.0% misadjustment at adaptation time n=153 samples, the second order method showed a faster convergence rate. The learning curves for this pseudo second order adaptive LSP scheme using $\mu = 0.015$ are plotted in Figure 4.13. The 'noisy' curve is the learning curve of a single Gaussian noise sequence and the smoother curve is an ensemble average of 50 learning curves. When compared with the fixed μ scheme as shown in Figure 4.9, the second order adaptive system shown in Figure 4.13 exhibits a faster and much smoother adaptation than the fixed μ scheme. In many digital signal processing application, a smooth adaptation algorithm is always desired.

Further enhanced schemes have also been investigated. It was found that if (4.72) is modified to:

$$a_{i,n+1} = a_i + \mu\, E[e_n^2]\, e_n\, r_{ii}\, (\partial e_n/\partial a_i) \qquad (4.74)$$

for $i = 1,2,.....,p$, where the effective adaptation coefficient is

$$\mu'' = \mu'\, E[e_n^2],$$

a remarkable improvement in performance over the fixed μ and the 2nd order method was obtained. The performance curve and learning curve of this modified 2nd order method are respectively plotted in Figure 4.14 and Figure 4.15 with curves of the other methods for comparison. The preferred μ of this method was

$$\mu_{opt} = 0.02 \qquad \text{(modified 2nd order LMS/LSP)}$$

that gave 0.83% misadjustment after 39 iterations. The improvement of this method may be due to two actions being taken: firstly, the simplified Hessian matrix assists the adaptive system to find a better path on the error surface towards the minimum; secondly, the adaptation step size is effectively controlled by (4.74) according to the decreasing energy of the error signal. When the adapting error signal energy is large, which means that the adaptation may be some distance away from the minimum, the effective step size is proportionally enlarged so that the adaptation rate is relatively large. When the error signal energy is relatively small, the effective step size is proportionally reduced so that the coefficients are being adjusted more slowly and smoothly.

There are many other well known LMS type adaptive filter schemes [69], for instances, adaptive ladder filters, adaptive lattice filters and adaptive IIR filters. Two common adaptive filter schemes, the adaptive ladder and adaptive lattice filters, have been simulated for comparison with the new adaptive LSP scheme.

The coefficient update equation for the traditional adaptive ladder as originally proposed by Widrow [67] is

$$\underline{a}_{n+1} = \underline{a}_n + \mu\, e_n\, \underline{x}_n \qquad (4.75)$$

where $\underline{a}_n$ is the filter coefficient vector, μ is the fixed adaptation coefficient, e_n is the instantaneous filter error output and $\underline{x}_n$ is the de-

layed signal vector. The performance curve for this adaptive filter was far away from the performance curves shown in Figure 4.14, and had a much larger misadjustment at larger adaptation times. Hence, the performance curve is not drawn in the graph as it is readily concluded that this adaptive filter is nowhere near as good as the adaptive LSP scheme.

For the simulated adaptive lattice scheme used by Davis [70], the reflection coefficients k_i are updated by the formula:

$$k_{i,n+1} = k_{i,n} + 2\,\mu_i f_{i,n}\, b_{i-1,n-1} \tag{4.76}$$

where $\mu_i = \mu/(12.r_{oi})$, and $r_{oi} = E[b_{i-1,n-1}{}^2]$. In (4.76), $k_{i,n+1}$ is the i-th stage reflection coefficient at time instant n+1, $f_{i,n}$ is the forward error, $b_{i,n}$ is the backward error, μ_i is the i-th stage convergence factor, and r_{oi} is mean square of i-th stage backward error. This scheme is referred to as the stage-by-stage LMS/lattice as it aims to minimise the total mean square error $E[e^2(n)]$ indirectly by successively reducing the sum of the mean-square values of the forward and backward errors at the output of each stage of a non-recursive lattice filter. The convergence rate of this scheme is supposed to be one of the fastest compared with many other adaptive filter schemes which use the LMS principle. To ensure stability of this adaptive lattice filter scheme, the values of the reflection coefficients must be bounded by ±1.

The performance curve for this stage-by-stage lattice adaptation scheme is plotted in Figure 4.14, from which the preferred μ is estimated as 0.003. When the lattice adaptation scheme is compared with the fixed μ LSP scheme, both techniques exhibit similar adaptation performance in the first 100 iterations, but for larger adaption time the LSP technique

achieves smaller misadjustment. Using the preferred μ, the learning curve of the stage-by-stage lattice method, as plotted in Figure 4.15b, shows a smoother adaptation than the learning curve for the fixed μ LSP technique in Figure 4.15a even when the preferred value of μ is used. Thus, the choice between these two schemes should depend on the design requirements.

If the adaptive lattice scheme is compared with the modified pseudo 2nd order adaptive LSP filter scheme as described above, the latter is superior in terms of misadjustment and convergence rate. However, the computational effort of the modified 2nd order LSP method is more than for the adaptive lattice method.

In this Section, a new adaptive filtering technique has been presented, and its performances has been compared with other well known schemes. The results shows this new technique to be promising, since it offers good adaptation and generates LSP coefficients directly, thus removing the need for conversion at a later stage.

4.5 EVALUATION OF THE ADAPTIVE LSP ALGORITHM FOR THE APC-AB CODER

The performance of the adaptive LSP algorithm was evaluated by applying it to the 16 kb/s APC-AB sub-band coder described in Chapter 3. The LSP coefficients used in the original APC-AB coder [57] were derived from the autocorrelation method via a transformation. This approach will be referred to as the AUTO-LSP method. The modified pseudo 2nd order adaptive LSP scheme was employed to replace the AUTO-LSP scheme without changing other aspects of the speech coding algorithm. As the adaptive LSP schemes described were based on a stationary signal with zero mean

and constant variance, the chosen preferred μ for the described adaptive system would be different for speech due to the speech signal having different variance. A method of estimating the preferred values of μ for adaptive LSP filters as used in the new APC-AB sub-band speech coder will be described in this section. The performance of this modified APC-AB coder implementation will then be compared with the AUTO-LSP based APC-AB coder by objective measurement and informal listening tests. The results of this comparison indicate that the adaptive LSP scheme can be made comparable to the AUTO-LSP scheme in terms of objective and subjective speech quality. A comparison of the computational effort required for the AUTO-LSP and the adaptive LSP method will also be given in this section.

4.5.1 ESTIMATION OF CONVERGENCE FACTOR FOR APC-AB SPEECH CODER

A method for choosing a convergence factor, μ, for a stationary signal has been described in the previous section. The convergence properties, however, depend not only on the value of μ, but also on the statistics of the incoming signal. To achieve a desired adaptive performance for a non-stationary input signal, e.g. speech, the value of μ should be carefully chosen. In this section, a simple strategy to estimate μ for speech is devised, which has been successfully applied to the APC-AB speech coder.

The APC-AB speech coder analyses speech in 3 subbands, each of which is applied to a fourth order adaptive predictive coder where the LPC parameters are represented by LSP coefficients. In the adaptive LSP scheme employed in this experiment, the LSP coefficients of each subband are extracted directly from blocks of subband samples, over which the

speech is assumed stationary. The adaptation commences with the LSP coefficients chosen to produce an all-pass frequency response. Each 20 msec subband sample block is successively applied to the corresponding adaptive LSP filters. The filter coefficients are adapted on successive samples from each subband block. Upon receipt of the last samples in a block, it is assumed that the filter coefficients are close to Wiener optimum values. The coefficients are then used in the short term predictors within the APC, and the resulting error signals, along with the LSP coefficients, are then quantised for transmission. As the speech characteristics change relatively slowly between 20 msec frames, the adaptation for a block can be started with the near optimum LSP coefficient set computed from the previous block.

The problem in estimating a convergence factor for each subband adaptive filter has been resolved by assuming that:

a) a 4-th order LPC model is suitable to model each subband speech signal,

b) the convergence factor is inversely proportional to the applied signal energy in each sub-band.

In choosing a convergence factor for a system adapting to speech, it is preferable to achieve a better performance for voiced than for unvoiced speech. The frequency spectrum of the 4-th order model used in the previous section is typical of voiced speech, and, therefore, assuming (a), the results obtained can be applied to the APC-AB coder. From previous experiments described in Section 4.4.2, the preferred μ for a speech signal with zero mean, and variance equal to 6.667 is 0.02 which gives 0.83% misadjustment and adaptation time of 39 iterations. Speech can be considered as a zero mean process even within a short period, but the

variation of the variance (or the energy) of speech can be very large between frames, especially, at the transition between silence and voiced speech. It has been observed that if a 12-bit A/D is used to sample the full band speech, the variations of the subband signal energy, s(n), are as follows:

$$\begin{aligned} 22\ \text{dB} &< E[s_1^2(n)] < 48\ \text{dB} && \text{for band 1} \\ 10\ \text{dB} &< E[s_2^2(n)] < 40\ \text{dB} && \text{for band 2} \\ 8\ \text{dB} &< E[s_3^2(n)] < 38\ \text{dB} && \text{for band 3 .} \end{aligned} \tag{4.77}$$

To achieve the desired convergence performance from the adaptive LSP scheme, the μ should therefore be varied with subband signal energy. The assumption (b) is based on the argument that if a higher energy signal, for example voiced speech, passed through the adaptive filter, the error output would also be increased. In this case, if the adaptation step size is too large due to a large μ, this may cause unstable adaptation and the residual signal would probably be larger than the input speech. This condition contradicts the minimum mean square error criterion and the speech quality of the speech coder would severely be degraded. Hence, a smaller convergence factor should be chosen for higher energy signals. This implies that the argument of inverse proportionality between μ and the speech energy, i.e.,

$$\mu \propto \frac{1}{E[s^2(n)]} \tag{4.78}$$

is appropriate. From (4.78) and the results taken in the previous experiment,

$$0.02 = \frac{k}{6.667} \quad \text{and} \quad k = 0.1333 \tag{4.79}$$

where k is a proportionality constant. Using k equal to 0.1333 and the range of subband signal energies from (4.77), the range of μ for each subband is

$$\begin{aligned} 0.00084 &< \mu_1 < 0.000002 && \text{for band 1} \\ 0.013 &< \mu_2 < 0.000013 && \text{for band 2} \\ 0.021 &< \mu_3 < 0.00021 && \text{for band 3.} \end{aligned} \tag{4.80}$$

Based on the above arguments, if the value of μ is fixed for each subband, the smallest μ should be chosen to prevent instability. The resulting convergence time for low energy signals would, however, be large, and therefore the corresponding adaptation would not reach a settling point within one block of the signal. The distortion resulting from the far-from-optimum prediction coefficients is, however, not perceptually important for low energy signals.

A difficulty with an adaptive algorithm arises during the analysis of voiced speech. For such signals, the adapting coefficients are maximally perturbed at instants of the speech pitch pulses. This effect causes temporarily retarded convergence, and poor signal modelling when perturbed coefficients are selected for the short term APC predictions. This problem, can be lessened by:

a) limiting coefficient changes between two iterations, and

b) using an appropriate average window size on the modified pseudo second order LSP method.

In the experiment described in this section, the maximum LSP coefficient change between two iterations was limited at 0.2, which is not so small as to affect the convergence properties. For the averaging window, the length should be short enough to avoid effects due to the non-stationary speech signal. The window should also be long enough to average out large

perturbations in the adapting coefficients due to rapid signal changes at pitch pulses. Window length of 20 msec was chosen.

The smallest values of μ from (4.80) were used initially. These values were tested with lengthy segments of speech and the effect on the adapting error signal in each subband of changing each value of μ was observed. It was found that the system showed better adaptation, with no instability, with the following values:

$$\begin{aligned} \mu_1 &= 0.000005 \\ \mu_2 &= 0.00003 \\ \mu_3 &= 0.00005 \quad . \end{aligned} \tag{4.81}$$

Thus, these values were adopted for the three adaptive subband filters in the modified APC-AB coder.

Histograms of the computed LSP coefficients of each subband are plotted in Figure 4.16, and their statistics are given in Table 4.1 as shown in Figure 4.18. This data can be used to compare the statistics of the LSP coefficients as derived by AUTO/LSP method with those obtained by the adaptive LSP filter, and also for designing a suitable quantizer for the LSP coefficients.

4.5.2 RESULTS AND DISCUSSION

The APC-AB speech coder was modified by replacing the AUTO/LSP technique by the adaptive LSP technique with no other changes. The computed LSP coefficients were encoded by differential encoding as described in Section 3.3.3.2. The statistics of the LSP differences obtained by analysing representative speech segment, i.e. 30 seconds of male speech and 30 seconds of female speech, are given in Table 4.2 (Figure 4.18). These were used to design the quantizers.

The modified coder was mainly tested by objective tests and informal listening. To compare the performance of the AUTO/LSP and the adaptive LSP algorithm, the long term SNR and segmental SNR were measured using 2 second segments of male speech. The results of this comparison for a typical segment were:

	AUTO/LSP	Adaptive LSP
SNR_{long}(dB)	19.31	18.79
SNR_{seg} (dB)	16.89	16.47

Thus, the objective performance of the adaptive LSP technique was only slightly worse than for the AUTO/LSP technique. The time variation of the segmental SNR, or the SNR contour, obtained using the adaptive LSP technique is plotted in Figure 4.17. This should be compared with the SNR contour for the AUTO/LSP technique shown in Figure 3.24. By comparing both diagrams, it is clear that for the adaptive LSP technique the SNR is lower during silence, but at some instants of high signal energy, the SNR is higher than for the AUTO/LSP technique. This implies that for some voiced speech, the adaptive LSP scheme could provide a better linear predictive performance than the AUTO/LSP method. Informal listening tests have been carried out on 30 seconds segments of male and 30 seconds segments of female speech. The speech quality produced by both methods was very similar.

To compare the computational complexity of the AUTO/LSP and adaptive LSP methods, programs for deriving LSP coefficients by both methods were carefully written so that the number of arithmetic operations were minimized. The approximate number of arithmetic operation required for both algorithms are given in Table 4.3 (in Figure 4.19).

For the AUTO/LSP method, the calculation of the LSP coefficients from the LPC coefficients is based on Soong's root finding algorithm [49], and the computational cost of this method exponentially increases with the accuracy required. Assuming that a 'multiply and add' or a 'cosine table lookup' is counted as one arithmetic operation, the total arithmetic operation count for the AUTO/LSP method can be obtained from Table 4.3a to give:-

$$\text{total operations}_{AUTO} = 3p^2/2 + p(N+4k-11/2) + 3(k+N-M-1) \quad (4.82)$$

for p even. The k and M are accuracy factor that depend on the root finding accuracy, TOL. If the LPC order, p, is 4 and the analysis frame length, N, is 32 as in the APC-AB coder, and M=128, TOL=0.005 and k=70, the total operations_{AUTO} is equal to 1169.

For the adaptive LSP algorithm, Table 4.3 only shows the required arithmetic operations for one iteration, which can be approximated as:

$$\text{total operations}_{adaptive} = 3p^2/4 + 9p/2 + 9 \quad (4.83)$$

for p even. If the LSP coefficients are calculated by the adaptive LSP scheme on a block by block basis, the total number of operations should be the product of the block length, N, and the the number of required operations for one iteration. For the APC-AB coder, if p=4 and N=32, total operations for the adaptive LSP scheme is 1248, which is slightly higher than the AUTO/LSP method as computed above. The computation cost of the adaptive LSP scheme is largely due to the gradient estimation, but for quasi-stationary signals, such as speech, the computation can be reduced

by updating the gradient estimates for every two samples. However, the adaptive LSP method does not require windowing, autocorrelation computation, matrix inversion and root finding as in the AUTO/LSP method. The implementation complexity of the adaptive LSP technique is therefore much less than the AUTO/LSP method.

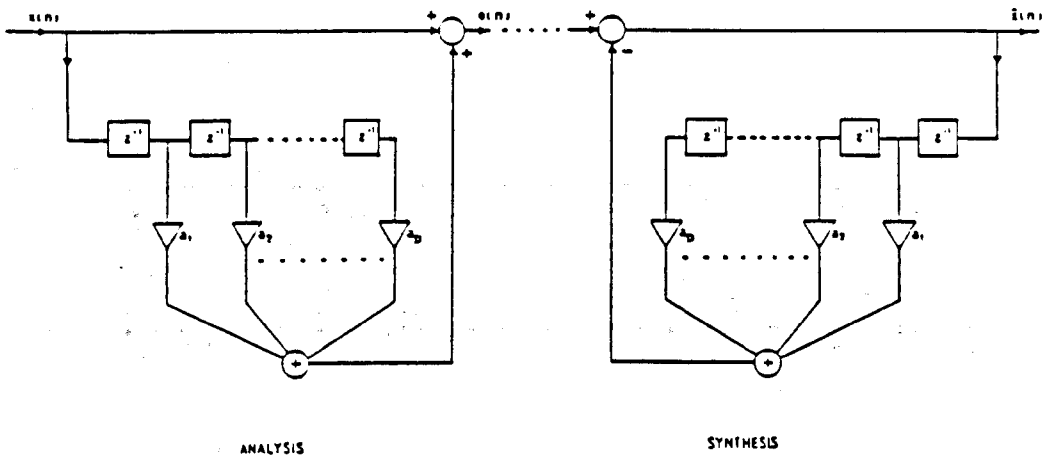
4.6 CONCLUSION

This Chapter has presented the basic principles of LSP coefficients and the conventional AUTO/LSP method for deriving them from segments of speech. Adaptive schemes based on LMS-type adaptation [67] were also described for calculating LSP coefficients directly from the analysis speech, without recourse to conversion from other parameters. Enhancement to the basic LMS adaptation technique were proposed using an approximation to the Hessian matrix.

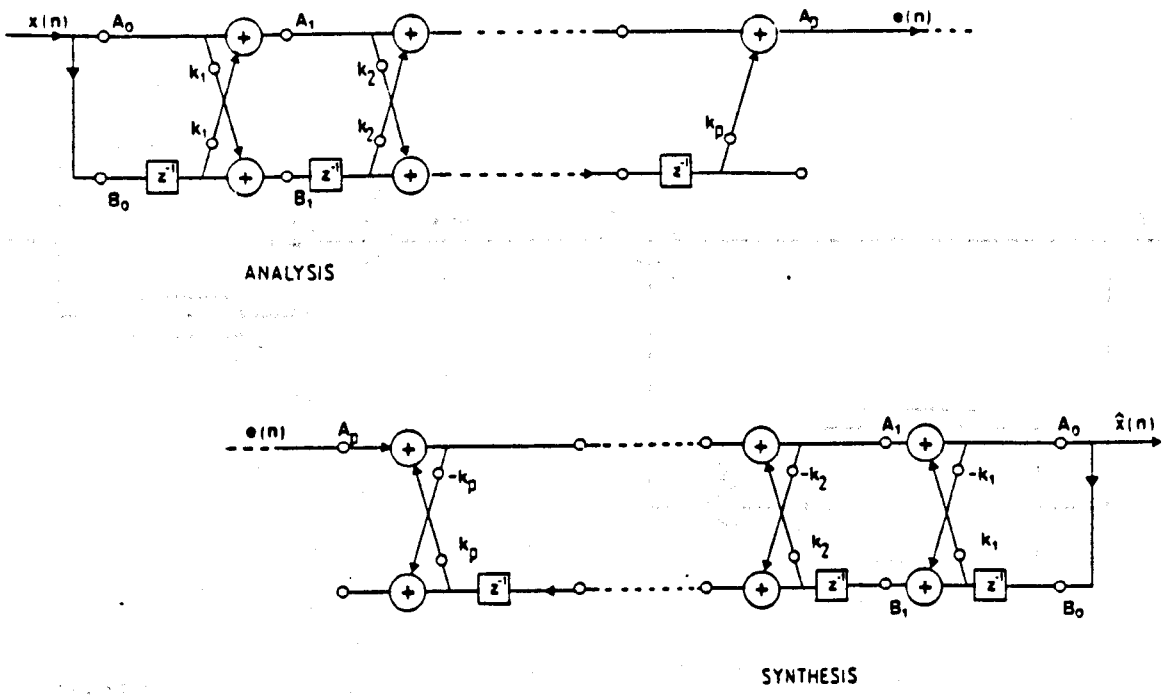
A method of assessing the performance of adaptation schemes was described. A performance curve for the 'LSP' adaptive filter was plotted and from this a preferred convergence rate factor, μ, was estimated. A drawback of this performance curve method is that it is based on an averaging process which smooths out the sample by sample variation in error signal power that occurs in each individual trial in the ensemble of trails. It is therefore necessary to examine the individual trails as well as the overall learning curve to provide a reliable means of comparison between different methods. The developed 'modified 2nd order' adaptive LSP method shows smooth adaptation as well as fast convergence as shown in Figure 4.15. Hence, the performance curve approach was found to produce an appropriate value of μ.

The adaptive LSP filter scheme has been compared with other known adaptive ladder filter and adaptive lattice filter schemes [63,70]. The results indicate that the adaptive LSP scheme performs much better than the other methods in terms of the convergence rate.

The 2nd order adaptive LSP scheme was applied to the APC-AB coder to replace the AUTO/LSP technique. Informal listening tests and objective measurements showed that the adaptive scheme can be made comparable to the AUTO/LSP scheme in terms of objective and subjective speech quality. Furthermore, the adaptive LSP scheme can be implemented rather more simply.



(a)



(b)

Figure 4.1 Linear prediction analysis/synthesis filters.
(a) Ladder filter, (b) Lattice filter.

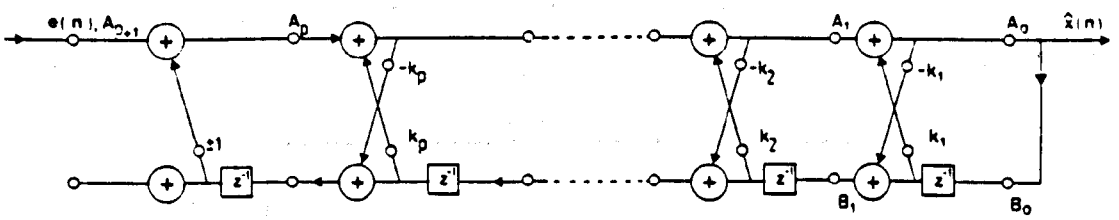


Figure 4.2 Argumented lattice filter.

