UNDERWATER ACOUSTIC VOICE COMMUNICATIONS USING DIGITAL TECHNIQUES

by

Hayri Sari, BSc, MSc

A doctoral thesis submitted in partial fulfillment of the requirements for the award of Doctor of Philosophy of the Loughborough University

April 1997

Supervisor: **Professor Bryan Woodward**, PhD, DIC, CEng, FIEE, FIOA, FRGS

Department of Electronic and Electrical Engineering

To my parents to whom I dedicate this thesis

Abstract

An underwater acoustic voice communications system can provide a vital communication link between divers and surface supervisors. There are numerous situations in which a communication system is essential. In the event of an emergency, a diver's life may depend on fast and effective action at the surface. The design and implementation of a digital underwater acoustic voice communication system using a digital signal processor (DSP) is described. The use of a DSP enables the adoption of computationally complex speech signal processing algorithms and the transmission and reception of digital data through an underwater acoustic channel. The system is capable of operating in both transmitting and receiving modes by using a mode selection scheme. During the transmission mode, by using linear predictive coding (LPC), the speech signal is compressed whilst transmitting the compressed data in digital pulse position modulation (DPPM) format at a transmission rate of 2400 bps. At the receiver, a maximum energy detection technique is employed to identify the pulse position, enabling correct data decoding which in turn allows the speech signal to be reconstructed.

The advantage of the system is to introduce advances in digital technology to underwater acoustic voice communications and update the present analogue systems employing AM and SSB modulation. Since the DSP-based system is designed in modular sections, the hardware and software can be modified if the performance of the system is inadequate. The communication system was tested successfully in a large indoor tank to simulate the effect of a short and very shallow underwater channel with severe multipath reverberation.

The other objective of this study was to improve the quality of the transmitted speech signal. When the system is used by SCUBA divers, the speech signal is produced in a mask with a high pressure air environment, and bubble and breathing noise affect the speech clarity. Breathing noise is cancelled by implementing a combination of zero crossing rate and energy detection. In order to cancel bubble noise spectral subtraction and adaptive noise cancelling algorithms were simulated; the latter was found to be superior and was adopted for the current system.

ACKNOWLEDGMENTS

I would like to express my gratitude to my supervisor Professor Bryan Woodward for his motivation and guidance, and for always being available as my diving test subject for speech recordings. I also offer my appreciation for his encouragement while I was learning to dive.

I am grateful to the Turkish Government for providing a scholarship; without its support, I would not have been able to study in the UK.

I would like to thank Professor Zheng Zhaoning (visiting scholar) of Southeast University, Nanjing, China for his valuable comments throughout the first year of this study. I would like to thank to Dr. Rambod Naimimohasses for sharing his valuable experience in writing this thesis, Dr. Robert Istepanian of the University of Portsmouth, Dr. Baochun Hou of the University of Hertfordshire and especially to Dr. Ourania Vrondou of the Loughborough University for their motivation. There are many other people whose friendship has kept me going throughout my years at Loughborough University. Special thanks are due to my colleagues Andrew Ng, Paul Lepper, Paul Connelly, Chris Richards, Hamid Reza-Alikhani and Yiliang Song.

Finally, I would like to thank to all my family, above all my mother and father for their love and support throughout many years.

List of Figures and Tables

Glossary Abbreviations

ADC	Analogue to Digital Converter
ADPCM	Adaptive Differential Pulse Code Modulation
AIC	Analogue Interface Circuit
AM	Amplitude Modulation
AMDF	Average Magnitude Difference Function
ANC	Adaptive Noise Cancellation
ASK	Amplitude Shift Keying
CELP	Code Excited Linear Prediction
CMRR	Common Mode Rejection Ratio
DAC	Digital to Analogue Converter
DI	Directivity Index
DM	Delta Modulation
DPPM	Digital Pulse Position Modulation
DPSK	Differential Phase Shift Keying
DSBSC	Double Sideband Suppressed Carrier
DSP	Digital Signal Processor
DT	Detection Threshold
EPROM	Erasable / Programmable Read Only Memory
FET	Field Effect Transistor
FM	Frequency Modulation
FSK	Frequency Shift Keying
GSM	Group Special Mobile
ISI	Intersymbol Interference
LMS	Least Mean Square
LPC	Linear Predictive Coding
NL	Noise Level
PCM	Pulse Code Modulation
PPM	Pulse Position Modulation
PSK	Phase Shift Keying

PTT	Press-to-Talk
PWM	Pulse Width Modulation
RL	Receiver Level
RPE-LTP	Regular Pulse Excited - Long term Prediction
SCUBA	Self Contained Underwater Breathing Apparatus
SL	Source Level
SRAM	Static Random Access Memory
SSB	Single Sideband Modulation
TDMA	Time Division Multiple Access
TL	Transmission Loss
VOX	Voice Activated Switch
VSELP	Vector Sum Excited Linear Prediction

TABLE OF CONTENTS

CHAPTER ONE

INTRODUCTION

1.1 General

From the early days of humans conducting underwater tasks, it has always been desirable to have an efficient means of communication for the transmission of scientific data and control signals and speech. To fulfill this desire, many systems have been designed and these are classified as wirelink and wireless systems. They have been successfully utilized for underwater communications for many years although they have been mainly used for military purposes, with limited civilian applications. The wirelink systems are preferred if an umbilical line is available. The wireless, i.e. acoustic, systems are chosen for applications requiring mobility, as in the case of underwater vehicles and SCUBA divers. In recent years, there have been significant developments in the field of underwater acoustic communications. These improvements are promoted especially by the emerging use of underwater communications for numerous commercial applications, such as pollution monitoring, collection of scientific data from an underwater instruments, remote control of underwater vehicles, and many others. While the number of applications has considerably increased, inevitably underwater acoustic communications have attracted the attention of researchers as a research area. However, the technological achievements in this field have not been as spectacular as those in wireless radio communications. This is simply due to the limitations of the underwater communication channel, in which acoustic signals are greatly absorbed at high frequencies and the received signals can be distorted severely by multipath reverberations. The latter is the most prominent and the most challenging problem which degrades performance of any underwater acoustic communication system. However, recent advances in the Digital Signal Processor (DSP) technology and digital signal processing algorithms have enabled researchers to overcome this problem in real-time applications. These technological developments, when combined with digital communications techniques, have provided transmission of high-rate data through-water.

In most commercial underwater work, a voice communication link is an essential requirement for underwater operators because it allows co-ordination with surface activities, warnings of dangers and calls to assist other divers. Although DSP-based hardware systems (interfaced to a computer) are utilised for underwater acoustic data transmission and reception, it is surprising that they are not implemented for voice communications. Commercially available through-water communication systems are based on analogue technology and this introduces practical limitations in the performance of the system. However, in the event of an emergency underwater, a diver's life may depend on fast and effective action at the surface, so good communication between a pair of divers or between the divers and the surface is essential.

In analogue systems, the multipath effect produces distortion of the received speech signals unless efficient precautions are taken to avoid it. Also, there are other limitations to any analogue system which a digital system can overcome. One of these is to have a private communication link between divers or between a diver and the surface, so that there is no unwanted cross-talk with any other divers in the same area.

Modern telecommunications systems are now very sophisticated, especially since the advent of mobile telephones, but most systems available for underwater voice communication are still comparatively archaic. This thesis presents a comprehensive new design methodology for underwater acoustic voice communications. The initial aims of the research are to design a digital system based on a DSP, to implement speech signal processing algorithms and to transmit and receive speech data in digital modulation format. The thesis presents the work done to develope such a system, together with the results in order to demonstrate the success of the project. The main consideration in the design of a digital underwater communication system is to provide a diver with a comparable level of communications capability as is provided by a digital mobile telephone.

1.2 Organization of the Thesis

Chapter Two provides a comprehensive review of underwater voice communications. Emphasis is given to currently available commercial acoustic voice communication systems. The technology utilized for them is summarized and its limitations are presented.

Chapter Three presents speech coding techniques. First, a brief description of currently available speech coding standards is given. A theoretical study is then presented for a suitable kind of coding to be implemented, and it is shown that low bit rate speech coders are essential for underwater acoustic communications because of the channel limitations. A Linear Predictive Coding (LPC) technique, a speech compression method which is one of the coding standards, for a the bit rate of 2.4 kbit/s, is considered. A theoretical study of the LPC technique is presented and quantization of the estimated LPC parameters is discussed. Decoding of the quantized speech parameters is also presented and synthesizing of speech signals is performed. These algorithms are all written in assembly language for TMS320C31 DSP and burnt into EPROMs.

Chapter Four considers the speech production environment, since talker and listener may be submerged in water. Effects of inhaled gas constituents and pressure on speech signals are studied and techniques to reduce these effects are described. The effects of diving masks on speech signals are also presented and suitable masks and microphones are considered to achieve good quality voice communications.

The main consideration is given to noise cancellation due to the existence of extraneous noise sources which decrease the intelligibility of speech. Amongst them, breathing and bubble noises are studied since they are found to be dominant. Time and frequency domain properties of these signals are presented. It is illustrated that zero crossing and energy measurements of breathing noise are sufficient to cancel breathing noise, while spectral subtraction and adaptive noise cancelling methods are investigated to eliminate bubble noise. Before implementing these algorithms, basic theories and the derivation of appropriate equations are given. The steepest descent gradient LMS adaptive noise-cancelling algorithm is then implemented.

Chapter Five presents a study of the underwater acoustic communication channel. Ambient noise conditions are studied and emphasis is given on the multipath phenomenon of the channel. The sonar equation is also briefly described and some criteria are presented for the voice communication system to achieve the required communication range. A review of through-water digital data transmission methods is also studied and importance is given to the transmission of quantized speech parameters using digital pulse position modulation (DPPM). Detection principles of the DPPM signals are also presented.

Chapter Six presents hardware implementation of a digital underwater acoustic communication system. A Texas Instrument Digital Signal Processor TMS320C31 DSP is utilized as the core of the design. Memory organization and interrupt management of the DSP are described and implemented. The system is analyzed in two sections as a transmitter and a receiver. In the transmitter section, keypad and power amplifier units are implemented, and a brief study of the specifications of the hydrophone are presented. A press-to-talk switch enabling half-duplex communication protocol is also considered. At the receiver section, a solid-state switching network to change the operation mode of the hydrophone is explained. Analogue signal conditioning is presented and analogue-to-digital conversion is performed to decode received DPPM signals. Moreover, analogue processing of speech signals and digitization are illustrated. Finally, several algorithms control the hardware and their operations are explained.

Chapter Seven describes the implementation of the DPPM technique for transmission and detection. Block coding of quantized speech parameters is illustrated and data transmission protocol is presented. Several experimental results are illustrated for DPPM signal transmission. Detection of DPPM signals is performed and experimental results are presented.

Chapter Eight concludes the thesis with summary and provides a number of suggestions to improve the features and performance of the digital underwater acoustic voice communication system.

CHAPTER TWO

REVIEW OF UNDERWATER VOICE COMMUNICATION SYSTEMS

2.1 Introduction

The last few years have witnessed a remarkable development in mobile telecommunication technology. Advances in low bit rate speech coding methods and digital signal processors (DSP) have extensively contributed to this achievement [1,2]. As a result of this progress, the number of mobile telephones has dramatically increased and for many people they have become a basic necessity with numerous advantages.

However, the technological development achieved in mobile telephony has not been paralleled with a corresponding advance in underwater voice communications and most systems available are still comparatively archaic [3]. This is due to several constraints introduced by the nature of the underwater acoustic channel. The most serious problems encountered are the reception of multipath interference and the limitation of usable a frequency spectrum; these are studied in detail in later chapters. Therefore, the purpose of this review is to focus on the underwater voice communication systems used so far and to provide a study of the different voice communication techniques available.

2.2 Underwater Voice Communications

The main limitation to man's effective operation in the sea in various conditions, as illustrated in Fig.2.1, continues to be his inability to communicate effectively underwater [4,5]. In order to provide this facility for both commercial and military divers, a number of relatively sophisticated underwater voice communication systems have been developed. In the design of these systems, two different technologies have

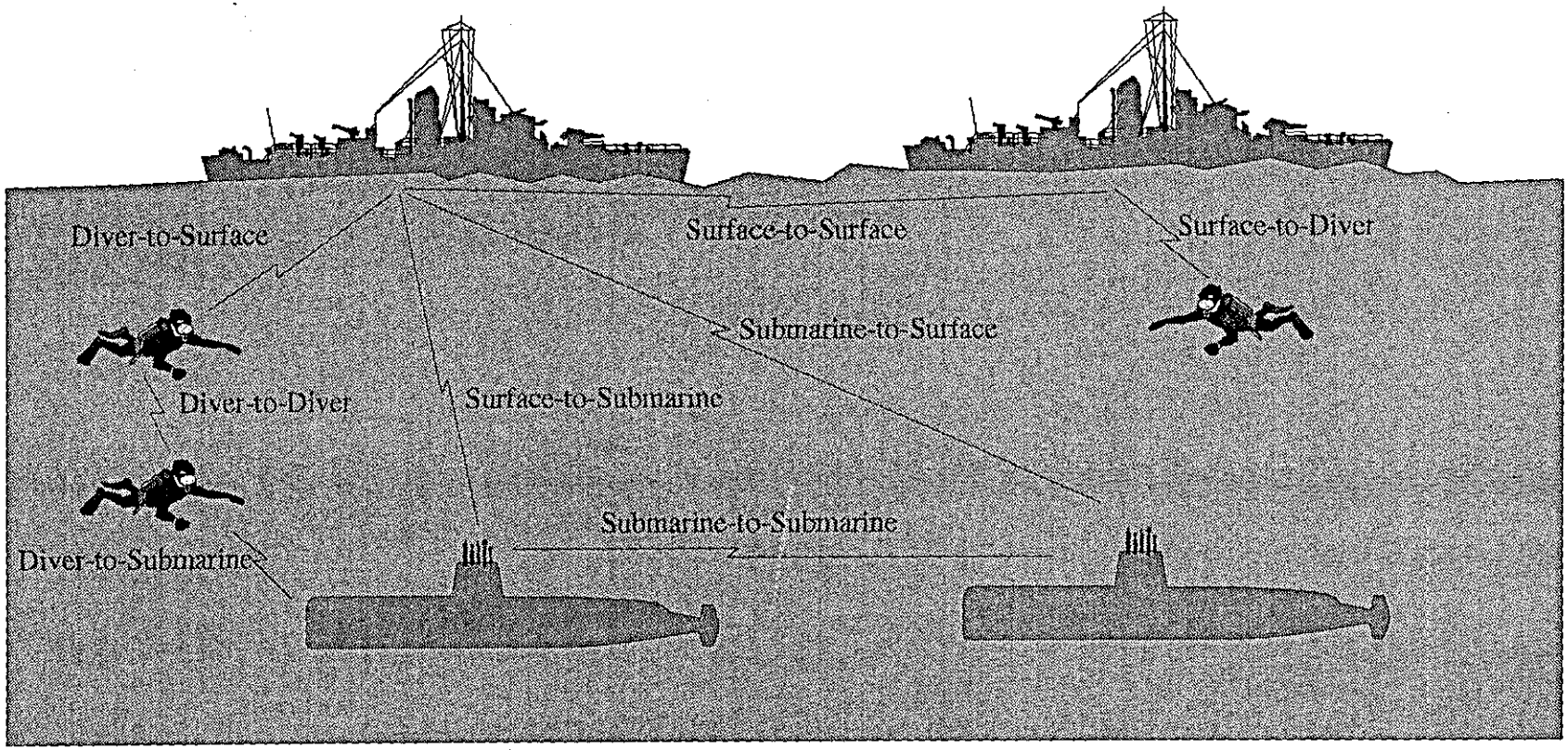


Figure 2.1 Underwater communications

been utilised; these are hard-wired and through-water transmission [6], which will be discussed in this chapter.

2.3 Hard-wired Systems

In deep water diving, the diver usually has an umbilical line which provides a physical connection to the surface or to a diving bell, therefore hard-wired communications between divers or between a diver and supervisor is preferred [6]. A hard-wired unit employs a closed system, comparable to a telephone, including a microphone, a cable and a receiver. Latest versions of these systems incorporate helium speech unscrambling units and speech recording facilities [7], e.g. *Model 3524* (Helle Engineering). Commercially available hard-wired communication systems are classified into two groups according to their configurations; these are *two-wire* and *four-wire* systems [8].

In the two wire configuration, such as *Model 3200B* (Helle Engineering), each diver is connected to the surface unit by two wires, which allow one-way communication, referred to as *half-duplex* operation. The supervisor's system is usually on receive mode, allowing him to monitor the divers' speech and breathing noise, which is a useful safety feature. To talk to them he uses a *press-to-talk* (PTT) switch which reverses the direction of communication. This means that while he is speaking he can not hear the divers. In the event of an accident this could be critical and an obvious potential hazard. Moreover, because he can not hear the diver he is unable to pace his speech to the diver's exhalation. This is important for efficient communication, otherwise bubble noise produced during exhalation may decrease his speech being understood. Because of the obvious limitations of the two-wire system, *full-duplex* systems with four-wires were developed.

In a four-wire configuration, such as *Model 3220C* (Helle Engineering), each diver is connected to the surface by four wires in a "round-robin" configuration which allow two-way communication, referred to as *full-duplex* operation. In this mode, all diver earphones are connected together, as are all diver microphones, but failure of one device will affect communications from the others. In this type of configuration, when two divers share a common microphone uplink, breathing noise from one diver also tends to

obscure the speech of the other. The quality of communication is also affected by noise from crackle due to ingress of water into connectors and umbilicals, and electrical noise from extraneous sources such as motors and switches [4, 6-8]. Although two-wire connection is utilised to provide *full-duplex* operation in telephony systems, which saves half amount of wire, this is not implemented in current underwater hard-wired systems. A possible explanation for this is that it obviates the necessity of the diver carrying electronic circuitry.

Moreover, fibre optic transmission and fibre optic microphones, in which an incident light beam is modulated by the diver's voice, are recently under active consideration [6,8]. Nevertheless, because of the obvious encumbrance of an umbilical line, especially in areas where it is difficult for a surface supervisor to operate, " wire-less" or "through-water" communication systems were introduced in the 1950s; these are more suitable for divers using Self Contained Underwater Breathing Apparatus (SCUBA) [9-15].

2.4 Through-water Communication Systems

Wireless communication between underwater operators has always been in high demand because it provides the advantage of freedom of movement. For this purpose, several signal transmission techniques, such as electromagnetic, conduction, optical and acoustic [9, 10] have been considered. Although acoustic transmission is found superior to the others, they were also experimented on for through-water voice communications. For example, it was reported in [10] that electromagnetic transmission of baseband speech signals was practiced and with a power level of 10 W, 80-90 m communication range was achieved. However, absorption of electromagnetic energy in sea-water is extremely high, about $45\sqrt{f}$ dB per km, where f is frequency in Hz [11], so this high attenuation restricts its use for underwater voice transmission. Conduction systems based on voltage gradients developed in water by a flow of electric current produced by a transmitter [10] were also designed, such as the Dipole Diver Phone of S.A. Laboratoires de Mécanique Appliquèe in France. However, performance of this method was limited by the safe level of electric current in water and the separation between electrodes. Optical systems using blue-green lasers for aircraft to submarine

communications underwater have also been experimented on for military purposes; however, the clarity of the water plays an important role in the efficiency of this system as stated in [10]. From the above discussions, it is evident that none of the electromagnetic, conduction and optical transmission techniques is suitable for underwater communications. However, the acoustic properties of water offer much better performance for signal transmission over long ranges [12], therefore acoustic transmission is widely used for underwater applications including voice communications.

Nowadays, there are many commercially available acoustic voice communication systems and some of them are listed in Table 2.1. Two basic methods of underwater acoustic voice communication exist. Like the hard-wired systems, the earliest versions used baseband methods in which signals were transmitted without modulation. They employed direct amplification of the audio signal and its subsequent transmission via a pressure-compensated electromagnetic or ceramic transducer. Several systems, *Hydrotalk* [16], Raytheon *Yack-Yack* and Bendix *Watercom*, based on this principle were designed and it was concluded that a range of 30 metre could be achieved [16]. But these systems were strongly affected by ambient acoustic noise which is dominant below 6 kHz [17]. Moreover, since baseband methods require more power at low frequencies, they are not suitable for mobile diver communications [18]. Therefore, in the design of later systems, the baseband spectrum was shifted to a higher band by modulating a carrier frequency with the speech signal. For through-water communications, Amplitude Modulation (AM) [19], Frequency Modulation (FM) [20] and Single Sideband (SSB) modulation [21-23] techniques have been implemented.

The early through-water systems employed AM as their transmission method and it was in use for many years [19]. As is well known in literature, in AM transmission, the carrier signal is continuously transmitted, which uses two-thirds of the total power, as well as two sidebands. The other one-third of the power is used by the transmission of the upper and lower sidebands, each of which carries the actual speech signal.

Manufacturer	Specifications		
	Modulation	Frequency(kHz)	Range(m)
Hydrotalk(USA)	Baseband	no carrier	30
Marconi (UK)	AM	31.5	1300
Reference [20]	FM	120	500
Graseby Dynamics (UK)	SSB	10,27,30,37,41,43	1500
Marconi (UK)	SSB	40	1000-1500
Divelink (CANADA)	SSB	31.25	50
Helle Engineering(USA)	SSB	25	1200
Orcatron (USA)	SSB	10, 27,30, 37.5	2000-3500
Slingsby Engineering(UK)	SSB	40.2, 42	1000
Sonardyne(UK)	SSB	30	1000
OTS(USA)	SSB	32.768	1000

Table 2.1 Underwater acoustic voice communication systems and their specifications.

Transmission of the carrier signal, even during periods when no speech signal is being sent, wastes power without any benefit. This is an obvious shortcoming of AM systems and will decrease effective operation time. Moreover, the AM receivers use twice as much bandwidth to accommodate the two sidebands, although they contain identical voice intelligence; this results in serious distortion at the receiving end, i.e. this wider bandwidth causes the receiver to pick up more noise, which is then reproduced at the speaker or head set in addition to the voice signal. The AM signal, which must maintain a certain relation between the carrier and the sidebands, also suffers from the phenomenon of fading in underwater communications. Fading is strongly frequency dependent, i.e. various frequency components suffer different attenuation and nonlinear phase shifts. Because of fading, only the carrier may be attenuated or there may be unequal attenuation and nonlinear phase shifts of the two sidebands and the carrier. This will contribute to distortion in the demodulated speech signal [23]. Although AM has several disadvantages, this technology was used for many years because other techniques, such as SSB, required the use of more components. That would increase the

size, weight and cost of the communication system. However, with the advent of surface mount technology, the implementation of SSB became more attractive [21-23].

The commonest diver communication systems now use the more convenient single sideband (SSB) modulation [21]. This effectively encounters many of the shortcomings of AM transmission because during periods when no speech is being transmitted, SSB dissipates no power. In SSB modulation, the energy in the carrier is suppressed and one of the sidebands (as both sidebands contains identical information) is eliminated at the transmitter before being sent. Hence only one sideband, usually upper sideband, is transmitted. Since no power is wasted on the carrier and lower sideband, the communication range is increased compared to AM. Also, the technique has the advantage of having half the bandwidth of earlier conventional AM systems and Double Sideband Suppressed Carrier (DSBSC) systems, hence it uses the available frequency spectrum more efficiently. Furthermore, as a result of SSB transmission, ambient noise in the range of the lower side band is attenuated, providing better speech quality in the receiver. When the SSB modulation technique is applied to underwater communications, experience has shown that the quality and intelligibility of the received speech are superior to that of the AM technique [23].

Apart from the above modulation schemes, FM transmission has also been tried as a means of implementing an underwater communications system [20]. Although FM provides much better speech quality in radio broadcasting, due to the use of a much wider bandwidth, the ambient noise may cause distortion in the demodulated speech signal. Moreover, speech quality is decreased by multipath propagation in the channel since the frequency and phase, and also magnitude, of the received signal may be affected, consequently it is seen of very little use.

In the modulation techniques discussed so far, the speech is transmitted as a band-limited signal that modulates a carrier frequency centered on the resonant frequency of an omnidirectional electrostrictive transducer. There is no legislation on the use of frequency bands underwater, although certain frequencies have been adopted unofficially for communications [24]. The lower the frequency the greater the range of transmission achievable, since range is inversely proportional to the square of

frequency. Another factor that governs the choice of frequency is the effect of ambient noise, which is noticeable at low frequencies, particularly in the audible range. While some systems use frequencies as low as 8 kHz and as high as 70 kHz, the most common frequency band used in commercially available systems are 30-42 kHz [21-23]. This is high enough to avoid excessive extraneous noise and low enough to give a range of about 1 km, depending on the power output of the transmitter.

For the above modulation techniques, as pointed out earlier, the multipath effect and channel fading produce distortion of the received signals unless, by appropriate signal processing, precautions are taken to avoid it. Also, there are limitations to any analogue system which a digital system can overcome. One of these is to have a private communication link between divers or between a diver and the surface, so that there is no unwanted cross-talk with any other divers in the same area. Although this is possible in principle with DSBSC and SSB, there is practical limit to the number of carrier frequencies that can be accommodated in the bandwidth of a transducer; by contrast this can be overcome by using digital techniques [3].

2.5 Ergonomic Design of Diver Communication System

Apart from the modulation techniques implemented for underwater communications, the ergonomic design of the system is also important [25]. This is associated with the ON/OFF function being implemented by means of an immersible switch. Most of the current acoustic diver communication systems use a press-to-talk switch when the diver wants to talk. Since the switch operation requires the use of one of the diver's hands, this is considered a disadvantage. An alternative is to use a voice-operated switch, which allows transmission automatically when the diver speaks, and recently designed systems tend to incorporate this technique. The voice activated (VOX) transmit system leads to either unintentional transmission when breathing heavily or transmission break up if the diver's speech is too quiet. To prevent this, some systems provide a VOX level control for the diver, but this does not always perform satisfactorily. The solution of a press-to-talk switch works well and is totally under the diver's control.

Some of the units are attached to the diver's aqualung in such a way as to give optimum acoustic coverage but without obstructing diver movement. This is especially pointed

out by experienced divers [25] and they suggested that the whole unit must be placed within the diving helmet or mask. This would obviate the need for long wires that tend to get in the way or get damaged. Advances in microelectronics technology now enable designers to consider this problem and most presently available commercial systems are manufactured under this guideline, as shown in Fig.2.2.

Although current acoustic communication systems tend to improve the communication quality between divers or between diver and supervisor, the diver units still lack facilities, such as a helium speech unscrambler (for oxy-helium diving) which is accompanied by most of the surface units, and private communications. These can be accomplished by means of digital technology. With the advance of digital signal processing technology, a diver's speech signal may be enhanced before transmission. However, the biggest challenge is to provide a private communication facility which is highly restricted due to the use of frequency spectrum. All this can be achieved by digital processing of the speech signal and in the following chapters, initial work on this topic is presented.