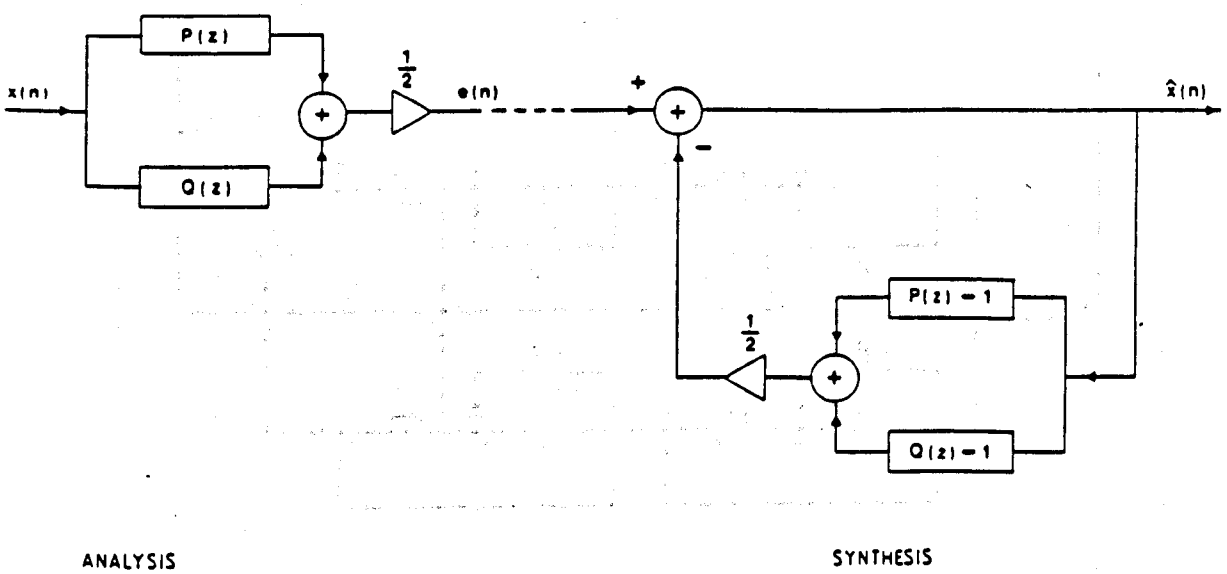
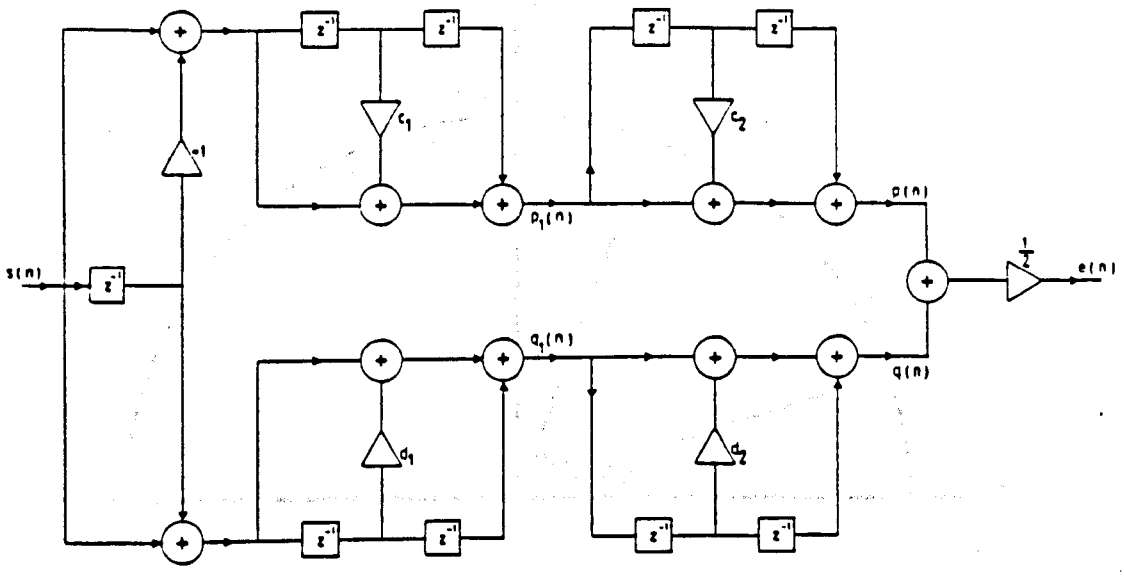
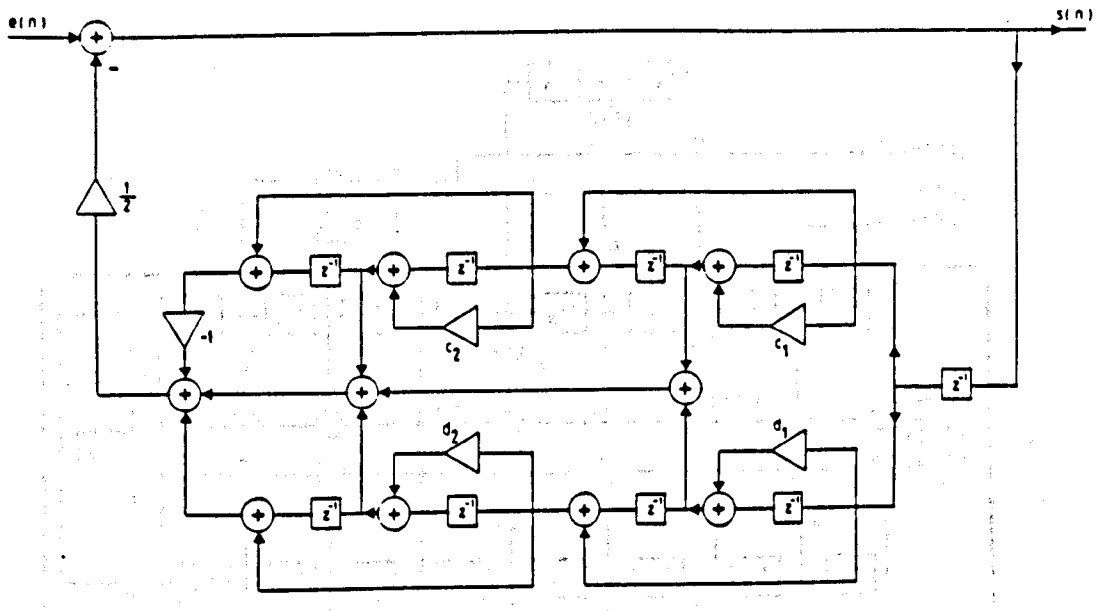


Figure 4.3 Linear Prediction in LSP filter structure.



(a) Analysis



(b) Synthesis

Figure 4.4 Fourth order LSP analysis/synthesis filters.

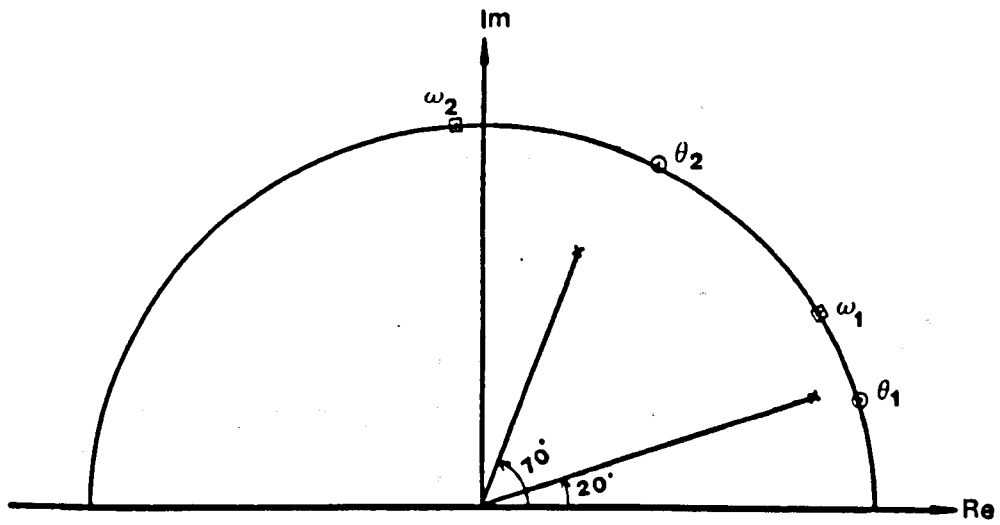


Figure 4.5 Plot of LSP's and poles on the unit circle.

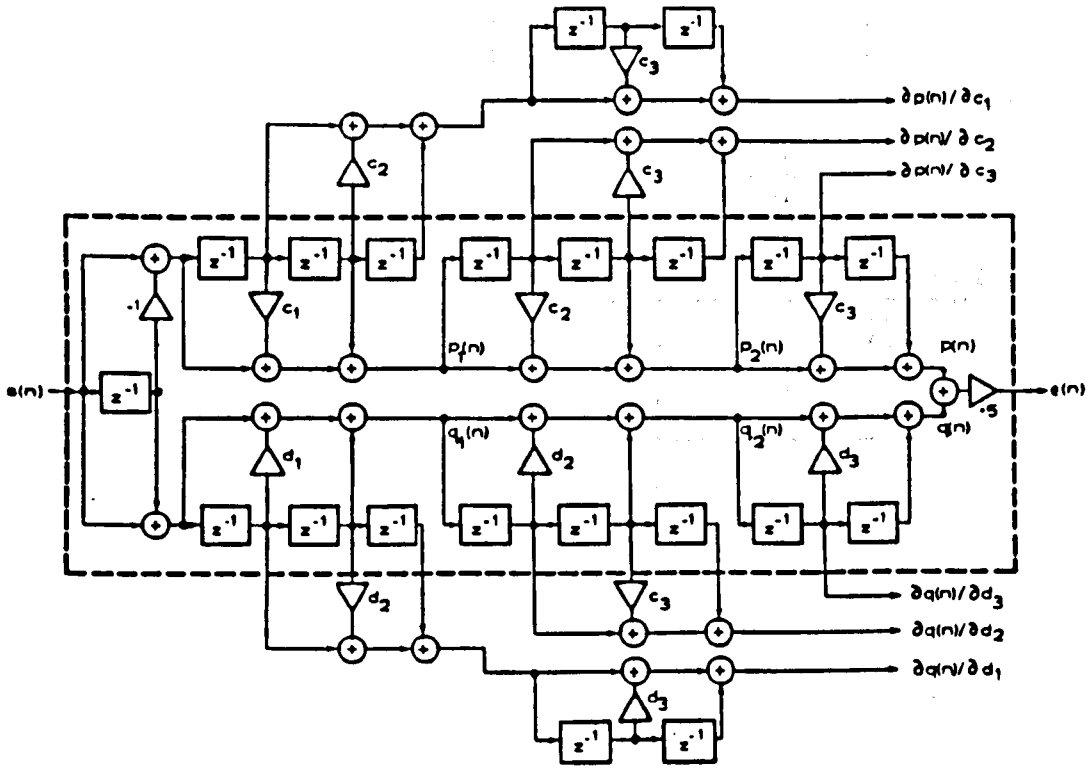


Figure 4.6 A sixth order adaptive LSP filter to generate gradient estimates.

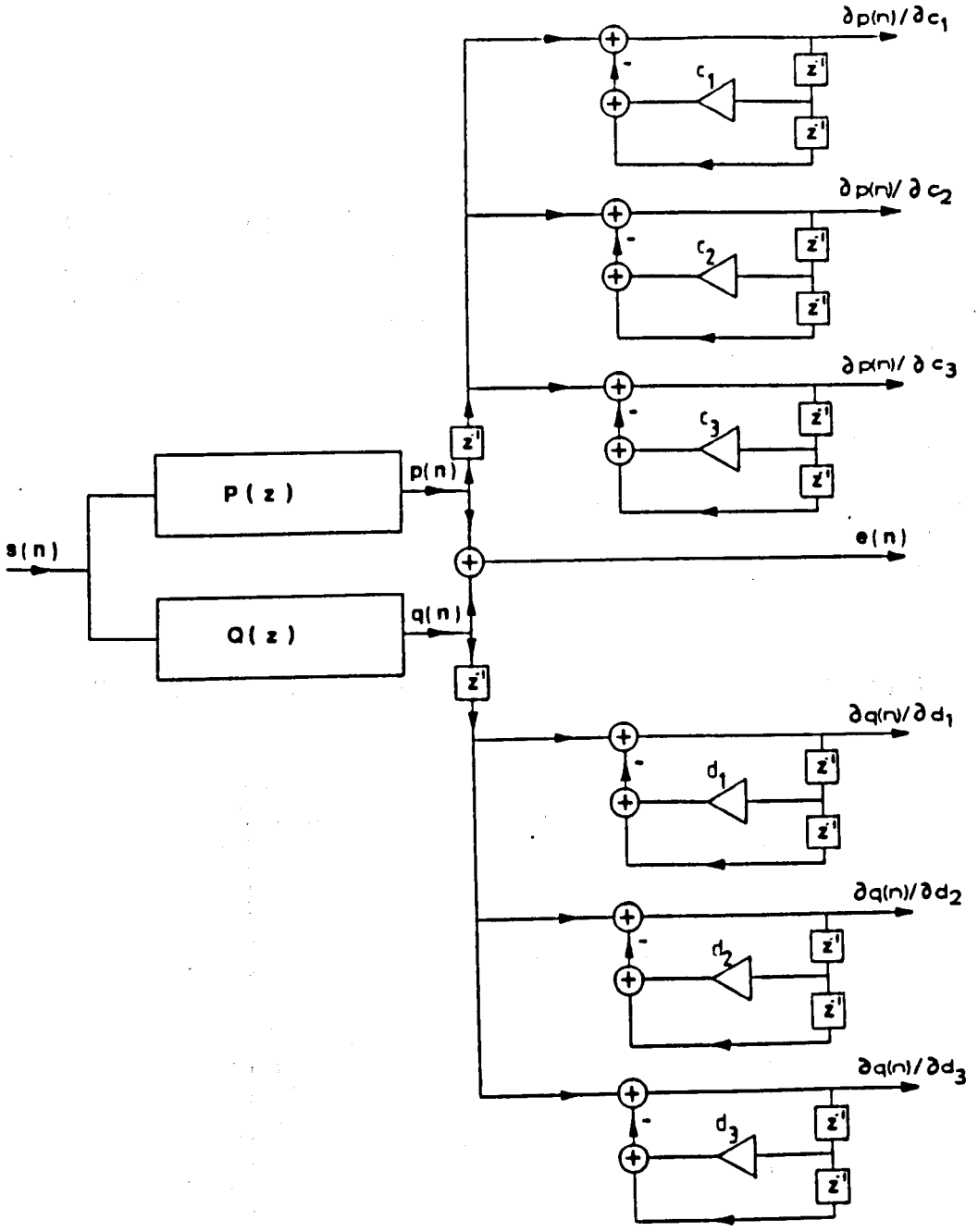

∂p(n)/∂c1
∂p(n)/∂c2
∂p(n)/∂c3
P(z)
p(n)
s(n)
e(n)
Q(z)
q(n)
∂q(n)/∂d1
∂q(n)/∂d2
∂q(n)/∂d3
c1
c2
c3
d1
d2
d3
z-1

Figure 4.7

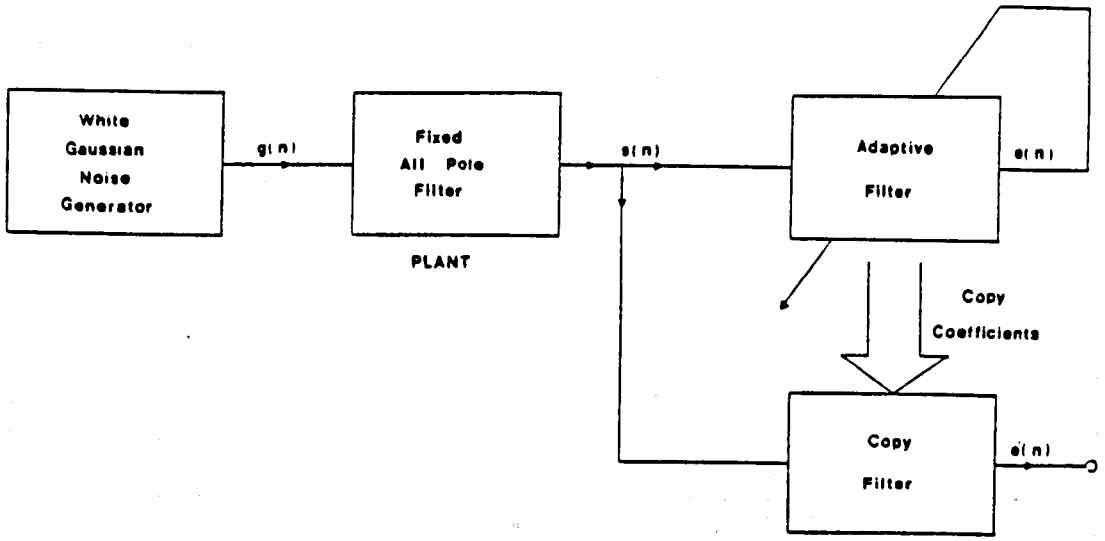


Figure 4.8 Block diagram of an adaptive predictive system.

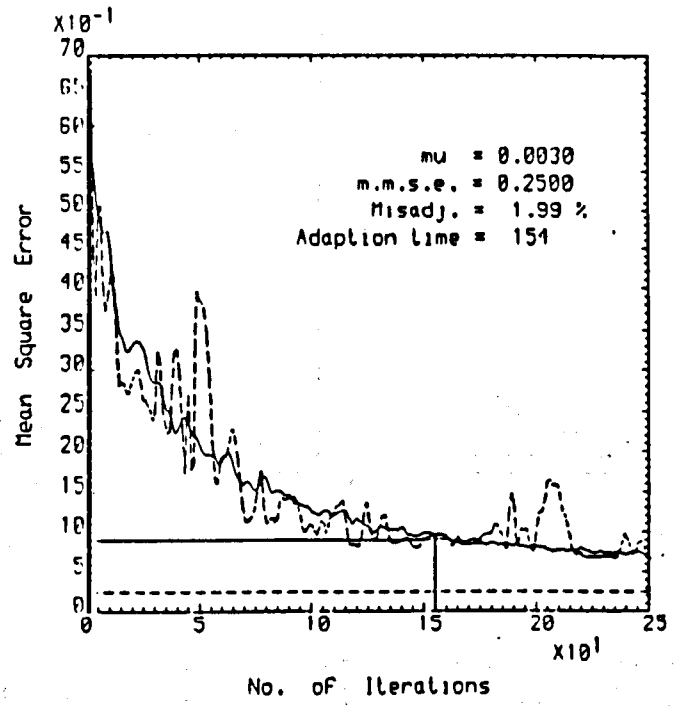


Figure 4.9 Learning curves obtained by an 4-th order predictor to a correlated Gaussian noise input signal generated by a four-pole recursive filter.

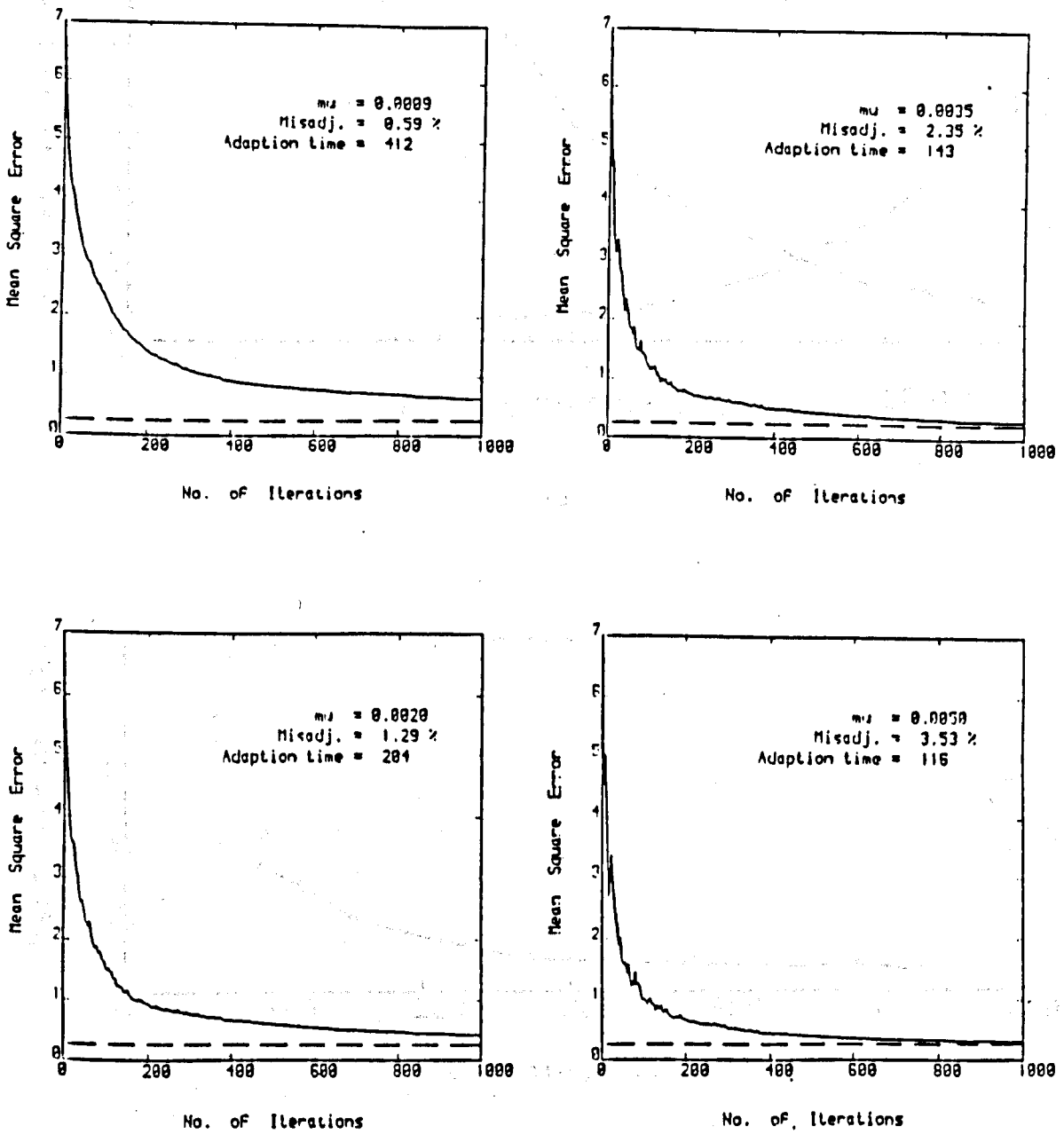
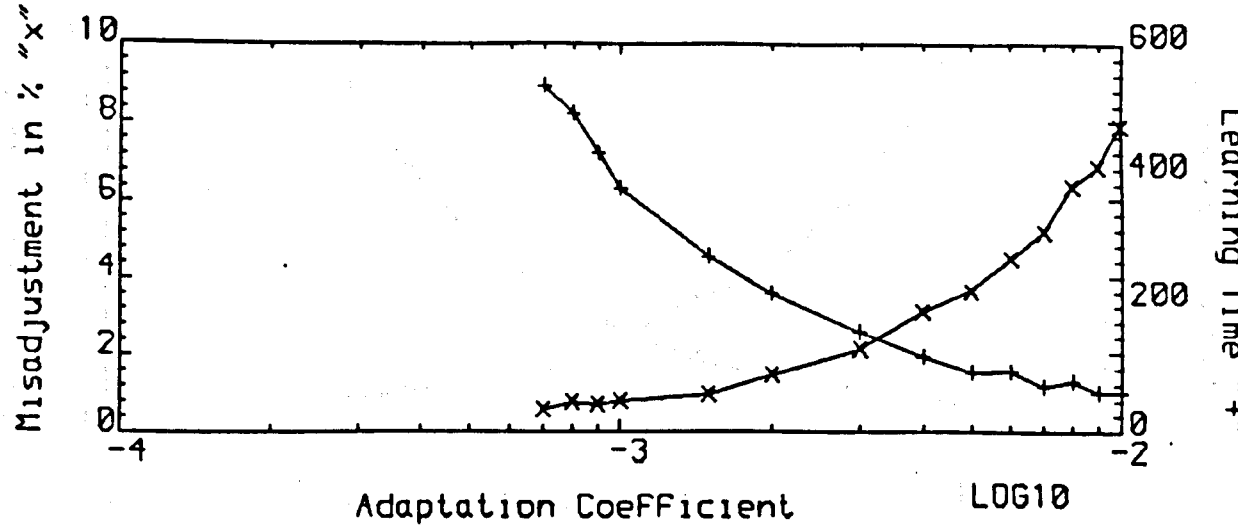
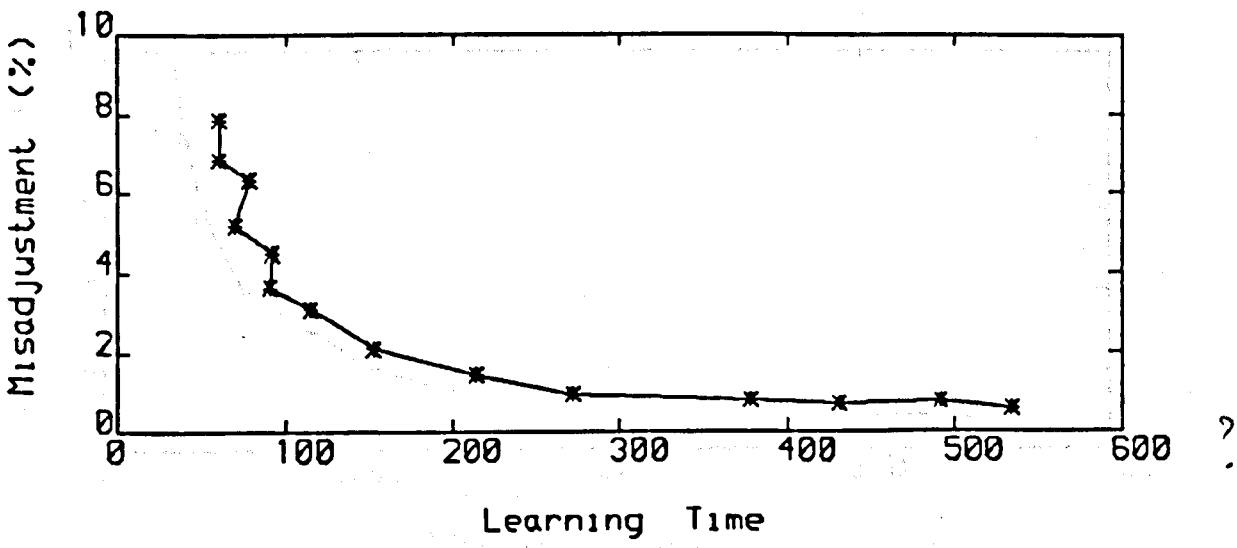


Figure 4.10 Ensemble learning curves for 4-th order adaptive LSP filter at different values of μ. The broken line shows the Wiener minimum m.s.e..

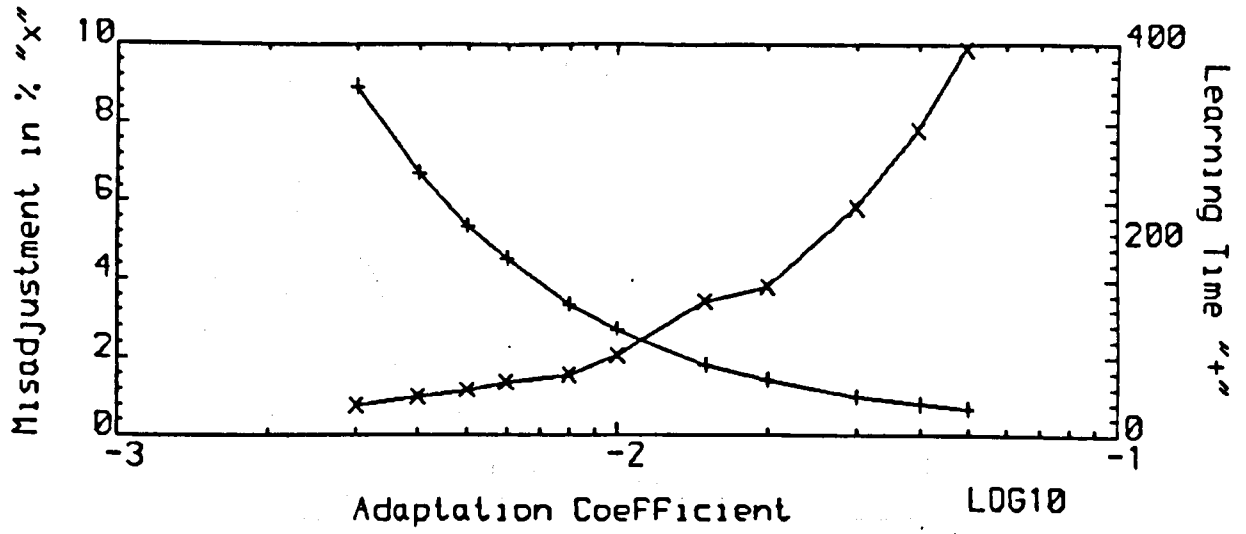


(a)

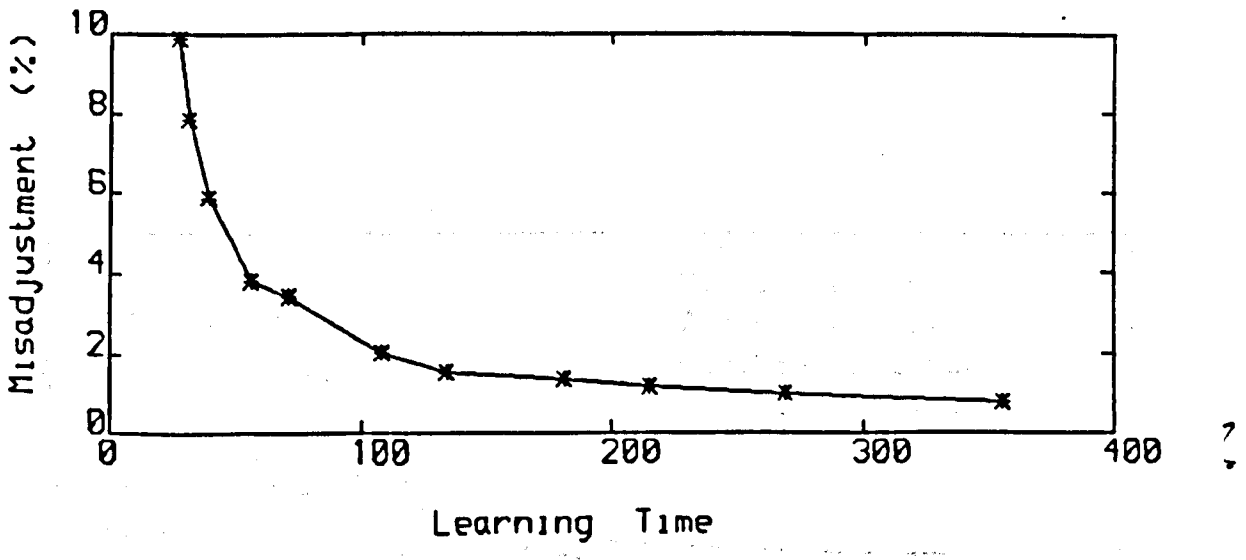


(b)

Figure 4.11 Performance measure for fixed μ adaptive LSP scheme.
(a) Variation of misadjustment and learning time with μ
(b) Performance curve.



(a)



(b)

Figure 4.12 Performance measure for 2nd order adaptive LSP scheme.
(a) Variation of misadjustment and learning time with μ
(b) Performance curve.

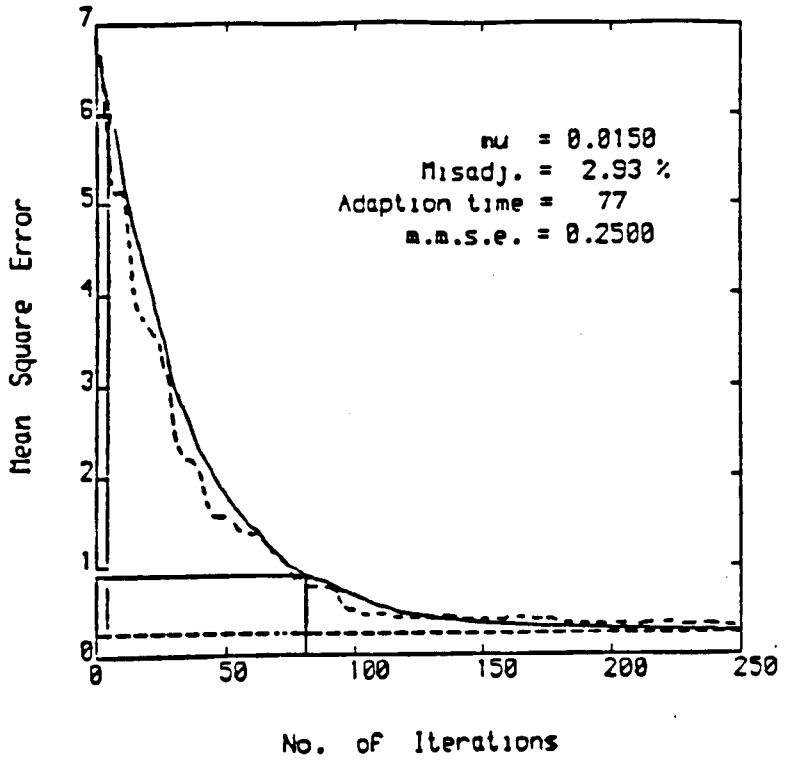


Figure 4.13 Learning curves of the 2nd order adaptive LSP method obtained by an 4-th order predictor to a correlated Gaussian noise input signal generated by a four-pole recursive filter.

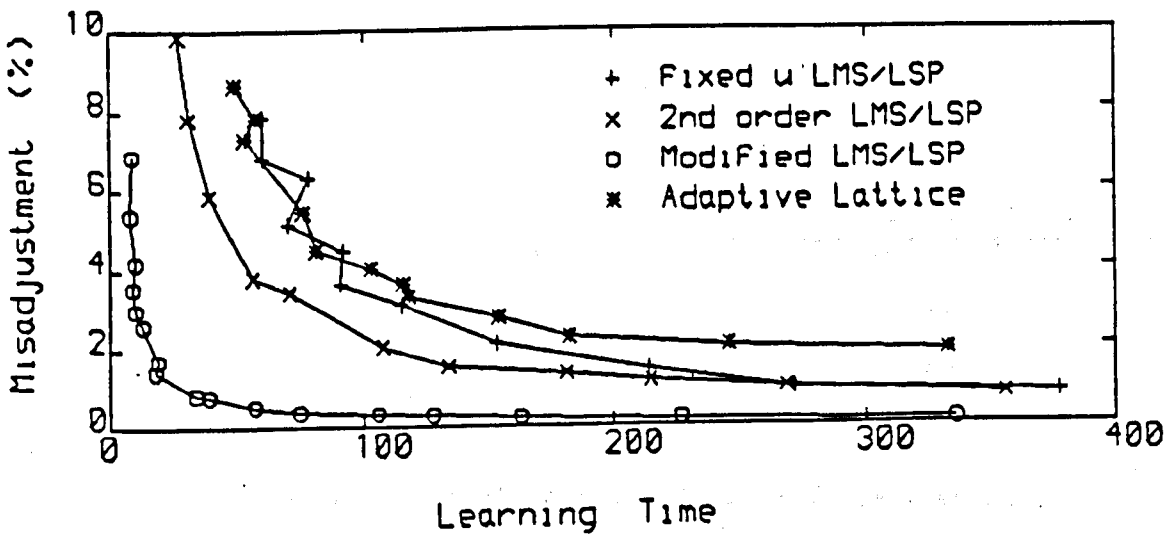


Figure 4.14 Performance curves of various adaptive methods.

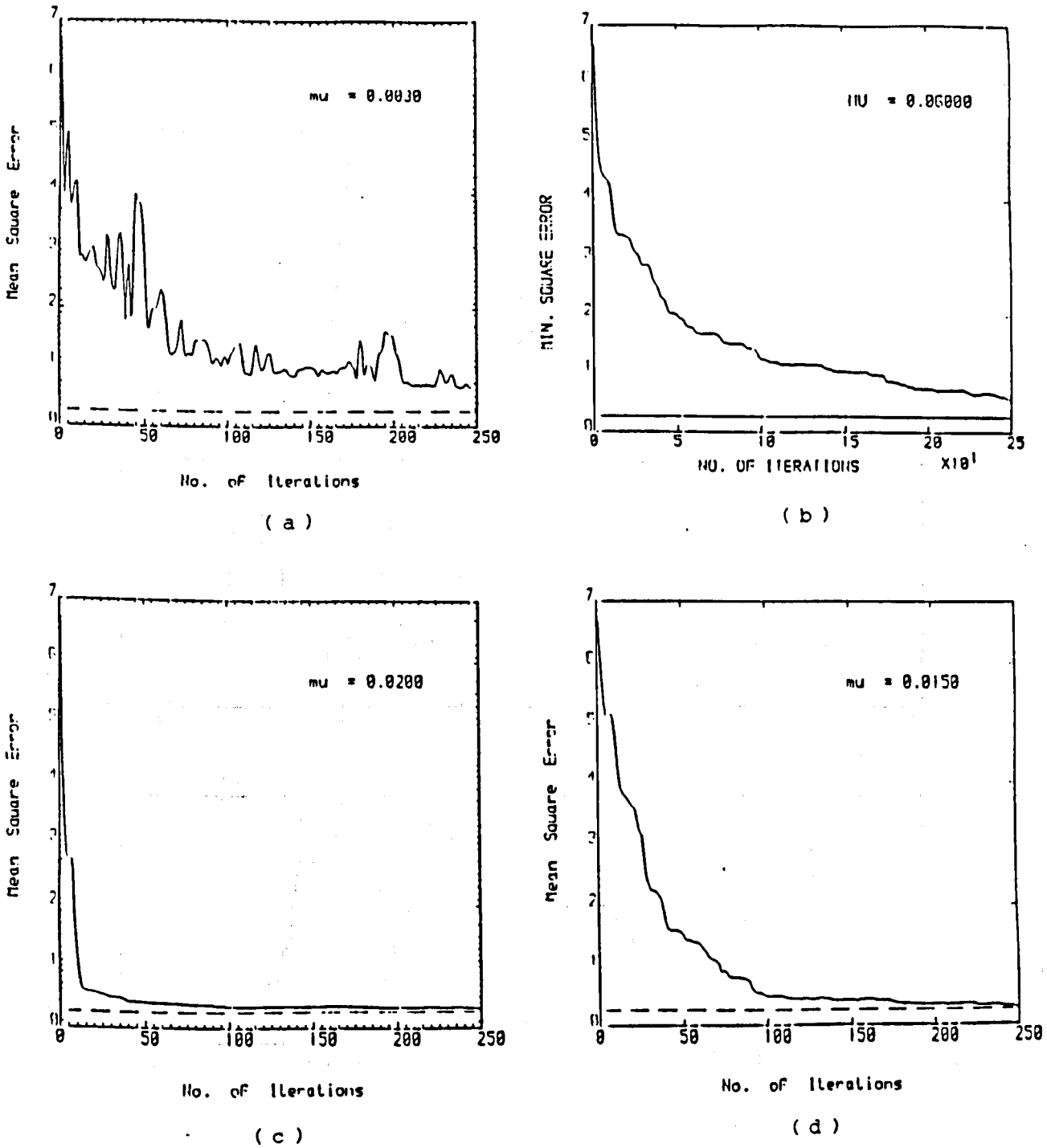


Figure 4.15 Learning curves of various adaptive schemes at their preferred μ values.

a) . fixed μ LSP filter, (b). adaptive μ lattice filter,
c) . modified 2nd order LMS-LSP filter,
d) . 2nd order LMS-LSP filter.

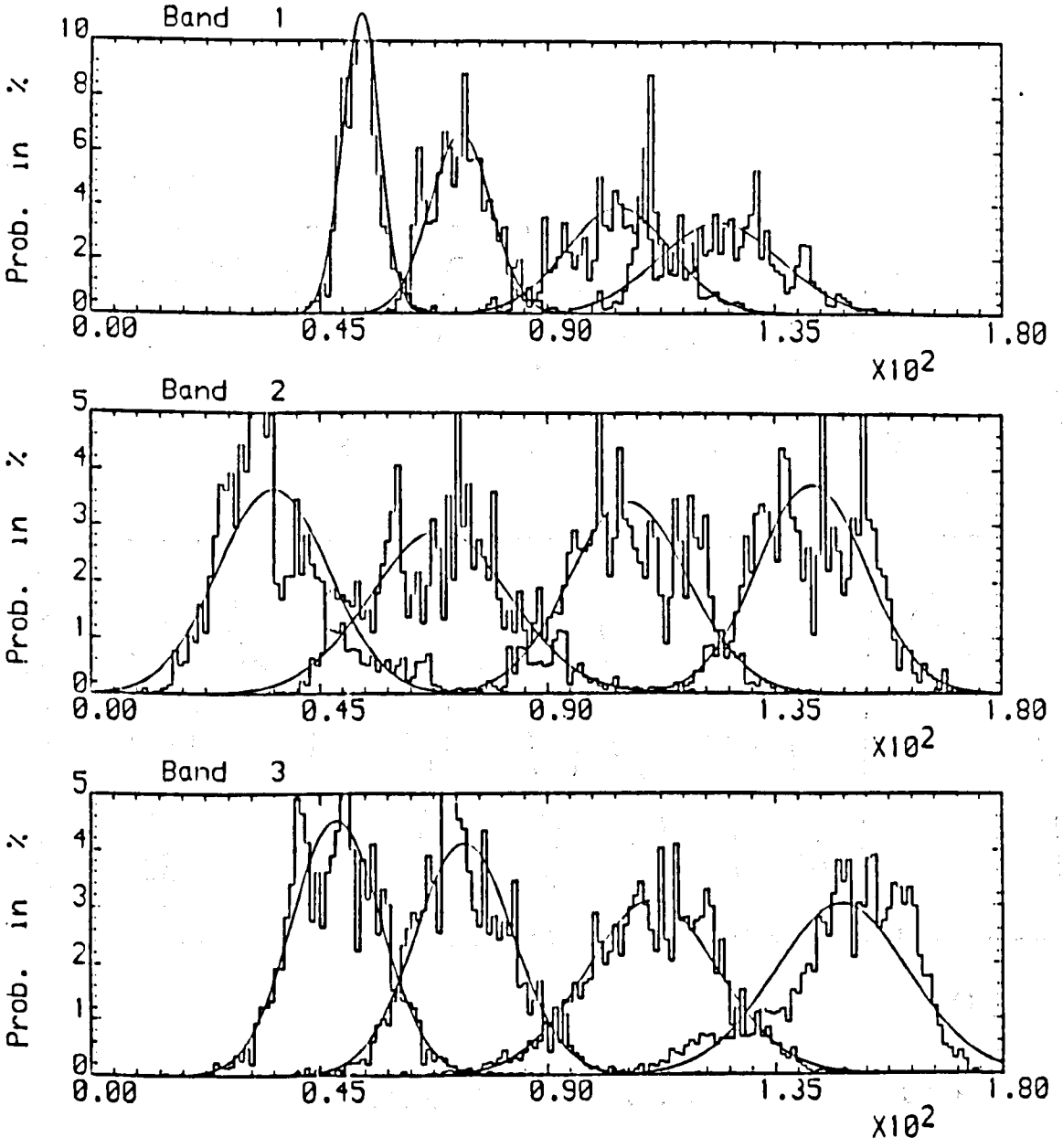


Figure 4.16 Histograms of LSP coefficients (derived from adaptive method) for the APC-AB coder.

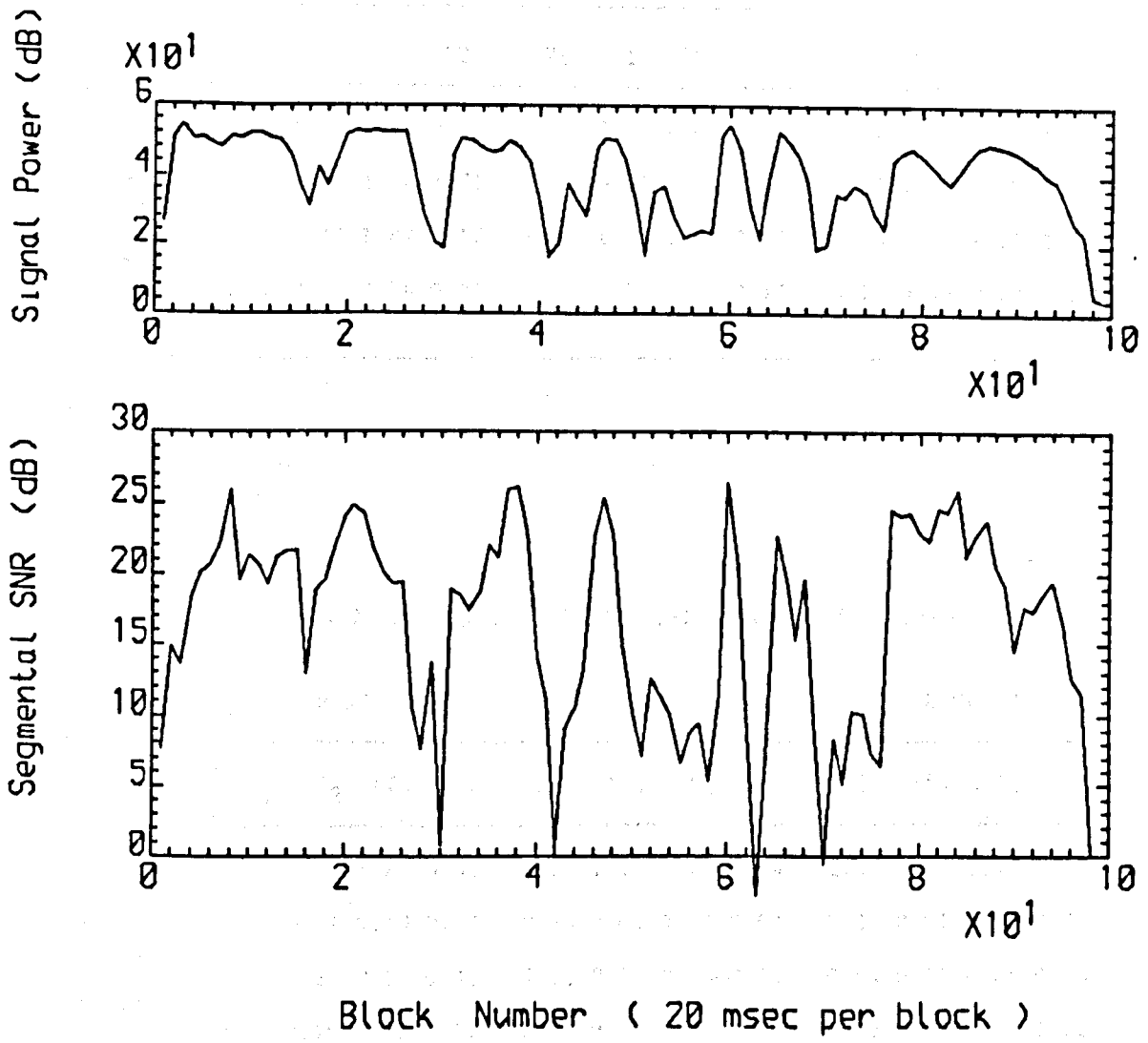


Figure 4.17 SNR_{seg} performance of the 'Adaptive LSP' based APC-AB coder.

Table 4.1 Statistics of LSP coefficients

	BAND 1				BAND 2				BAND 3			
	θ_1	ω_1	θ_2	ω_2	θ_1	ω_1	θ_2	ω_2	θ_1	ω_1	θ_2	ω_2
Maximum (in rad.)	1.18	1.67	2.32	2.75	1.87	2.06	2.48	3.00	1.50	2.09	2.70	3.04
Minimum (in rad.)	0.63	1.01	1.34	1.71	0.10	0.61	1.24	1.91	0.25	0.68	0.90	1.55
Mean (in rad.)	0.93	1.27	1.82	2.17	0.62	1.20	1.85	2.49	0.84	1.28	1.93	2.58
Std. dev. (in rad.)	0.06	0.10	0.18	0.21	0.19	0.24	0.20	0.19	0.15	0.17	0.23	0.23

(a)

Table 4.2 Statistics of LSP coefficient differences

	BAND 1				BAND 2				BAND 3			
	Δf_1	Δf_2	Δf_3	Δf_4	Δf_1	Δf_2	Δf_3	Δf_4	Δf_1	Δf_2	Δf_3	Δf_4
Maximum (in rad.)	1.18	0.82	1.27	0.68	1.87	1.34	1.23	1.54	1.50	1.42	1.48	1.47
Minimum (in rad.)	0.63	0.01	0.10	0.07	0.10	0.02	0.02	0.03	0.25	0.01	0.03	0.03
Mean (in rad.)	0.93	0.34	0.55	0.35	0.62	0.58	0.65	0.63	0.84	0.44	0.65	0.65
Std. dev. (in rad.)	0.06	0.10	0.18	0.11	0.19	0.19	0.20	0.17	0.15	0.17	0.22	0.22

(b)

Figure 4.18 Statistics of LSP coefficients computed by adaptive LSP scheme.

AUTO-LSP

	⊗	⊗ & ⊕	⊕ / ⊖	Others	Remarks
Window	N	-	-	-	Eqn(3.61)
Auto-function	-	(N-p/2)(p+1)	-	-	Eqn(3.62)
Matrix equation	2(p-1)	p(p-1)	(p^2-1)	p ⊖	Eqn(3.59)
A to LSP	k(2p+3)-M+2	-	(7+k)p-k	k(p+1)(TLU)	Durbin Ref.[49]

(a)

Adaptive-LSP (Eqn 4.74)

	⊗	⊕ / ⊖	others	Remarks
Error Signal	(p+1)	(2p+3)	-	Eqn(4.48)
Gradient Vector update	p(p/2-1)/2	p(p/2-1)	-	Eqn(4.54)
Coefficient Vector update	(p+3)	(p+2)	p ⊖	Eqn(4.73)

(b)

p - order of LPC model.
⊗ - multiplication operation, ⊗ & ⊕ - multiply & add operation,
⊕ - division, ⊕ / ⊖ - addition or substraction,
TLU - cosine table look-up.
k - [M+p(J+ι)]/2,
M - number of spectral points(typically M=128 for p=10),
J - INT[$\log_2[2\pi/(M.tol)]$]-1, INT - integer of -.
ι - INT[$\log_2[2\pi/(p.tol)]$-1],
tol - an accuracy factor (typically tol=0.005 for p=10).