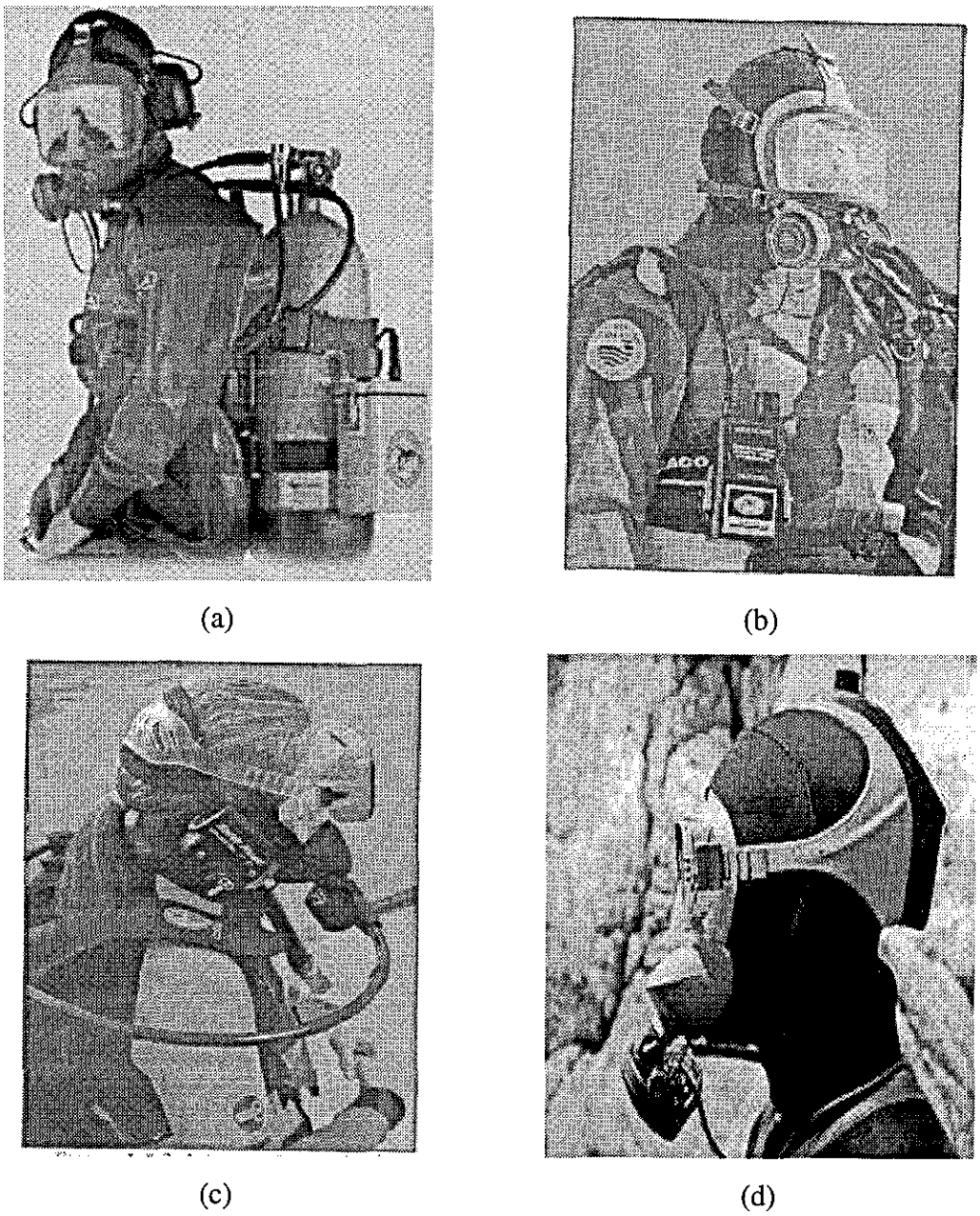


Figure 2.2 Illustration of several diver communication systems:

(a) Orcatron diver unit

(b) Ocean Technology System (OTS) diver unit,

(c) Recently designed diver communication unit of OTS,

(d) Divelink diver unit.

CHAPTER THREE

SPEECH SIGNAL CODING FOR UNDERWATER ACOUSTIC VOICE COMMUNICATIONS

3.1. Introduction

It is evident from present technology that commercial and private usage of speech communication systems continues to grow. This fast growing demand is mostly accomplished by transforming analogue systems into digital systems. The analogue systems have served the needs of underwater communications remarkably well considering its technological simplicity. However, modern information technology requirements have introduced the need for a more robust and flexible alternative to the analogue systems. The attraction of digitally encoded speech is obvious since it has the advantage of ease of regeneration, signaling, flexibility and security although having the disadvantage of needing extra bandwidth for transmission if it is directly applied (without compression) [26-33].

3.2 Feasibility of Speech Coding Techniques

In principle, digital encoding of speech can be achieved by any of the well-known modulation techniques. These include pulse width modulation (PWM), pulse position modulation (PPM) and delta modulation (DM) and there are several internationally recognized speech encoding techniques listed in Table 3.1 whose performance is illustrated in Fig3.1. Amongst them, 64 kbit/s Log-PCM (Pulse Code Modulation) [26] and 32 kbit/s ADPCM (Adaptive Differential Pulse Code Modulation) [28-32] which are classified as waveform coders, served digital systems for many years and have certain advantages and disadvantages. Although they produce high quality speech and low complexity coders, they have been found inadequate in terms of spectrum efficiency when applied to newer bandwidth limited communications [27], i.e. satellite

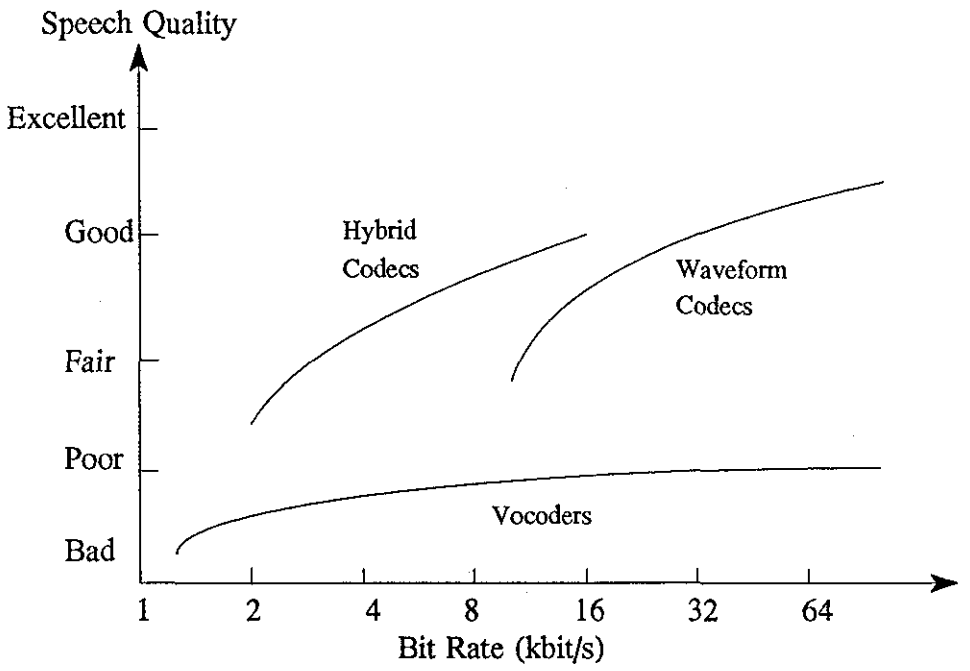


Figure 3.1 Comparison of speech encoding techniques

communications, digital mobile radio, private networks and especially underwater communications.

These are perfectly feasible for hard-wired underwater communication systems, although they do not seem to have been adopted in any presently available product [3] and also not applied for through-water communications. There are severe problems that need to be addressed in addition to the phenomenon of multipath interference for underwater acoustic communication systems. One of this is related to the sampling rate requirement of speech signals. Since the speech bandwidth for telecommunication applications (200 to 3400 Hz) is about 3200 Hz, according to Nyquist sampling theorem the minimum sampling rate needed for recovery of a digital signal back to its original analogue form is at least twice this, i.e. 8 kHz used for voice communications. The corresponding sampling period is therefore 125 μs so in the case of using waveform quantization techniques, e.g. PCM, DM, ADPCM, this is the maximum time allowed to transmit encoded information from an electrostrictive transducer. Because of the limitation of the transducer used in this study, with a resonance frequency of 70 kHz and quality factor of Q=4.8, amongst the waveform coding techniques only ADPCM with the 16 kbit/s, i.e. 62.5 μs bit period, may be worth considering. A necessary condition

for the bit period, which will be discussed in detail in Chapter 6, is that it should be about Q times greater than the carrier waveform period.

It is evident from these considerations that waveform coders are not suitable for underwater acoustic voice communications, other than if a very high carrier frequency is used for a very short range [34]. Since bandwidth is severely restricted in underwater acoustic communications, low rate transmission is essential, hence speech signal compression is vital. For digital speech, previously infeasible, the signal compression is achieved via digital signal processing techniques, which is facilitated by rapid development in digital hardware. Many different strategies for suitably compressing speech for bandwidth restricted applications have been developed [33, 35, 36]. These speech coding techniques, classified as parametric speech coders that rely on a speech production model, compress the speech information into a few parameters which are then encoded for transmission.

Amongst the parametric coding techniques listed in Table 3.1, the FS1015 LPC10 speech coding (Linear Prediction Coding) standard at the bit rate of 2.4 kbit/s [37] is used for military communications although the speech quality sounds far from natural. However, this is acceptable for secure communications. Because of inbuilt limitations of this voice coder, i.e. simplified model of speech production, good quality speech communication is not achievable. In order to improve speech quality and keep the bit rate lower than waveform coders, other speech coding techniques are developed and these are called *hybrid codecs* since they utilize speech production models and waveform information. The FS1016 CELP (Code Excited Linear Prediction) coding method is mainly used for secure military communications at a rate of 4.8 kbit/s [38], and the RPE-LTP (Regular Pulse Excited - Long Term Prediction) coding method at a rate of 13 kbit/s [39] is implemented by GSM (Group Special Mobile) as a standard in Europe. There is also half-rate GSM utilising VSELP (Vector Sum Excited Linear Prediction) coding at a rate of 5.6 kbit/s [33]. These hybrid codecs produce good quality speech and have been successfully applied in mobile communications. Details of these speech coding techniques can be found in [40-47].

Standard	*Year*	*Coding Type*	*Bit Rate (kbit/s)*	*Class*
ITU-G.711	1972	PCM	64	Waveform Coding
ITU-G.721	1984	ADPCM	32	
ITU-G.727	1991	VBR-ADPCM	16,24,32 ,40	
US DoD FS1015	1984	LPC-10	2.4	Parametric Coding
GSM Full-rate IS54	1989	RPE - LTP	13	
US DoD FS1016	1990	CELP	4.8	
IS-96	1991	Qualcomm CELP	1,2,4,and 8	
GSM Half-Rate	1994	VSELP	5.6	

Table 3.1 Several well-known speech coding standards

The parametric speech coding techniques are feasible to accommodate underwater acoustic voice communications due to their low data rate, i.e. high compression rate. In this study, LPC10 vocoder is chosen, although having poor quality speech since it has the lowest complexity and a low data rate. Furthermore, its performance can also be increased by introducing extra functions that are utilized in the hybrid coders. In the following section, this speech coding method is studied and its implementation is described. However, before that, it is appropriate to describe the mechanism of human speech production since parametric speech coders are based on this mechanism.

3.3 Speech Production

The characteristics of a particular speech sound are determined by the shape of the vocal tract and the excitation mechanism [40-47]. The vocal tract is considered to extend from the vocal cord to the lips and the nose as illustrated in Fig. 3.2. The shape of vocal tract determines the gross frequency spectrum of a speech sound, although the effect of a sound radiating from the lips also modifies its frequency spectrum [43-46]. The resonances introduced by the vocal tract are called formants and their frequencies carry information. The formants in the spectrum are denoted F1, F2, F3, F4, F5 beginning with the lowest frequency. Although there are an infinite number of formants in a given sound, usually 3-5 formants are found in the speech band [41]

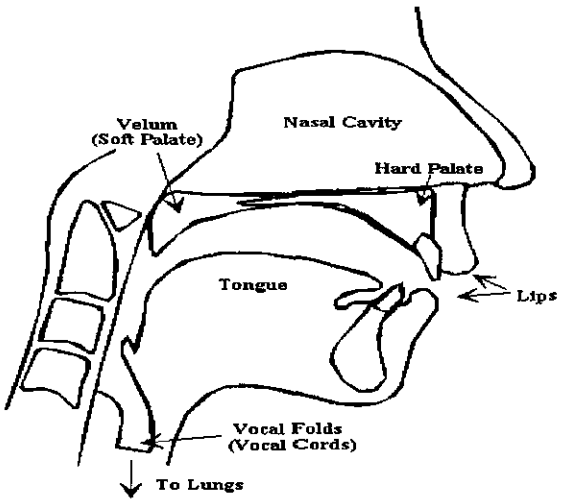


Figure 3.2 A schematic diagram of the human speech production mechanism

There are three basic types of excitation in speech production, voiced, unvoiced and plosive[40-47]. Voiced sound, e.g. */i/* in *seem,* is generated by periodic opening and closing of the vocal cords and its frequency is referred to as the pitch of the speech. Unvoiced sound, e.g. */f/* in *face*, is generated by forcing air past a constriction in the vocal tract and has the characteristics of noise. Plosives are generated by creating a closure in the vocal tract, increasing the air pressure behind it and then releasing the pressure suddenly. The points of closure in the vocal tract determines the voiced, e.g */b/* in *boot*, and unvoiced, e.g. */p/* in *puff*, plosive that is produced. Furthermore, sounds can be generated from a mixture of excitation mechanism, for example the */z/* in *maze* (mixed voiced and unvoiced).

From the above definitions, the speech production mechanism may be modeled as shown in Fig. 3.3 [48], which consists of a time-varying filter and two possible excitation inputs. The excitations are impulse inputs for voiced sounds and random noise for unvoiced sounds; and the time-varying filter represents the response of the vocal tract. In parametric coding of speech signals, these two components are extracted and encoded in the transmitter, and decoded to form synthesized speech in the receiver. Linear prediction plays an important role in this type of speech coding and will now be described in detail.

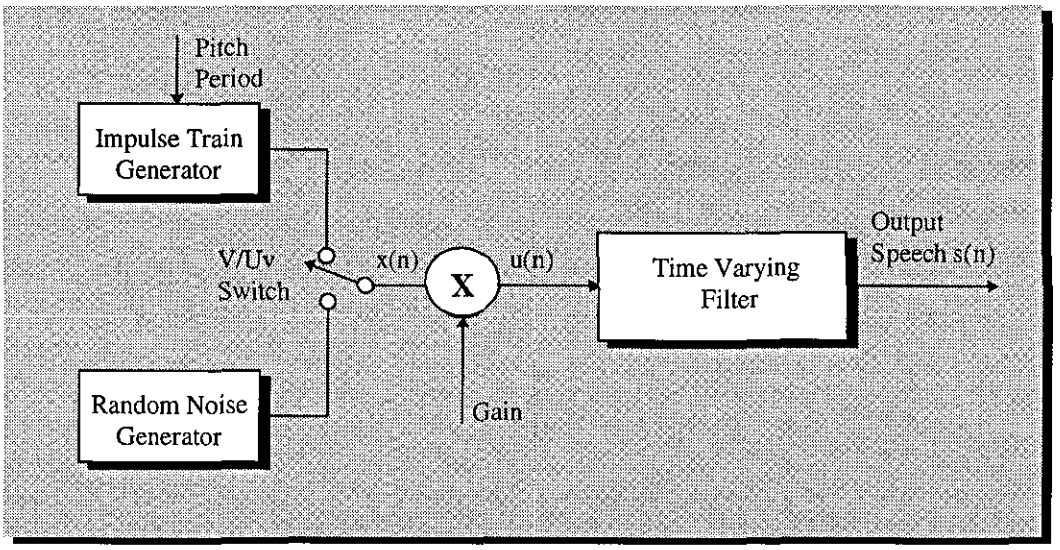


Figure 3.3 Block diagram of the simplified source filter model of speech production

3.4 Linear Prediction Coding of Speech Signal

A linear speech production model, as illustrated in Fig. 3.3, has been developed and analysed in [45]. The vocal tract is modelled as a time-varying digital all-pole filter which is formed by cascading a small number of two-pole resonators representing the formants and defines the envelope of the speech spectrum [43]. The system is excited by two types of inputs, as mentioned previously (impulse for voiced sounds and random noise for unvoiced sounds). In linear prediction coding of speech signals, the excitation and all-pole linear filter coefficients and gain parameters are estimated from a given speech signal sequence and quantized to achieve the transmission rate of 2.4 kbit/s [37, 44]. A mathematical definition of the output of this model may be defined as:

$$H(z)=\frac{S(z)}{X(z)}=\frac{Gain}{1-\sum_{j=1}^{p}a(j)\,z^{-j}}=\frac{Gain}{A(z)} \tag{3.1}$$

where

$$A(z) = 1 - \sum_{j=1}^{p} a(j) z^{-j} \tag{3.2}$$

S(z) and X(z) correspond to the speech and excitation signals. Eq. 3.1 represents the combined spectral contributions of the glottal flow, vocal tract and the radiation of the lips during a stationary frame of speech. Transforming Eq. 3.1 into the sampled time domain, we obtain

$$s(n) = Gain.x(n) + \sum_{j=1}^{p} a(j) s(n-j) \quad , \qquad n=0,1,2,..... \tag{3.3}$$

where $s(n)$ is the present speech signal, $s(n-j)$ is the $(n-j)^{th}$ past speech sample, $a(j)$ is j^{th} coefficient of the filter, $x(n)$ is the excitation input and p is the number of coefficients. The excitation inputs, $x(n)$, are assumed to be statistically independent and therefore have a flat frequency distribution.

As can be seen from Eq.3.1, the speech spectral distribution is the product of the frequency distributions of the filter, $H(z)$ and $X(z)$, and the spectral shape information is contained in the filter $H(z)$. The all-pole filter $H(z)$ can model accurately the spectral resonances (formants) which are characteristics of most speech sounds, but can not model well sounds that contain spectral antiresonances (such as nasal sounds, */ng/* sing) [45, 46].

The filter coefficients are estimated using the Linear Prediction Coding (LPC) method [44]. As illustrated in Fig.3.4, the linear prediction estimates the values of the current speech sample based on a linear combination of the p past samples,

$$\hat{s}(n) = \sum_{j=1}^{p} a(j) s(n-j) \tag{3.4}$$

The error between the current and estimated input speech sample may be defined as:

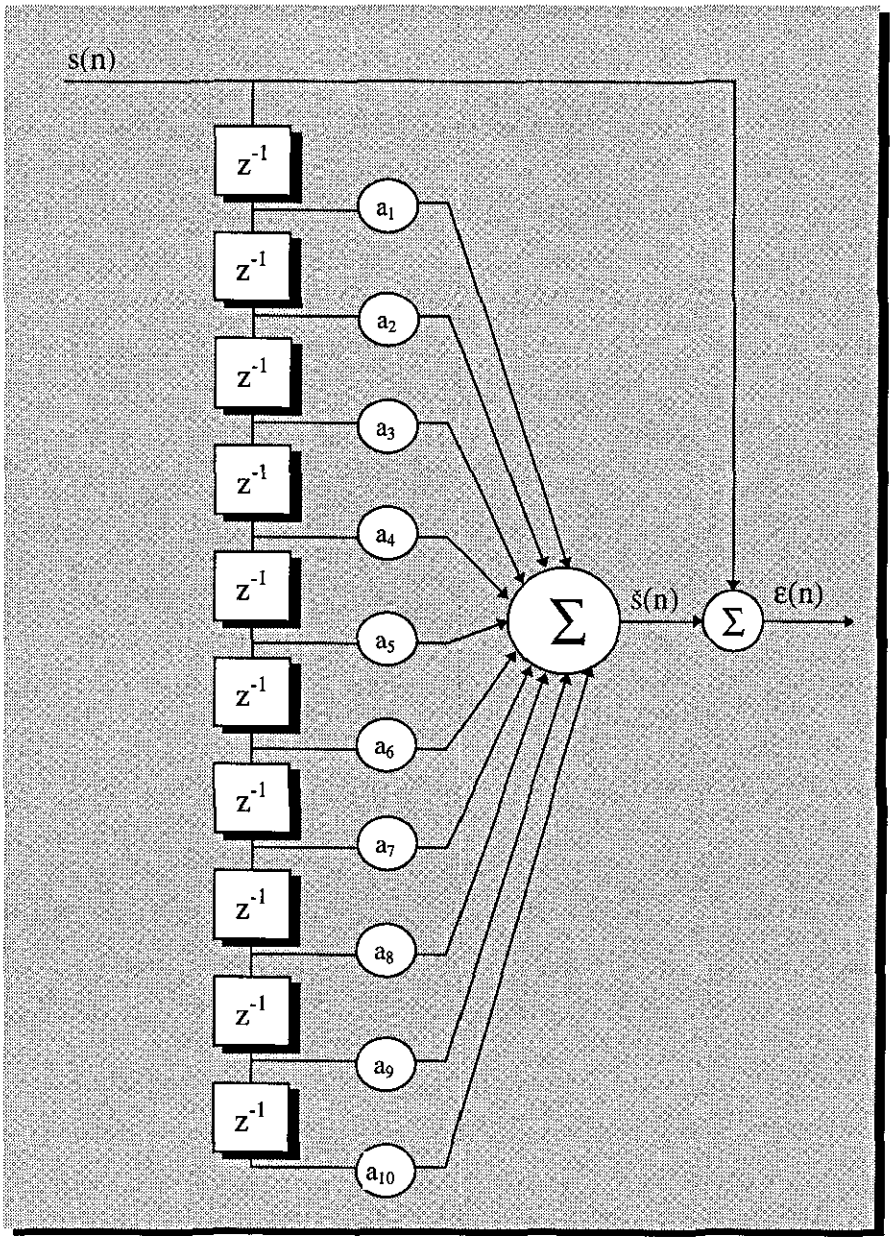


Figure 3.4 Prediction error interperetation of the inverse all-pole filter

$$\varepsilon(n) = s(n) - \sum_{j=1}^{p} a(j)s(n-j) \tag{3.5}$$

Hence the summed squared error, E, over a finite window of length N is

$$E = \sum_{n=0}^{N-1} \varepsilon(n)^2 = \sum_{n=0}^{N-1} \left(s(n) - \sum_{j=1}^{p} a(j)s(n-j) \right)^2 \tag{3.6}$$

By setting the partial derivative of E with respect to $a(i)$ to zero for $j=1,2,...,p$, we get

$$\frac{\partial E}{\partial a(i)} = 0 = -2\sum_{n=0}^{N-1}\left(\left(s(n)-\sum_{j=1}^{p}a(j)s(n-j)\right)s(n-i)\right) \quad , i=1,2,...,p \tag{3.7}$$

Then the Eq. 3.7 can be rearranged to give:

$$\sum_{j=1}^{p}a(j)\sum_{n=0}^{N-1}s(n-j)s(n-i) = \sum_{n=0}^{N-1}s(n)s(n-i) \quad , i=1,2,...,p \tag{3.8}$$

then, it may be simply defined as:

$$\sum_{j=1}^{p}a(j)R(i-j) = R(i) \tag{3.9}$$

where

$$R(i) = \sum_{n=0}^{N-1-i}s(n)s(n+i) \tag{3.10}$$

is the autocorrelation function of the signal $s(n)$ and $R(i)$ is an even function of i, i.e., $R(-i)=R(i)$, and

$$R(i-j) = \sum_{n=0}^{N-1}s(n-i)s(n-j) \tag{3.11}$$

or

$$R(i-j) = \sum_{n=0}^{N-1-(i-j)}s(n)s(n+i-j) \tag{3.12}$$

is the autocorrelation function. Hence, this method is called the *autocorrelation method.* Eq.3.9 may be given in normal matrix form as:

$$\begin{bmatrix} R(0) & R(1) & . & . & R(p-1) \\ R(1) & R(0) & . & . & R(p-2) \\ . & . & . & . & . \\ . & . & . & . & . \\ R(p-1) & . & . & . & R(0) \end{bmatrix} \begin{bmatrix} a(1) \\ a(2) \\ . \\ . \\ a(p) \end{bmatrix} = \begin{bmatrix} R(1) \\ R(2) \\ . \\ . \\ R(p) \end{bmatrix} \tag{3.13}$$

The above autocorrelation matrix has the property that is symmetrical and all the elements along a given diagonal are equal, i.e. a Toeplitz matrix, and can be solved by the simple matrix inversion of the p x p matrix. However, this is not usually performed because of computational complexity. By exploiting the Toeplitz characteristic, very efficient recursive procedures have been devised and the most widely used is Levinson - Durbin algorithm [43], i.e:

$$E_p(0) = R(0) \tag{3.14}$$

$$k(i) = \frac{\left[R(i) - \sum_{j=1}^{i-1} a(j) R(i-j) \right]}{E_p(i-1)} \qquad 1 \le i \le p \tag{3.15}$$

$$a(i) = k(i) \tag{3.16}$$

$$a(j) = a(j) - k(i)a(i-j) \qquad 1 \le j \le i-1 \tag{3.17}$$

$$E_p(i) = \left(1 - k(i)^2\right) E_p(i-1) \tag{3.18}$$

After solving Eq. 3.15 to Eq. 3.18 recursively for $i = 1, 2, ...,p,$, the $k(j)s,$ known as reflection coefficients, and $a(j)$s are found and processed for transmission. In the linear prediction coding, as previously described, gain and the state of the excitation input must be known, as well as the pitch period for voiced sounds, in order to synthesise

speech signals in the receiver. Computation of these parameters is described in the following section.

3.5 Pitch Period Calculation

Pitch period is one of the essential parameters for speech synthesis and must be known in the voiced excitation state. The pitch frequency or fundamental frequency varies from 50 Hz for males to 400 Hz for females [43, 45, 46] and its accurate estimation in speech coding is desired. In the literature, there are several techniques available to carry out pitch period calculation [49-57] and their performance alters in different applications. It is reported that background noise affects the performance of pitch estimation [49, 50]. Since the speech signal produced in underwater is corrupted by bubble noise it is expected that there will be a deviation in pitch period between clear speech and noisy speech; this is illustrated in Fig.3.5.

The main principle of pitch detection algorithms relying on waveform similarities is to find the pitch by comparing the similarity between the original signal and its shifted version. If the shifted distance is equal to the pitch, the two signal waveforms should have the greatest similarity. The autocorrelation method, which is a measure of signal agreement [54], and the average magnitude difference function (AMDF), which is a measure of the disagreement [55, 56], are the two most widely used techniques. The advantage of AMDF is the computational simplicity, as the subtraction and magnitude computation structures are much faster than the multiply-add structure, as in the autocorrelation method. Therefore, it is accommodated in real time applications of speech communications [37] and recommended for underwater acoustic voice communications. The AMDF is defined as:

$$AMDF(\tau) = \sum_{n=0}^{N-1} |s(n) - s(n+\tau)| \quad ; 20 \le \tau \le 156 \tag{3.19}$$

where τ is delay and when the pitch is equal to the value of τ, the result is a minimum.