Table 4.3 Comparison of the computation complexity.
(a) for a N-sample block of speech,
(b) for each adaption of the adaptive LSP filter.

Figure 4.19

CHAPTER 5 AN ERROR PROTECTED SPEECH CODER FOR MOBILE RADIO CHANNEL

5.1 INTRODUCTION

Low bit-rate digital speech coders suit the current trend of ISDN development [2,3,4], and are required for increasing channel capacity in Land Mobile Radio (LMR) systems. The advantages of such coders have been briefly described in Chapter 1. However, speech coding transmission over LMR channels must be robust to channel errors, especially errors due to Rayleigh fading. As the amount of information contained in each bit is inversely proportional to the speech coding rate [35], each encoding bit, therefore, becomes more important as the bit-rate is reduced. Thus, a very small number of bit errors can corrupt the speech very severely and cause roughness or missing information in the reconstructed speech. Channel coding is a possible way of mitigating the harsh effect of frequent errors in the LMR channel, but the added redundancy contradicts the aim of high spectrum efficiency. Although the speech coding rate can be reduced to balance the added redundancy, this in turn imposes more distortion in the original speech quality. To be acceptable for an effective LMR system, the digital voice transmission must produce speech quality better than or at least comparable to that of present analogue systems, and must also have high spectral efficiency. To design a coding technique for such systems, an understanding of speech coding schemes and channel coding capability is essential to compromise between the degradation in speech coder voice quality itself and the error correction effect of the channel coding.

The thesis so far has described an LMR channel simulation, the theory of a speech coding algorithm (APC-AB) and the properties of speech

coding parameters including LSP coefficients. Using this knowledge, a channel coding method based on the APC-AB speech coder is investigated in this Chapter.

In our study, a 900 MHz FDMA single channel per carrier (SCPC) LMR system is chosen. The gross bit-rate of the coding system for transmission is assumed to be 16 kb/s. Consideration is given to transmission with zero errors and to transmission in the presence of errors with error ratios of 1:100 and 1:1000. For comparison with analogue FM system currently in use in Western Europe, the bit error rates (BERs) of 1:100 and 1:1000 were equivalent to carrier to noise ratios, C/N, of 18 and 26 dB which give 90% and 50% area coverage respectively [72]. With these assumed conditions, the target is to develop a combined speech and channel coding technique to achieve a satisfactory average performance with BER up to 1 %.

The organisation of this Chapter is as follows : Section 5.2 is concerned with measures of channel distortion caused by random and LMR fading burst errors. Speech coding distortion measurements at various speech coding rates are also described. In Section 5.3, two forward error correction (FEC) coding schemes, using Reed-Solomon codes and Hamming codes, will be investigated. Based on the distortion measure results and the error correction capability, a bit-by-bit error protection method is developed. In Section 5.4, it will be shown that more reliable speech transmission can be achieved if a "zero-redundancy" error correction scheme is used, this being based on a good understanding of the properties of LSP coefficients.

5.2 CHANNEL AND SPEECH CODER DISTORTION MEASUREMENTS ON APC-AB

An LMR channel is a time varying channel which is affected by mobile speed, ambient environment, motor noise etc. Analysis of the effect of channel distortion on a speech coding system can be simplified by using a model as shown in Figure 5.1, from which the total speech distortion, D, over the channel can be expressed as

$$D = E[\ (\ s(n) - s_r(n)\ )^2\ ] \qquad (5.1)$$

where $s(n)$ is the original speech sample and $s_r(n)$ is the reconstructed speech sample at the receiver. Assuming $s_t(n)$ is the reconstructed speech over an error free channel, (5.1) can then be written as

$$\begin{aligned} D &= E[\ (\ s(n) - s_t(n) + s_t(n) - s_r(n)\ )^2\ ] \qquad (5.2) \\ &= \sigma_s^2 + \sigma_c^2 + 2\rho\sigma_s\sigma_c \end{aligned}$$

where $\sigma_s^2 = E[(\ s(n) - s_t(n)\ )^2]$ and

$\sigma_c^2 = E[(\ s_t(n) - s_r(n)\ )^2]$.

σ_s^2 is the speech coder distortion, and σ_c^2 is the channel distortion due to channel errors. ρ is the cross-correlation coefficient between σ_s and σ_c, which can be approximated to be zero [33]. The total distortion, D, is therefore simplified as

$$D = \sigma_s^2 + \sigma_c^2 \ . \qquad (5.3)$$

Once a speech coding scheme has been chosen for the LMR system, the total distortion may be minimised by using antenna diversity, advanced modulation schemes and error control coding in order to lower the probability of channel errors. Error control coding is a powerful and flexible adjunct to speech coding. However, due to the high channel error rates which occur in mobile radio channels, only high redundancy error correction codes are effective and these will decrease the channel capacity considerably.

To minimise the total speech distortion at a fixed transmission rate, there is a trade off between σ_s^2 and σ_c^2. Speech quality degradation, σ_s^2, can be decreased by using higher bit-rate speech coding method. However, higher channel distortion would then be a problem as high redundancy channel coding could not be accommodated. Contrarily, increasing the number of redundancy bits for channel coding would reduce σ_c^2, but this would have to be done by reducing the speech coding bit-rate thus reducing the speech quality.

In various low bit rate speech coding schemes [11], such as APC-AB, the transmission code consists of many different speech parameters. The speech distortion which occurs due to channel errors will depend on which parameters are most susceptible to corruption. Some speech parameters may produce large speech distortion if received in error, whereas others may produce relatively minor effects. Hence, an effective error control coding scheme could be achieved by protecting the individual speech parameters according to their sensitivity to channel errors. It may be possible to have no protection for parameters for which the sensitivity to channel errors is very low. Channel coding design using this approach must be based on measures of the effect of channel errors on each type of speech parameter. In our study, two time domain distortion measures, the long term signal to noise ratio, SNR_{long}, and the segmental signal to noise ratio, SNR_{seg}, which have been defined in Section 3.4, will be used to measure the performance of APC-AB in the presence of transmission errors.

5.2.1 CHANNEL DISTORTION ASSESSMENTS ON APC-AB

The APC-AB speech coding system proposed by Itakura [31] has been described in Chapter 3. Briefly, input speech is sampled at 6.4 kHz and each 20 msec segment of 128 samples is encoded by 320 bits using natural binary coding of a range of different parameters. Figure 5.2 shows a 320-bit transmission frame format of the 16 kb/s APC-AB coder. Each frame contains an 8-bit synchronisation codeword for frame synchronisation, 80 bits of side information that consists of 6 different speech parameters, and 232 bits of subband residual data. Before proceeding to consider any error correction technique, we first investigate the performance of the 'unprotected' APC-AB coder in the presence of transmission errors. In the following, the results of three different channel assessments are presented from which a better understanding of the effects of channel errors on the decoded speech can be gained.

The anticipated performance of a speech coder when used in a noisy channel may be assessed by simulating the coder in software and introducing errors in the transmission code. In the first assessment, the effect of random channel error and LMR Rayleigh channel error are investigated. For the Rayleigh fading channel assessment, noncoherent frequency shift keying (NCFSK) is adopted as the mode of transmission. The simulation of this system has been described in Chapter 2. At mobile speeds of between 50 km/h to 100 km/h, bit error rates of 1:100 and 1:1000 are equivalent to CNR values of about 20 and 30 dB respectively. Figure 5.3 illustrates the performance of the APC-AB coder in those channels by plotting the SNR_{seg} against bit error rate (BER). The speech material used is a 30 sec section of a low pitched male voice and a 30 sec section of high pitched female speech.

Figure 5.3 shows that the performance of the coder (measured in terms of SNR_{seg} in the random error channel was worse than its performance in the Rayleigh fading channel. This is because at fixed BER, the random errors are evenly distributed over a given speech packet, whereas the fading channel errors tend to be clustered together. As bit errors produced by LMR fading channels tend to be concentrated in a smaller number of speech parameter codewords, this tends to cause less erroneous speech parameters than occur in purely random error fading channels. Hence, the coder degrades more severely in the random error channel than in the Rayleigh fading channel.

By comparing the coder performance at two different vehicle speeds as shown in Figure 5.3, it may be seen that the severity of the Rayleigh fading depends on the mean CNR and vehicle speed. Table 5.1 (in Figure 5.13) gives the level crossing rate and the mean fade duration for a few mean CNR values and mobile speeds which can be computed by Equation (2.23) and (2.26). From Table 5.1, it is seen that at 900 MHz and 50 km/h, if the mean CNR is 15 dB, the carrier signal falls below the noise level about 17 times per second and remains below that level for about 2 msec for each fade. Hence, the signal is in a fade approximately 10% of a speech frame. The bit error rate in the fade depends on the modulation technique and the depth of the fade. If the carrier frequency and the mean CNR remain unchanged, and the mobile speed is increased to 100 km/h the signal is in a fade approximately 5% of a speech frame. However, the fading rate increases to 35 fades per second and causes more erroneous speech parameters. Thus, the speech distortion is higher and the SNR performance curve at 100 km/h becomes closer to the random error channel performance curve.

Informal listening tests were carried out on the noisy channel speech at BER 1:100. The quality of the reconstructed speech was severely corrupted, and had substantial roughness, frequent pops and clicks, and occasional reverberant sounds. Attempts were then made to identify the nature of the distortion which is caused by errors in specific APC-AB parameters, the aim being to discover which parameters were most sensitive to channel errors. Random errors at BER 1:100 were introduced in a single specific APC-AB parameter on a frame by frame basis. All the other APC-AB parameters remained intact for each frame. For example, the effect of errors in the pitch period parameter was evaluated by introducing random errors in the 7-bit pitch period codeword as it occurs frame by frame. This evaluation was repeated for all other APC-AB parameters. Figure 5.4 illustrates the results of these evaluations in the form of bar graph, which shows the influence of errors in each parameter on SNR_{seg} and SNR_{long}. The capital letters underneath the horizontal axis in Figure 5.4 indicate the speech parameters indexed as denoted in Table 5.1. The letters F1, F2 and F3 represents the 1st, 2nd and 3rd subband LSP coefficient sets; and G1 to G7 represents the first seven KLT coefficients. As the distortion of all the other KLT coefficients was somewhat similar to G6 and G7, they are not shown in the bar graph. The dotted line represents the performance of the coder at 0% BER. It may be deduced that from the graph that different parameters are error sensitive to different degrees as measured in either SNR_{seg} or SNR_{long}.

From informal listening tests, the speech distortion due to error in C, F2, F3 or H was less annoying than the distortion caused by errors in the other parameters and was more uniformly distributed throughout the duration of the speech. The difference between the effect on SNR_{seg} and

the effect on SNR_{long} for these parameters was relatively small. The effect of errors in the pitch gain parameter was relatively unnoticeable. When random errors were introduced into parameters F2 or F3, low level whistling noise was heard in the speech background. The effect of introducing errors into the coded residual (H) was to produce natural sounding speech but with some roughness. Although Figure 5.4 is indicating a large effect on SNR due to errors in H, the distortion appeared less distributing than that caused by errors in other parameters, for examples, F2 and F3. It was found that when the difference between SNR_{seg} and SNR_{long} was large, for the parameters B, D, E, F1 and G, the distorted speech had annoying features, such as frequent pops and clicks, and reverberant sounds. The speech distortion due to errors in the pitch period parameter (B) or errors in the sub-interval position (D) was reverberant sounding because of the abnormal periodic propagation by the pitch predictors and incorrect recovery of the adaptive bit allocations. Errors in the first subband LSP coefficient (F1) caused frequent pops and clicks, which produced a lower SNR_{seg} and larger difference between SNR_{seg} and SNR_{long}. Perceptually this was the most annoying form of distortion as compared with the others. Errors in the first two KLT coefficients caused the distorted speech to have a very loud warbling sound (as if spoken in water), and errors in all other KLT coefficients resulted in low level warbling sounds. When random error was introduced in the average energy (D), the distorted speech had pops and clicks which was due to abnormal amplification of some speech segments.

From the results of informal listening tests, speech distortion due to the noisy channel can be categorised into two types: (i) transient distortion perceived as pops, clicks and reverberant sounds, (ii) uniform

distortion and general roughness distributed throughout the speech signal. In speech transmission, although the underlying speech quality may be very good, the effect of transient distortion, causing occasional strong pops and clicks can be to severely degrade the subjective quality of the speech coder performance. From the previous two measurements, it may be observed that for speech containing the more subjectively annoying transient distortion, the difference between the SNR_{seg} and SNR_{long} is large. It is known that [33,73] in general in an error-free channel, the objective performance of a speech coder as measured by SNR_{seg} is in better agreement with subjective assessments than would be the case with SNR_{long}. This is because when the distortion is not very large, the SNR_{seg} is equally weighted for very large and small energy parts of a speech signal. However, it was found that when the performance of a coder is measured in a noisy channel and the distortion in some of the decoded signal segments is very large, the SNR_{long} seems to reflect high levels of transient distortion better than the SNR_{seg}. The speech parameters which tend to produce transient distortion in noisy channels are B, D, E, F1 and G. It is confirmed in Figure 5.4 that the SNR_{seg} values seem less sensitive to transient distortion than SNR_{long}. Therefore, it is concluded that both objective distortion measures are useful in assessing the performance of speech coders over noisy channels. SNR_{seg} is more useful when the speech distortion is small, and SNR_{long} is more useful when the speech distortion is large. Ultimately, performance must be judged by formal subjective listening tests and there may be some discrepancies between what is predicted by either objective measurement and what is actually heard. Objective measurements are nonetheless useful as they are predictable, easy to obtain and may be used for automated

optimisation procedures which can continually monitor the system performance.

From the previous channel distortion measurements, it is found that errors in the side information tend to result in transient distortion in speech; that is, subjectively disturbing pops and clicks or warbling sounds tend to occur. A set of measurements referred to as the bit inversion test is therefore introduced to test the importance of each bit in the side information. In this test, a single bit in the side information of every speech frame is inverted so that the distortion in the reconstructed speech is only due to the inverted bit. This test was carried out for each bit of the binary code for pitch period (B), pitch gain (C), sub-interval position (D), average energy (E), KLT coefficients (G) and the most significant bits of every LSP coefficients (F'). The results of this measurement are expressed in terms of SNR_{seg} and SNR_{long}, and are illustrated in Figure 5.5.

Figure 5.5 shows that B, D, and G are equally important, probably because these parameters are used to recover the bit allocation scheme for the sub-band residual. If any bits of these parameters become erroneous, the bit allocation calculation at the decoder would be severely corrupted. When bit inversion was tested on most significant bits (MSBs) of the second and third LSP coefficients of the 1st sub-band, and MSBs of the first two LSP coefficients of the second sub-band, high levels of pops and clicks were detected. This is because the LSP coefficients are implicitly controlling the formant location and formant bandwidth. A bit inversion on the MSB of an LSP coefficient would be very likely to cause two or more LSP coefficients to become too close to each other and hence would tend to over-emphasise a speech formant.

It was also found that when the most significant bits of the first two LSP coefficients in the second sub-band were inverted, frequent loud squeaking noises were heard. This kind of deterioration in the decoded speech seems to be shown more prominently in the long term SNR measure.

Errors in the average energy code (E) have different effect. Bit inversion tests on the first two MSBs resulted in unintelligible speech, but the degradation in the reconstructed speech is barely noticeable when the last 2 bits are inverted.

When errors were introduced into the pitch gain (C), distortion was barely noticeable in the reconstructed speech.

By comparing the conclusions of informal listening tests conducted to assess the effects on bit inversion and objective measurements, it was concluded that the SNR_{long} again indicates the severity of the transient distortion more reliably than SNR_{seg}.

5.3 AN ERROR PROTECTED METHOD FOR APC-AB CODER

LMR channels can show very poor error rate performance. When the transmitted data is received in an error burst, the bit error rate (BER) can be as high as 1:2. From the channel distortion measurements described in Section 5.2.1, it is apparent that the performance of the 16 kb/s APC-AB coder degrades very rapidly when the mean channel BER is higher than 1:1000. Especially, at BER of 1:100, the decoded speech becomes very distorted and unnatural sounding and intelligibility is lost occasionally. Methods of using error correction coding and interleaving has been investigated in various publications [e.g.74,75,76], for mitigating the fading channel effects. With a view to improving the coder performance

at fading BERs of the order of 1:100, this section presents a bit-by-bit error control method, using error correction coding and interleaving, for the APC-AB coder under the constraint of keeping the channel transmission rate at 16 kb/s.

5.3.1 ERROR CONTROL CODING AND INTERLEAVING

There are two major types of error control coding methods: block coding and convolution coding. A block encoder produces a block of n output symbols which depend only on k input symbols in any given time unit. In convolutional encoders, however, each encoded block depends not only on the corresponding k message symbol block at the same time unit, but also on previous message blocks.

Considering the APC-AB coder, the binary output data stream has a well defined block structure with block length of 320 bits. Each of these blocks may be further divided into subblocks representing, for example different side information parameters and residual parameters. These sub-blocks pertain to information of unequal relevance to the process. It seems reasonable to favour block codes for encoding APC-AB digitized speech because this will make it easy to assign different degrees of protection to different blocks. We have therefore adopted block coding in this work.

The set of all possible encoded messages output from the block encoder is called a codebook, and each encoded message is called a code-word. The purpose of the code is to maximise the probability of correct decoding. It is thus desirable to ensure, as far as possible, that the transmitted codeword is not converted to another during transmission

[77]. This can be achieved by making the codewords as different as possible.

Error control coding generally takes on one of two forms:

(i) Automatic Repeat Request (ARQ). If the decoder detects errors in a block, it asks for a re-transmission of that block. ARQ systems become less efficient as the channel error rate increases.

(ii) Forward Error Correction (FEC). The decoder corrects a certain number of errors per block, thus eliminating the re-transmission time delay and obviating the need for a return channel. However, greater redundancy is needed and the decoding equipment is more complex in general.

As the ARQ process requires additional delays and feedback information for re-transmission, and this, especially, will not be very efficient for high error rate channels, FEC methods are, therefore, to be adopted to protect the APC-AB coder.

The error structure of the LMR channel tends to be in bursts as described in Chapter 2 and in Section 5.2.1, because the fading affecting the channel is of long enough duration to affect several bits of data. Simple FEC systems such as Hamming code designed to detect or correct errors may not able to deal with burst behaviour, so some degree of burst error correction is needed. An effective technique for applying FEC on a burst error channel is to use interleaving. A simple interleaving technique for use with block codes can be visualised as a rectangular array with I rows and N columns. The vertical dimension of the array, I, is called the interleaving degree. If an error burst lasts for b code symbols and b is less than I, the interleaver is required to produce at most a single error in b consecutive codewords. With this approach, a burst error channel is transformed into a random error channel, which may allow many effective FEC coding techniques to be applied. Selection of I usually

depends on the expected error burst lengths, and care must be taken when channel disturbances are expected to produce periodic errors.

The use of interleaving can totally randomise a burst error channel, but at the expense of considerable delay which may be unacceptable for speech channels. For the APC-AB coder, it is therefore necessary to limit the degree of interleaving, and this will mean that some burst errors will remain in the interleaved code - but to a lesser degree than before interleaving. Because of the existence of these remaining burst errors, it is advantageous to adopt a more sophisticated error protection scheme than the simple single error correction code referred to earlier. The scheme to be used is the well known Reed-Solomon coding scheme which has the advantage of correcting multiple-bit errors.

In the following section, a well known single error correction code, the Hamming code, and an optimal burst error correction code, the Reed-Solomon code, will be described. They were applied for the APC-AB coder to protect the speech code from fading channel error. Then, the use of interleaving techniques will be described in greater length in Section 5.3.3.

5.3.2 FORWARD ERROR CORRECTION CODING

Before proceeding to describe the function of block coding methods, it is necessary to define a useful parameter known as Hamming distance. The Hamming distance between two codewords is defined as the number of symbols in which they differ. If d is the minimum distance between any pair of codewords in a codebook, then the maximum number of correctable

errors, t, and detectable errors, e, must obey the following inequalities [77]:

$$e \leq d - 1 \tag{5.4}$$
$$t \leq (d - 1)/2 \; . \tag{5.5}$$

Clearly, $t \leq e$, since an error must be detected before it can be corrected. Normally, an error correction code capable of correcting t errors is denoted by (n,k,t), where n is the overall block length and k is the information block length. The efficiency or the rate of the code is given by

$$R = k/n \tag{5.6}$$

which indicates the degree of required redundancy for the code.

Possibly the simplest and most efficient error correcting code is the Binary Hamming code [78] which has a minimum distance of 3, and can therefore correct one bit error in each n-bit block. If the added number of "parity" check bits, m=n-k, is greater than or equal to 3, there exists [79] a Hamming code with the following parameters:

$$n = 2^m - 1 \tag{5.7}$$
$$k = n - m$$
$$t = 1 \; .$$

Then, the rate of the Hamming code is:

$$R = (n-m)/n = 1 - m/n \; . \tag{5.8}$$

As the error pattern produced by an LMR channel tends to be in bursts, the use of Hamming code may not be powerful enough for this channel since it can correct only one error per block even after a limited degree of interleaving. A more sophisticated "burst error correction" coding scheme is therefore required. In theory [77], the number of parity check symbols that must be introduced to allow a single burst of b consecutive symbol errors within a block of n symbols to be corrected is at least 2b. Therefore,

$$n - k \geq 2b \tag{5.9}$$

Any burst error correction code which satisfies the inequality: $b \leq INT[(n-k)/2]$, where INT[.] is the integer of, are said to be optimal. Reed Solomon (RS) code [80] is considered as the best type of block code for dealing with bursts of errors in codewords. RS code is a symbol error correction code. Instead of coding blocks of individual bits, as the Hamming code does, the RS code encodes blocks of symbols each symbol representing a group of say ℓ bits. An RS code is denoted by the parameters (n,k,t) when n is the block length, i.e., the number of ℓ-bit encoded symbols to be transmitted, and k is the number of ℓ-bit symbols. The encoded RS code is a block sequence of Galois field (or finite field) symbols [77], each symbol being represented by a sequence of ℓ bits. Each symbol is an element of a finite field of 2^{ℓ} symbols referred to as a Galois field or $GF(2^{\ell})$. The encoding and decoding process for RS code requires finite field arithmetic, and therefore the complexity of the RS codec is higher than the Hamming codec. For t-symbol error correcting RS code over $GF(2^{\ell})$, as denoted by (n,k,t), with each symbol represented by ℓ bits, it follows that [77]

$$\begin{aligned} n &= 2^{\ell} - 1 \text{ symbols} = \ell(2^{\ell}-1) \text{ bits} \\ k &= n - 2t \text{ symbols} = \ell(n-2t) \text{ bits} \\ d &= 2t + 1 \text{ symbols.} \end{aligned} \tag{5.10}$$

Therefore, a burst error of length of t symbols, or t successive erroneous symbols, can be corrected. For example, if (15,9,3) RS code over $GF(2^4)$ is considered, 36(=k.ℓ) bits of information are grouped into 9 4-bit symbols, where each symbol formed is an element of an alphabet of 2^4 symbols. Then, 9 symbols are encoded to 15 channel symbols, and 3 successive symbol errors (=12 bit error burst) can be corrected. However, a burst error of length $(t-1)\ell+1$ bits or less, which is equal to 9 bits in

this case, can affect t(=3) successive symbols. Hence, the maximum length of the correctable bit error bursts is not strictly $\ell.t(=12)$ bits but in the worst case is $(t-1)\ell+1(=9)$ bits. The rate of the R-S code is given by

$$R = (n-2t)/n = 1 - 2t/n . \quad (5.12)$$

Normally, the rate of an RS code is lower than that of a Hamming code but the Hamming code can only correct one error in a n-bit block.

Methods for implementing Hamming codes and the RS codes are described in many standard text books [77,78] and no attempt is made to describe them here. For the block coding methods described above, if t or less errors are detected, at the decoder, the received codeword can be perfectly corrected. When a codeword is received with more than t errors, the decoding process is not only unable to correct errors but will also erroneously decode codewords received correctly and may cause more errors in the decoded data. For some systems, additional error detection capability is desired to alleviate the erroneous decoding conditions. It can be shown that [77] most block code can be extended by adding a minimal amount of redundancy to provide additional error detections capability. For example, a (15,11,1) binary Hamming code can be extended to (16,11,1) by adding an overall parity check bit to provide error correction of all single bit errors plus detection of all double errors (detect 2,4,6,8,10,12,14,16 errors). After error detection, the received data may be discarded or some other action may be taken.

When using error correction codes, an important fact must be taken into account [77]. If the channel error rate exceeds the correction capability of the code, the error correction coding technique will adversely affect the transmission quality. The transmission quality will generally

be worse than without error correction coding. Hence, the degree and type of error correction coding should ideally depend on the channel error characteristics. As mentioned in Section 5.2.1, LMR speech coders are often required to operate in conditions where fade duration can extend up to 10% of a speech frame (320 bits). This means that there could be up to 32 bits in a fade. Thus, the error correction coding system is required to be capable of correcting a burst error with length of at least 32 bits.

Consider the use of a Hamming code to cope with error burst of 32-bits or longer in the 320-bit frame length of an APC-AB codec. The 320-bit frame must be split into smaller blocklength for convenient coding. The choice of the Hamming code block-size has an important influence; the smaller the block length the less probable is the occurrence of an error in a block, but the coding of small blocks by the addition of redundant bits is less efficient than the equivalent redundancy coding spread over a large block. To correct error bursts of at least 32-bits within a 320-bit frame, the frame must first be randomised by an interleaving technique so that the errors are uniformly distributed within the frame and are isolated so that the probability of having groups of consecutive bits with more than one error remains small until the group size approaches 10(=320/32). Clearly, at least 32 Hamming coded blocks will be needed and therefore a (7,4,1) binary Hamming code (see Equation 5.7) is an appropriate choice. If a 320-bit frame is divided among 45 7-bit subblocks and each subblock is protected by a (7,4,1) Hamming code, more than 32 bit errors can be corrected when they are evenly distributed between each block. The burst behaviour of the LMR channel can be randomized by interleaving the Hamming encoded speech frame at a depth of 45. Hence,

in principle one channel error is intended to be trapped in a Hamming encoded subblock after de-interleaving, and can be corrected by Hamming decoding. As the number of parity check bits is equal to three for each subblock, the total number of redundant bits per frame is 135 bits. To maintain the transmission rate at 16 kb/s, the net speech coding rate has to be reduced from 16 kb/s to about 9.25kb/s. Speech quality at this rate will be degraded.

From the channel error distortion measurements for the APC-AB coder as described in Section 5.2.1, it was deduced that some of the APC-AB speech parameters are less sensitive to channel errors than others, as measured by the effect on the final speech quality. Hence, instead of protecting the entire bit stream of a speech frame, excessive overhead for redundancy can be avoided if error protection bits are only allocated to protect the bits or speech parameters which are most error sensitive. Based on the channel error distortion measurements in Section 5.2.1, a bit-by-bit protection scheme is described in the following section.

5.3.3 AN ERROR CONTROL METHOD ON APC-AB

From informal listening tests it was found that the most undesirable effect in the decoded speech was the transient distortion. The effect is fairly prominent when some side information is affected by channel errors; especially errors on the LSP parameters. Bit error rates of 1:100 on the residual bits only cause general roughness distributed throughout the decoded speech. To minimize the transient distortion high redundancy error correction coding can be applied to minimize the error probability on the error sensitive speech parameter. However, to keep the transmission

rate constant at 16 kb/s, the error correction code must be carefully applied to the error sensitive bits or speech parameters so that the amount of additional redundancy can be kept to a minimum.

To balance the additional error correction coding redundancy, the net speech coding rate must be reduced. The net speech coding rate can easily be changed by adjusting the mean residual bit rate, R, as given in (3.82). A simple experiment was performed to investigate the performance of the coder when R is varied from 0.875 to 1.88; equivalent to net speech coding rates from 9.6 kb/s to 16 kb/s. The objective performance is shown in Figure 5.6. Then, informal listening tests showed that good quality speech was still maintained at 12 kb/s (R=1.25). Below this rate, slight roughness was heard in the decoded speech. Thus, it is desired that the additional redundancy should not more than 4 kb/s.

LSP coefficients change relatively slowly between frames [61], which means that the LSP information has a fair amount of redundancy when compared over a number of frames. Thus, instead of using highly redundant FEC (forward error correction) codes, less redundant error detection codes may be used. If errors are detected in the LSP coefficient set for a frame, the previous LSP coefficient set may be used without incurring excessive distortion. For the 'protected' APC-AB coder being investigated, two (16,11,1) extended binary Hamming codes are used to protect the three sub-band LSP coefficient sets. The LSP coefficients of the APC-AB codec are quantized at 3 bits per coefficient. Figure 5.4 indicates that the overall speech quality is more sensitive to the first sub-band LSP coefficient, F1, than to the second and third sub-band LSP coefficient, F2 and F3 respectively, because F1 usually contains more important speech information than other sub-bands. The bit inversion test

results as depicted in Figure 5.5 suggests that the following scheme for protecting the most sensitive LSP coefficient bits will be reasonably efficient.

(i) A first (16,11,1) Hamming code applied to all the LSP coefficients in band 1 except the last bit of the last LSP coefficient;

(ii) A second (16,11,1) Hamming code applied to

a) the two most significant bits of the first three LSP coefficients and the most significant bit of the last coefficient in band 2,
b) the most significant bit of all the LSP coefficient in band 3.

If errors are detected in the first Hamming code, the erroneous first sub-band LSP coefficients are replaced by the corresponding coefficients from the previous frame. If errors are detected in the second Hamming code, band 2 and band 3 LSP coefficients from the previous frame are used. All the other LSP bits are left unprotected as they are not too sensitive to channel errors.

Apart from the LSP coefficients, there are another 44 bits of side information including pitch period, pitch gain, sub-interval position, average energy and KLT coefficients. Figure 5.4 and Figure 5.5 shows that speech quality is not very sensitive to channel errors in the pitch gain, the least two significant bits of the average energy and the least significant bit of the pitch period. Errors in the remaining 38 bits seem to cause a similar degree of distortion as measured objectively and subjectively, because all these bits are very important for the residual bit allocation computation at the decoder. To achieve smaller error probability on these inter-related bits, a burst error correction code Reed-Solomon (15,9,3) code is found to be appropriate. The details of this code were briefly described in a previous section. The side information bits to be protected by the RS code are as follows:

a) the first 6 bits of the pitch period,

b) the 5-bit sub-interval position,
c) the first 3 bits of the average energy, and
d) all the KLT coefficients except the least significant bits of the first two KLT coefficients.

The total number of bits for the side information, the redundancy bit and the synchronisation codeword is 122 bits (=80+34+8 bits). Then, the total number of bits that can be allocated for the residual samples is 320-122=198 bits. Hence, the mean residual bit rate is 1.547 (=198/128) bits/sample; this is equivalent to a net speech coding rate equal to 12.9 kb/s. Then, the encoded speech format can be as shown in Figure 5.7.

As given in (5.11), the (15,9,3) RS code can correct a b_{min}(=9) bit burst. The Hamming code can correct 1 error and detect every double error. A typical burst length in LMR channels is 32 bits, and therefore this long burst error can easily destroy the function of the error correction code. Hence, an interleaving technique must be used to break up the burst to enhance reliability of the FEC decoding. Various interleaving structures were tested. The one that was found to be more efficient is shown in Figure 5.8. Basically, the interleaver is designed to separate the RS code symbols as far as possible within a speech frame. Also the error burst should be randomized to avoid burst error in the de-interleaved Hamming code. To achieve this aim, the width of the interleaver is set at a length of 15 symbols. In the interleaving process, the frame synchronisation codeword is not interleaved and is always situated at the beginning of a frame. The 15 4-bit symbols are first read into the interleaver by rows, and all the remaining bit symbols fill up the interleaver by rows. After all the symbols have filled up the interleaver, the interleaved symbols are read out by columns for transmission. The transmission sequence for a frame can be depicted as follows:

SYNC R1 69 84 .... 309 R2 70 85 ..... 307 R15 83 308

where R stands for the RS code symbols. In this way, the RS code can correct an error burst of at least 43 bits, and burst errors are less likely to occur in the de-interleaved Hamming code.

The 'protected' APC-AB coder was tested using the simulated Rayleigh fading channel for mobile speeds of 50 km/h and 100 km/h. The performance was measured in SNR against bit error rates (BER) in the range of 1:1000 to 1:100, as the SNR_{long} was shown to be a better performance measure than SNR_{seg} for testing the coder in a noisy channel. The SNR_{long} and SNR_{seg} measures are plotted in Figure 5.9 and Figure 5.10 for comparisons.

Figure 5.9 also shows the performance of the protected coder in the simulated channel when interleaving is not applied. By comparing Figure 5.9 and Figure 5.3, it may be seen that the error-free performance of the protected coder was reduced by about 1 dB, because the net speech coding rate was reduced to 12.9 kb/s. However, the SNR_{seg} performance of the protected coder was improved by about 1.5 dB at mean BER of 1:100. From the SNR_{long} performance graph in Figure 5.9, it may be observed that the coder behaves better at higher mobile speed when the speech code is not interleaved. This may be because in an LMR channel, as the mobile is moving slower, the mean burst duration will be longer. At mobile speeds of about 50 km/h, the mean fade duration is equivalent to about 32-bit duration, but the RS code can only correct up to 12 bits of burst error. Hence, the RS correction is very likely to be destroyed in this environment. When the mobile speed is higher and the mean fade duration is correspondingly shorter, the RS code is likely to have a higher tolerance. The protected coder was also tested in a random error channel. The performance is objectively and subjectively better than in an LMR channel.

Its SNR measures are also plotted in Figure 5.9. This random error test can be considered as indicative of the performance of a protected coder when it is perfectly interleaved. To attain this performance in an LMR channel, the transmission speech code may need to be interleaved between many speech frames which would then increase the codec delay dramatically.

As interleaving between frames introduces more codec delay, in our experiment interleaving was only performed within each individual speech frame as described previously. Figure 5.10 shows the SNR measurements of the protected coder with the use of interleaving. Comparing the performance curves in Figure 5.9 at BER of about 1:100, the effect of interleaving is improved by the introduction of error correction coding, and this improvement is more pronounced at lower mobile speed; at 50 km/h the SNR_{long} value is improved by about 4 dB. However, at BER of around 1:1000, the SNR_{long} value is reduced by 1 dB by the use of interleaving. Referring to Table 5.1, a BER of 1:1000 corresponds to a CNR about 30 dB. The mean fade duration is about 0.303 msec at 50 km/h which is approximately equivalent to a 5-bit time duration. Therefore, error bursts of about 5-bits are to be expected. The RS code being employed is capable of coping with 5-bit error bursts even without interleaving, and the combined Hamming code will perform better with interleaving. Hence, it appears that the 1 dB degradation at BER of 1:1000 is caused mainly by errors in the residual data rather than by errors in the APC-AB side information. This is because at lower BER the occurrence of transient distortion will become very infrequent as the side information is well protected by error correction code. However, an error burst will be randomly dispersed in its effect on the residual bits after the de-interleaving process. This causes a higher word error rate on the unprotected residual samples. For example,

one short burst (affecting say 5 consecutive bits) may originally affect one residual sample, whereas the same burst error dispersed by interleaving may generally cause more erroneous residual samples.

The residual signal has higher resistance to channel errors and, perceptually, errors in the residual should not cause disastrous effects in the final speech quality. When Figure 5.9 and Figure 5.10 are compared at mean BER 1:100, the SNR_{long} values are seen to be greatly improved after the interleaving process. This may imply that there is less transient distortion in the decoded speech. From informal listening tests for the 'protected' APC-AB coder, it was concluded that the speech quality at BER 1:100 was improved compared with the 'unprotected' coder. The decoded speech became more stable and intelligible and the transient distortion effects were greatly reduced. Therefore, it is concluded that error correction coding and interleaving are very useful in speech coding for transmission in LMR channels..

5.4 'ZERO-REDUNDANCY' ERROR CONTROL METHOD FOR APC-AB

Although error correction coding and interleaving techniques can improve the quality of digitized speech transmitted in Rayleigh fading environments, the degree of improvement is limited by the amount of additional redundancy that can be accommodated and the amount of codec delay, as needed by the error correction coding and interleaving schemes, that can be tolerated. However, with an understanding of the properties of the transmitted speech parameters, it is possible to devise a scheme which achieves more robust speech transmission in high BER environments at no additional cost in terms of bit-rate and delay. This section pre-

sents an example of such a scheme which may be applied to LSP coefficient transmission. The scheme will be referred to as the 'zero-redundancy' error control method and is based on the properties of LSP coefficients which allow a degree of error detection and 'correction' without introducing additional redundancy.

5.4.1 LSP QUANTIZATION METHODS

LSP coefficients provide an alternative representation of the short term spectra of speech segments and may be extracted from blocks of speech according to a minimum mean square error (m.m.s.e.) criterion as described in Chapter 3. Using these coefficients, a block of speech samples is transformed into a block of residual samples which can then be more efficiently encoded. If the LSP coefficients are corrupted by channel errors, the m.m.s.e. criterion is destroyed, and also the inverse transformation from the residual signal back to speech using erroneous LSP coefficients may cause a range of different effects such as transient distortion in the reconstructed speech. The LSP encoding technique used so far is based on Soong's LSP difference encoding technique [49], which was described in Section 3.3.3.2. The advantages of this method are:

a) the dequantized LSP coefficients are always in stable order,

b) it is more efficient than the direct LSP quantization method with respect to quantization noise.

The direct LSP quantization method individually encodes each LSP coefficient by a scalar quantizer. The major drawback of the direct method is that the histograms of the LSP coefficients overlap each other as shown

in Figure 3.14 so that the de-quantized LSP coefficients may cross over and thus fail to obey the ordering condition for stability.

In noisy channel environments, the disadvantage of the LSP difference encoding method is that if an LSP difference value is corrupted by channel errors, this error will propagate and effect all the LSP coefficients up to and including the last of the frame. One very simple way of detecting such error propagation is to examine the value of the last LSP coefficient. If this value is greater than or equal to 180°, it is clear that channel errors have seriously affected the LSP difference code. In such a case, the previous stable set of LSP coefficients may be used to replace the erroneous set. However, in many cases, the last coefficient will not exceed 180° even when channel errors have occurred. An alternative method is therefore required to cover a wide range of errors.

Returning to the direct LSP quantization method, each coefficient may be quantized by Max's quantizer [34] whose design is based on a knowledge of the mean and standard deviation of each individual LSP coefficient as given in Figure 3.13. If each coefficient is finely quantized, the quantized coefficients will not cross over and will satisfy the stability criterion of LSP coefficients as:

$$0 < f_1' < f_2' < .... < f_n' < .... < f_p' < \pi \qquad (5.13)$$

where p is the LPC order in each sub-band, and an odd index and even index of f' corresponds respectively to a θ_n and a ω_n coefficient as given in (3.64). However, if channel errors occur, the corrupted LSP coefficients may cross over after decoding. Fortunately, this crossing over of coefficients can be detected, and corrective action can be taken without adding any transmission redundancy redundancy. This approach is further investigated and described below.

For the APC-AB coder, it was found that [31] it is sufficient to encode each coefficient by 3 bits. However, problems can occur with direct quantization the distance between a pair of adjacent coefficients is very small. In this case, the quantized LSP coefficients may fail to be correctly ordered for stability due to the effects of quantization. LSP coefficients are highly correlated with speech formant frequencies and bandwidths. A slight change in one of the LSP coefficients changes the spectral shape of the speech mainly around its particular frequency. The introduction of unstable ordering by quantization at the encoder must be avoided by perturbing the crossed over coefficients slightly until the LSP coefficient set is stable.

Assume that p scalar quantizers for the p LSP coefficients are predesigned for the direct LSP quantization method, and let $f_{i,j}'$, $i=1,2,..,p$, $j=1,2,..,s^{b}i$, denote the jth reconstruction level of the quantizer operating on the ith LSP coefficient. b_i is the number of bits allocated to the ith coefficient. A typical vector $(f_1, f_2, \ldots, f_p)$ of LSP coefficients should be quantized to a vector $(f_{1,j}', f_{2,j}', \ldots, f_{p,j}')$ of reconstruction levels. After quantization, the stable ordering of the quantized LSP coefficients should be tested. If a pair of coefficients fails to satisfy the stable order condition:

$$f_{i,j}' > f_{i+1,j}' \tag{5.14}$$

then f_{i+1}' is taken as a reference coefficient. The distances between f_{i+1}' and f_{i+2}', and between f_{i+1}' and f_{i-1}' are measured as

$$\begin{aligned} d_u &= f_{i+2}' - f_{i+1}' \\ d_\ell &= f_{i+1}' - f_{i-1}' \end{aligned} \tag{5.15}$$

According to the LSP property, if d_ℓ is smaller than d_u, the speech formant peak near f_{i-1}' is higher than near f_{i+2}'. To avoid speech formant

over-emphasis after reordering, the reconstruction level of $f_{i+1,j}'$ is moved up one level higher to $f_{i+1,j+1}'$. On the other hand, if d_u is smaller than d_ℓ, the reconstruction level of $f_{i,j}'$ is moved one level lower to $f_{i,j-1}'$. After the re-ordering of the quantized LSP coefficients, the ordering check is repeated until all the quantized LSP coefficients are in a correct order for stability. Then, the code for each re-ordered quantized value will be sent to the receiver. Figure 5.11 presents a comparison of the two quantization methods: the LSP difference method (red curve) and the direct method (green curve), for each subband of the APC-AB coder for a 2 second segment of speech. The black curve illustrates the unquantized LSP coefficients. It may be seen that the red curves follow the black curve better than the green curves, which implies that, quantization-wise, the LSP difference encoding method performs better than the direct LSP quantization method. However, in informal listening tests, the direct method was still able to maintain good quality speech for the 16 kb/s APC-AB coder, and only slight degradation was heard compared with the LSP difference method. The speech quality was also measured in terms of SNR_{seg}, and it was found that only 1 dB degradation was caused by using the direct LSP quantization method for a 16 kb/s 'protected' coder.

5.4.2 'ZERO-REDUNDANCY' ERROR CONTROL METHOD FOR LSP COEFFICIENTS

The above method guarantees that the quantized LSP coefficient set is in stable order before it is transmitted to the channel. If 'crossover' coefficients are, however, detected after de-quantization at the decoder, this indicates that the received LSP parameters must have been affected by channel errors. A simple technique for ensuring LSP parameter stability

is to use the previous 'stable' set of LSP coefficients, once 'crossover' is detected. Figure 5.11 shows that the LSP coefficients change relatively slowly between frames, which means that the LSP coefficients may be strongly correlated between frames. Hence, a better method for improving the 'crossover' situation is to replace just the individual erroneous LSP coefficient by the same coefficient from the previous frame instead of replacing the whole set. However, the erroneous LSP coefficient may not correctly be identified, as a crossover condition involves at least two coefficients. A better detection method is to use the knowledge of LSP coefficients gained from previous frame.

Assume that a vector of LSP coefficient (f_{1n}', f_{2n}', ..., f_{pn}') is received, where n is the frame index. If a pair of decoded LSP coefficients, $f_{i,n}$ and $f_{i+1,n}$ fails to satisfy the stable ordering condition:

$$f_{i,n}' > f_{i+1,n}' \qquad \text{for some } i,$$

the distances between the n-th frame coefficients and the (n-1)th frame coefficients are then measured as

$$d_{i,n} = | f_{i,n}' - f_{i,n-1}' |$$

$$d_{i+1,n} = | f_{i+1,n}' - f_{i+1,n-1}' | .$$

With the assumption that the LSP coefficients change fairly slowly between frames, if $d_{i,n}$ is larger than $d_{i+1,n}$, the i-th LSP coefficient is more likely to be corrupted by channel errors than the (i+1)th coefficient. In such a case, $f_{i,n}'$ is set equal to $f_{i,n-1}'$. On the other hand, if $d_{i+1,n}$ is larger than $d_{i,n}$, then $f_{i+1,n}'$ is set equal to $f_{i+1,n-1}'$. This 'zero-redundancy' algorithm is implemented as follows, assuming that p is the order of the LPC analysis:

Step 0 Initialisation : Set i = 1.

Step 1 If $i \geq p$ then EXIT (error 'correction' finished).

Step 2 Error detection :
IF $f_{i,n}' \geq f_{i+1,n}'$ then goto Step 3,
else increment i by 1 and goto Step 1.

Step 3 Error 'correction':
If $|f_{i,n}'-f_{i,n-1}'| \geq |f_{i+1,n}'-f_{i+1,n-1}'|$ then
set $f_{i,n}' = f_{i,n-1}'$
else set $f_{i+1,n}' = f_{i+1,n-1}'$.
Set i equal to 1 and go back to Step 1.

This algorithm was applied to the APC-AB coder simulation, and its performance was tested at a random bit error rate of 1:100. Figure 5.12 illustrates the performance of the algorithm (green curve), and compares this with the LSP difference encoding technique (red curve). The black curves show the unquantized LSP coefficients. The red curves exhibit a number of 'spikes' when compared to the unquantized LSP coefficients these being caused by channel errors. With the help of the 'correction' strategy adopted for the direct quantization method, the green curves are closer to the unquantized LSP coefficients. The modified APC-AB coder was tested over a simulated Rayleigh fading channel at mean bit error rate 1:100 at mobile speed 50 km/h. The speech quality was improved and transient distortion was reduced. When compared with the LSP difference encoding method under the same channel conditions, the direct LSP quantization method improved the SNR_{long} value by approximately 1 dB. Hence, it is concluded that for high error rates the derived direct LSP quantization process as described above performs better than the LSP difference method.

5.5 CONCLUSION AND DISCUSSIONS

A bit-by-bit error control method for the APC-AB coder has been presented in this Chapter. The protected coder was tested over simulated LMR channels and random error channels. In designing error control coding for the coder, mean level crossing rate and mean fade duration of the Rayleigh type LMR channel were considered. From informal listening tests and objective measurements for the 'protected' APC-AB coder, it was concluded that forward error correction (FEC) coding is able to enhance robustness of the coder in high BER a LMR channel, and also interleaving is particularly useful when high redundancy FEC coding becomes difficult to apply.

The three channel distortion measurements described in Section 5.2.1 are found to be useful, especially for designing FEC code for a low bit-rate speech coder. The information conveyed by individual bits becomes very important for a low bit-rate coder and the selection of a bit-selective error protection process is crucial and should be based on the minimisation of some channel distortion criterion. An optimization algorithm has been used to decide how much FEC redundancy should be applied to each bit in a speech frame according to an objective test measurement considered suitable for testing low-bit coded speech transmitted over noisy channels.

It was found that at high bit error rates (BER), the decoded APC-AB speech will generally suffer from severe transient distortion which would lead to very poor performance in subjective tests. Elimination of transient distortion is, therefore, essential for a robust speech coder. From our tests, SNR_{long} seemed to provide a better indication of the severity of transient distortion than the SNR_{seg} measurements. This is because pops

and clicks in some badly corrupted speech segments do not show up prominently in SNR_{seg} measurements. The SNR_{long} measure is thus a better measure than SNR_{seg} for noisy channel speech tests.

FEC codes can alleviate but not totally eliminate the occurrence of transient distortion. Zero-redundancy error correction methods, by exploiting the properties of speech parameters, can greatly reduce and even eliminate transient distortion without increasing the required bit-rate. In this thesis, a 'zero-redundancy' error control method was derived for LSP coefficients. This method was applied to each APC-AB subband residual. If a pair of 'crossed over' LSP coefficients are detected, the previous frame of LSP coefficients may be re-used or a simple correction strategy can be applied. As LSP coefficients are closely related to speech formant locations and bandwidths [81], an even better correction strategy can be devised, based on a formant estimation method [81].

From experiments with the 'zero-redundancy' method, it was found that error detection plays a very important part in robust speech coder design, because a 'post-detection' strategy such as error correction or muting, can avoid the danger of decoding sets of severely corrupted speech parameters. Hence, if the Reed-Solomon (RS) code used to encode the pitch period, subinterval position and KLT coefficients is extended [77] by adding one more RS code symbol (4-bit) to gain additional error detection capability, a more robust speech coder is obtained by using a suitable 'post-detection' strategy. For example, a correction scheme based on a knowledge of pitch synchronous features in speech could be devised.