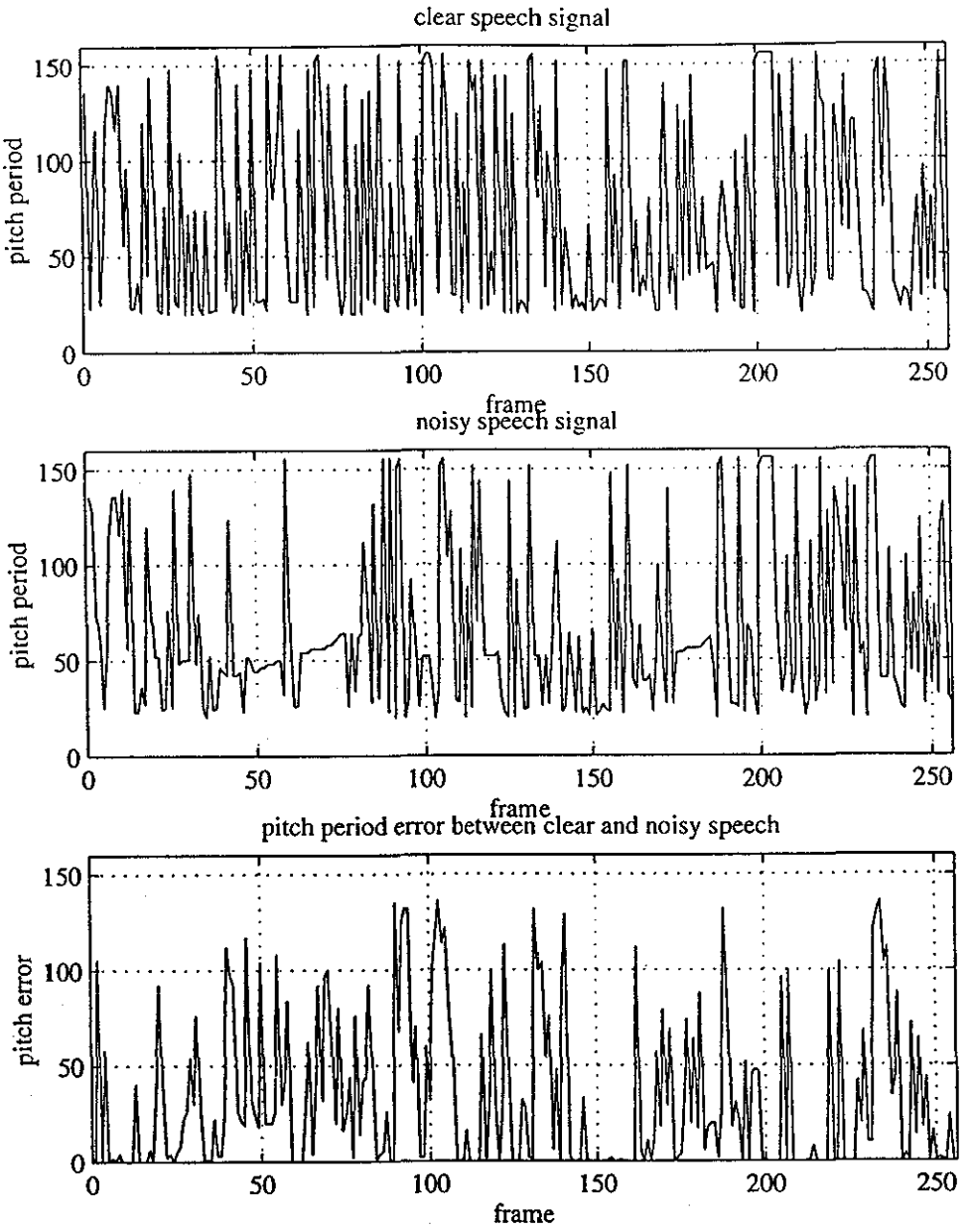


Figure 3.5 An illustration of pitch period of clear and bubble noise added speech signals and variation of the pitch period (speech signal is from the thesis author and bubble noise is recorded during diving trials with a *wet mask*).

3.6 Estimation of Excitation Mode

A particular difficulty encountered in all low bit rate voice coders, i.e. parametric coders, is the reliable and accurate measurement of the voiced/unvoiced nature of speech [37, 49]. Obviously, human speech violates the simple binary voiced/unvoiced hypothesis by being a product of mixed excitation or by changing the classes over the classification interval. This is the main limitation degrading the speech quality and intelligibility of the LPC method [45, 46]. Therefore, the correct voicing decision is crucial to the perceived quality and naturalness of LPC systems and an improved voicing classification method is developed for LPC applications [58].

Since the LPC method is employed for our underwater acoustic voice communications, extraction of voicing information is also essential. Since the speech signal is processed in frames, a voicing decision is made for each frame as in [58]. The voicing detector in the system may utilize the ratio of the maximum and minimum values of the AMDF function, i.e. increases if the frame is voiced. The zero crossing rate indicates the dominant spectral concentration in the speech signal. Lowband energy of the input speech signal (most obvious and simple indicator of voiced sounds) can also be employed to improve accuracy in the voicing decision. The first two parameters are found sufficient to define a two-state excitation model as given:

$$if \begin{cases} \dfrac{max(AMDF))}{min(AMDF)} \geq 1.8 \quad \text{and} \quad \text{zero\_crossing} \leq 90\,; & \text{voiced excitation} \\ \text{otherwise} \qquad\qquad\qquad\qquad\qquad\quad ; & \text{unvoiced excitation} \end{cases}$$

3.7 Implementation of LPC10 Algorithm for Underwater Acoustic Communications

The LPC10 algorithm, i.e. with $p=10$ filter coefficients, as illustrated in Fig.3.6, is implemented for underwater acoustic voice communications in order to achieve low bit rate transmission. An accepted operational rule for choosing the order of the all-pole digital time-varying filter is as defined in [43]:

$$p = \begin{cases} f_s + 4, & \textit{voiced} \\ f_s, & \textit{unvoiced} \end{cases} \tag{3.20}$$

where f_s is the sampling frequency of the speech signal in kHz. This provides one pole pair per kHz of speech bandwidth, which assumes that its spectrum contains one formant per kHz.

The speech signals are bandpass filtered at frequencies of 200 - 3400 Hz, sampled at a rate of 8 kHz, then quantized to 14-bit resolution. The pre-processing of speech signal is done in hardware as described in Chapter 6. The digitized speech samples are stored in a buffer reserved for use by LPC algorithm and the size of the buffer is selected as *N=180* samples (22.5 ms), on the assumption that the speech is stationary. This interval is chosen sufficiently long to compose at least one pitch period, i.e. approximately 20 to 160 sample intervals. In the first stage of LPC implementation, the speech signal is pre-emphasised, as illustrated in Eq.3.21, so that the energy of the high frequencies in the spectrum is increased by 6 dB per octave to compensate for the spectral effect of glottal poles [46].

$$P(z) = 1 - 0.95z^{-1} \tag{3.21}$$

A windowing operation is then applied to the pre-emphasised speech signal. There are various types of windows available for various applications and their shape and length can affect the frequency representation of the speech [59]. Amongst them, the Hamming window, as shown in Eq.3.22, is the most commonly used one having lower spectral leakage than a rectangular window outside the frequency band of interest; hence it was chosen for this study.

$$w(n) = \begin{cases} 0.54 - 0.46\cos\dfrac{(2\pi n)}{N-1} & ;\ 0 \le n \le N-1 \\ 0 & ;\ \text{otherwise} \end{cases} \tag{3.22}$$

The next step is computation of the filter coefficients from Eq. 3.13, therefore firstly the autocorrelation coefficients are calculated from the speech frame by using Eq.3.10 and reflection coefficients, *k(j)* and all-pole filter coefficients, *a(j),* are estimated using the Levinson-Durbin algorithm, Eq.3.15 - Eq.3.18.

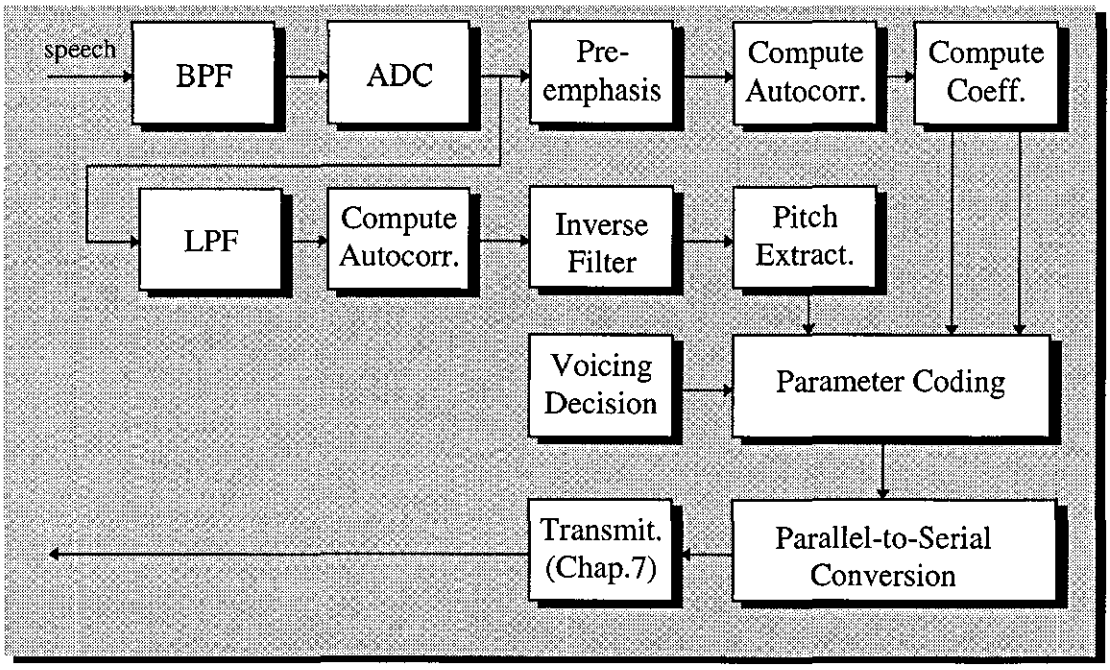


Figure 3.6 Linear predictive speech coder

Once the filter coefficients are found, the pitch period is estimated by using the AMDF algorithm, where $20 \leq \tau \leq 156$ has *sixty pitch values*, as shown in Table 3.2. Before applying Eq.3.19 to find the pitch period, since the pitch frequency is over the range of 50 to 400 Hz, the frame of the input speech signal is lowpass filtered with a cut-off frequency of 900 Hz so that second and higher formants are attenuated [55-57]. Since the filtered speech signal is still composed of the first and probably the second formant frequencies, it is passed through a second-order all-zero filter (also called inverse filter whose coefficients are calculated using autocorrelation matrix, i.e. Eq. 3.13) in order to remove the formant frequencies and improve the performance of the pitch extractor, as implemented in [53]. Then Eq.3.19 is applied to this signal and when the minimum value of the AMDF is found, index of the corresponding τ value in the Table 3.2, which defines the pitch period, is transmitted.

index	τ	*index*	τ	*index*	τ	*index*	τ	*index*	τ	*index*	τ
0	20	*10*	30	*20*	40	*30*	60	*40*	80	*50*	120
1	21	*11*	31	*21*	42	*31*	62	*41*	84	*51*	124
2	22	*12*	32	*22*	44	*32*	64	*42*	88	*52*	128
3	23	*13*	33	*23*	46	*33*	66	*43*	92	*53*	132
4	24	*14*	34	*24*	48	*34*	68	*44*	96	*54*	136
5	25	*15*	35	*25*	50	*35*	70	*45*	100	*55*	140
6	26	*16*	36	*26*	52	*36*	72	*46*	104	*56*	144
7	27	*17*	37	*27*	54	*37*	74	*47*	108	*57*	148
8	28	*18*	38	*28*	56	*38*	76	*48*	112	*58*	152
9	29	*19*	39	*29*	58	*39*	78	*49*	116	*59*	156

Table 3.2 Pitch period delays

The excitation mode and gain of the excitation signal must be also defined before transmission. The excitation mode decision is made from the zero crossing and the ratio of maximum and minimum values of AMDF as described earlier. To define the gain parameter, from Eq. 3.1 to Eq.3.3, we can say that the vocal tract model is excited by the error signal, *u(n),* hence its energy measurement for the given speech frame will provide this as defined in [44]. After computing the required parameters, to produce speech, they must be quantized to transmit them through an underwater acoustic communication channel. This process is described in the following section.

3.8 Encoding and Decoding of Speech Parameters

In the LPC technique, information related to a frame of speech signal is extracted. As described before, these are filter coefficients, pitch period (which is already coded at 6 bits), energy of the excitation input, and the excitation mode, i.e. voiced/unvoiced and 1 bit. Since a transmission rate of 2.4 kbit/s is intended the reflection coefficients and excitation gain must be quantized accordingly.

3.8.1 Quantization of Filter Coefficients

In LPC implementation, stability of the analysis filter in Eq. 3.2 must be provided and this condition may not be guaranteed if the LPC filter coefficients, *a(j),* are directly quantized since they are very sensitive to quantization noise [43-47] and if they were to

be coded, they would require 8-10 bits per coefficients for accurate representation. However, there are several other quantization techniques available to achieve a low quantization rate, such as vector quantization [60], PARCOR (partial correlation) coefficients or reflection coefficients, *k(j),* [61] and line spectral frequencies [62, 63]. The last two are most often used in speech communication applications. However, in this study, reflection coefficients, *k(j),* are coded to overcome the stability problem of the LPC filter and the filter is said to be stable if |k(j)| ≤1. The distribution plots of reflection coefficients for a long sequence of speech with a 10th order LPC filter are shown in Fig. 3.7 and utilized to define the quantization of the reflection coefficients. Step size for uniform quantization of the reflection coefficients, Δ, is described by

$$\Delta(j) = \frac{k(j)_{max} - k(j)_{min}}{2^{b_j} - 1} \quad ; \quad j=1,2,....,p \tag{3.23}$$

where b_j is the quantization rate for the *jth* reflection coefficient and an important parameter because of its influence in data transmission rate. The number of bits used by each coefficient results from the coefficient's relative importance in the resolution of speech. In this application, reflection coefficients, k(1) - k(4), k(5) - k(8), k(9) and k(10), are quantized at the rate of 5 bits, 4 bits, 3 bits and 2 bits, respectively, in a similar manner in [37] and a look-up table as illustrated in Table 3.3 is generated. During the encoding process of these coefficients, the closest match is found in the look-up table and its *index* is encoded to be transmitted corresponding to the appropriate coefficient.

index	*k(1,2,3,4)*	*k(5,6,7,8)*	*k(9)*	*k(10)*
0	0.966500	0.823100	0.619200	0.653100
1	0.903642	0.717547	0.454171	0.341153
2	0.840784	0.611993	0.289143	0.029166
3	0.777926	0.506440	0.124114	-0.282800
4	0.715068	0.400887	-0.040914	
5	0.652210	0.295333	-0.205943	
6	0.589352	0.189780	-0.370971	
7	0.526494	0.084227	-0.536000	
8	0.463636	-0.021327		
9	0.400778	-0.126880		
10	0.337919	-0.232433		
11	0.275061	-0.337987		
12	0.212203	-0.443540		
13	0.149345	-0.549093		
14	0.086487	-0.654647		
15	0.023629	-0.760200		
16	-0.039229			
17	-0.102087			
18	-0.164945			
19	-0.227803			
20	-0.290661			
21	-0.353519			
22	-0.416377			
23	-0.479236			
24	-0.542094			
25	-0.604952			
26	-0.667810			
27	-0.730668			
28	-0.793526			
29	-0.856384			
30	-0.919242			
31	-0.982100			

Table 3.3 Quantization of reflection coefficients

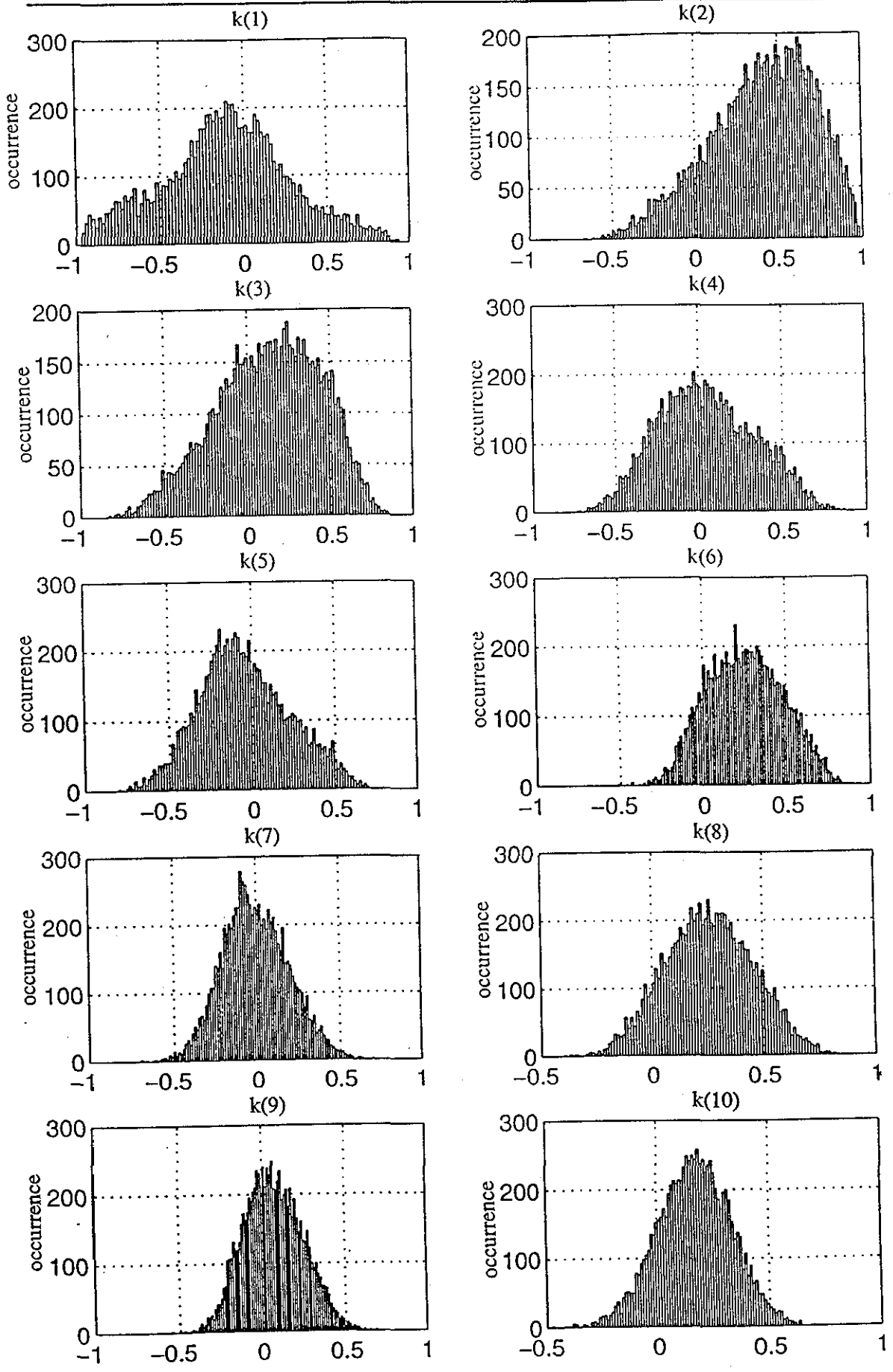


Figure 3.7 Distributions of reflection coefficients for a sequence of speech signal from the thesis author.

3.8.2 Quantization of gain parameter

As defined earlier, one of the parameters required to be transmitted in the LPC technique is the gain of the excitation signal which may be may be computed by utilizing Eq. 3.24 [43, 44].

$$gain = \sqrt{E_p} = \sqrt{R(0) + \sum_{j=1}^{p} a(j)R(j)} \tag{3.24}$$

Since during the implementation of the Levinson-Durbin algorithm, the error signal energy, E_p (see Eq. 3.18), is calculated and minimized, only its square root is computed and used for quantization. As illustrated by the distribution of gain of clear and noisy speech signals in Fig 3.8, the gain parameter may have a maximum value of 1024. Although this is not measured it must be considered while quantizing. For this purpose, 6 bits are utilized and the look-up table shown in Table 3.4 is generated at a step size of 0.4778 dB to encode this parameter. When the gain input matches the closest value in the look-up table, its index is utilized for transmission as in reflection coefficient transmission.

index	*gain*	*index*	*gain*	*index*	*gain*	*index*	*gain*
63	1024	*47*	181.0193	*31*	32	*15*	6.303923
62	918.891	*46*	162.4385	*30*	28.71535	*14*	5.656854
61	824.571	*45*	145.7649	*29*	25.76785	*13*	5.076204
60	739.9326	*44*	130.8028	*28*	23.12289	*12*	4.555155
59	663.9819	*43*	117.3765	*27*	20.74943	*11*	4.087589
58	595.8271	*42*	105.3284	*26*	18.6196	*10*	3.668016
57	534.6682	*41*	94.51687	*25*	16.70838	*9*	3.291511
56	479.7869	*40*	84.81515	*24*	14.99334	*8*	2.953652
55	430.539	*39*	76.10926	*23*	13.45434	*7*	2.650473
54	386.3461	*38*	68.29699	22	12.07332	*6*	2.378414
53	346.6894	*37*	61.28661	*21*	10.83404	*5*	2.134281
52	311.1033	*36*	54.99582	*20*	9.721979	*4*	1.915207
51	279.17	*35*	49.35075	*19*	8.724062	*3*	1.718619
50	250.5144	*34*	44.28512	*18*	7.828576	2	1.542211
49	224.8003	*33*	39.73945	*17*	7.025009	*1*	1.38391
48	201.7255	*32*	35.66038	*16*	6.303923	*0*	1.241858

Table 3.4 Quantization of gain of excitation input

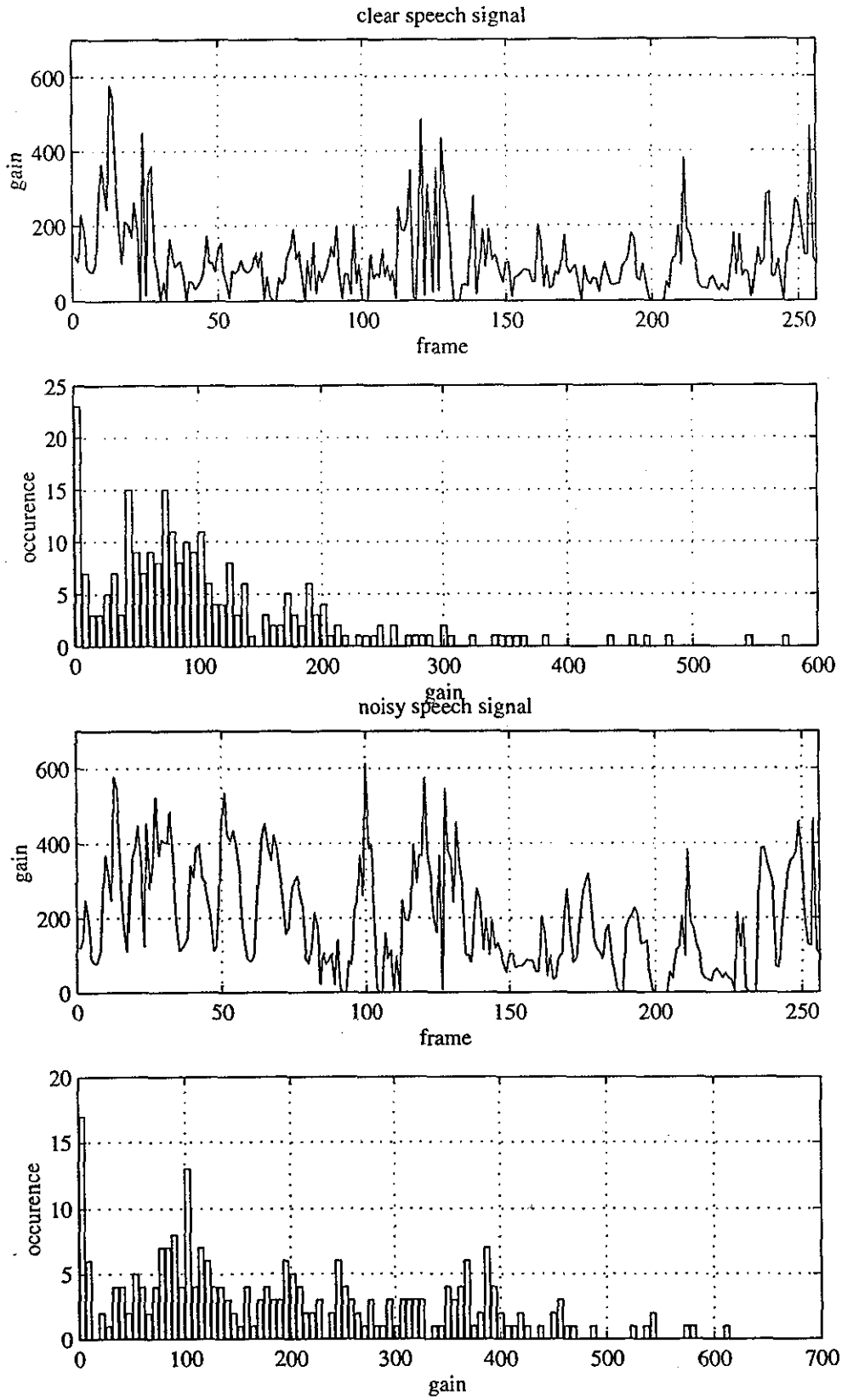


Figure 3.8 Measurement of excitation gain estimated from clear and noisy speech signal and their distributions.

Before discussing the decoding process, we may summarize the number of bits allocated for transmission and this is illustrated in Table 3.5. As seen from the table, a 22.5 ms speech frame is encoded into a total number of 54 bits. This suggests that the compression ratio is 26.66 per second in case of quantizing the sample of speech at 8 bits and 46.66 per second in the case of 14 bits, which is employed for the system. These parameters are transmitted, as will be discussed in Chapter 7.

Parameter	*Bits*
gain	6
pitch	6
k(1)	5
k(2)	5
k(3)	5
k(4)	5
k(5)	4
k(6)	4
k(7)	4
k(8)	4
k(9)	3
k(10)	2
voice/unvoiced	1
total	54

Table 3.5 Bit allocations for LPC parameters

3.8.3 Decoding of Speech Parameters

After receiving the transmitted parameters, assuming that an error-free channel, and completing the channel decoding process, i.e. block decoding of received data studied in Chapter 7, the parameters must be decoded so that the speech signal is reconstructed. The decoder utilizes the same tables used for encoding, and the decoding is accomplished by table look-up. Once the parameters, reflection coefficients, pitch period, gain and voiced/unvoiced decision, are decoded in the receiver, they are passed to a speech synthesising stage to produce the speech signal, as illustrated in Fig.3.9 and discussed in the following section.