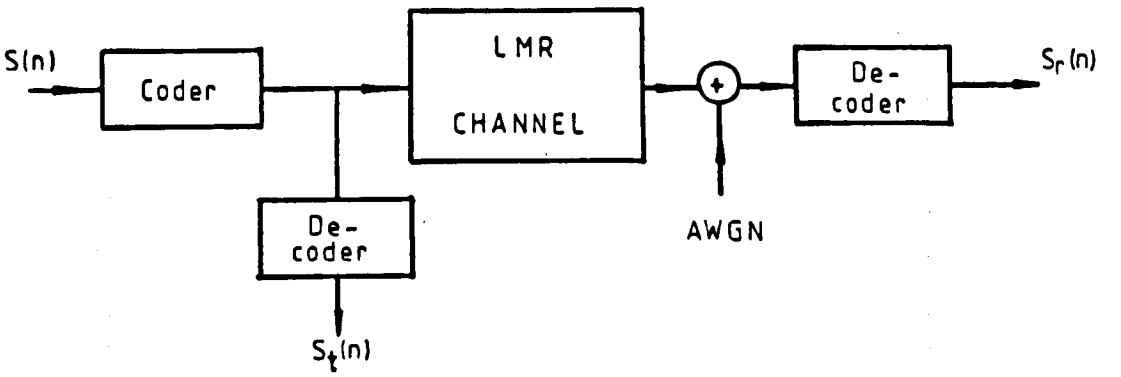


Figure 5.1 Block diagram for channel distortion analysis.

A)	SYNC (synchronisation codeword)	8 bits
B)	Pitch period	7 bits
C)	Pitch gain	3 bits
D)	Sub-interval position	5 bits
E)	Average energy	5 bits
F)	LSP coefficients	36 bits
G)	KLT coefficients	24 bits
H)	Sub-band residual	232 bits

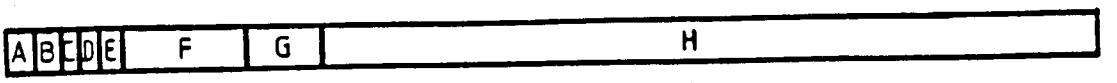


Figure 5.2 Frame format of 20 msec speech code.

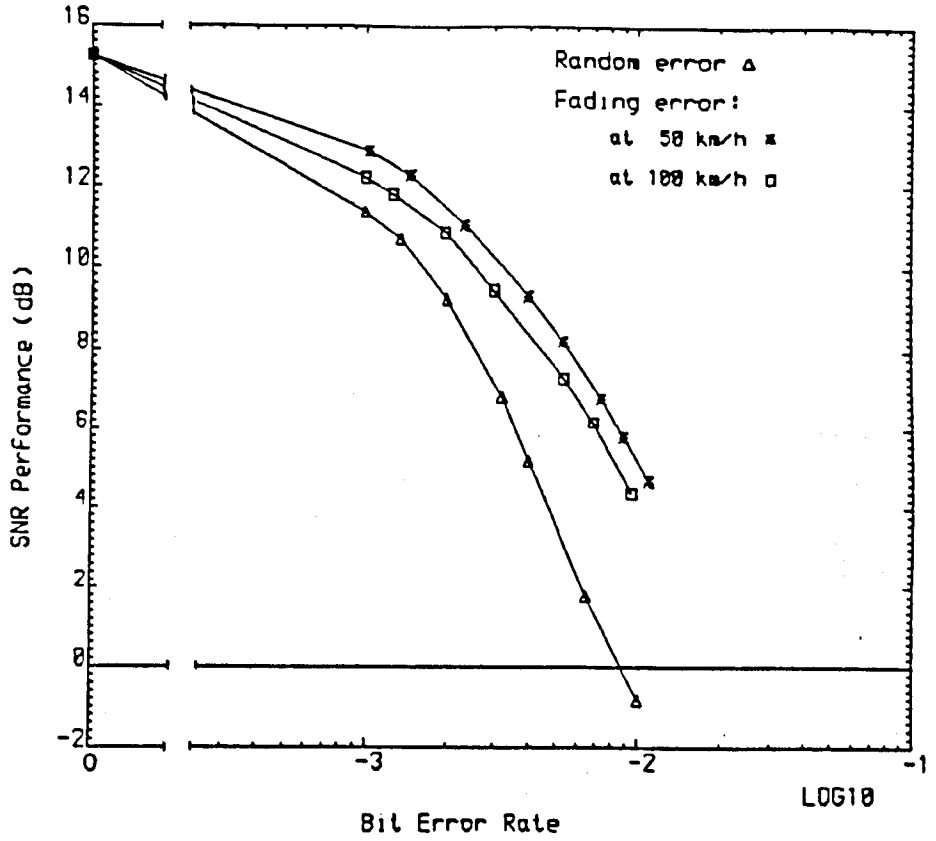


Figure 5.3 Illustration of the performance of the 16 kb/s APC-AB speech coder in presence of transmission errors.

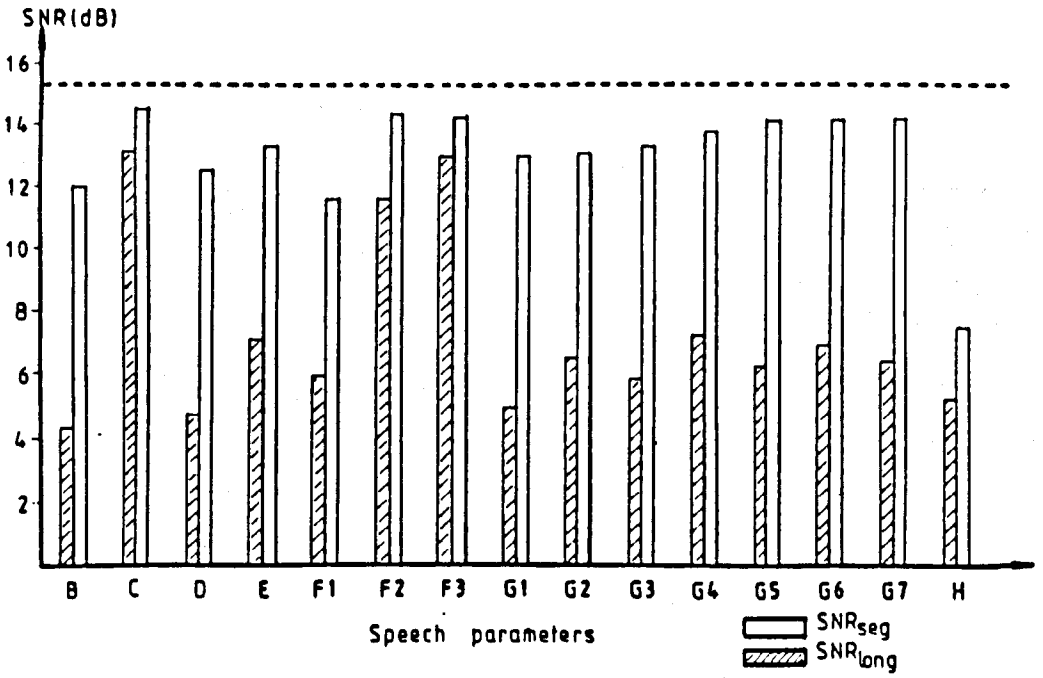
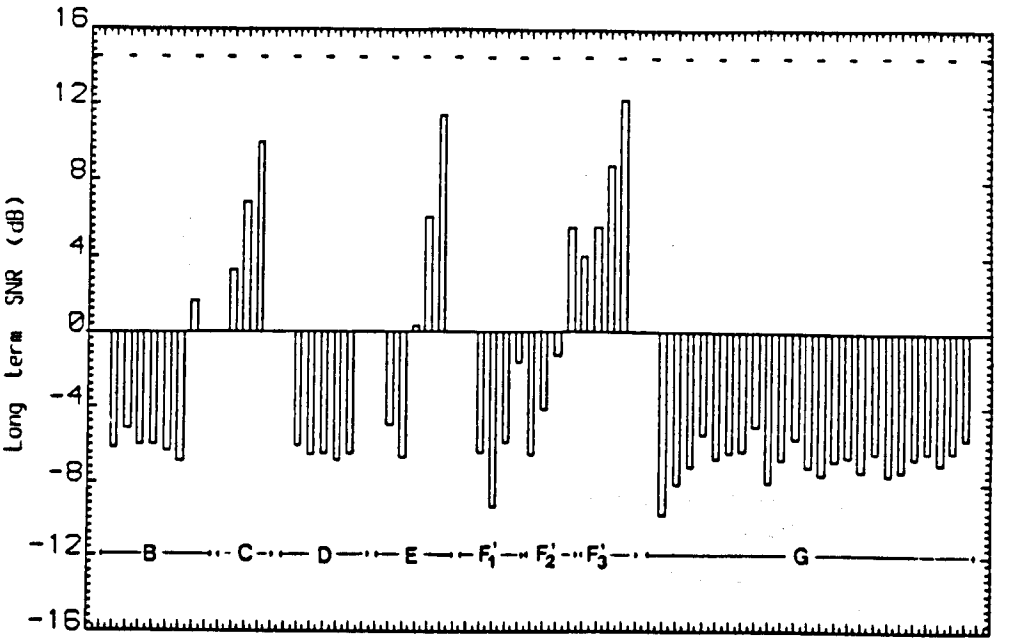
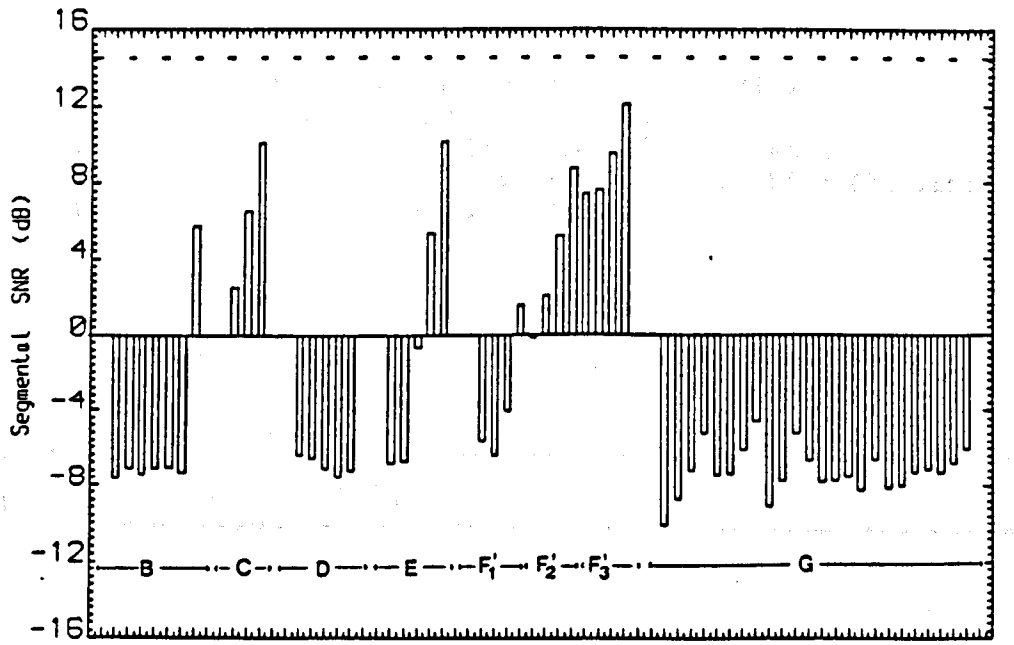


Figure 5.4 Illustration of distortion measurements on speech parameters at bit error rate 1:100.



a



b

Figure 5.5 SNR performance of the bit inversion tests.

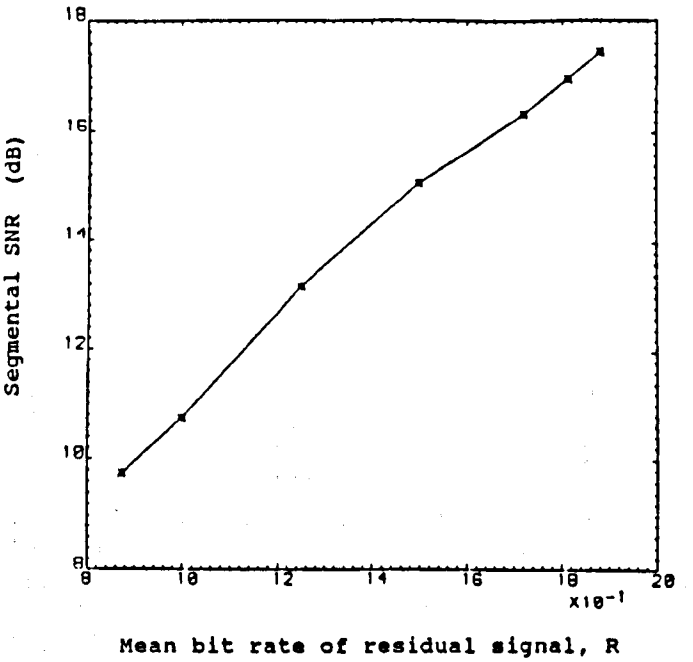


Figure 5.6 Rate-Distortion curve of APC-AB coder.

S :- synchronisation codeword (8 bits).
RS :- Reed Solomon code (60 bits).
H_1 & H_2 :- the two Hamming codes (16 bits each).
U :- 'unprotected' side information bits (22 bits).
R :- residual bits (198 bits).

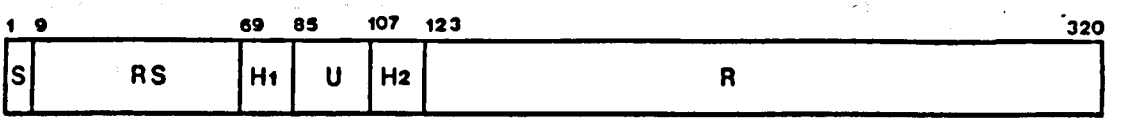


Figure 5.7 Format of a 'protected' APC-AB speech frame.

SYNC														
R1	R2	R3	R4	R5	R6	R7	R8	R9	R10	R11	R12	R13	R14	R15
69	70	71	72	73	74	75	76	77	78	79	80	81	82	83
84	85	86	87	88	89	90	91	92	93	94	95	96	97	98
99	100	101	102	103	104	105	106	107	108	109	110	111	112	113
114	115	116	117	118	119	120	121	122	123	124	125	126	127	128
129	130	131	132	133	134	135	136	137	138	139	140	141	142	143
144	145	146	147	148	149	150	151	152	153	154	155	156	157	158
159	160	161	162	163	164	165	166	167	168	169	170	171	172	173
174	175	176	177	178	179	180	181	182	183	184	185	186	187	188
189	190	191	192	193	194	195	196	197	198	199	200	201	202	203
204	205	206	207	208	209	210	211	212	213	214	215	216	217	218
219	220	221	222	223	224	225	226	227	228	229	230	231	232	233
234	235	236	237	238	239	240	241	242	243	244	245	246	247	248
249	250	251	252	253	254	255	256	257	258	259	260	261	262	263
264	265	266	267	268	269	270	271	272	273	274	275	276	277	278
279	280	281	282	283	284	285	285	286	287	288	289	290	291	292
294	295	296	297	298	299	300	301	302	303	304	305	306	307	308
309	310	311	312	313	314	315	316	317	318	319	320			

SYNC is the 8-bit frame synchronization codeword.
R1 - R15 are Reed-Solomon code symbols (4-bit per symbol).

Figure 5.8 Structure of an interleaver for APC-AB coder. Speech codes are read into the interleaver by rows, and read out by columns for transmission.

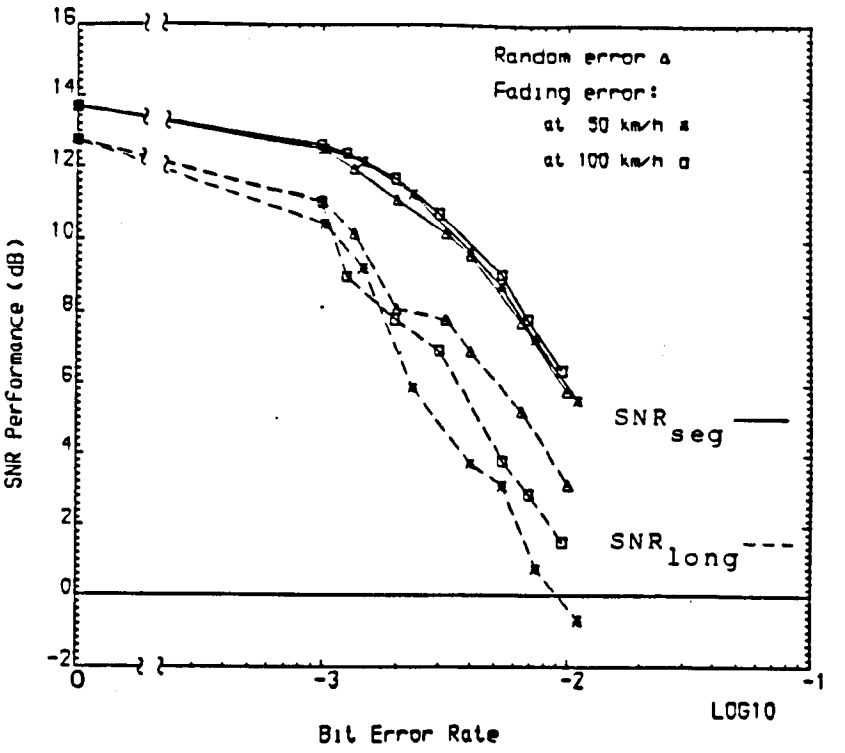


Figure 5.9 SNR performance of the 'protected' 16 kb/s APC-AB coder.

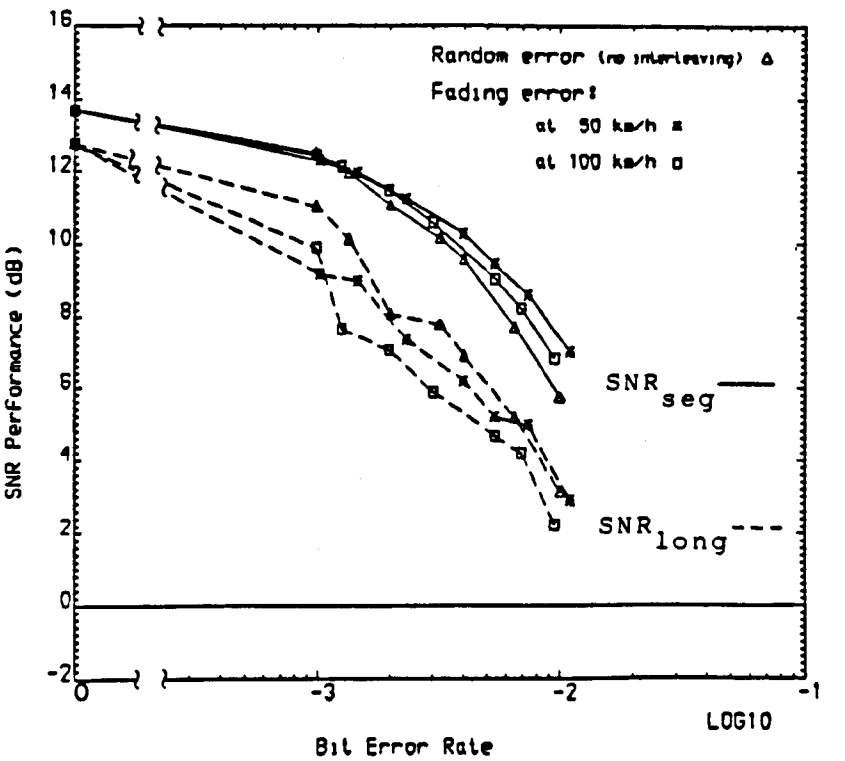


Figure 5.10 SNR performance of error correction scheme and interleaving technique on the 16 kb/s APC-AB coder.

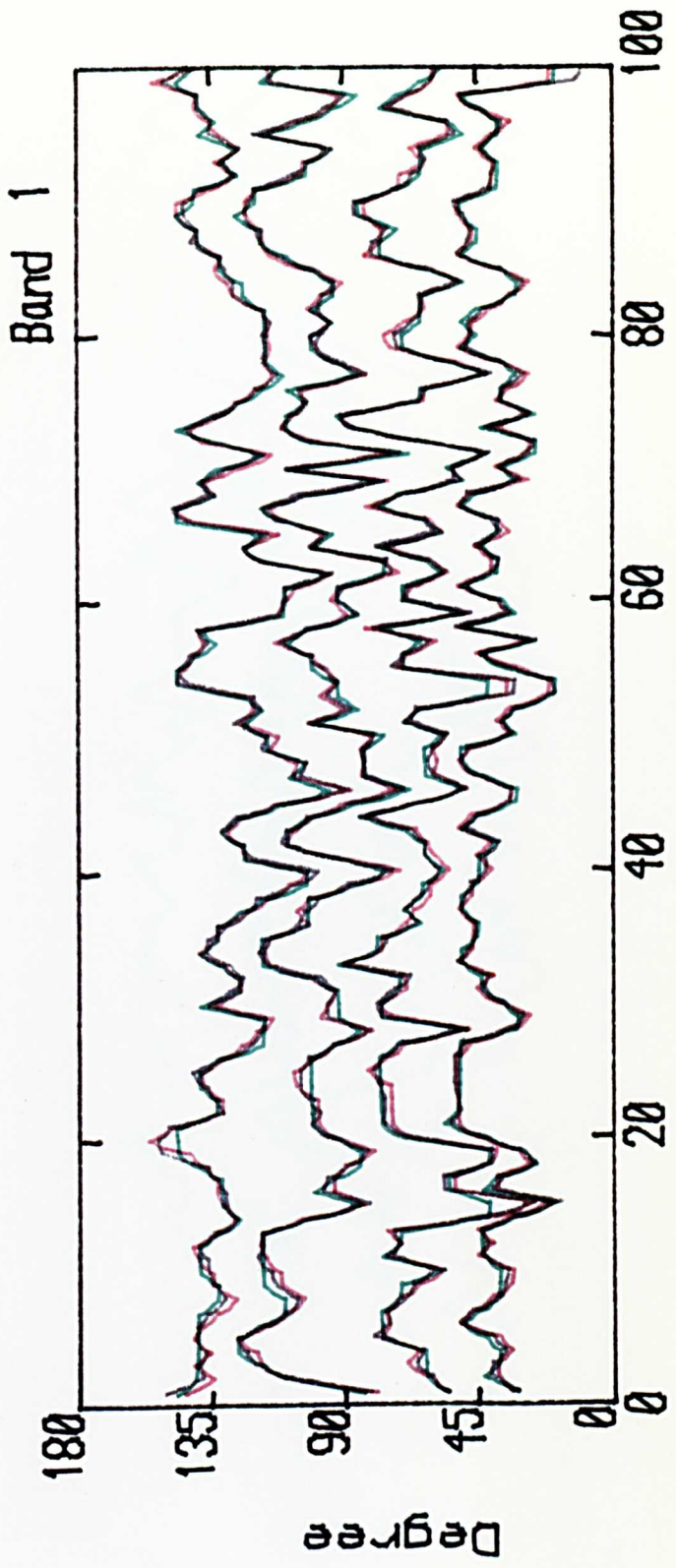


Figure 5.11a LSP quantization performance

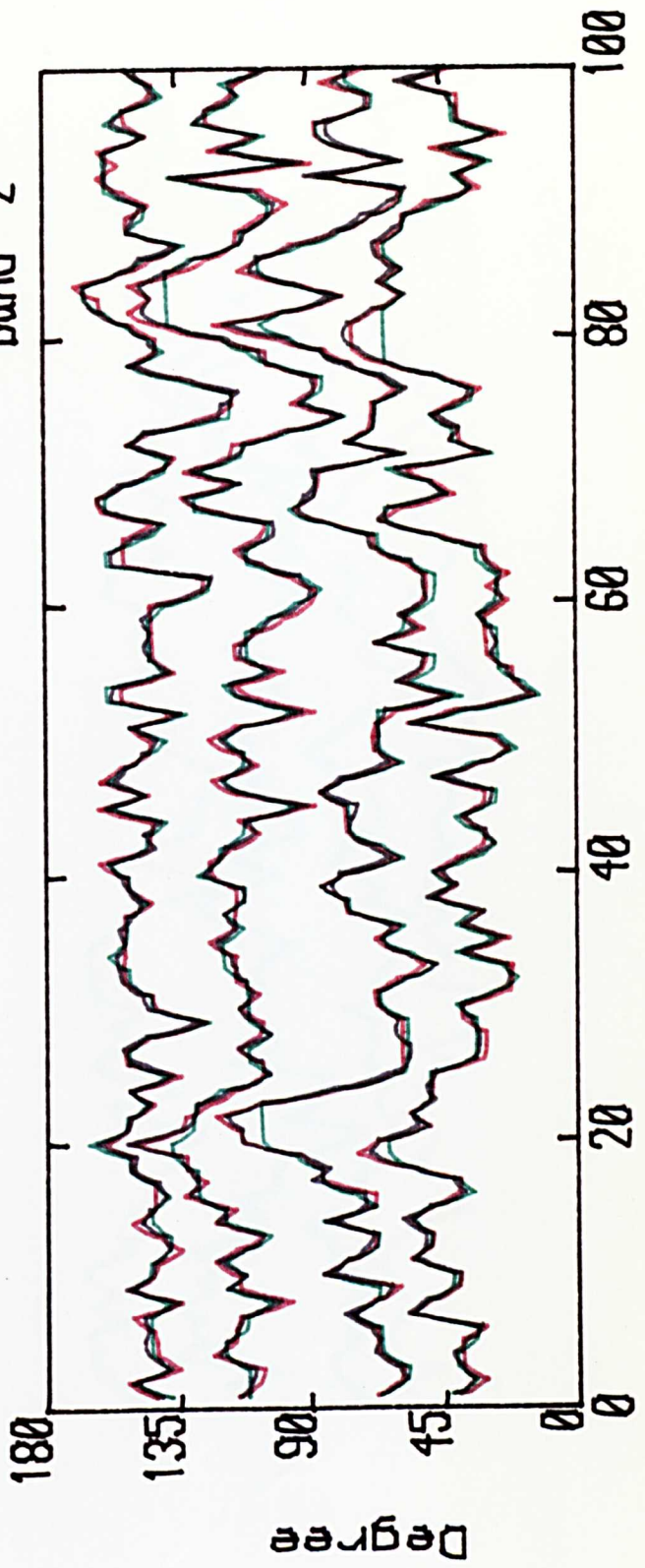


Figure 5.11b LSP quantization performance

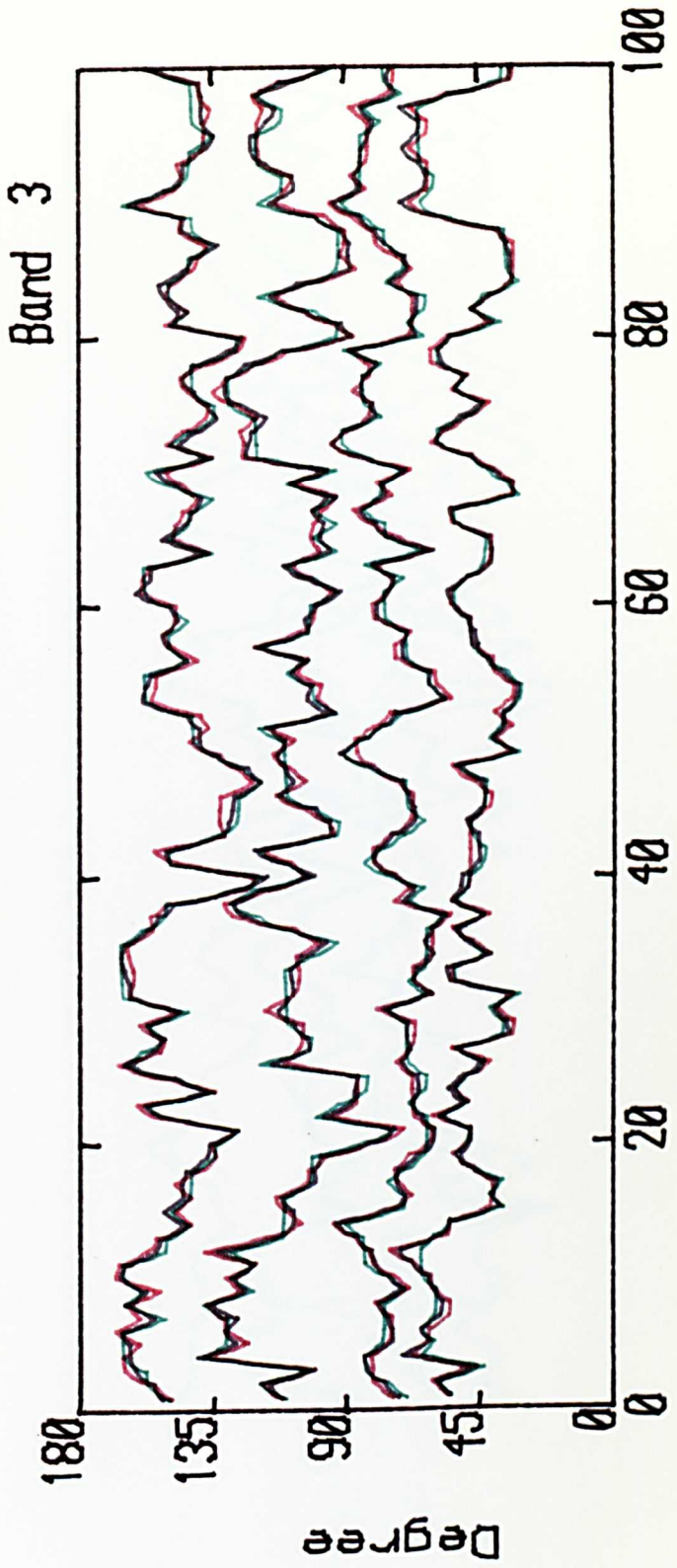


Figure 5.11c LSP quantization performance

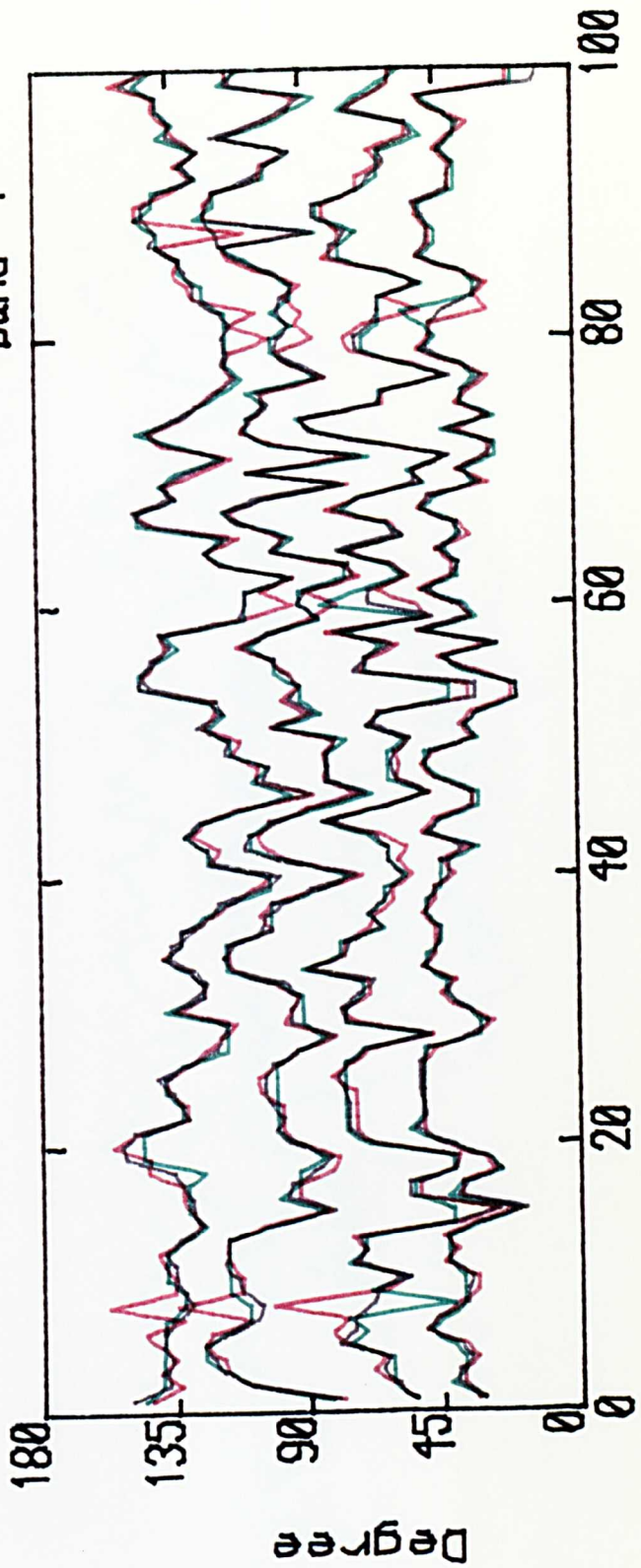


Figure 5.12a LSP quantization performance at a random bit error rate of 1 in 100.

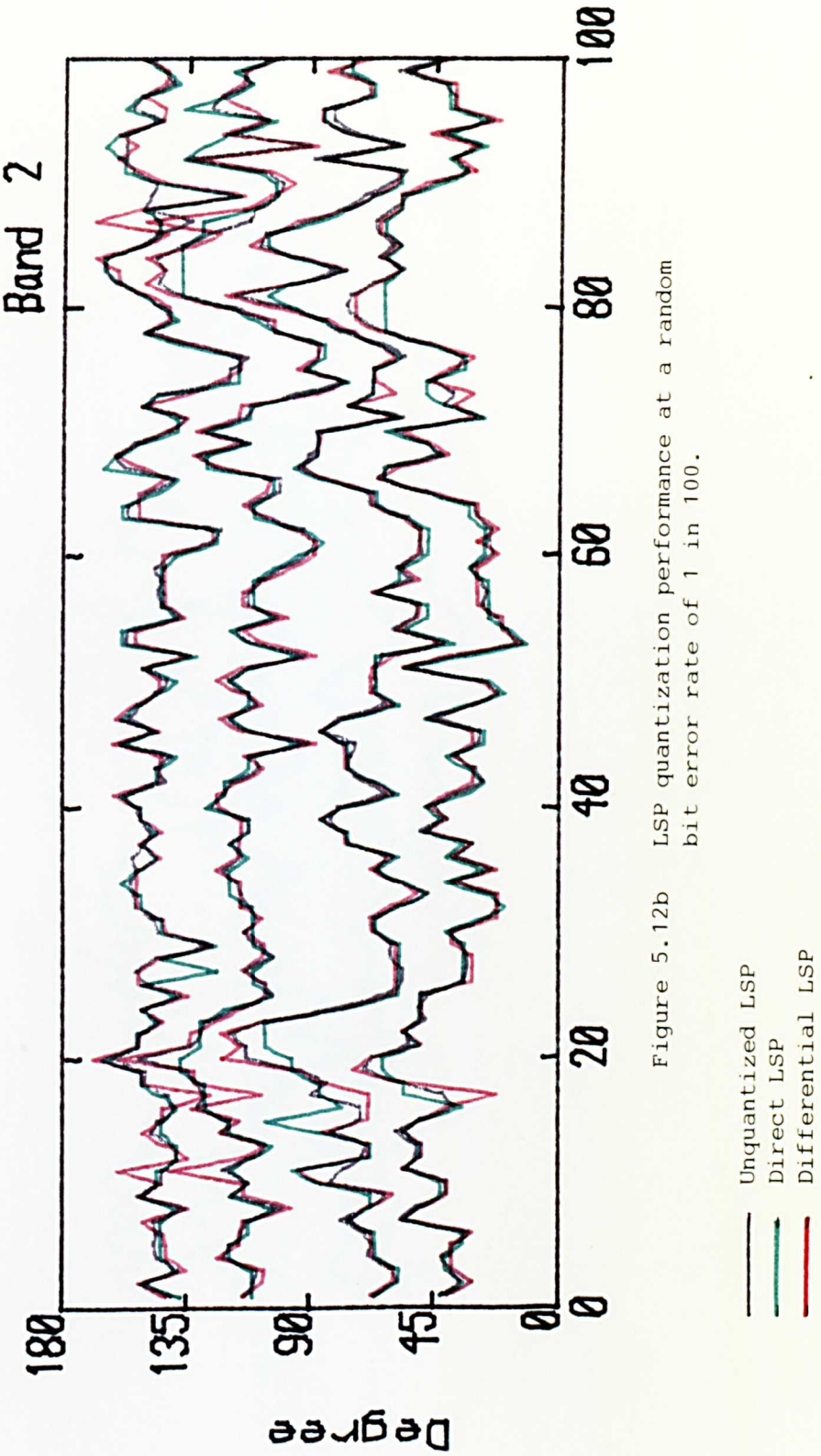


Figure 5.12b LSP quantization performance at a random bit error rate of 1 in 100.

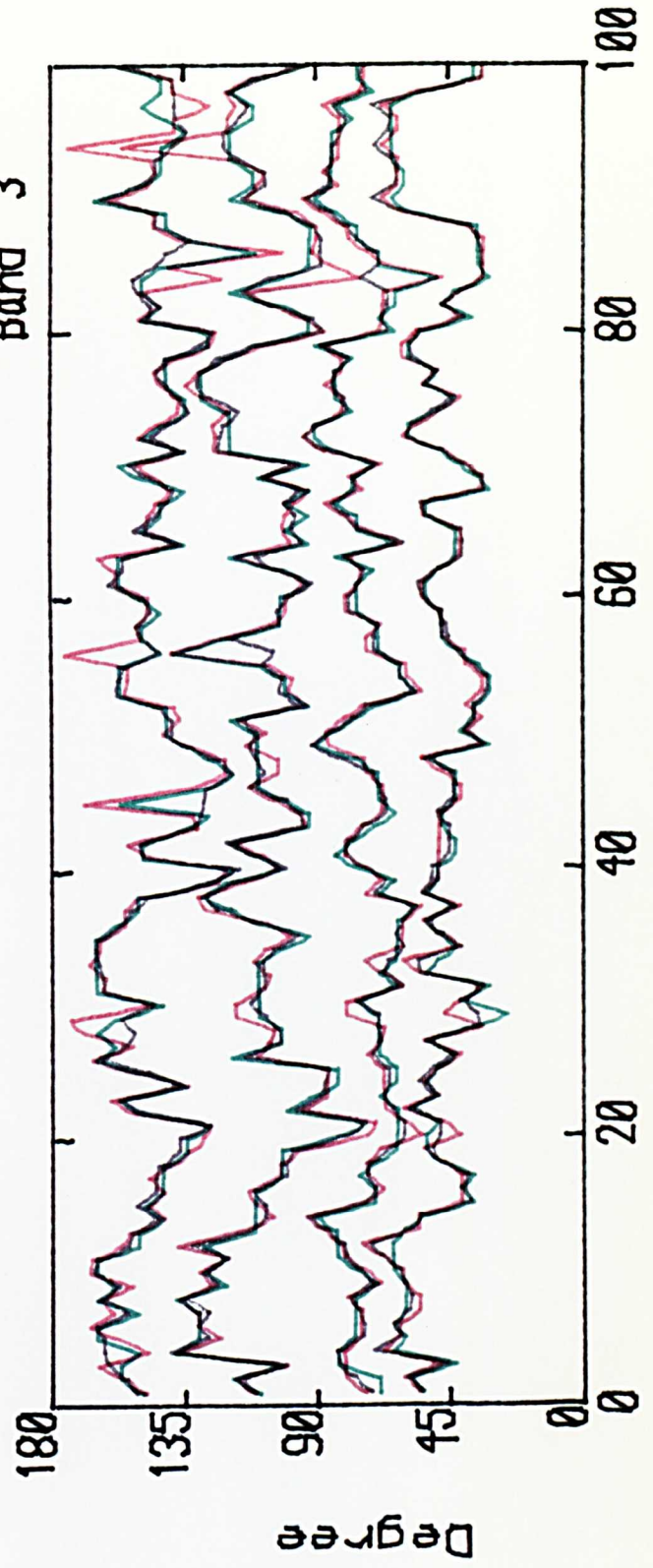


Figure 5.12c LSP quantization performance at a random bit error rate of 1 in 100.

mobile speed (km/hr)	mean level crossing rate (Hz)		mean fade duration (msec)	
	15 dB	30 dB	15 dB	30 dB
50	17.99	44.35	1.73	1.10
100	35.98	88.70	0.86	0.55

Table 5.1 Fade statistics for two mean CNR and mobile velocities at 900 MHz carrier frequency.

Figure 5.13.

CHAPTER 6 CONCLUSION AND FUTURE WORK

6.1 CONCLUSION

In this thesis a detailed study of APC-AB subband coding is presented: the aim of this work was to produce a robust 16 kb/s speech coder for mobile radio applications. The work involved mobile radio channel simulation, speech coding and error control coding analysis. Through the study of mobile radio channel characteristics, a LMR channel error simulator was developed based on a theoretical knowledge of the Rayleigh fading effect, the random FM effect and the effects of space diversity. The theory of the APC-AB coder, as published by Honda and Itakura [31], was presented, a new adaptive LSP filter algorithm was described and a 'zero-redundancy' error correction scheme was developed for the transmission of LSP coefficients.

Traditionally, speech coding and error control coding have been considered as related but essentially separate problems. When separately optimised speech coding and error control coding schemes are used together to achieve robust digitized speech transmission over a noisy channel, the combination is normally overdesigned or underdesigned in some aspects. An optimal error control coding scheme may use an interleaving algorithm which introduces an unnecessarily long delay and which may only slightly improve the decoded speech quality when the transmitted bit stream is subject to burst errors. A high redundancy error correction code may also result in poor speech coding performance at low error rates. An optimal speech coding scheme may involve parameters which are highly sensitive to channel errors especially as the object of speech coding is to remove as much redundancy from the speech as possible.

A common approach to error protection for speech coding parameters [33] is to assume that the highest sensitivity lies in the most significant bits of each speech parameter and therefore to protect these bits more heavily than less significant bits. This simple approach is not always optimal and a better allocation of error protection bits can often be devised depending on the coding algorithm. This was demonstrated by the result of the bit inversion test reported in Chapter 5. Figure 5.5 illustrates that all the bits of the pitch period parameters and sub-interval position parameters are equally important. This is because a bit error may corrupt the whole bit allocation calculation at the receiver.

The work presented in this thesis attempts to bridge the gap, to some extent, between speech coding and error control coding. It has been shown that simple channel parameters such as mean level crossing rate and mean fade duration statistics provide sufficient information for choosing appropriate error control schemes and interleaving depth. Instead of using high redundancy error correction coding schemes to combat high error rate channels, low redundancy error detection schemes with appropriate 'post-detection' was found to be efficient for high error rate LMR channels. As some error correction codes or some speech parameters may be sensitive to either random errors or burst errors, interleaving can be operated in mixed mode: bit interleaving and symbol interleaving can be combined to achieve a better interleaving efficiency. Coding techniques which are considered efficient for noise free channel transmission are not necessary well suited to noisy channels. Alternative speech coding techniques can often be proposed which are better suited. For example. the LSP difference encoding method is an efficient way of quantizing LSP coefficients, but the direct LSP quantization method has been found to

be more robust because a degree of error detection or even 'correction' is an inherent property of this technique.

This thesis has suggested various techniques for combining speech coding and error control coding. It is believed that with fine tuning and the introduction of appropriate 'post-detection' techniques, a more robust APC-AB coder can be expected. Furthermore, the design criterion can also be applied to other speech coding systems.

6.2 FUTURE WORK

As digital techniques will dominate future LMR telecommunication networks, there will be ever increasing interest in digitized speech transmission, especially for the second generation of digital mobile radio systems, for which the speech coding bit rate, including error correction coding should be around 11 kb/s. Much more research for lower bit rate and more robust coding schemes will certainly be necessary. Further to the work reported here, there is scope for deriving a general optimisation scheme for determining the optimal speech coding rate and error correction coding rate subject to a channel distortion minimization criterion.

As well as the improvements in speech quality obtained by using error control coding, the transmission of speech signals digitised by methods such as subband APC-AB coding, can be further improved by measuring the characteristics of the channel and using these measurements to decide how severely degraded the received data is likely to be. For example, channel field strength information can be used for deciding when to use interpolation of the residual signal during unreliable periods of

transmission. Interpolation can be performed between 'reliably' received segments, or an elegant but complex 'waveform substitution technique' can be used [83,84].

The work extended from the adaptive LSP filtering scheme [64] discussed in Chapter 4 opens a new area for adaptive filter algorithms. More research can be carried out in extending the work perhaps to higher order filtering systems. Improved methods for calculating adaptation coefficients for the adaptive LSP filter would be useful.

The Rayleigh fading channel simulator was employed to test the performance of the APC-AB coder. This channel simulation can, in fact, be further used to investigate the fading channel behaviour such as the statistics of burst error length and interburst length, or the relationship between other burst error channel models [85,86,87]. This knowledge will be useful in gaining a better understanding of the LMR channel characteristics, or in selecting suitable error correction schemes to protect the transmitted code.

APPENDIX

APPENDIX 1 QUADRATIC INTERPOLATION FOR PITCH DETECTION ALGORITHM

As the autocorrelation function (ACF) calculation in the pitch detection process requires a number of multiply and add operations, the pitch period is, therefore, estimated after the decimation process to reduce the computational cost. In this case, the computation is reduced approximately by the decimation factor, D, but unfortunately the time resolution of the estimated pitch period is also reduced by D. To achieve a better time resolution of the pitch period estimation from low time resolution, quadratic interpolation can be applied at points near the autocorrelation peak. This section describes a quadratic interpolation technique which requires only minimal computation to recover a better resolution of the pitch period estimation.

An example of 2 to 1 interpolation on a time scale is illustrated in Figure A.1.1. A more accurate estimation of the pitch period may be obtained by a quadratic polynomial approximating the autocorrelation function for values of t close to t_0,

$$f(t) = a.t^2 + b.t + c \qquad (A.1.1)$$

To solve for the polynomial coefficients, a, b and c, the un-interpolated pitch peak estimate at t_0 and adjacent terms at t_{-2} and t_2 from the ACF calculation are used to form the following set of equations.

$$\begin{aligned} a.t_{-2}^2 + b.t_{-2} + c &= f(t_{-2}) \\ a.t_0^2 + b.t_0 + c &= f(t_0) \\ a.t_2^2 + b.t_2 + c &= f(t_2) \end{aligned} \qquad (A.1.2)$$

Since the original signal is decimated by a factor of 2, the corresponding re-scaled pitch period at t_0 is equal to $2.I_p'$, where I_p' is the pitch period estimated from the decimated signal. Hence, the re-scaled times at t_{-2} and t_2 are $2.I_p'-2$ and $2.I_p'+2$ respectively. They can be written as:

$$t_{-2} = 2.I_p' - 2 \qquad (A.1.3)$$
$$t_0 = 2.I_p'$$
$$t_2 = 2.I_p' + 2.$$

By using the relations in (A.1.3), (A.1.2) can be solved as:

$$a = f(2.I_p'-2)/8 - f(2.I_p')/4 + f(2.I_p'+2)/8 \qquad (A.1.4)$$
$$b = -(2.I_p'+1).f(2.I_p'-2)/4 - I_p'f(2.I_p') - (2.I_p'-1).f(2.I_p'+2)/4$$
$$c = I_p'(I_p'+1).f(2.I_p'-2)/2 - (I_p'^2-1).f(2.I_p') + I_p'(I_p'-1).f(2.I_p'+2)/2$$

To find the value of t when f(t) reaches maximum, f(t) is differentiated with respect to t and the resulting expression is set to zero. Hence, the differential coefficient of f(t) is

$$(t-2.I_p'-1).f(2.I_p'-2)/4 + (I_p'-t/2).f(2.I_p') + (t-2.I_p'+1)/4 = 0 \qquad (A.1.5)$$

Rearranging (A.1.5), the maximum point of f(t) is

$$t_{pitch} = 2.I_p' + \frac{f(2.I_p'-2) - f(2.I_p'+2)}{f(2.I_p'-2) - 2.f(2.I_p') + f(2.I_p'+2)} \qquad (A.1.6)$$
$$= 2.I_p' + \Delta$$

where Δ is a pitch period correction factor. Hence, (A.1.6) gives a better estimation for the pitch period than is obtained without interpolation. As an integer value of pitch period is required, the corrected pitch period, I_p is therefore written as

$$I_p = 2.I_p' + 1 \quad \text{if } \Delta \geq 0.5 \qquad (A.1.7)$$
$$I_p = 2.I_p' - 1 \quad \text{if } \Delta \leq 0.5$$
$$I_p = 2.I_p' \quad \text{otherwise.}$$

The division in calculating Δ can be avoided by putting

$$\Delta = D_d / N_d$$

Then,

$$2.D_d \geq N_d \quad \text{for} \quad \Delta \geq 0.5, \text{ and}$$
$$2.D_d \leq -N_d \quad \text{for} \quad \Delta \leq -0.5.$$

Hence, a more accurate pitch period estimation can be written as in (3.27).