3.9 Speech Signal Synthesis

The speech synthesising process operates as illustrated in Fig.3.9, and the model requires filter coefficients, i.e. reflection coefficient, state of excitation and gain, and pitch period parameters. Since they are already extracted during the LPC analyses, then transmitted by the transmitter and correctly decoded in the receiver, enough information is provided to reconstruct the speech signal.

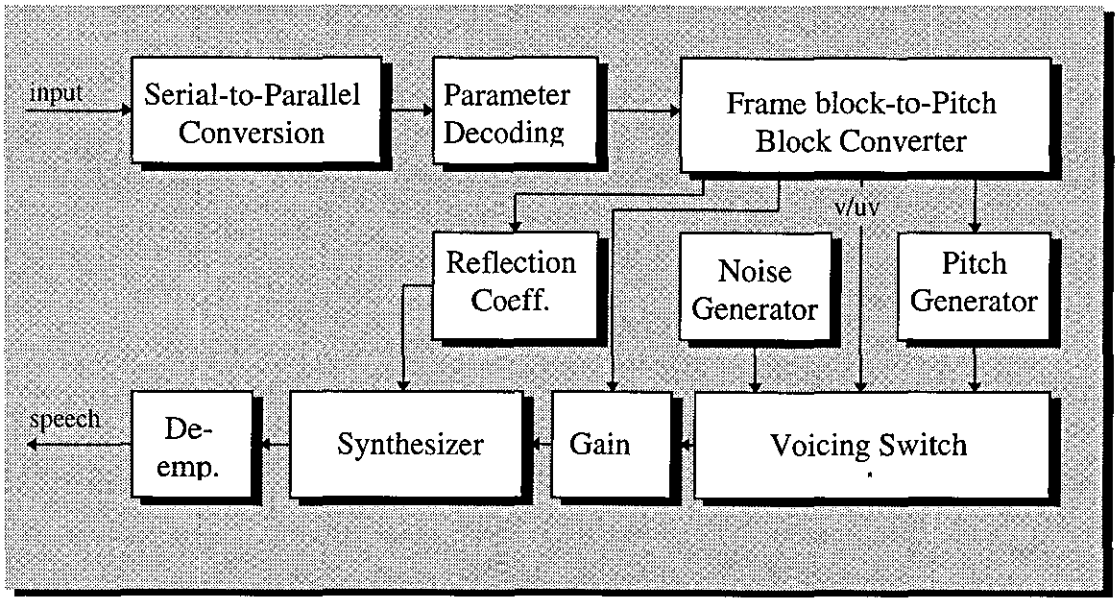


Figure 3.9 Linear predictive code synthesizer

The synthesis algorithm follows from the description of the all-pole model as defined in Eq. 3.3. In the synthesizer illustrated in Fig.3.10, the lattice filter structure [43] using $p=10$ reflection coefficients is excited by a periodic impulse train for voiced speech or white noise for unvoiced speech. The white noise unvoiced excitation is generated by reading random numbers from a table with a random starting point and with a random incrementing index for each call to the subroutine [64]. The excitation gain is defined for unvoiced and voiced speech in [43] as:

$$Excitation\ Gain = \begin{cases} \dfrac{Gain}{\sqrt{\dfrac{N}{3}}} & unvoiced \\ Gain\sqrt{\dfrac{Pitch}{N}} & voiced \end{cases} \tag{3.25}$$

A frame of delay is added in the receiver so that smoothing can be applied to the present frame. The executive functions in the receiver are accomplished by the frame-block-to-pitch-block conversion and interpolation.

The reconstructed signal at the output of the synthesis filter is de-emphasised before application to the D/A converter. This filter, whose transfer function is given in Eq.3.26, reverses the effect of the original pre-emphasis filter [44-47]. Then the speech signal is applied to a D/A converter and low pass filter (with a cut-off frequency of 3.4 kHz) in the Analogue Interface Circuit (AIC) of the hardware.

$$H(z) = \frac{1}{1 - 0.95z^{-1}} \tag{3.26}$$

The LPC method described in this chapter is written in TMS320C31 DSP assembly language and stored in the memory of the prototype underwater acoustic voice communication system. To test the algorithms, samples of the voiced part of a speech waveform, illustrated in Fig. 3.11, are stored in system's memory and analysed. The analysing process is done as shown in Fig.3.6 and the estimated speech parameters are coded for internal transmission, i.e. modeling a noise-free channel, as will be defined in Chapter 7. The received parameters are decoded accordingly and the speech signal is synthesised in a manner shown in Fig.3.9. The synthesised speech samples are sent to the output port of the AIC and shown in Fig. 3.11.

When the LPC technique is employed for speech compression, the quality of the synthesised speech signal is significantly decreased, although speech intelligibility is maintained. Indeed, this result is expected due to the limitation of the algorithm which

only employs a binary voicing decision and two state of excitation. However, several speech coding techniques which improve the excitation input to the synthesis filter are available and these are implemented in mobile telephony, e.g. CELP, RPE-LTP. Although these methods will increase the transmission rate, they are always lower than PCM or ADPCM, and for the sake of achieving the same communication standard as mobile telephony, they must be utilized in underwater acoustic voice communications.

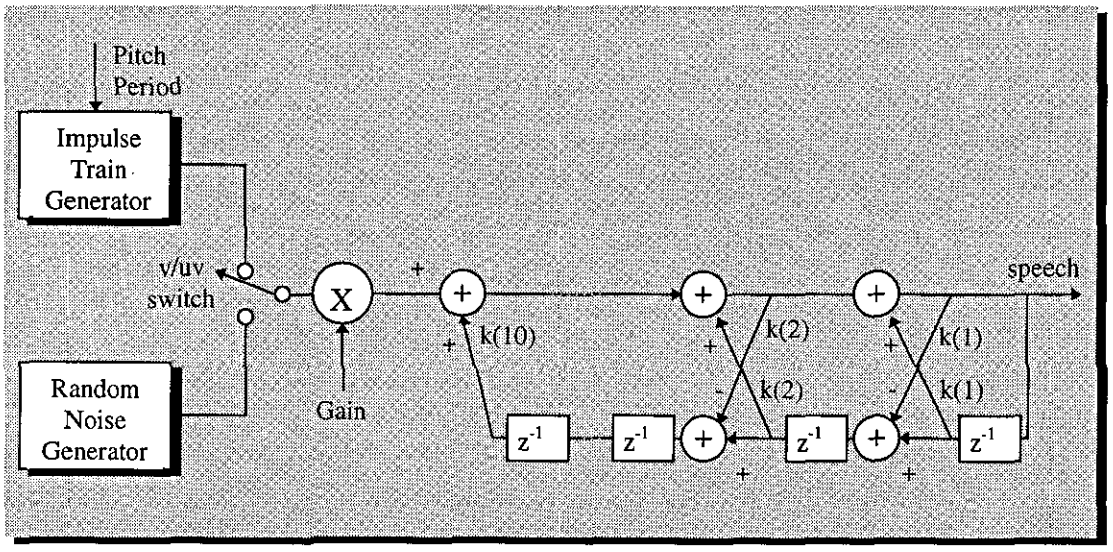


Figure 3.10 Synthesis lattice structure

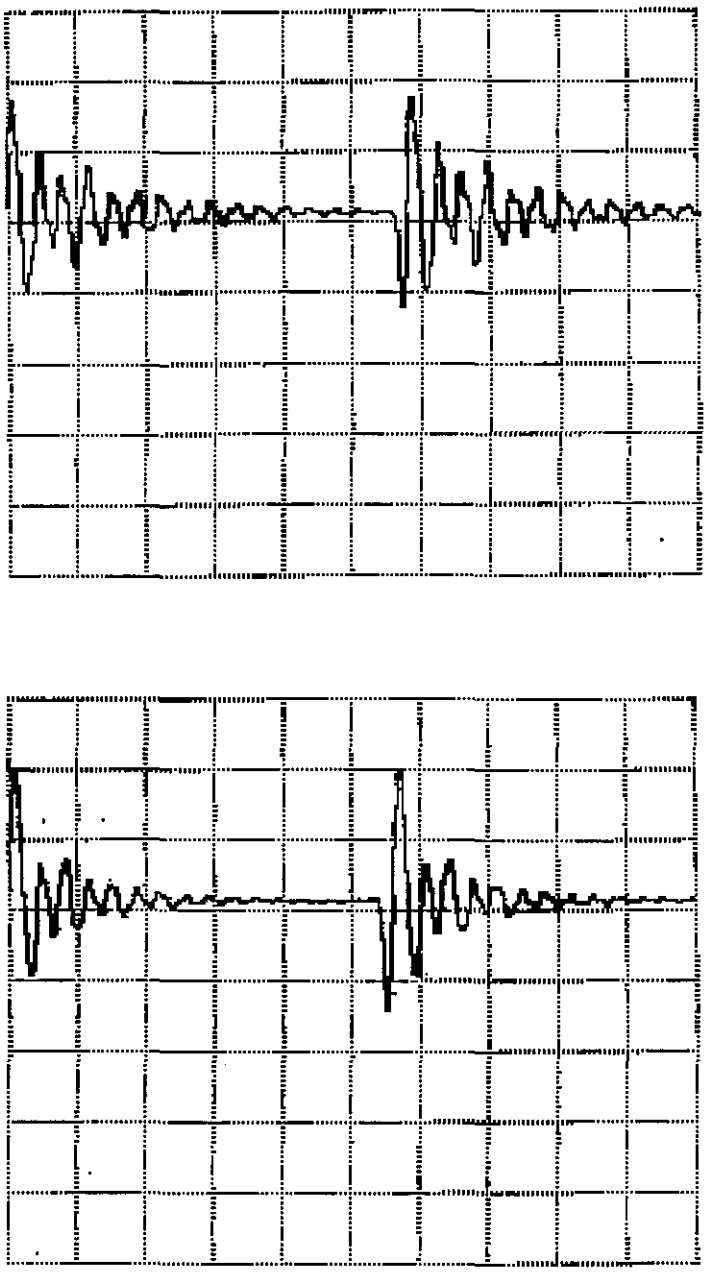


Figure 3.11 Analysed and synthesised speech signal.

CHAPTER FOUR

SPEECH QUALITY ENHANCEMENT FROM DIVER'S MASK

4.1 Introduction

In diver communications, transmission of clear speech signals is essential for reliable communication between divers, or between a diver and a supervisor at the surface, for the safety reasons and operational efficiency. Since the 1940s, researchers initiated studies in order to investigate factors influencing diver's speech [3-8], mainly noise and distortion these are introduced at the source, in the transmission channel (whose effect is elaborated in Chapter 5) and at the receiver. Distortion introduced at the source depends on the type of microphone employed for the system, the shape of the diver's mask and the breathing gas constituents. In addition, bubble noise, breathing noise and other external noises, which must be cancelled or at least minimised before transmission, have a significant effect in speech quality. In this chapter, these effects are studied and several noise cancelling methods are experimented in order to enhance speech signal quality whilst leaving the diver's speech undistorted.

4.2. Speech Environment

The main scope for improving the underwater communication systems is by improving the reliability of components, especially microphones and earphones [8]. A diver's microphone has three serious constraints. Firstly, it must resist water damage and corrosion. Secondly, due to pressure changes, i.e. compression and decompression, enclosed gas volumes are not permissible. Taken in combination, these first two factors cause many of the problems peculiar to underwater microphones. Thirdly, in the case of diver breathing heliox, the gas increases the frequency range of human speech up to around 8-10 kHz [7], which is well above the response range of ordinary communications microphones.

Despite the fact that the diver's microphone and earphones are placed within the mask, they are subject to the full ambient pressure. Depending on the form of construction, adequate steps must be taken either to pressure proof or pressure balance these devices. Either way, the frequency response which may be achieved at the normal pressure is likely to be adversely affected at elevated pressure. During diving trials, two different electret microphones, CF-2949 and MR-8406, manufactured by Knowles Ltd were utilised. It was noticed that use of the waterproof MR-8406 microphone dramatically decreased the speech quality when the ambient pressure was increased and noise was generated. Moreover, since it is used in a confined environment and not designed for the purpose of close range applications, the diver's mask also has an influence on its performance. Therefore, in the case of using this type of microphone, precautions must be taken for compensating for the overall effect of these factors, otherwise a different type of microphone must be accommodated. When the CF-2949 microphone was utilized, its superiority was observed in the clarity and intelligibility of the speech. This is specifically designed to operate in near-field and manufactured in such a way as to compensate the change in ambient pressure.

In addition to the limitations introduced by the microphone, the choice of earphones employed is important since they will also be used underwater. They must be coated with water proof material and put in a suitable protective enclosure. Piezo-ceramic speakers are reported to couple well into water [8] and are recommended for use in underwater voice communications systems since they are low cost devices and have a response of 500 Hz to 20 kHz. Although the system described here has not been tested underwater, to improve performance of the system the same type earphones must be utilised.

4.2.1 Effect of Ambient Pressure on Speech Production

In underwater communications, pressure, breathing gas constituents and obstruction of the vocal processes all compound to create considerable difficulties in the establishment of clear and intelligible speech communication. When the depth increases, the pressure of inhaled gas must be equal to the ambient pressure level. The consequence of this, according to one report is that at a depth of more than 30 m, i.e. pressure in excess of 4 bar, the overall sound pressure level of voiced sounds increases and a relative loss of spectrum intensity at high frequency occurs, resulting in speech having a nasal quality [65]. An explanation given for this is the reaction of nasal cavity, i.e. the velum (see Fig.3.2 in Chapter 3), which is opened to allow for sound propagation. However, as described in [66], an x-ray picture of the velum during speech production illustrated that the velum was in normal status at 6 bar pressure and it was concluded that the nasal quality was caused by the functioning of the velum. Other studies on the effects of high air pressures on speech in a decompression chamber [66, 67] shows a non-linear shift of low-frequency vocal resonances and it has also been observed that speech signal is subjectively perceived as "nasality". This effect is primarily related to vocal cavity wall vibrations resulting from a reduction in the impedance mismatch between the air and the cavity walls. The reduced impedance mismatch causes the cavity walls to absorb energy which results in their oscillation. The effect of cavity wall vibrations is to increase the radiation of sound from the skin of the head and throat and to add reactive elements participating in the tuning of vocal resonances. It was concluded that the latter effect is the main cause of this distortion [67]. In the present system, compensation for the effects of ambient pressure was not attempted, since the clarity and intelligibility of the speech has been the main interest.

4.2.2 Gas Constituents

Deep sea diving is a specialist industry which has grown dramatically as a result of oil exploration. For the safety of divers, when diving depth is in excess of 50 metres, air as the breathing gas is replaced by a mixture of helium and oxygen (95 % helium, 3 % nitrogen and 2 % oxygen), the ratio of oxygen to helium varying as a function of depth[68]. This inevitably results in variation in the speed of sound, which directly

influences the speech signal spectrum. The speed of sound in a gas is related to its physical properties by the following relationship [69]:

$$c = \sqrt{\frac{\gamma . P}{\rho}} \tag{4.1}$$

where P pressure of the gas, ρ is density of the gas and γ is the ratio of the specific heats, i.e C_p/C_v at constant pressure, p, and constant volume, v. Due to the increase in the speed of sound in the helium gas, being nearly three times higher than in air, voice distortion occurs, and speech becomes garbled, often referred to as the "Donald Duck" effect [65]. The effect is complex, resulting in an upward shift in the resonant frequencies of the speech signal except for the fundamental frequency of voiced sounds. Based on empirical and analytical results shown in Fig.4.1, Fant and Lindqvist [67] proposed the following equation to relate formant frequencies of speech signals in a heliox gas mixture

$$f_h^{\,2} = k^2 f_a^{\,2} + k^2 \left(\frac{\rho_h}{\rho_a} - 1 \right) f_o^{\,2} \tag{4.2}$$

where

f_h is formant frequency in the heliox mixture,

k is c_h/c_a where c_a=*340m/s,* the speed of sound in air at 1 bar, and c_h=*967 m/s,* the speed of sound in heliox mixture

f_a is associated the formant frequency in air at 1 bar,

ρ_h is density of heliox mixture at a pressure of 0.179 kg/m^3.

ρ_a is density of air at 1 bar, 1.27 kg/m^3.

f_o is the closed tract resonance in air at 1 bar and should be constant for a given individual and varies from 150 to 180 Hz.

In order to compensate for the effect of the heliox environment and ensure intelligible speech, a number of researchers have studied this issue [70-74]. Several algorithms have been developed to correct only the formant frequencies of the helium speech,

implementing both time domain processing [70] and frequency domain processing [71]. However, since it is not within the scope of this study, helium speech correction has not been attempted. Nevertheless, with the aid of DSP hardware, it is possible to implement any of the helium speech enhancement algorithms derived from Eq.4.2 and this should be included in future designs.

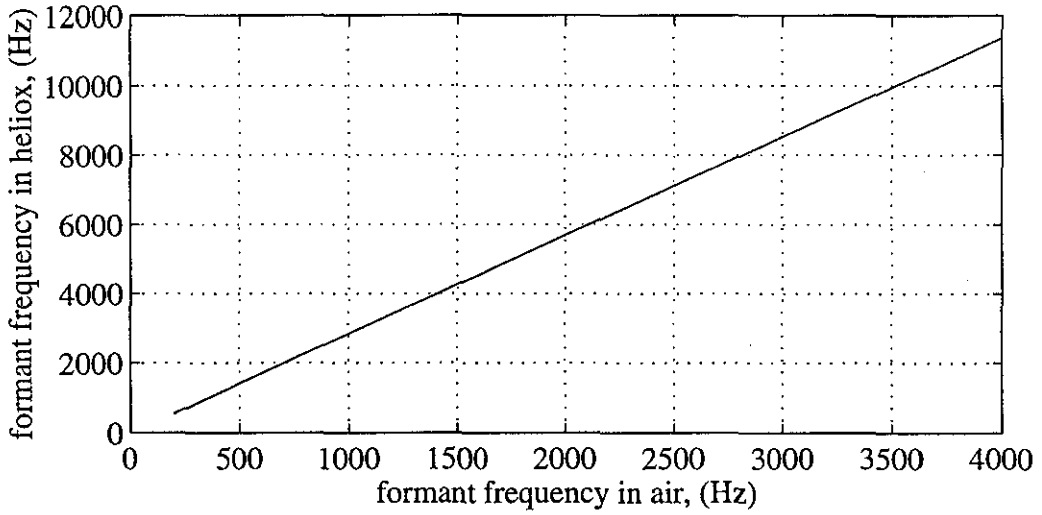


Figure 4.1 Formant frequency in air and in heliox mixture

4.2.3 Effects of Diving Mask on Speech Signal

One of the main problems associated with underwater speech communication between divers is the production of recognisable sounds while wearing SCUBA equipment. It is virtually impossible to form words when speaking directly into water (when the mouthpiece is removed); utilising an aural mask or using a full-face mask allows divers to speak into an air cavity. As anticipated, the breathing apparatus limits the movement of the lips and tongue, making clear pronunciation difficult. The precise way in which mask is worn, particularly regarding the force with which the webbing holds the mask against the diver's face, also limits speech production. In addition, there is the effect of the acoustic nature of the helmet or mask cavity which may introduce alteration in the speech signal spectrum.

It has been demonstrated that speaking in a non-radiating enclosure, such as diving mask, can result in first formant shifts for speech [75, 76]. A possible explanation for

this involves two effects of wearing a mask. The first effect is that the addition of a non-radiating enclosure over the mouth effectively increases the length of the vocal tract by approximately 4 cm over the nominal length of 17 cm for a male. The effective lengthening of the portion of the vocal tract forward of the point of constriction would result in an increase of the first and second formant frequencies (F_1 and F_2) [77], as illustrated in Fig. 4.2.

The second effect is the restriction of jaw motion provided by the mask, a fact that has been commented by divers. The restriction of jaw movement by the presence of the mask means that the size of the mouth opening is reduced compared to when no mask is worn. Theoretically, this results in a lowering of the formant frequencies when the mask is worn [77]. Moreover, it is also reported that the presence of a mask results in increases in vowel duration, fundamental frequency and total energy. It is concluded that there is a frequency compression in the first formant frequency, as illustrated in Fig. 4.2, due to the reasons explained above. However, although many researchers have predicted that there is a formant frequency shift due to the mask, this does not always occur, and sometimes the formants becomes indistinct rather than shift in either direction[76]. In the case of wearing a diving mask, there is no useful radiation to the outside, and communication becomes dependent on the type of microphone placed inside the cavity. Morrow and Brouns [78] investigated the acoustics of a diving mask-cavity, of a similar size to the wet mask used for speech recording experiments in this study, by employing different microphones. It is reported that good speech intelligibility was obtained with a pressure-compensated microphone, essentially independent of cavity volume and shape. The use of this type of microphone close to the lips tends to be insensitive to speech waves reflected by the mask, as experimented in [79, 80].

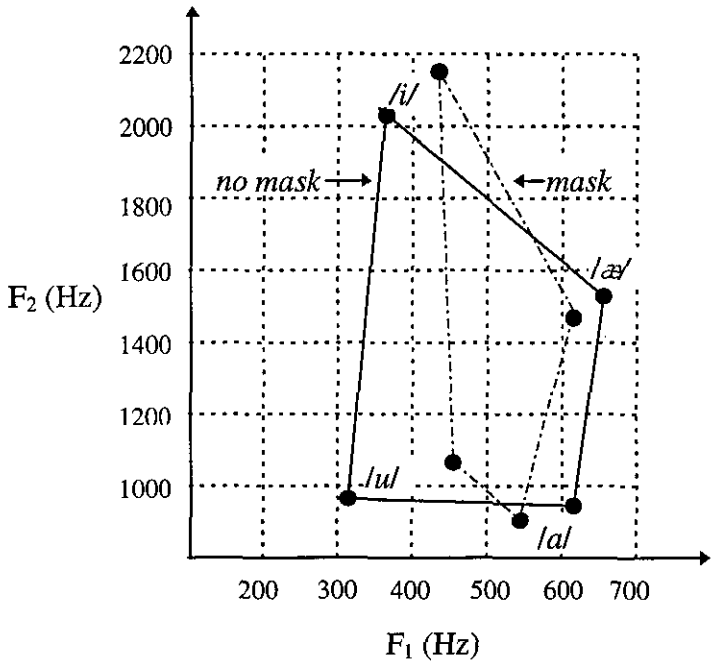


Figure 4.2. The shift in average centre frequencies of F_1 and F_2 for four vowels, for the phonemes of */i/, /æ/, /a/* and */u/,* as a function of wearing a mask [77].

4.3. Response of a Diver Mask on Surface and Underwater

Although in this study, any effects of the diver's mask on speech signal are not aimed to be compensated, it is important to examine the distortion introduced by the mask so that useful information can be provided for the future development of the system. However, depending on the results of this analysis, a digital filter to compensate the mask response may be implemented.

For this purpose, four different full-face masks were used: *wet mask, Aga, EXO-26* and *Aqua Lung,* each with a CF-2949 microphone, as shown in Fig.4.3. Speech signals were recorded while the masks were worn, and later processed. The response of a mask was easier to ascertain when the diver was on the surface, because of the absence of external noise sources. To carry out the response test, the diver read a sentence and list of words from Table 4.1, with and without the mask. As illustrated in Fig. 4.4, the waveform and spectrogram of the speech signals suggest that there is an increase in the energy in the low frequency band, i.e. lower than 1000 Hz. There are two possible explonation for this: The first is that since the diver hardly hears his own voice due to the mask, i.e.

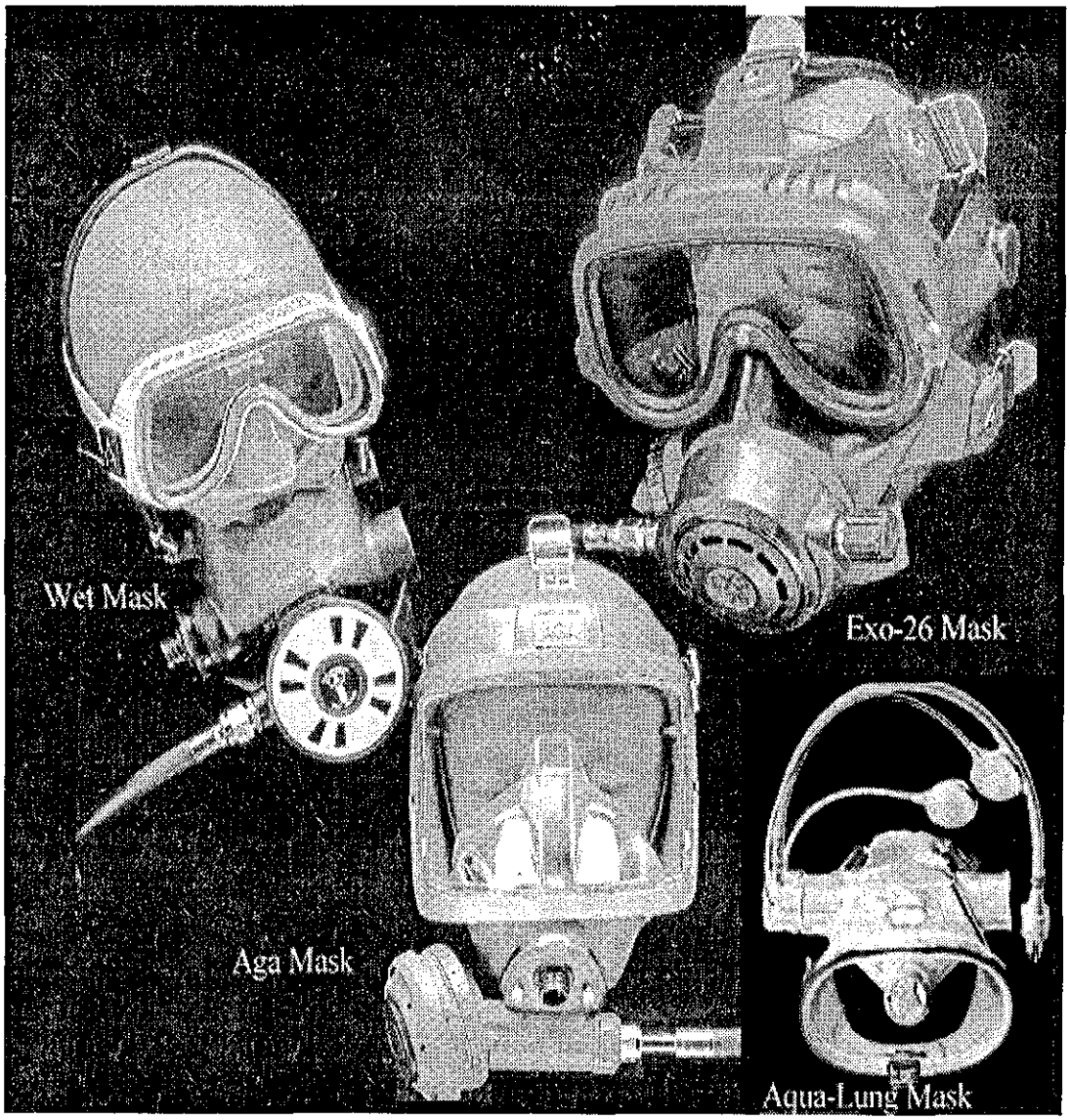


Figure 4.3 Illustration of divers' mask used during underwater speech processing

limited feedback is available, it is difficult to control his speech intensity, hence, he increased his speech intensity. The other is that since the the diver speaks into a non-radiating small cavity, short-time arrival of echos contributes amplification in speech intensity. In order to estimate the response of the mask, voiced parts of the speech signals were analysed in frequency domain by using the MATLAB programming language. In the frequency domain analysis, the response of the mask may be defined as:

$$H(f) = \frac{Y(f)}{S(f)} \tag{4.3}$$

where $H(f)$ is the response of the mask, $Y(f)$ represents frequency spectrum voiced speech sounds with the mask on and $S(f)$ is frequency spectrum of the same voiced sounds without the mask. The frequency spectra of these signals are calculated as explained in Section 4.5.3.1 and frequency response of the mask cavity is estimated as illustrated in Fig. 4.5. This suggests that with the specified mask the frequency band of approximately 2000-3200 Hz, i.e. F3 and higher formant frequencies, is attenuated due to the standing wave phenomena inside the mask. However, at the lower frequencies amplification is introduced because of the resonances of the mask cavity. These resonances will occur approximately at 137 Hz, 276 Hz, 448 Hz, 586 Hz, 758 Hz and 862 Hz. As the high frequency band relates to unvoiced sounds and the low frequency band relates to voiced sounds, the speech becomes bass heavy, as described in [81], which is confirmed by listening to the recorded speech. Moreover, from the experiments conducted with the mask, there is no evidence to prove that formant frequency shifts occur.