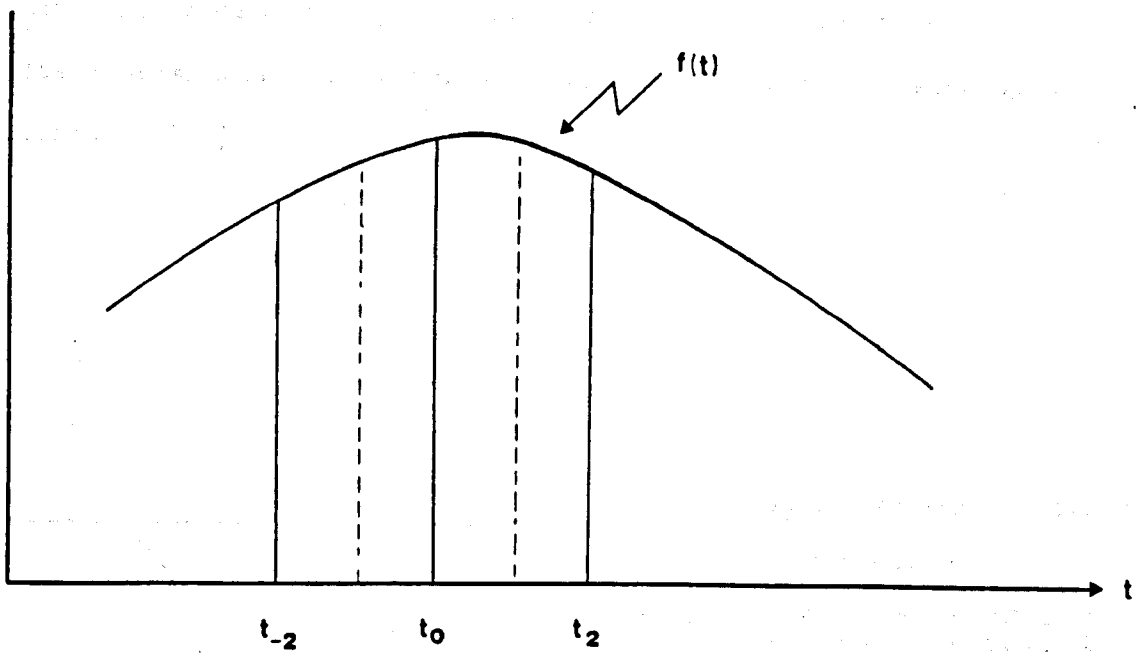
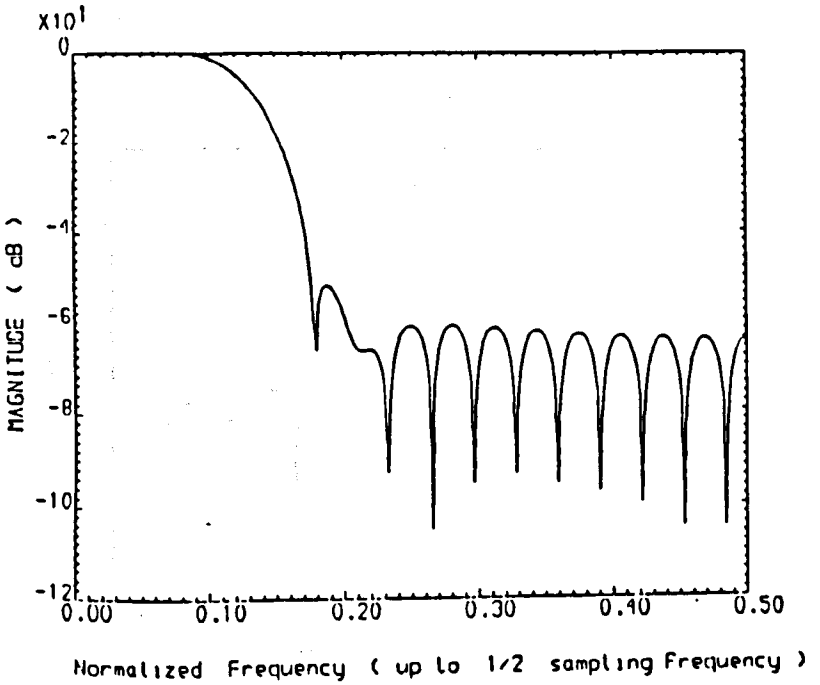


Figure A.1.1 Illustration showing 2 to 1 interpolation about the un-interpolated pitch peak estimate, t_0, and adjacent terms t_{-2} and t_2 of the ACF function.

APPENDIX 2 FILTERS OF PHASE-SHIFTERS FOR APC-AB CODER

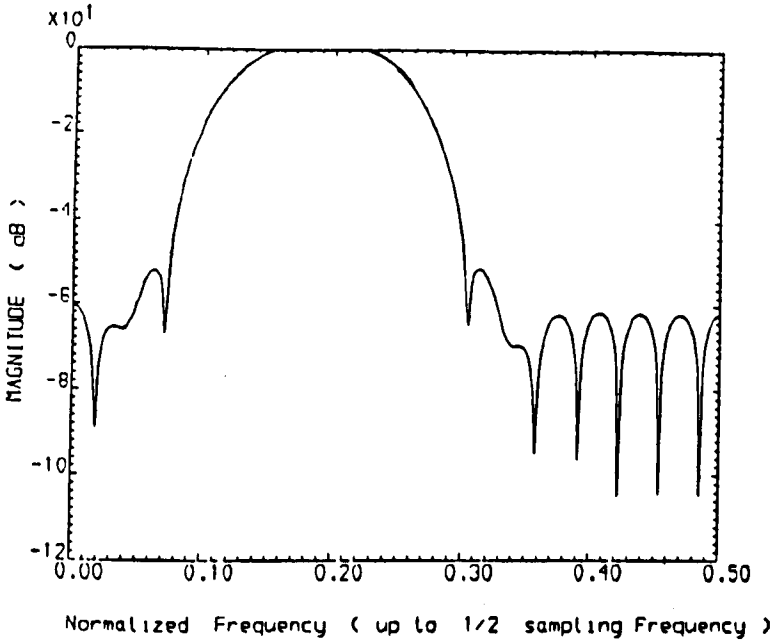
Referring to Section 3.3.3.1, the non-integer delay line of each sub-band pitch predictor can be implemented by a phase shifting circuit [45]. Due to different interpolation and decimation processes for the phase shifters in different sub-bands, different phase shifter filters are required. The following presents the filter coefficients and their equivalent frequency responses used in the APC-AB coder simulation. These filters were designed by the Windowing method (Hamming window) as described in [82].



Band 1 phase-shifter filter

h(0) = h(32) = 0.0
h(1) = h(31) = -0.001330
h(2) = h(30) = -0.002610
h(3) = h(29) = -0.002722
h(4) = h(28) = 0.0
h(5) = h(27) = 0.005809
h(6) = h(26) = 0.011563
h(7) = h(25) = 0.011238
h(8) = h(24) = 0.0
h(9) = h(23) = -0.020209
h(10) = h(22) = -0.037912
h(11) = h(21) = -0.035743
h(12) = h(20) = 0.0
h(13) = h(19) = 0.069074
h(14) = h(18) = 0.153280
h(15) = h(17) = 0.222651
h(16) = 0.249509

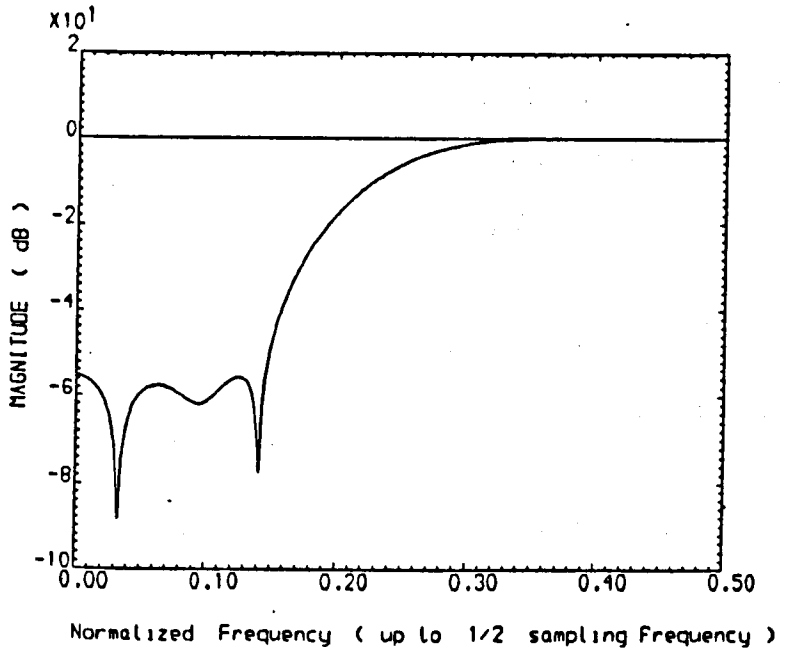
Figure A.2.1.a



Band 3 phase-shifter filter

h(0) = h(32) = 0.0
h(1) = h(31) = -0.000550
h(2) = h(30) = 0.002604
h(3) = h(29) = 0.006556
h(4) = h(28) = 0.0
h(5) = h(27) = -0.013990
h(6) = h(26) = -0.011535
h(7) = h(25) = 0.0046440
h(8) = h(24) = 0.0
h(9) = h(23) = -0.008351
h(10) = h(22) = 0.037822
h(11) = h(21) = 0.086085
h(12) = h(20) = 0.0
h(13) = h(19) = -0.166363
h(14) = h(18) = -0.152916
h(15) = h(17) = 0.92006
h(16) = 0.248916

Figure A.2.1.b



Band 2 phase-shifter filter

h(0) = h(16) = 0.0
h(1) = h(15) = 0.005221
h(2) = h(14) = 0.0
h(3) = h(13) = -0.023132
h(4) = h(12) = 0.0
h(5) = h(11) = 0.075847
h(6) = h(10) = 0.0
h(7) = h(9) = -0.30665
h(8) = 0.499168

Figure A.2.1.c

REFERENCES

1. Webster, E.M., "Utilization and Expansion of Vehicular Radio Communications," IEEE Trans. Veh. Comm. 1, February, 1952, p.66.

2. Lomer, G.J., "Telephoning on the move-Dick Tracy to Captain Kirk," IEE Proc., Vol. 134, Pt. F, No. 1, February 1987, pp.1-8.

3. MacDonald, V.H., "Advanced Mobile Phone Service: The Cellular Concept," B.S.T.J., Vol. 58, No. 1, January 1979, pp.15-41.

4. Matthews, P.A., "Communications on the move," Electronics & Power, July 1984, pp.513-518.

5. Gans, M.J., "A Power-Spectral Theory of Propagation in the Mobile-Radio Environment," IEEE Trans. VT-29, February 1972, pp.27-38.

6. Cox, D.C., & Leck, R.P., "Correlation Bandwidth and Delay Spread Multipath Propagation Statistics for 910-MHz Urban Mobile Radio Channels," IEEE Trans. COM-23, November 1975, pp.1271-1280.

7. Bajwa, A.S. & Parsons, J.D., "Large area characterisation of urban UHF multipath propagation and its relevance to the peformance bounds of mobile radio systems," IEE Proc., Vol. 132, Pt. F, No. 2, April 1985, pp.99-106.

8. Jakes, W.C., "Microwave Mobile Communications," John Wiley & Sons, 1984.

9. Proakis, J.G., "Digital Communication," McGraw-Hill, 1983.

10. Lee, W.C.Y., "Mobile Communications Engineering," McGraw-Hill, 1982.

11. Jayant, N.S., "Coding Speech at Low Bit Rates," IEEE spectrum, August 1986, pp.58-63.

12. Haskell, B.G., & Steele, R., "Audio and Video Bit-Rate Reduction," Proc. of IEEE, Vol. 69, No. 2, February 1981, pp.252-262.

13. Sluyter, R.J., "Digitization of Speech," Philips Tech. Rev. 41, No. 7/8, pp.201-223.

14. French, R.C., "Error Performance in Mobile Radio Data Transmission in the Urban Environment," Nachrichten Technische Zeitschrift, Vol. 31, March 1978, pp.200-203.

15. Dempsey, H.E., "Digital Message Reliabilty and Integrity in Land Mobile Communication Systems," IEEE 1979 International Electrical, Electronics Conference and Exposition, Toronto, 2-4 October 1979.

16. Barnett, W.T., "Multipath Fading Effects on Digital Radio," IEEE Trans. COM-27, No. 12, December 1979, pp.1842-1848.

17. Arredondo, G.A., Chriss, W.H. & Walker, E.H., "A Multipath Fading Simulator for Mobile Radio," IEEE Trans. VT-22, No. 4, November 1973, pp.241-244.

18. Comroe, R., "An All Digital Rayleigh Fading Simulator," Proceedings of the NEC, Vol. 32, October 1978, pp.136-139.

19. Bello, P.A., "Characterization of Randomly Time-Variant linear channels," IEEE Trans. CS-11, 1963, pp.360-393.

20. Turin, G.L., Clapp, F.D., Johnston, T.L., Fine, S.B. & Lavry, D., "A Statistical Model of Urban Multipath Propagation," IEEE Trans. VT-21, No. 1, 1972, pp.1-9.

21. Papoulis, A., "Probability, Random Variables, and Stochastic Processes," McGraw-Hill, New York, 1965.

22. Schwartz, M., Bennett, W.R. & Stein, S., "Communication Systems and Techniques," McGraw-Hill, New York, 1966.

23. Dorf, R.C., "Modern Control Systems," Addison-Wesley, 1980.

24. Stremler, F.G., "Introduction to Communication Systems," Addison-Wesley, 1982.

25. Pierce, J.N., "Theoretical Diversity Improvement in Frequency Shift Keying," Proc. IRE, Vol. 46, May 1958, pp.903-910.

26. Parsons, J.D., Henze, M., Ratcliff, P.A. & Withers, M.J., "Diversity Techniques for Mobile Radio Reception," IEEE Trans. VT-25, August 1976, pp.75-84.

27. Crochiere, R.E., Webber, S.A. & Flanagan, J.L., "Digital Coding of Speech in Sub-bands," B.S.T.J., Vol. 55, No. 8, October 1976, pp.1069-1085.

28. Atal, B.S. & Remde, J.R., "Split-Band APC System for Low Rate Encoding of Speech," Proc. ICASSP, April 1981, pp.599-602.

29. Galand, C.R. & Esteban, D.J., "Multirate Sub-Band Coder with Embedded Bit Stream," Proc. ICASSP, 1983, pp.1284-1287.

30. Heron, C.D., Crochiere, R.E. & Cox, R.V., "A 32-Band Sub-band/Transform Coder Incorporating Vector Quantization for Dynamic Bit Allocation," Proc. ICASSP, 1983, pp.1276-1279.

31. Honda, M. & Itakura, F., "Bit Allocation in Time and Frequency Domains for Predictive Coding of Speech," IEEE Trans. ASSP-32, No. 3, June 1984, pp.465-473.

32. Esteban, D. & Galand, C., "Application of Quadrature Mirror Filters to Split Band Voice Coding Schemes," Proc. ICASSP, May 1977, pp.191-195.

33. Jayant, N.S. & Noll, P., "Digital Coding of Waveforms," Prentice-Hall, 1984.

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

35. Berger, T., "Rate Distortion Theory," Englewood Cliffs, NJ: Prentice-Hall, 1971.

36. Habibi, A. & Hershel, R.S., "A Unified Respresentation of Differential Pulse-Code Modulation and Transform Coding Systems," IEEE Trans. COM-22, May 1974, pp. 692-696.

37. Rabiner, L.R., Cheng, M.J., Rosenberg, A.E. & McGonegal, C.A., "A Comparative Performance Study of Several Pitch Detection Algorithms," IEEE Trans. ASSP-24, No. 5, October 1976, pp. 399-418.

38. Witten, I.H., "Principles of Computer Speech," Academic Press, London, 1982.

39. Bristow, G.J. & Fallside, F., "An Autocorrelation Pitch Detector with Error Correction," Proc. ICASSP, 1982, pp. 184-187.

40. Rabiner, L.R., Sambur, M.R. & Schmidt, C.E., "Applications of a Nonlinear Smoothing Algorithm to Speech Processing," IEEE Trans. ASSP-23, No. 6, December 1975, pp. 552-557.

41. Sondhi, M.M., "New Methods of Pitch Extraction," IEEE Trans. AU-16, June 1968, pp. 262-266.

42. Dubnowski, J.J., Schafer, R.W. & Rabiner, L.R., Sambur, "Real-Time Digital Hardware Pitch Detector," IEEE Trans. ASSP-24, No. 1, February 1975, pp. 2-8.

43. Rabiner, L.R., "On the Use of Autocorrelation Analysis for Pitch Detector," IEEE Trans. ASSP-25, No. 1, February 1977, pp. 24-33.

44. Markel, J.D. & Gray, A.H., "Linear Prediction of Speech," Springer-Verlag, New York, 1976.

45. Crochiere, R.E., Rabiner, L.R. & Shively, R.R, " A Novel implementation of Digital Phase Shifters," BSTJ, Vol. 54, No. 8, October 1975.

46. Gray, A.H., Gray, R.M. & Markel, J.D., "Comparison of Optimal Quantization of Speech Reflection Coefficients," IEEE Trans. on ASSP-25, No. 1, February 1977, pp. 9-23.

47. Tohkura, Y. & Itakura, F., "Spectrum Sensitivity Analysis of PARCOR Parameters for Speech Data Compression," IEEE Trans. on ASSP-27, No. 3, June 1979, pp. 273-280.

48. Gray, A.H. & Markel, J.D., "Quantization and Bit Allocation in Speech Processing," IEEE Trans. on ASSP-24, No. 6, December 1976, pp. 459-473.

49. Soong, F.K. & Juang, B.H, "Line Spectrum Pair (LSP) and Speech Data Compression," IEEE ICASSP 84, 1984, pp. 1.10.1-1.10.4.

50. Jayant, N.S., "Digital Coding of Speech Waveforms: PCM, DPCM, and DM Quantizers," Proc. IEEE, Vol. 62, May 1974, pp.611-632.

51. Honda, M., Kitawaki, N. & Itakura, F., "Real Time Realization of APC-AB Codec on Digital Signal Processor," Trans. of the Committee on Speech Research, The Acoustical Society of Japan, S82-44, October, 1982, pp. 345-352, (in Japanese).

52. Campanella, S.J. & Robinson, G.S., "A Comparison of Orthogonal Transformations for Digital Speech Processing," IEEE Trans. Communications, COM-19, 1971, pp. 1045-1050.

53. Kitawaki, N., Honda, M. & Itoh, K., "Speech Quality Assessment Methods for Speech Coding Systems," IEEE Comm. Magazine, Vol. 22, No. 10, 10/1984, pp. 26-33.

54. Sugamura, S. & Itakura, F., "Speech Data Compression by LSP Speech Analysis-Synthesis Technique," Trans. IECE, J64-A, 8, Aug. 1981.

55. Itakura, F. & Saito, S., "Speech Information Compression based on the Maximum Likelihood Spectrum Estimation," Journal of Acoust. Soc., (in Japanese), 27-463, 1971.

56. Wakita, H., "Linear Prediction Voice Synthesizers: Line Spectrum pairs is the newest of several techniques," Speech Technology, Fall 1981, pp. 17-22.

57. Itakura, F., Kobayashi, T. & Honda, Masaaki, "A Hardware Implementation of a New Narrow to Medium Band Speech Coding," IEEE ICASSP 1982, pp. 1964-1967.

58. Widrow, B. & Hoff, M.E., "Adaptive Switching Circuits," IRE 1960 WESCON Conv. Rec., Part 4, pp. 96-104.

59. Sugamura, N. & Itakura, F., "Speech Analysis and Synthesis Methods developed at ECL in NTT -from LPC to LSP-," Speech Communication 5 (1986), pp. 199-215.

60. Tobayashi, T. & Itakura, F., "An Algorithm to Extract Line Spectrum Pair Parameter (in Japanese)," Nippon Acoustic Conv., pp. 599-600.

61. Mantila-Montaivo, R., "On the Computation of LSP Parameters," IEEE Sixteenth Asilomar Conference on Circuit and Computers, 1983, pp. 481-484.

62. Kabal, P. & Ramachandran, R.P., "The Computatin of Line Spectral Frequencies Using Chebyshev Polynomials," IEEE Trans. on ASSP-34, No. 6, December 1986, pp. 1419-1425.

63. Freij, G.J., "Enhanced Sequential Adaptive Linear Prediction for Speech Encoding," Ph.D. thesis, 1985, The University of Liverpool, England.

64. Cheetham, B.M.G., "Adaptive LSP Filter," Electronics Letter, Vol. 23, No. 2, 1987, pp. 89-90.

65. Daniell, T.P. & Brown, J.E., "Adaptation in Non-stationary Applications," Proc. 1970 IEEE Symp. Adaptive Processes (9th), Austin TX., December 1970.

66. McCool, J.M., et al, "Stationary and Nonstationary Learning Characteristics of the LMS Adaptive Filter," Proc. IEEE, Vol. 64, No. 8, August 1976.

67. McCool, J.M. & Widrow, B., "Principles and Applications of Adaptive Filters: A Tutorial Review," IEE Conf. The Impact of New Technology in Signal Processing, 1976.

68. Abey, P.R. & M.A.H. Dempster, "An Introduction to Optimazation Theory," Chapman.

69. Haykin, S., "Introduction to Adaptive Filters," Macmillian Inc., 1984.

70. Davis, A., "Waveform Coding of Speech and Voiceband Date Signals," Ph.D. thesis, The University of Liverpool, 1988.

71. Wong, W.T.K. & Cheetham, B.M.G., "Adaptive LSP filter for Sub-band Speech Coding," IEE Colliqium on DSP, Jan. 1987.

72. Norway: Some Consideration on medium bit rate speech coding for digital mobile radio, CCITT Study Group XII, Contribution 79.

73. Scagliola, C, "Evaluation of Adaptive Speech Coders under Noisy Channel Conditions," BSTJ, Vol. 58, 1979, pp. 1369-1394.

74. Sampei, S & Kamio, Y., "Performance of FEC with Interleaving in Digital Land Mobile Communications," Trans. of IECE of Japan, Vol. E-68, No. 10, 10/1985, pp. 651-652.

75. Jayant, N.S., Schafer, R.W. & Karim, M.R., "Step-size Transmitting Differential Coder for Mobile Telephony," Conf. Rec., 1975 IEEE Int. Conf. on Commun., Vol. 11, 6/1975, pp. 3006-3010.

76. Goodman, D.J. & Sundberg, C.E., "Combined Source and Channel Coding for Matching the Speech Transmission Rate to the Quality of the Channel," Rec. IEEE GLOBAL Telecommun. Conf., Nov/1982, pp. 316-321.

77. Balhut, R.E., "Theory and Practice of Error Control Codes," Addison-Wesley, 1984.

78. Williard, M.W., " Introduction of Redundancy Coding," IEEE Trans. on Vehicular Techn., Vol. VT-27, No. 3, 8/1978, pp. 86-98.

79. Hamming, R.W., "Coding and Information Theory," Prentice-Hall, 1980.

80. Wiggert, D., "Error-Control Coding and Applications," Artech House, 1978.

81. Cheetham, B.M.G. & Hughes, P.M.,"Formant Estimation from LSP Coefficients," IERE, Fifth International Conference on Digital Processing of Signals in Communications, No. 82, Sept., 1988, pp. 183-190.

82. Bozic, S.M., "Digital and Kalman Filtering," Arnold, 1979.

83. Wong, W.C., Steel, R., Glance, B. and Horn, D., "Time Diversity with Adaptive Error Detection to Combat Rayleigh Fading in Digital Mobile Radio," IEEE COM-31, No. 3, March 1983, pp 378-387.

84. Goodman, D.J., Lockhart, G.B., Wasem, O.J. and Wong, W.C., "Waveform Substitution Techniques for Recovering Missing Speech Segments in Packet Voice Communication," IEEE ASSP-34, No. 6, December 1986, pp 1440-1447.

85. Gilbert, E.N., "Capacity of a Burst Noise Channel," BSTJ, September 1960, pp 1253-1265.

86. Elliot, E.O., "A Model of the Switched Telephone Network for Data Communications," BSTJ, January 1963, pp 80-109.

87. Berger, J.M. and Mandelbrot, B., "A New Model for Error Clustering in Telephone Circuits," IBM Journal, July 1963, pp 224-235.