Once the response of the mask cavity is estimated, a digital filter with fixed coefficients may be designed in order to compensate for the effect of the cavity. Due to the lack of time, an optimum filter design has not been achieved. However, the results show that this is possible and may be implemented with a further study.

SUD	SUM	SUB	SUN
FILL	FIG	FIN	FIZZ
BUST	JUST	RUST	GUST
HIP	RIP	TIP	DIP
LED	SHED	RED	WED
PEAK	PEAS	PEAL	PEACE
DONE	DUD	DUNG	DUB
TEN	PEN	DEN	HEN
DIG	DIN	DID	DIN
BAT	BATCH	BASH	BASS
TEETHE	TEER	TEASE	TEEL
WE'RE	WEAL	WEAVE	WED
WILL	HILL	KILL	TILL
PICK	PIT	PIP	PIG
MAT	MAD	MATH	MAN
RENT	BENT	WENT	DENT
KICK	CHICK	THICK	PICK
TOP	HOP	POP	COP
BARK	DARK	MARK	LARK
ZIP	LIP	NIP	GYP
SAD	SAT	SAG	SACK
LAWS	LONG	LOG	LODGE
NEST	BEST	REST	WEST
PIN	SIN	TIN	WIN
TAB	TAN	TAM	TANG
THEEE	DEE	LEE	KNEE
SHEATH	SHEAVE	SHEATHE	SHEEN
CUFF	CUB	CUT	CUP
LASS	LACK	LASS	LAUGH
SOLD	COLD	HOLD	TOLD
FELL	REEL	SEAL	ZEAL
TOSS	TAJ	TONG	TALKS
SING	SIP	SIN	SIT
GALE	PALE	TALE	BALE
SHAME	GAME	CAME	SAME
PUP	PUFF	PUB	PUCK
DUMP	DUB	DUTH	DUFF
DIG	WIG	BIG	RIG
PASS	PATH	PACK	PAD
HATH	HASH	HALF	HAVE
PEEL	FELL	ELL	HELL
WIG	WITH	WIT	WITCH
VIE	THY	FIE	THIGH
FIN	TIN	SHIN	KIN
MAT	FAT	VAT	THAT
BEIGE	BASE	BAYED	BATHE
THIN	TIN	CHIN	SHIN
LEAVE	LEIGE	LEACH	LEASH
WAY	MAY	GAY	THEY
YORE	GORE	WORE	LORE

Table 4.1 Word list used for intelligibility test [70].

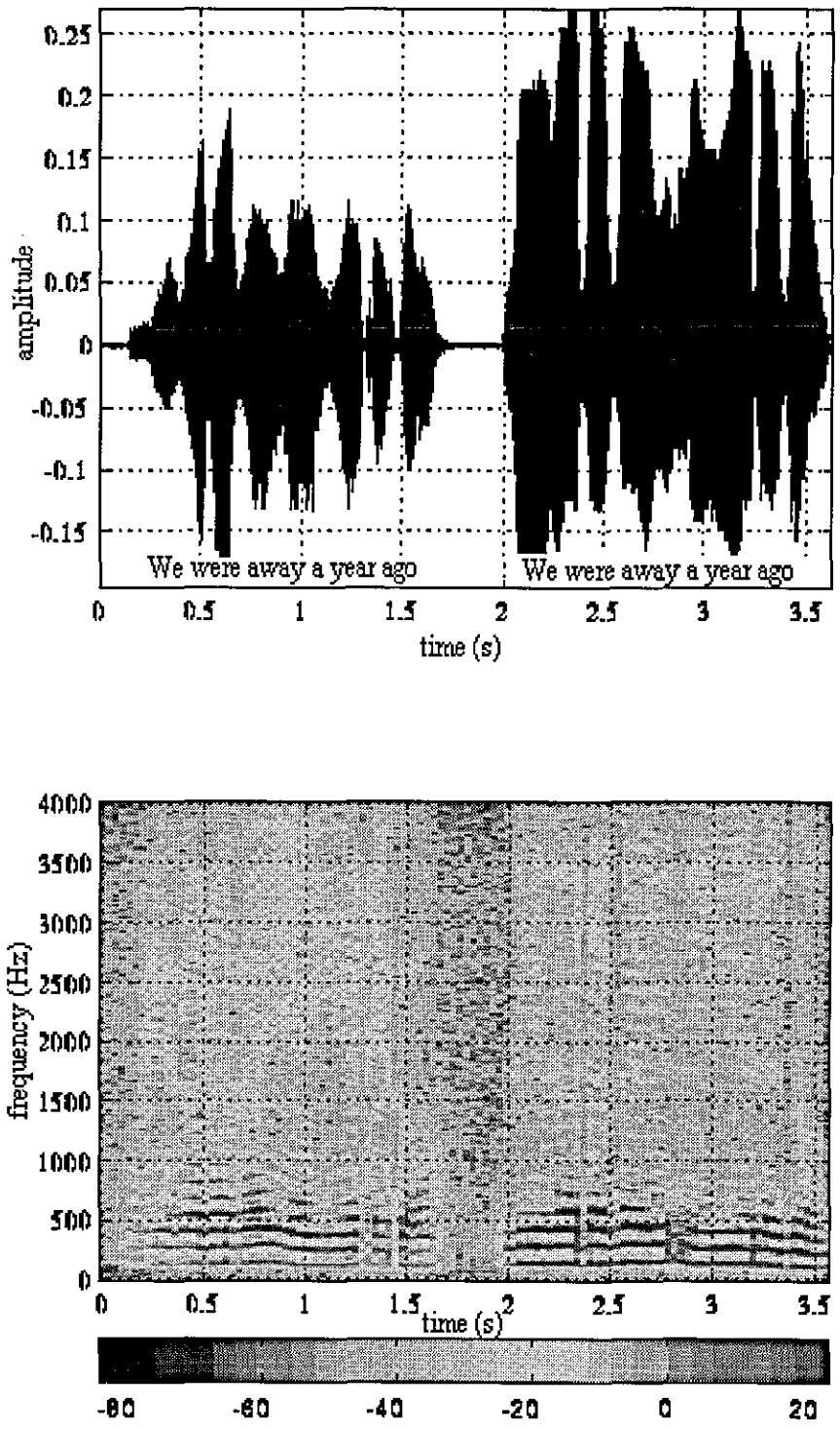


Figure 4.4 Speech waveforms and spectrogram (first part is without mask, second part is with mask) recorded on the surface (cont.).

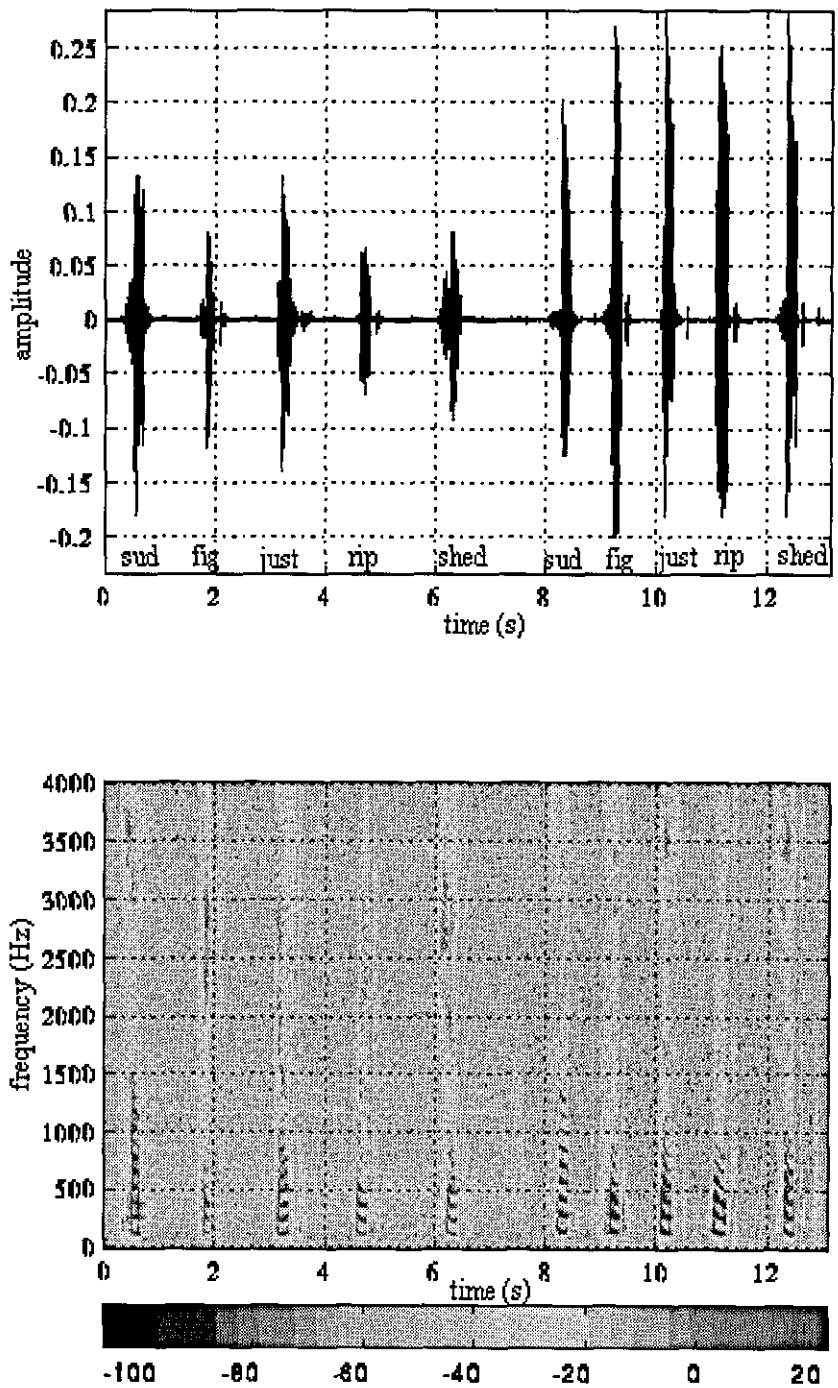


Figure 4.4 Speech waveforms and spectrogram (first part is without mask, second part is with mask) recorded on the surface.

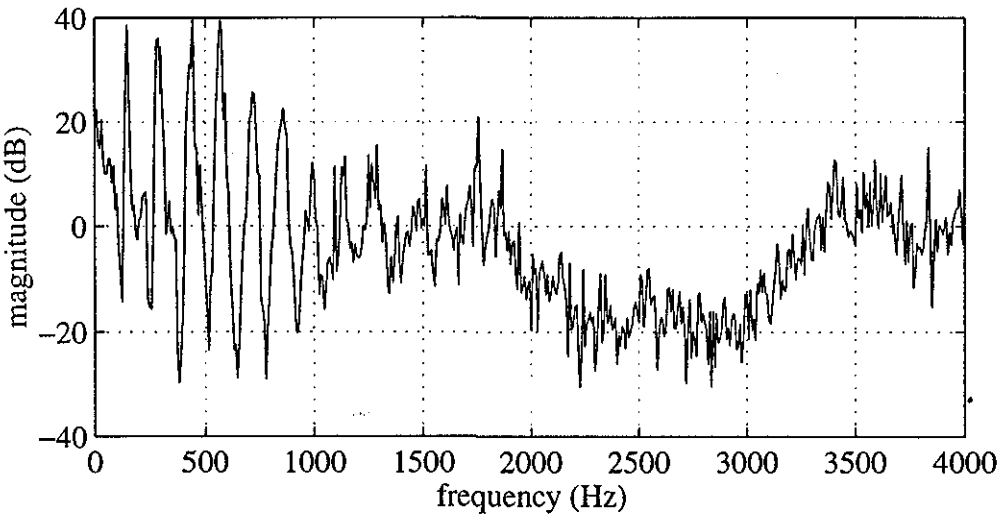


Figure 4.5 Response of a diver's mask on speech signals.

The results of the previous study were obtained with a diver wearing the mask on the surface. Since divers perform underwater they are subjected to high ambient pressure and this will affect the speech signal, as described earlier in this chapter. Therefore, it is essential to observe the effect of the mask on speech signals while the diver is underwater, although the speech will be distorted by external noises. For this experiment three full-face masks, *Aga*, *EXO-26* and *Aqua Lung*, provided by Graseby Dynamics Ltd, U.K., were worn by the diver, who was asked to read words from Table 4.1, first underwater then on the surface without the mask. The spectrograms of these speech signals are illustrated in Fig. 4.6 and it is obvious that the speech signal spectrum is different in free space from that underwater. However, the change in the spectrum is not only due to the mask cavity but also due to the combination of ambient pressure, bubble noise and the difficulty of producing speech underwater.

Although the analysis suggests that there are variations in speech spectra compared to free space, when listening to speech signals from these mask, the intelligibility of the speech is preserved in spite of bubble noise. This is due to the efficient performance of the CF 2949 noise-cancelling electret microphone used in the mask.

4.4 Underwater Hearing

It is obvious that human speech communication involves a listener and the hearing mechanism of that listener, which is a limiting factor in such communication. Having prior knowledge of underwater hearing thresholds is important since this defines the minimum output power requirement of any speech communication system that requires the listener to be directly in the water. While much is known about the limitations of the ear in air, little is known about its limitations underwater. The audibility threshold performance of the human ear in water has been experimented in [82, 83] and it was concluded that the underwater hearing thresholds are not affected by the ambient water pressure, are from 18 dB to 56 dB higher than the air-conduction thresholds, the difference increasing with frequency. One of the general conclusions drawn from the studies [84] is that the middle-ear mechanism is not used in underwater hearing but is bypassed by the mechanism of bone conduction. Therefore, currently available underwater voice communication systems utilise piezo-ceramic bone conduction hearing devices. Although the system developed here has not been tested extensively underwater, the same type of listening devices have been employed as would be used underwater.

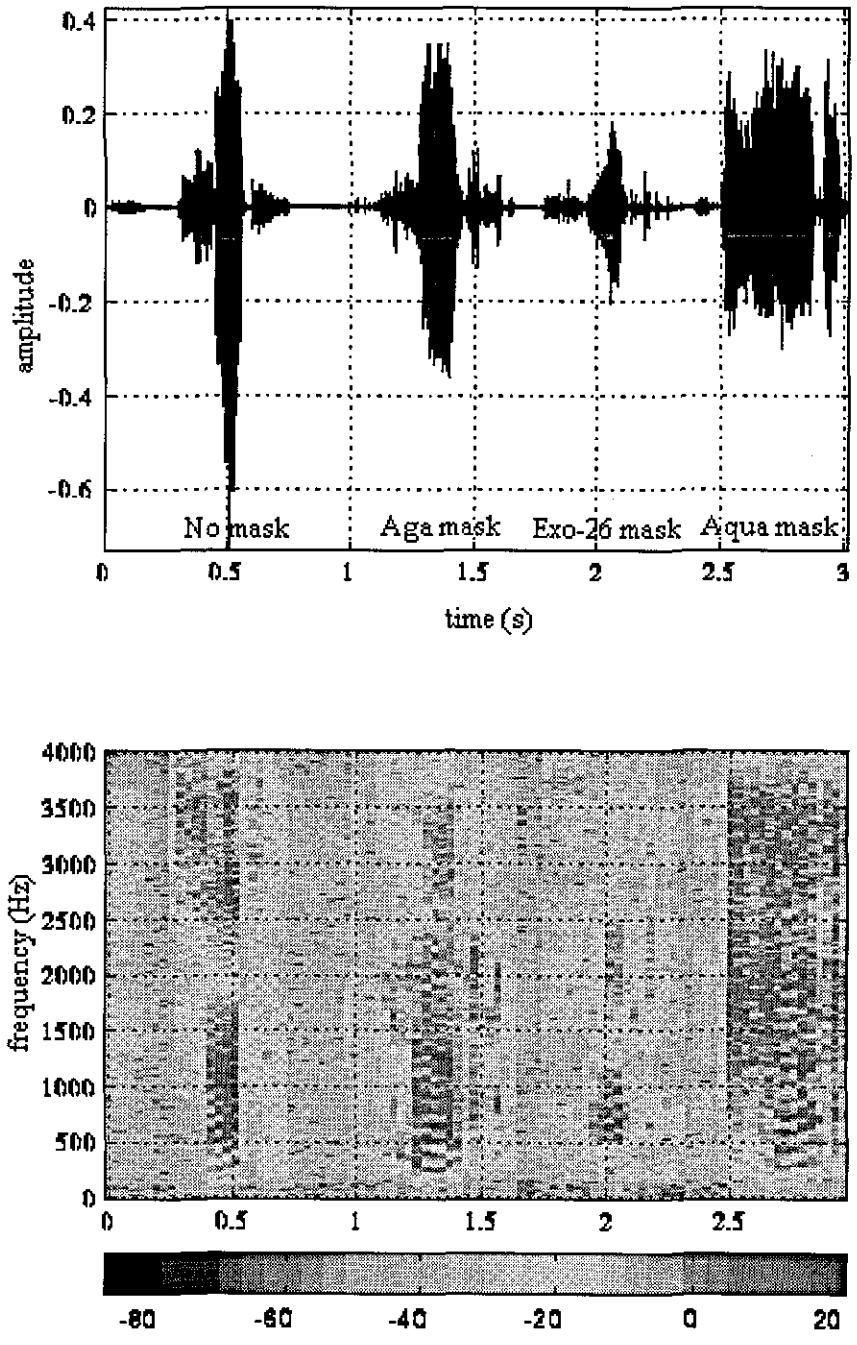


Figure 4.6 Speech waveforms (sud) recorded underwater with different diving masks, together with their spectrogram(cont.).

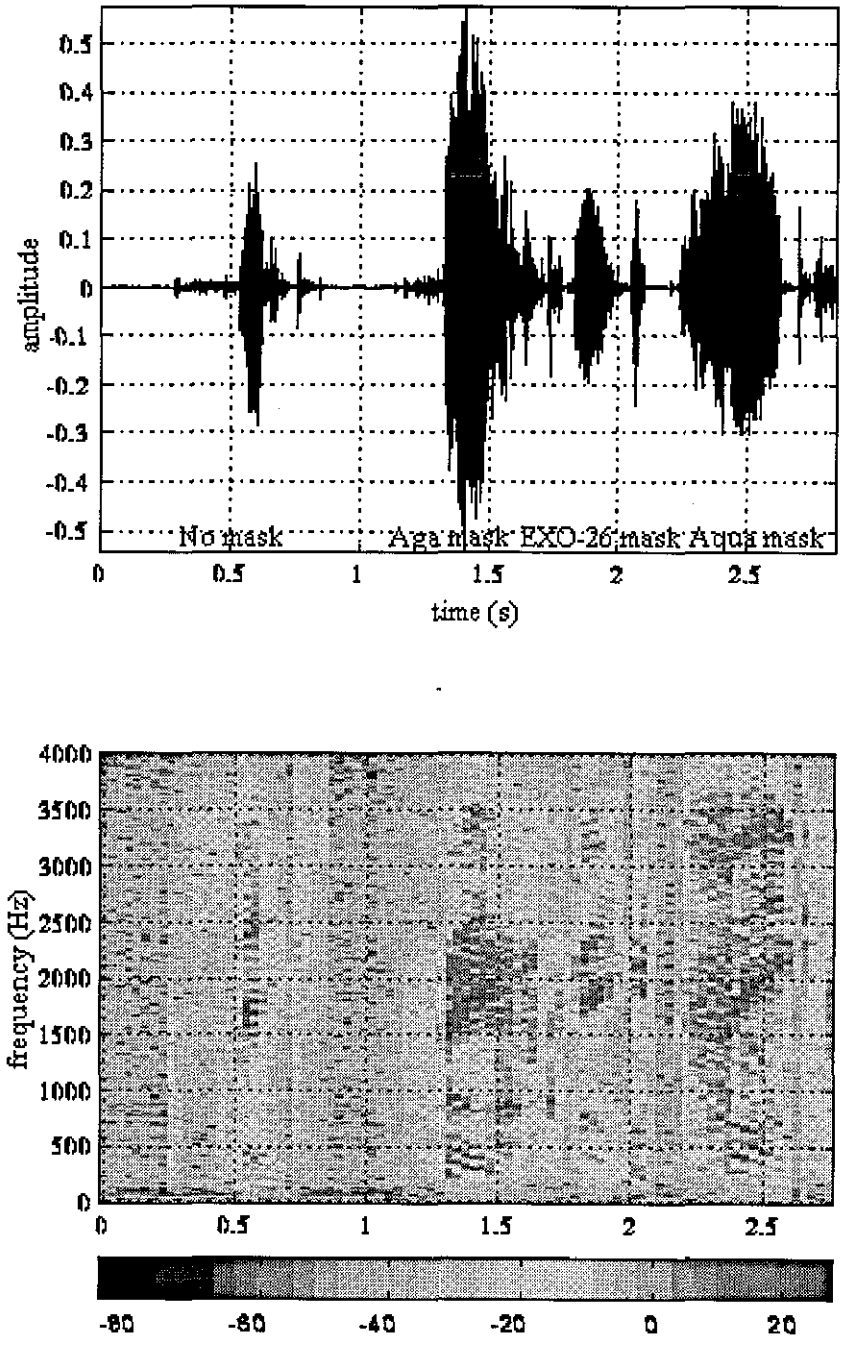


Figure 4.6 Speech waveforms (fig) recorded underwater with different diving masks, together with their spectrogram(cont.).

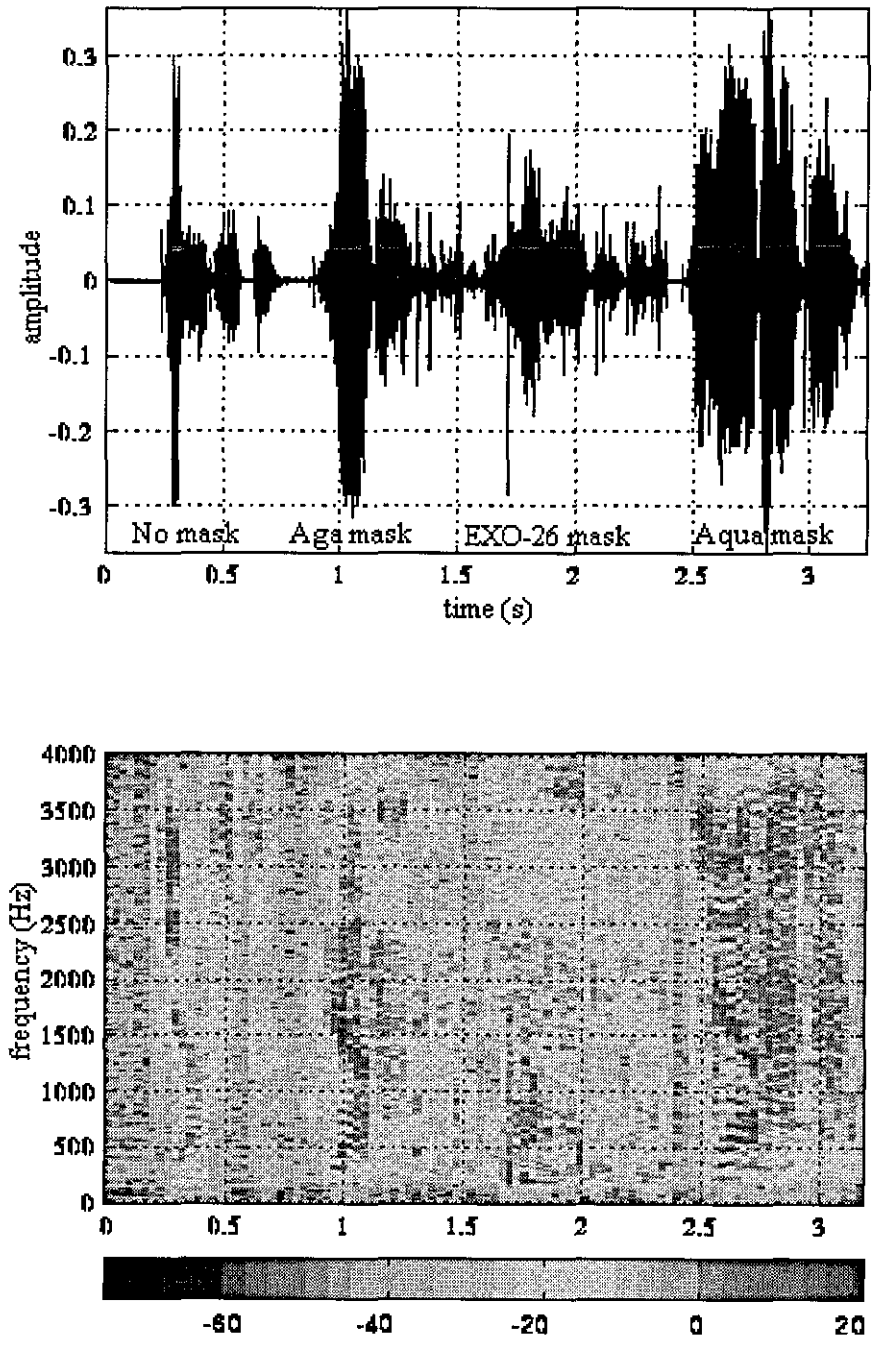


Figure 4.6 Speech waveforms (just) recorded underwater with different diving masks, together with their spectrogram.

4.5 Enhancement of Noise Corrupted Speech Signal Quality from a Diver Mask

In many speech communication systems, the presence of background interference causes the quality or intelligibility of speech to degrade. When a speaker and a listener communicate in a quiet environment, information exchange is easy and accurate. However, a noisy environment reduces the efficiency of communications.

For effective through-water voice communications between divers, it is also essential that a clear, intelligible speech signal is transmitted. However, in a diving environment, the speech is accompanied by extraneous noise, such as that from various items of machinery distributed throughout an underwater work site. Moreover, there are two types of significant noise affecting speech quality which should be eliminated before transmission to improve intelligibility at the receiver. One is breathing noise, produced during inhalation by air flow through the regulator of a conventional SCUBA (**S**elf **C**ontained **U**nderwater **B**reathing **A**pparatus) or by free-flowing air through a helmet or full-face mask. Due to its high signal amplitude, the breathing noise has a significant effect on the clarity of communications. The other is bubble noise generated during exhalation, including during speech periods, by escaping air from the regulator. Although the bubble noise occurs outside the mask cavity and can be detected by a microphone inside the mask it is also a problem and will be considered in detail here. A scheme which is highly desirable is to reduce or cancel bubble and breathing noise whilst leaving the diver's speech uncorrupted and undistorted.

4.5.1 Speech recording and pre-processing

For speech recording trials, four diving masks, as shown in Fig. 4.3, and their associated breathing regulators (demand valves) were used by two divers submerged in a laboratory tank (9m long, 5 m wide, 2m deep). The CF-2949 and tape recorder (Nagra type III) were used for recording. To process the recorded speech signals, a SUN work station is utilised and speech signals are loaded through its audio input port, which includes a low pass filter with a cut-off frequency of 3400 Hz and an

Analogue-to-Digital Convertor operating at a sampling rate of 8000 Hz with 8-bit resolution.

4.5.2 Breathing noise cancellation

When the human speech production system is studied, as is well known, no speech sounds are produced during inhalation period. It is evident that the speech signal, the breathing noise and the bubble noise can be considered to be mutually exclusive in time, although the speech signal includes some bubble noise because the diver exhales small amounts of air while speaking. Therefore, it is possible that the speech signals and breathing noise can be distinguished by employing an appropriate signal processing technique. To cancel breathing noise, several analogue noise suppression methods have been applied in *hardwired* communication systems [81, 85]. In the system described in [81], the input signal is passed through a band pass filter tuned to the central frequency of the breathing noise then a rectifier and integrator. The output is then compared to a pre-set threshold. When the input signal energy level exceeds the threshold level, i.e. breathing noise, an attenuater is activated and the signal is attenuated. This mode of operation obviously depends on the breathing noise being louder than any speech signal. Therefore, the threshold level must be set above the normal speech level but below the breathing noise levels. Moreover, it is also reported that the breathing noise magnitude and its frequency spectrum vary during a dive according to the ambient pressure and also the pressure of breathing air in the cylinder, hence affecting the system's performance. Another system, implemented in [85], approaches the breathing noise cancellation by employing two microphones; one is the speech microphone, the other is the noise microphone placed inside demand valve. The noise microphone, encapsulated so as not to pick up speech and bubble noise, monitors whether the diver breaths in or not. When the diver inhales, the noise microphone output activates an attenuater similar to the one previously explained. Implementation of these methods requires modification of the diver's mask with extra hardware. Since the system introduced in this study employs digital signal processing methods for speech coding, a digital noise cancelling technique is recommended. In the following, we describe how breathing noise can be cancelled by implementing a

combination of two techniques, leading to the achievement of better communication quality.

In order to cancel breathing noise, it is important firstly to find out about its characteristics so that a suitable method can be implemented. For this purpose, breathing noise shown in Fig.4.7 was recorded during diving trials using *wet mask*, *EXO-26*, *Aga* and *Aqua Lung* masks. When processed, the noise amplitudes appears to be Laplacian distributed, as shown in Fig.4.8. As illustrated the frequencies of these signals are predominant in the range of 1000 Hz to 4000 Hz, hence these spectral differences suggest a method of cancelling the breathing noise.

The same bandwidth corresponds to unvoiced sounds in the speech signals. This implies that a method used for estimation of unvoiced sound intervals may be implemented to recognize breathing noise. Measurement of zero crossings in a frame of speech signal is a common method of unvoiced sound recognition [46], therefore implementation of this technique may be considered. Furthermore, as defined earlier, during the inhalation period, a diver consumes at least 15 litres of air per minute and the velocity of air in the demand valve is very high, generating high sound pressure levels inside the mask cavity, hence the energy magnitude of the breathing noise becomes relatively high compared with that of the speech signal. This suggests that energy measurement of the input signal may also provide sufficient information about breathing noise. However, implementation of this for noise cancellation is not a robust enough method since it may intercept with voiced sounds having higher energy measurements than unvoiced sounds.

For discriminating breathing noise intervals from the speech signal, neither zero crossing nor signal energy itself provides an adequate solution. In the case of using only zero crossing measurements for this purpose, unvoiced sounds are also recognized as breathing noise. Additionally, when energy measurement is employed by itself, voiced sounds may be also identified as breathing noise. However, a combination of these two techniques may be applied to separate the wanted speech from the unwanted noise. Before measuring these two parameters, since the breathing noise occupies a high speech frequency band as defined earlier, the input signal is fed

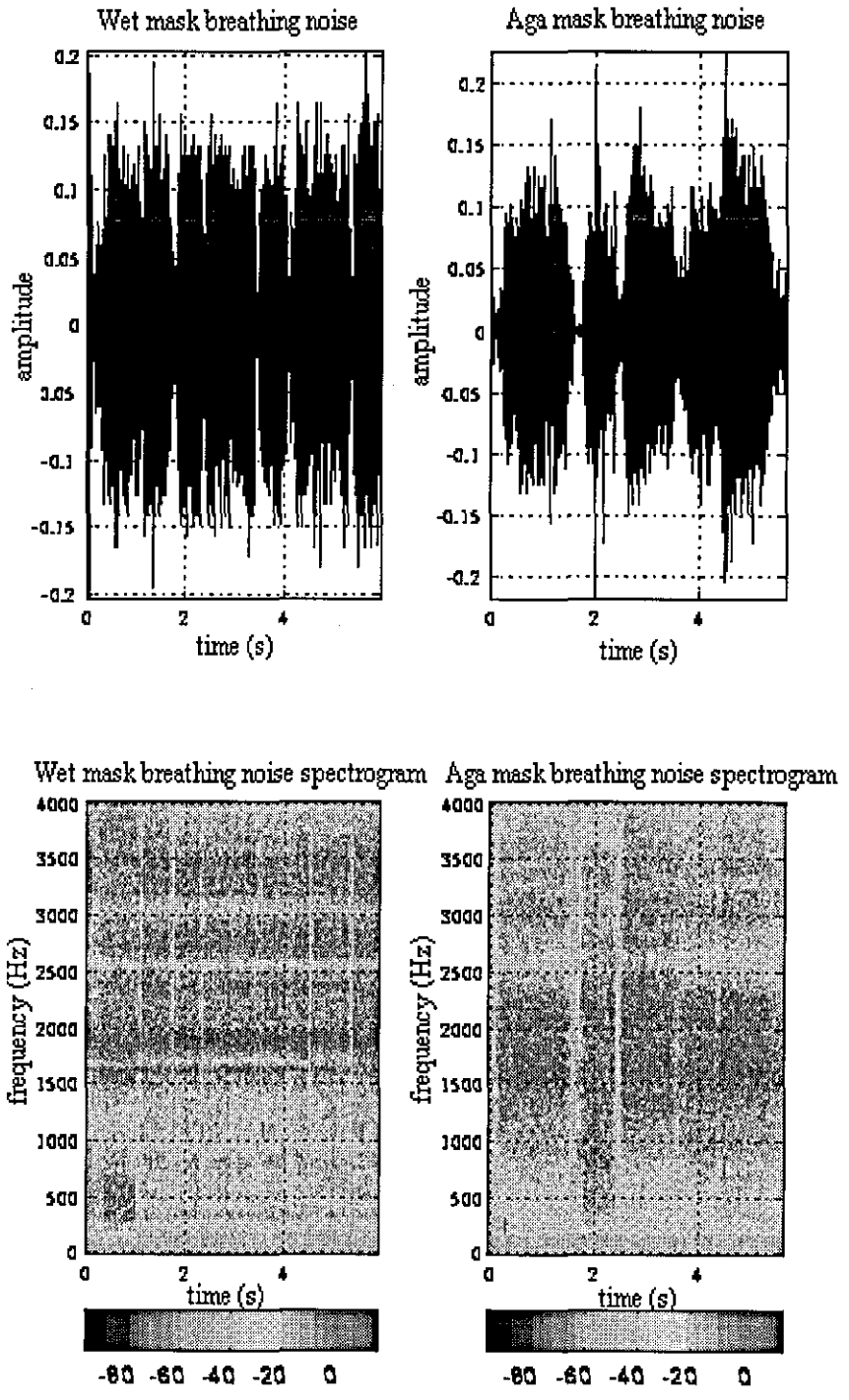


Figure 4.7 Illustration of breathing noises and spectrograms from divers' masks (cont.)

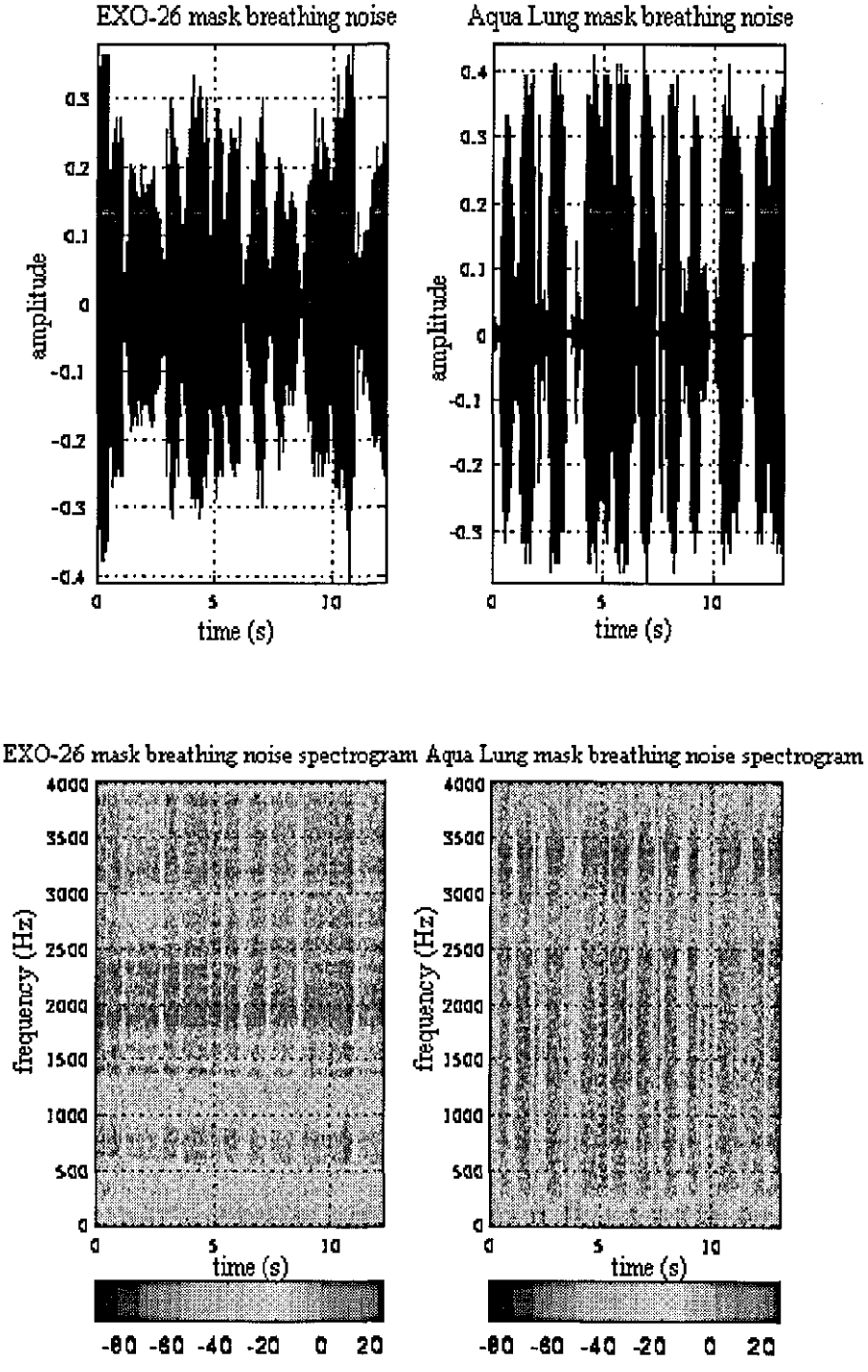


Figure 4.7 Illustration of breathing noises and spectrograms from divers' masks.

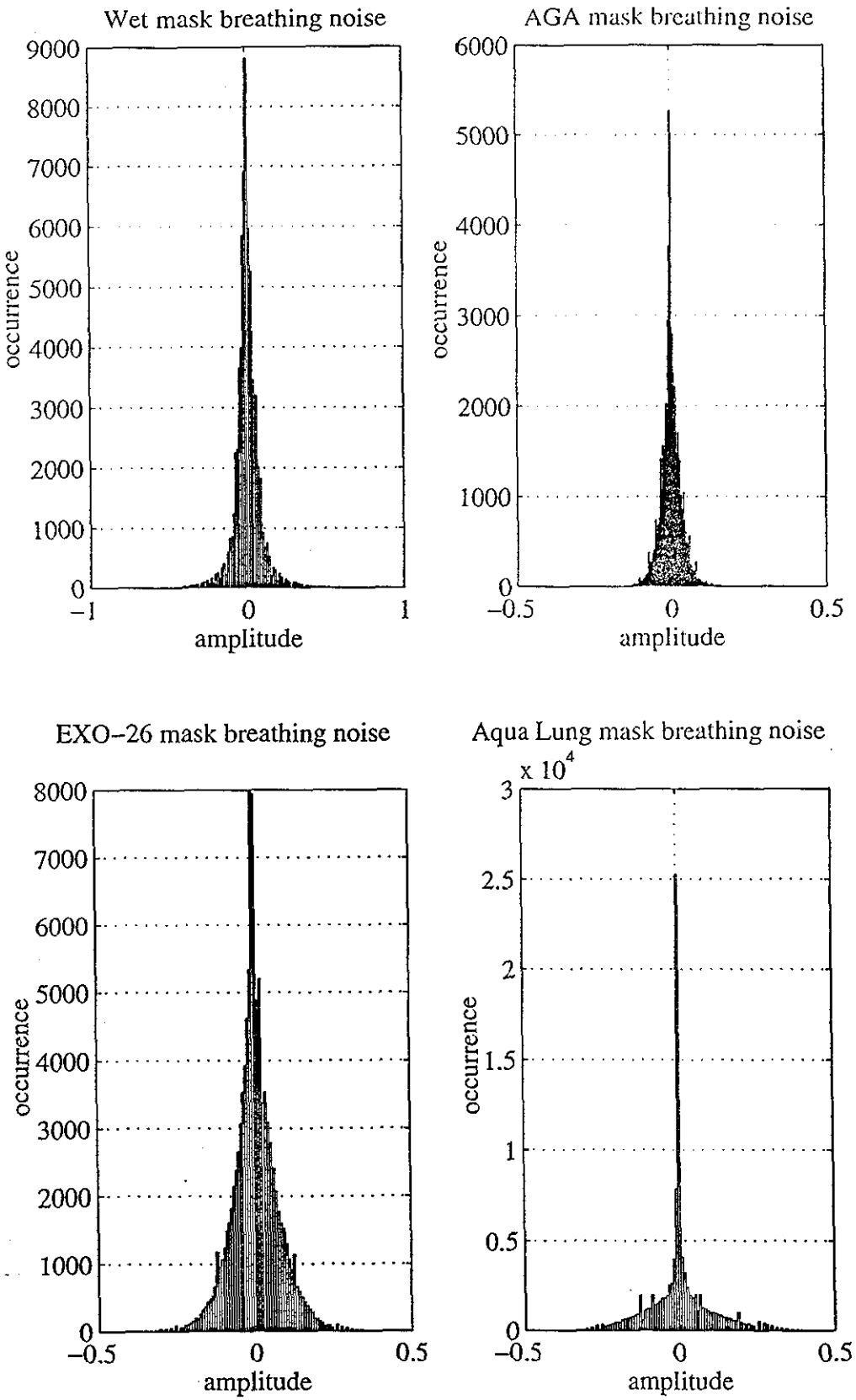


Figure 4.8 Amplitude distributions of breathing noise signals from divers' masks.

to a bandpass filter with a frequency band of 1000 - 4000 Hz. For this purpose, a 20-tap FIR structure bandpass filter was designed and employed. Hence, voiced sounds are attenuated and detection errors introduced by them are minimized.

In order to distinguish breathing noise from speech signals or bubble noise, decision levels for the above parameters must be defined. Therefore, the noise signals illustrated in Fig. 4.8 are processed to find appropriate parameters after passing through the bandpass filter. After that, zero crossing, *ZC(k)* and energy, *E(k)*, measurements of these noises are calculated by splitting them into 22.5 ms frames, as defined in Eq. 4.4 and Eq. 4.5 respectively,.

$$ZC(k) = \sum_{n=kN+1}^{(k+1)N-1} \left| sign(y(n)) - sign(y(n-1)) \right| \tag{4.4}$$

where

$$sign(y(n)) = \begin{cases} 1 \ \ if \ y(n) \geq 0 \\ -1 \ \ if \ y(n) < 0 \end{cases}$$

and

$$E(k) = \sum_{n=kN}^{(k+1)N-1} y(n)^2 \tag{4.5}$$

where $y(n)$ is the input signal, N is the number of samples (180) in a frame, and k is the number of frames.

Choosing appropriate threshold values is an important factor for successful cancellation of breathing noise. If they are inappropriately defined, there is a risk of misidentifying the voiced or unvoiced sounds as breathing noise. In order to estimate threshold levels for zero crossing and energy measurements, statistical analysis is applied. The distribution of these measurements are found, as illustrated in Fig.4.9, for zero crossings and energy measurements, both assumed to be Gaussian-like

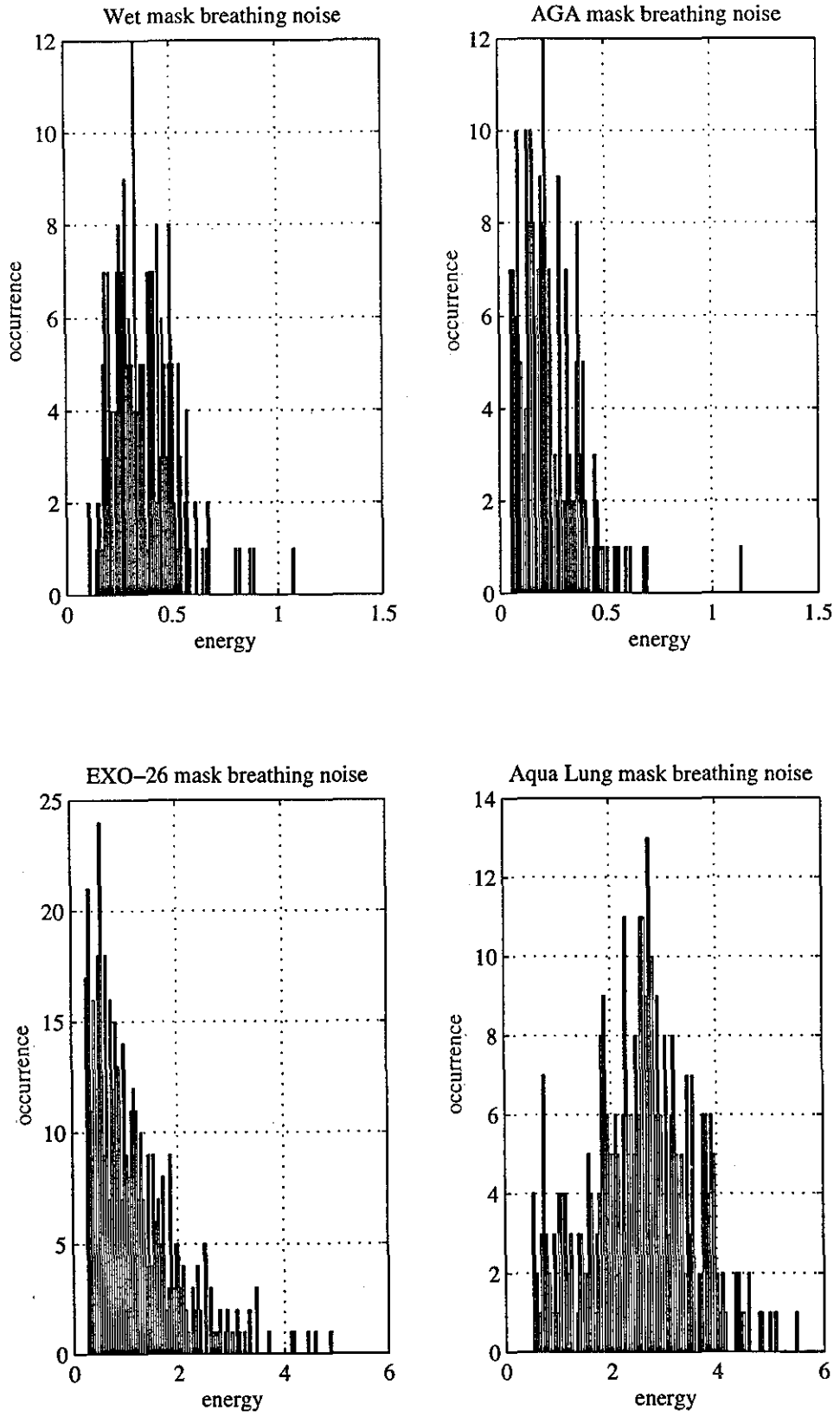


Figure 4.9 Distributions of energy and zero crossing measurements for breathing noise signals (cont.)

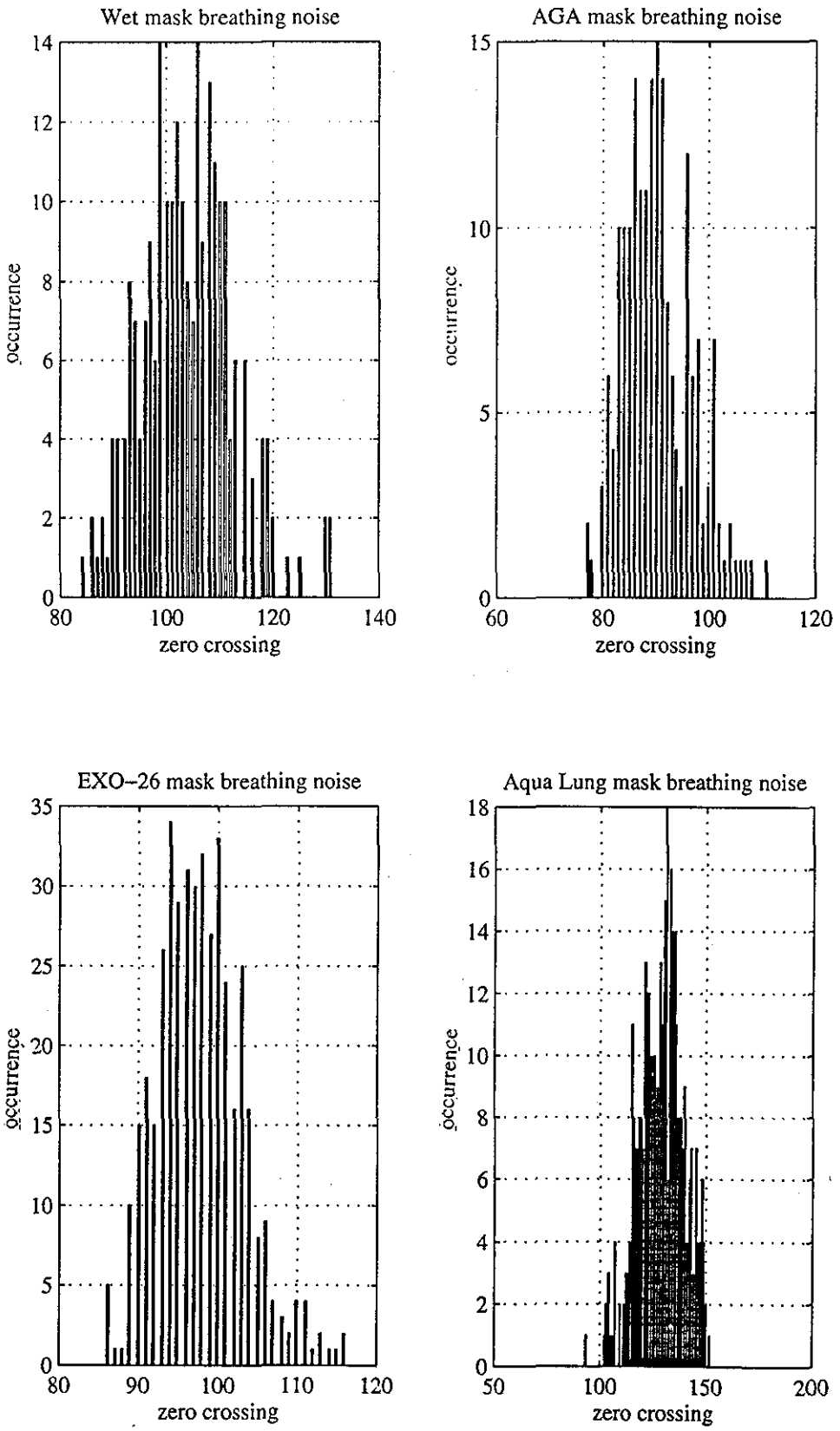


Figure 4.9 Distributions of energy and zero crossing measurements for breathing noise signals

distribution although the latter is more likely Rayleigh distribution. Then the threshold values for these measurements for the *wet* mask, *Aga*, *EXO-26* and *Aqua Lung* masks are defined for a 90% acceptance region, as shown in Table 4.2.

Mask	*Measurements*				***Threshold***	
Type	*mean E*	*std E*	*mean ZC*	*std ZC*	***thr E***	***thr ZC***
Wet	0.3230	0.1281	103.923	8.5724	**0.1590**	**93**
Aga	0.1462	0.0924	90.8178	6.6431	**0.0279**	**82**
EXO-26	1.0446	0.6904	97.8298	5.4013	**0.1608**	**91**
Aqua Lung	2.2614	0.8491	128.7286	10.6116	**1.1747**	**115**

Table 4.2 Estimation of threshold levels for zero crossing and energy measurements

The breathing noise cancellation method is based on energy and zero crossing measurements, implemented as shown in Fig. 4.10, by processing speech signals shown in Fig.4.11. In order to illustrate operation of this method, speech signals were recorded during diving trials and shown in Fig.4.11, where Br.N and Bu.N, representing breathing and bubble noise respectively, are applied to the breathing noise canceller.

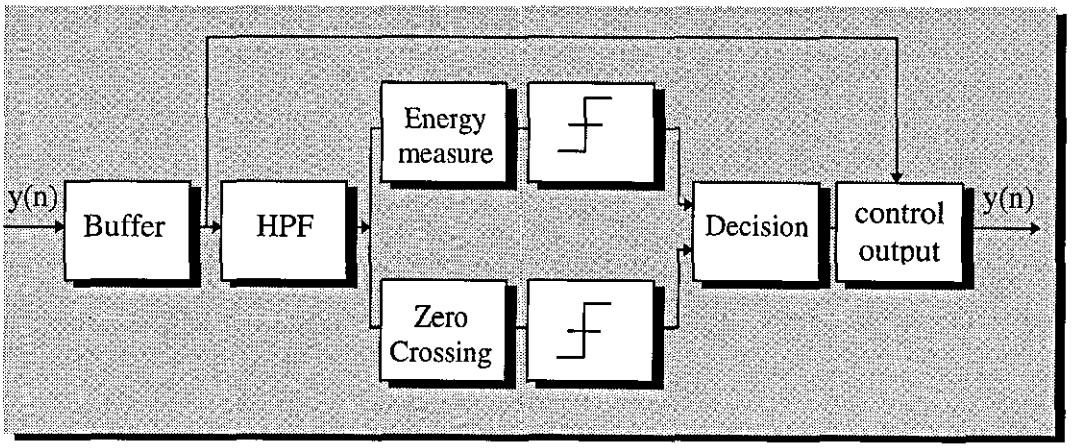


Figure 4.10 Breathing noise cancellation process

The speech signal is processed by splitting it into 22.5ms frames due to the reasons stated in Chapter 3. Frames of input speech signals are bandpass filtered so that the effect of voiced sounds in detection is minimized. This is an essential and beneficial process since the breathing noise spectrum occupies the frequency band of 1500 - 4000 Hz, as explained earlier. After that the zero crossing rates and energy are computed in each frame by applying Eq.4.4 and Eq.4.5 respectively. When both the energy and the zero crossing values for each speech frame are above or equivalent to the threshold levels, the frame is assumed to be breathing noise and the speech frame is set to zero (along with the following several input signal frames). When this method is implemented, breathing noise is cancelled, as illustrated in Fig.4.11.

The technique described here is well performed when utilized with the *wet mask*, *EXO-26* and *Aqua Lung* mask since the breathing noise signal magnitude is high. Nevertheless, when it is used with *Aga* mask, the breathing noise does not provide the condition for detection due to its low magnitude. This results in the misidentification of the speech signal and in the case of employing the cancellation algorithm it will cause an interruption in the processed speech. A possible explanation to this is air pressure in the air cylinder. Indeed, having used the same breathing equipment and the same size of air cavity as the other masks, the breathing noise magnitude is normally expected to be as high as in the other masks.

A breathing noise cancellation technique has been presented as a way of improving the subjective quality of a digital voice communication system used by divers. Tests have demonstrated that breathing noise can be cancelled by means of measuring the zero crossing rate and energy of the speech signal using carefully chosen threshold values. Further quality enhancement may be achieved by cancelling bubble noise.

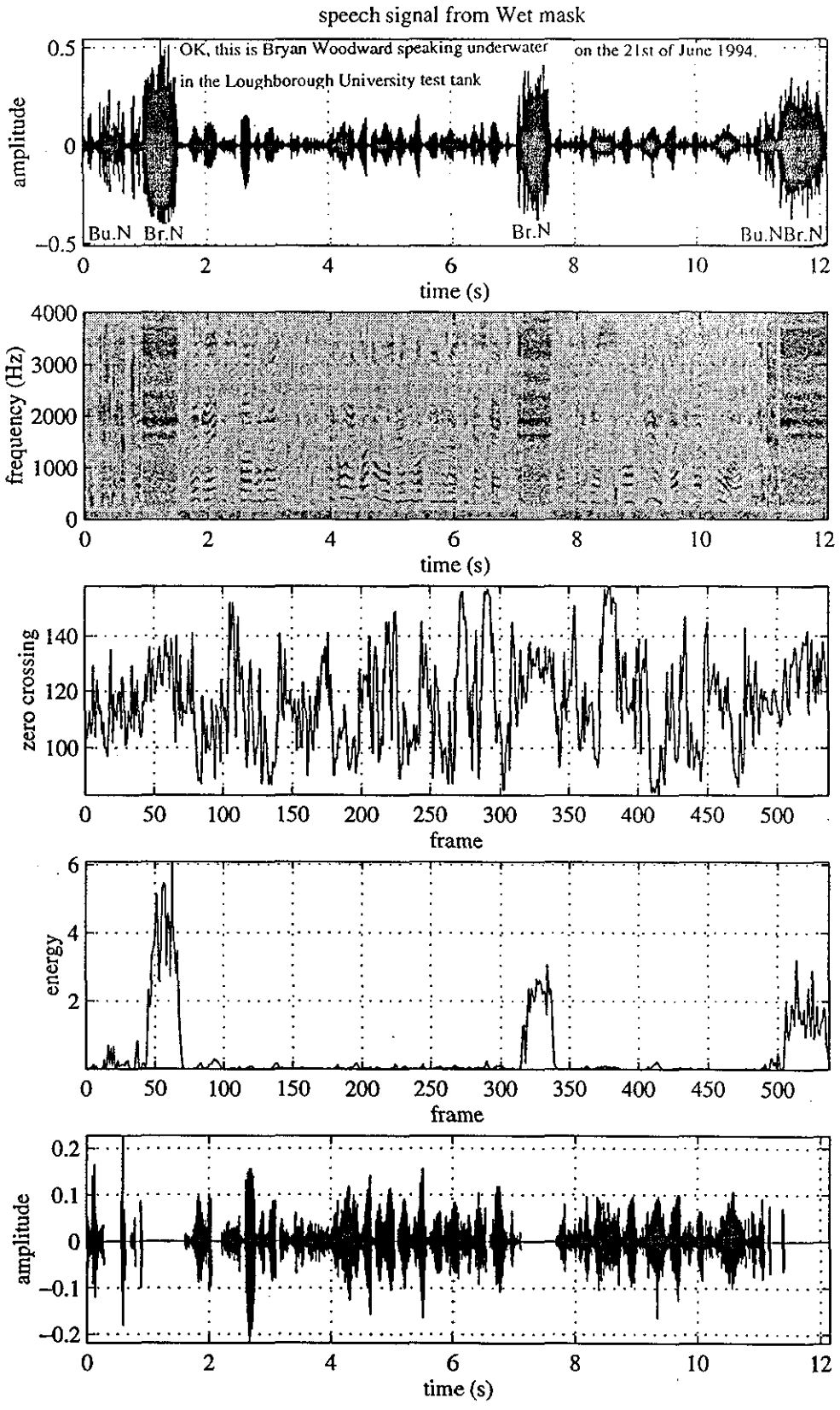


Figure 4.11 Illustration of breathing noise cancellation (cont.)

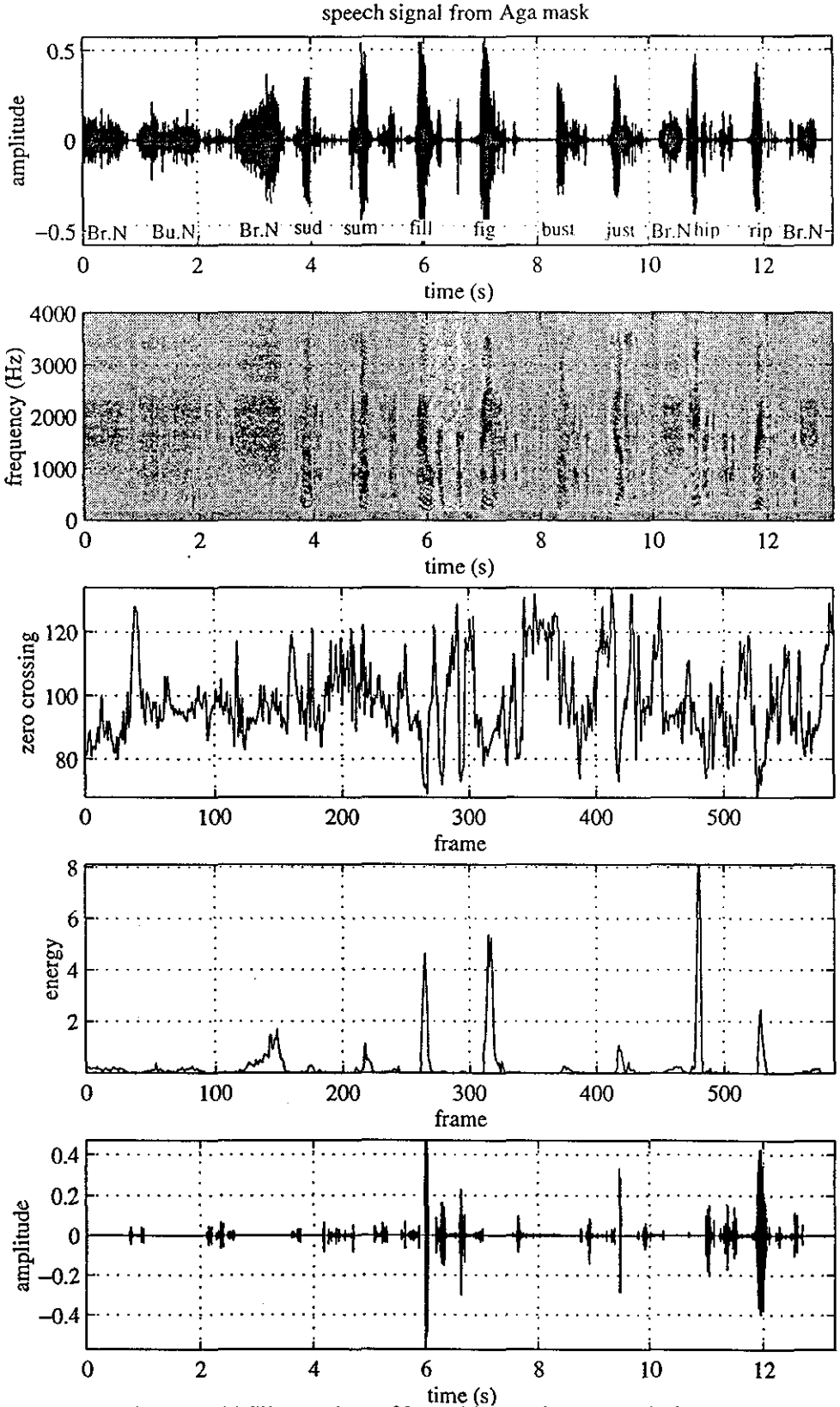


Figure 4.11 Illustration of breathing noise cancellation (cont.)

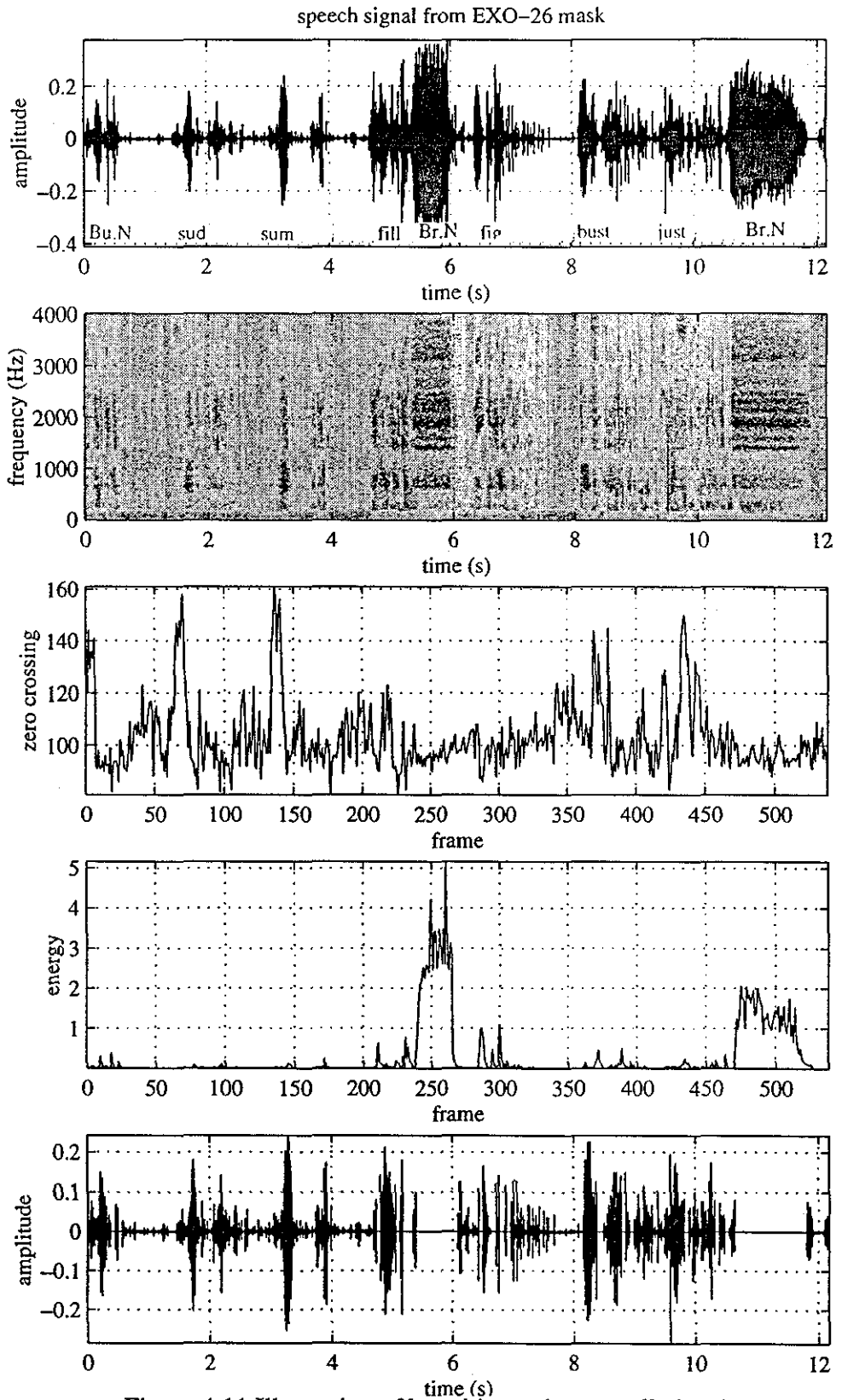


Figure 4.11 Illustration of breathing noise cancellation (cont.)

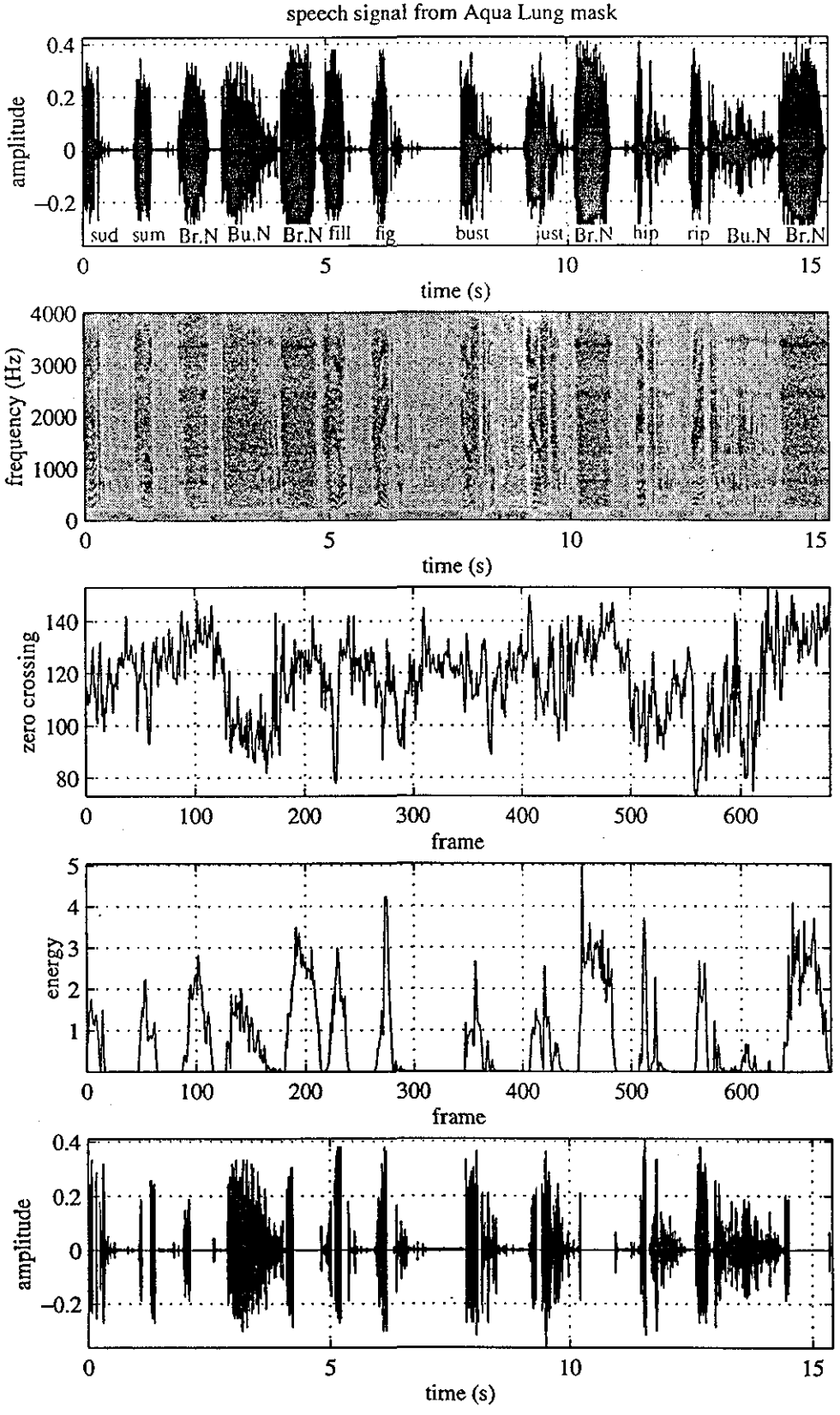


Figure 4.11 Illustration of breathing noise cancellation

4.5.3 Bubble Noise Cancellation

Background noise acoustically added to speech can degrade the performance of a digital underwater voice communication system and it must therefore be maintained at a level near that measured using noise-free input speech. In order to ensure continued reliability, the effects of background noise must be suppressed by applying noise cancelling algorithms or by employing a suitable microphone, capable of attenuating background noise or simply using a different type breathing apparatus, such as a re-breather. As mentioned earlier, bubble noise is generated due to the air escaping from the diver's demand valve when he is exhaling or speaking. Since a diver breathes high pressure air underwater and consumes more air during a dive than on the surface, for example at 30 m depth the air consumption will be typically 60 litre/min or more, this will increase the amount of air bubbles released into the water and bubble noise may have a more significant effect on speech signal.

Listening tests conducted on recorded speech show that the quality of speech signals significantly depends on the mask type utilised, e.g. location of the demand valve, and position of the microphone inside the mask. The closer the microphone to the demand valve, the more serious the bubble noise becomes. In this concept, the *Aga* mask offers much better speech quality performance. Owing to the design of the regulator and the microphone's position inside the mask, the bubble noise was hardly audible during tests. The *Aga* masks are designed for the purpose of improving the vision of a diver underwater. Therefore, the demand valve is placed on one side of the mask, as can be seen in Fig.4. 3 so that the air bubbles do not obstruct the vision of the diver. Another equipment-dependent solution for bubble noise cancellation would be the use of a re-breather system which recycles the exhaled air by a special process and does not release as much air outside the mask as in its conventional counterparts.

Moreover, during listening test of the speech signals, it was noticed that bubble noise affects the speech signal just after articulating voiced sounds. This is certainly due to the amount of air expelled by the diver. As mentioned earlier, in order to produce voiced sounds, the pressure of air leaving the mouth is higher than unvoiced sounds, hence producing higher signal energy. Therefore, we can describe bubble noise as non-

stationary and speech-dependent noise. To study bubble noise and observe its properties, bubble noise was "manually" extracted from speech sequences recorded while wearing the *wet mask*, *Aga*, *EXO-26* and *Aqua Lung* masks underwater. As illustrated in Fig.4.12 and Fig.4.13, bubble noise from these masks does not have dominant frequency components, as in breathing noise, and their amplitude distributions appear to be Laplacian. Since bubble noise and the speech signal occupy the same frequency band, the use of conventional filters with fixed coefficients for cancellation is infeasible.

The technique applied for breathing noise cancellation does not provide any spatial information about bubble noise so that it can be separated from speech signal. Therefore, it is necessary to seek alternative noise cancelling methods. There are two types successfully employed in speech signal processing applications. These are spectral subtraction and adaptive noise cancelling techniques [86]. In the following sections, we will examine the performance of these methods for bubble noise cancellation and the one performing best will be included in the underwater communication system. For all processing, the speech waveforms illustrated in Fig.4.11 are utilised.

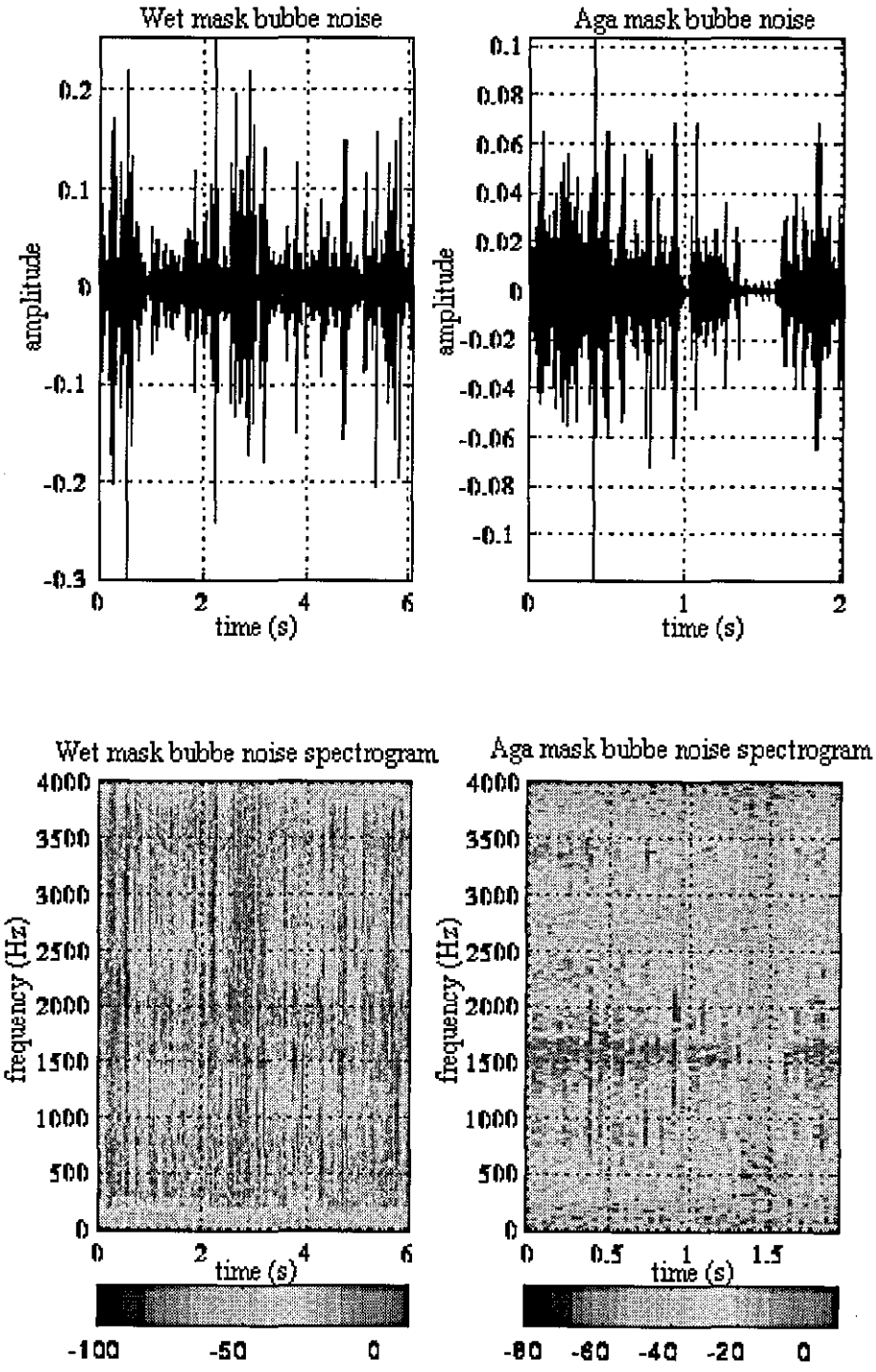


Figure 4.12 Bubble noise signals and spectrograms from different divers' mask (cont.).

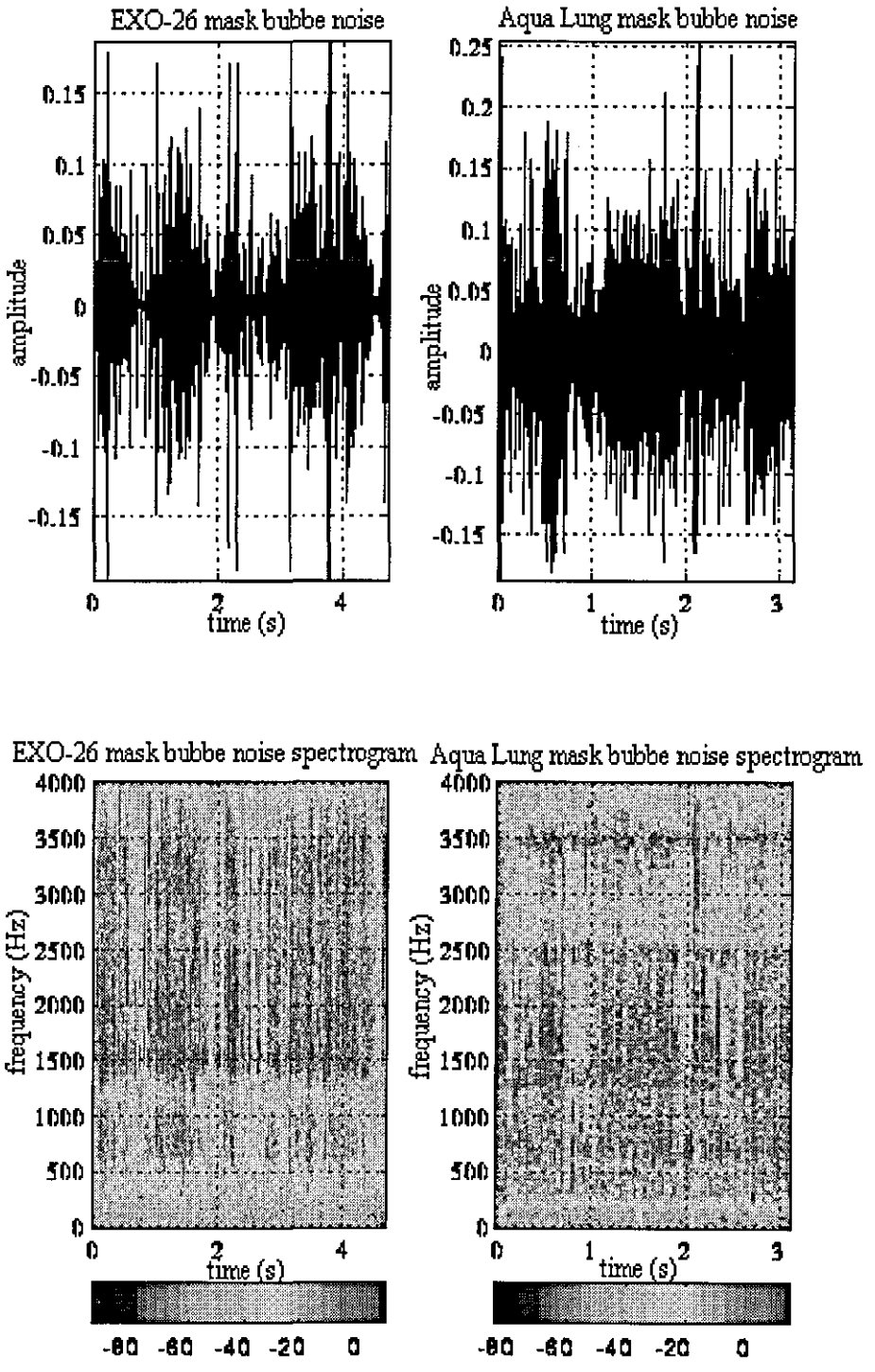


Figure 4.12 Bubble noise signals and spectrograms from different divers' mask.

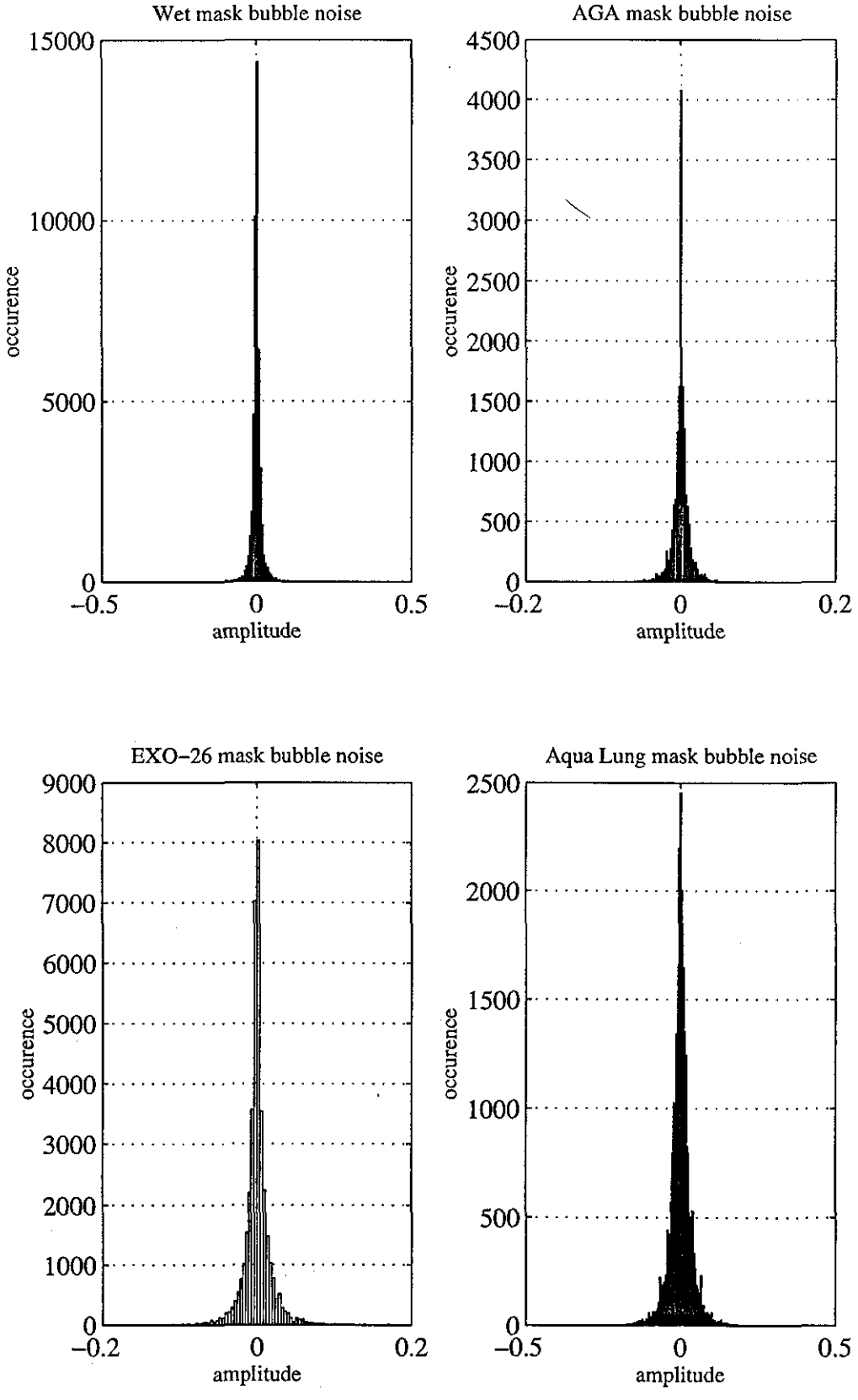


Figure 4.13 Illustration of amplitude distributions of bubble noise signals.

4.5.3.1 Spectral Subtraction Noise Cancelling

Spectral subtraction is a frequency domain method of noise reduction which is based on direct estimation of short-time spectral magnitudes, as shown in Fig. 4.14, and it is assumed that the speech signal is uncorrelated with the noise. Hence, the power spectrum of the noisy signal is the sum of the speech and noise power spectra. The spectral subtraction algorithm divides the speech into frames and subtracts the short-time spectral magnitude of noise from that of the noisy speech. Such a technique requires an estimate of the spectral magnitude of the noise signal [87-89]. In practice, this is extracted from preceeding silent intervals of speech, but when the speech is noisy it becomes almost impossible to detect the silent intervals [88]. In underwater voice communications, this is the major difficulty since bubble noise is nonstationary and speech dependent, as mentioned earlier. The only possible solution for detecting bubble noise intervals is by searching for exhalation intervals.

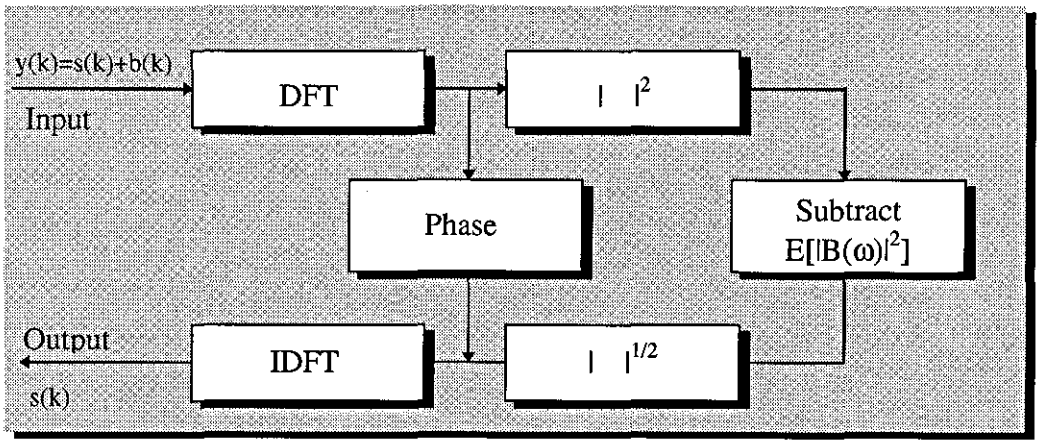


Figure 4.14 Illustration of power spectral subtraction method for bubble noise cancellation.

In order to illustrate the operation of the spectral subtraction method, assume that the noisy input speech signal, *y(k)* is defined as:

$$y(k) = s(k) + b(k) \tag{4.6}$$

where *s(k)* and *b(k)* represent clear speech signal and additive noise signal, respectively. Since the signal is processed frame by frame, a Hamming window having

better sidelobe attenuation as described in Chapter 3 is employed. The spectral magnitude of the windowed signal of interest is calculated by using the Discrete Fourier Transform (DFT) [46], formulated as:

$$Y(k) = \sum_{m=0}^{N-1} y(m)\, e^{-j\frac{2\pi km}{N}} \qquad k = 0,1,\ldots\ldots,N-1 \tag{4.7}$$

where $Y(k)$ is DFT of input signal $y(k)$, N is the number of samples in the process and $\omega=(2\pi/N)m$ for $0 \le m \le N-1$. The power spectral density of this signal is derived in [90] as:

$$|Y(\omega)|^2 = |S(\omega)|^2 + |B(\omega)|^2 + S(\omega).B^*(\omega) + S^*(\omega).B(\omega) \tag{4.8}$$

where $|Y(\omega)|^2$, $|S(\omega)|^2$, $|B(\omega)|^2$ and $\{S^*(\omega).B(\omega)+S(\omega).B^*(\omega)\}$ are power spectral density of noisy speech, clear speech signal and noise signal, and cross correlations between speech and noise signals, respectively. Assuming that the speech signal and additive noise are uncorrelated, the cross correlation term becomes insignificant, hence the new definition for the estimation of the clear speech signal is

$$|\hat{S}(\omega)|^2 = |Y(\omega)|^2 - E\left[|B(\omega)|^2\right] \tag{4.9}$$

where $E[|B(\omega)|^2]$ is the estimated power spectral density of the noise signal. In order to recover the speech signal from the power spectral density, phase information is also required. This increases computational load in processing. However, for all practical purposes, it is defined that the use of the noisy phase spectrum is sufficient [87]. Therefore, the noisy phase spectrum is utilized in order to estimate the enhanced speech spectrum. Finally, the clear speech signal is reconstructed by implementing the Inverse DFT (IDFT) [46] as defined:

$$\hat{s}(k) = \frac{1}{N} \sum_{m=0}^{N-1} \hat{S}(k)\, e^{j\frac{2\pi km}{N}} \quad , \quad m = 0,1,\ldots,N-1 \tag{4.10}$$

4.5.3.2 Implementation of Power Spectral Subtraction Method for Bubble Noise Cancellation

Due to the nature of bubble noise, i.e. non-stationary and speech dependent signal, estimation of bubble noise intervals in speech sequences is a difficult task and distinguishing it from the speech signal may be impossible. Here, for the illustration of this technique, bubble noise is "manually" extracted during exhalation intervals, in the absence of a speech signal shown in Fig. 4.12. This approach is essential since a single channel input is favoured for the sake of system complexity, hence there is no direct access only to bubble noise. The sequence of bubble noise may not be the same as one generated during speech activity, but this is the closest representation of bubble noise added to the speech signal.

The spectral magnitude of bubble noise is calculated by applying Eq. 4.7 and the ensemble average of the power spectral density is taken so that it can be used as input to the noise canceller. This is assumed to be stationary to the degree that it equals its expected value in speech intervals. If a satisfactory level of speech enhancement is achieved, then the estimated bubble noise spectral density may be assumed to be a stationary and deterministic input to the noise canceller. Therefore, the process of searching bubble noise intervals, which will increase the complexity of the algorithm, is eliminated. Moreover, since the power spectrum of the noise is already computed during the noise cancelling process, this is not calculated, i.e. computational complexity is decreased.

The power spectral subtraction technique, based on Eq.4.9, was applied to noisy speech signals from four diver's masks. Since fast processing is demanded in real time applications, the Fast Fourier Transform (FFT) [91], a computationally efficient form of the DFT, is utilised. The inverse FFT is also implemented to obtain an estimated clear speech signal in the time domain. During this processing, the estimated speech magnitude spectrum may be negative, but forcing this to zero introduces a "musical" tone artifact into the reconstructed speech, as shown in Fig.4.15, and to eliminate this effect, full-wave rectification is applied. There are other algorithms developed to overcome the musical effect, but they do not improve speech signal quality [89].

Moreover, when a listening test is conducted on the processed speech signals it is observed that by using the spectral subtraction method, the speech quality is decreased compared to the noisy speech signal. The obvious reason for this is the incorrect estimation of bubble noise spectral density. Because this is non-stationary and stochastic in nature, the assumptions made for the bubble noise spectrum in noise cancelling is not the same as in the noisy speech spectrum. Therefore, in order to achieve better performance in improving speech quality, adaptive noise cancelling methods may be considered [92] and in the following section, implementation of this technique will be studied.

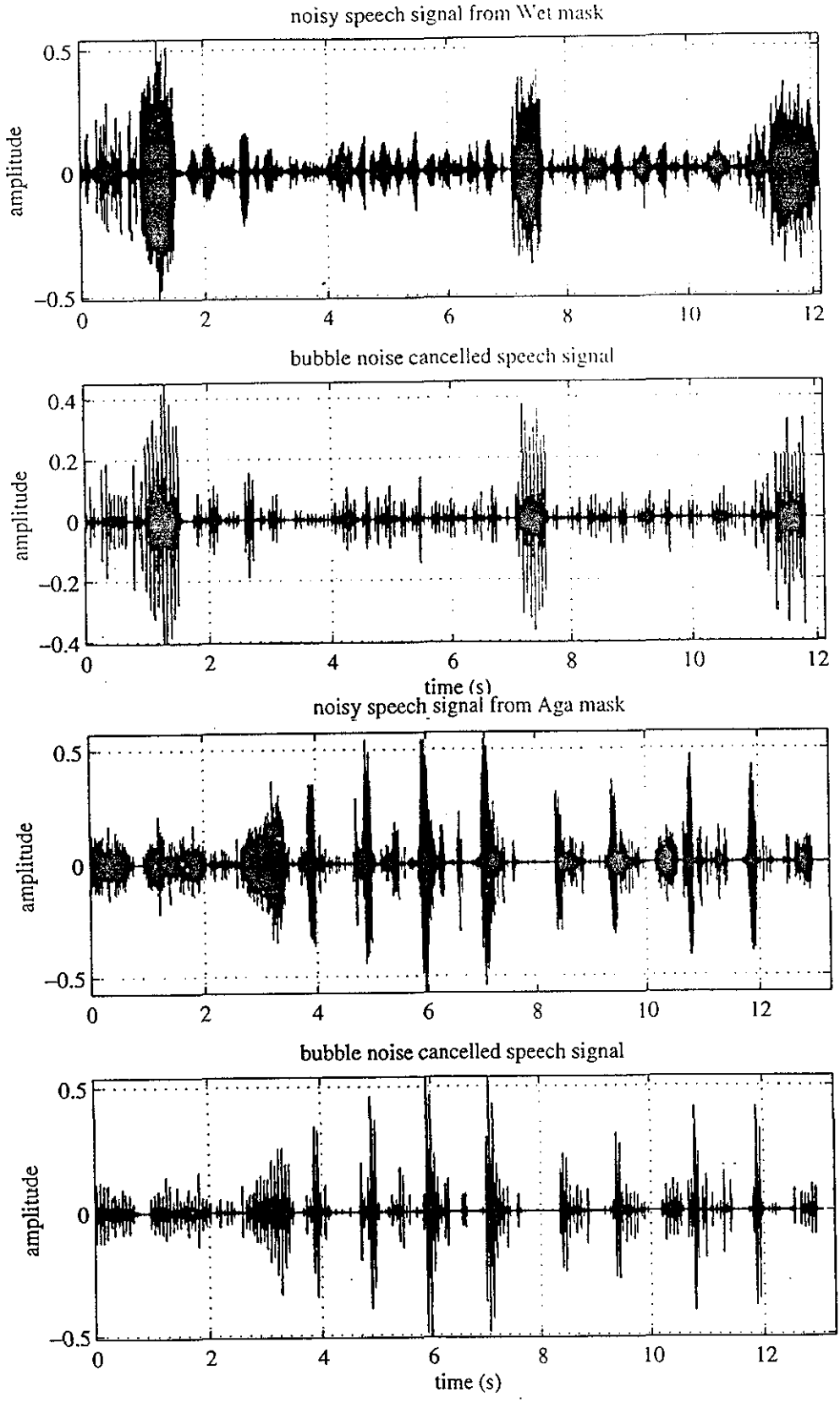


Figure 4.15 Bubble noise cancellation based on power spectral subtraction.

4.5.3.3 Concept of Adaptive Noise Cancelling

The second method considered for enhancing speech quality is the adaptive noise cancelling method [92, 93]. This is an alternative method of estimating signals corrupted by additive noise and extensively applied in speech processing. The technique employs two channels; a primary input containing the corrupted signal, *y(k),* and a reference signal containing noise, *b(k),* correlated in some unknown way with the primary noise *n(k)* as shown in Fig.4.16. In underwater voice communications, the primary input may be provided by putting a microphone inside the mask, and the other is outside the mask cavity to pick the bubble noise [94]. This type of noise canceller consists of an adaptive filter, which is a digital filter, that acts on the bubble noise to produce an estimate of the bubble noise, *x(k),* which is then subtracted from the noisy speech signal. The overall output of the noise canceller, *ε(k),* is used to control the coefficients of the adaptive filter so that a minimum mean square error is achieved, i.e. minimising the mean square error results in minimising the effect of bubble noise in the speech signal. In the literature, there are many algorithms to perform minimum mean square error estimation [93] and the one performing well in real-time applications must be considered for the underwater voice communications. In the following sections, we will discuss the principles of this method and its implementation in bubble noise cancelling.

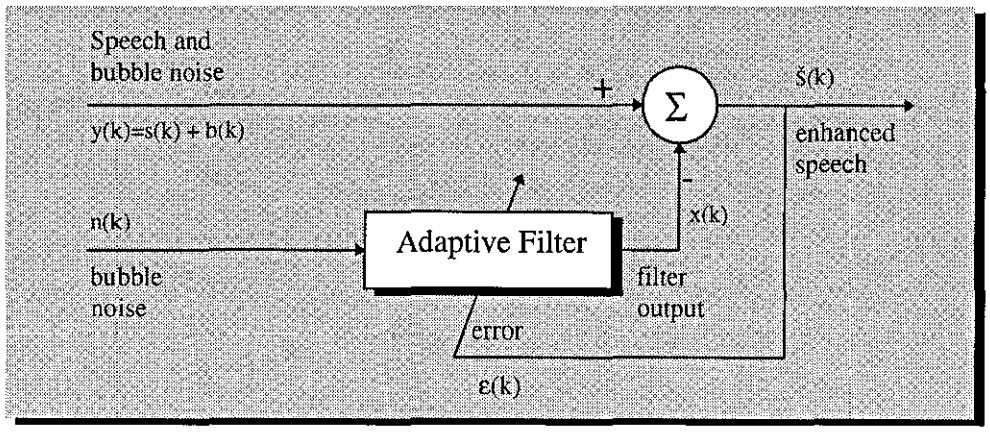


Figure 4.16 Adaptive bubble noise cancelling for diver's speech

4.5.3.4 Derivation of Optimum Filter Coefficients

When an adaptive filter is implemented as a tapped delay line with the length of M shown in Fig.4.17, the output of the filter which represents an estimation of bubble noise added to the speech signal is defined as:

$$x(k) = \sum_{i=0}^{M-1} w(i)\ n(k-i) \tag{4.11}$$

The error signal between the filter output and noisy speech is given by

$$\varepsilon(k) \ = \ y(k) - \sum_{i=0}^{M-1} w(i)\ n(k-i) \ = \ y(k) - \mathbf{W}^T\ \mathbf{N}(k) \tag{4.12}$$

where **N**(k) and **W** are the noise vector and the coefficient vector respectively, given by

$$\mathbf{N}_k = \begin{bmatrix} n(k) \\ n(k-1) \\ . \\ . \\ n(k-(M-1)) \end{bmatrix} \qquad \mathbf{W} = \begin{bmatrix} w(0) \\ w(1) \\ . \\ . \\ w(M-1) \end{bmatrix} \tag{4.13}$$

The square of the error is given as

$$\varepsilon(k)^2 = y(k)^2 - 2y(k)\mathbf{N}(k)^T\mathbf{W} + \mathbf{W}^T\mathbf{N}(k)\mathbf{N}(k)^T\mathbf{W} \tag{4.14}$$

The mean square error is obtained by taking the expectated values E[.] of both sides of Eq.4.14,

$$E\left[\varepsilon(k)^2\right] = E\left[y(k)^2\right] - 2E\left[y(k)\mathbf{N}(k)^T\mathbf{W}\right] + E\left[\mathbf{W}^T\mathbf{N}(k)\mathbf{N}(k)^T\mathbf{W}\right] \tag{4.15}$$

which can be further simplified by defining an M x M autocorrelation matrix, **R**, and an M length cross-correlation vector, **P**, as

$$R = E\left[\mathbf{N}(k)\mathbf{N}(k)^T\right] \quad , \quad P = E\left[y(k)\mathbf{N}(k)\right] \tag{4.16}$$

hence

$$E\left[\varepsilon(k)^2\right] = E\left[y(k)^2\right] - 2\mathbf{P}^T\mathbf{W} + \mathbf{W}^T\mathbf{R}\mathbf{W} \tag{4.17}$$

Gradient methods are commonly used for adjusting the weights to minimise the error. The gradient ∇ of the error function is obtained by differentiating Eq.4.17 as:

$$\nabla = \frac{\partial E\left[\varepsilon(k)^2\right]}{\partial \mathbf{W}} = -2\mathbf{P} + 2\mathbf{R}\mathbf{W} \tag{4.18}$$

The optimum weight vector, **W**, is obtained by setting the gradient of the mean square error function to zero:

$$\mathbf{W}_{opt} = \mathbf{R}^{-1}\,\mathbf{P} \tag{4.19}$$

This equation is known as the Weiner-Hopf equation [92, 93]. Implementation of this in real time systems is computationally expensive since it involves calculation of the autocorrelation matrix and the cross-correlation vector as well as matrix inversion. For practical systems, another way of obtaining the optimum coefficients is required so that the computation load is decreased, hence adaptive algorithms are used to achieve this. There are many adaptive algorithms that can be applied to perform adaptive filtering, such as Least Mean Square (LMS) and Recursive Least Mean Square (RLS) [92]. They introduce certain advantages and disadvantages in practical systems. For example, RLS produce optimum results but are computationally complex, while LMS is less complex but the coefficients are not optimum. Due to the fast performance of the LMS method, this will be implemented for bubble noise cancellation.

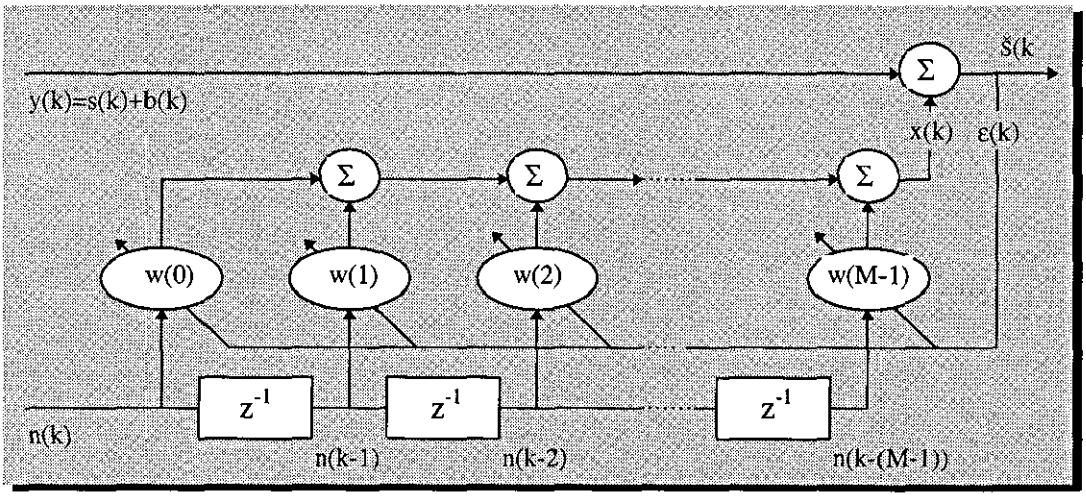


Figure 4.17 LMS Adaptive filter structure

4.5.3.5 LMS Adaptive Algorithm

The LMS algorithm is a practical method for finding close approximate solutions to filter coefficients defined by the Weiner solution in real time [92]. It does not require explicit measurements of correlation functions, nor does it involve matrix inversion. The LMS is based on the steepest descent algorithm where the coefficient vectors are updated from sample-to-sample as:

$$\mathbf{W}(k+1) = \mathbf{W}(k) - \mu\nabla(k) \tag{4.20}$$

where ∇(k) is the estimated gradient vector and μ is the factor that controls stability and rate of convergence and defined in practical applications by considering signal power, P_k and filter length as:

$$0 < \mu < \frac{1}{(M+1)P_k}$$

The LMS algorithm estimates an instantaneous gradient in a crude but efficient manner by assuming that $\varepsilon(k)^2$ is an estimate of the mean square error and then by differentiating $\varepsilon(k)^2$ with respect to **W**.

$$\nabla(k) = \begin{bmatrix} \frac{\partial \varepsilon(k)^2}{\partial w(0)} \\ \cdot \\ \cdot \\ \frac{\partial \varepsilon(k)^2}{\partial w(M-1)} \end{bmatrix} = 2\varepsilon(k) \begin{bmatrix} \frac{\partial \varepsilon(k)}{\partial w(0)} \\ \cdot \\ \cdot \\ \frac{\partial \varepsilon(k)}{\partial w(M-1)} \end{bmatrix} \tag{4.21}$$

Since in the LMS algorithm, instantaneous estimates are used for ∇(k), Eq.4.21 can be redefined as:

$$\nabla(k) = -2\varepsilon(k)\mathbf{N}(k) \tag{4.22}$$

Substituting this equation in the Eq.4.20 for the steepest descent algorithm, the Widrow-Hopf algorithm is extracted:

$$\mathbf{W}(k+1) = \mathbf{W}(k) + 2\mu\varepsilon(k)\mathbf{N}(k) \tag{4.23}$$

This is the Widrow-Hopf algorithm, having the advantage of not requiring knowledge of signal statistics (the correlations of noisy speech signal and bubble noise) as in Wiener algorithm [93], but uses their instantaneous values. However the coefficients are only estimates and improve gradually with time as the filter learns the characteristics of the signals.

A practical consideration when using the adaptive noise cancellation method is the length of the adaptive filter needed for good noise cancellation [95], and separation between the noise and speech microphones requires the adaptive filter to be able to estimate the delay between these input signals [96]. The small separation between the microphones significantly shortens the filter length required for noise cancellation [97]. For bubble noise cancellation, the separation should be small since the bubble noise source is close to the speech microphone, hence the size of the filter is arbitrarily defined as 10 taps.

4.5.3.6 Implementation of Adaptive Noise Cancelling for Bubble Noise

In this section, adaptive noise cancellation is implemented for enhancing speech quality from a diver's mask. Two approaches are considered in this study in terms of input signals; the first one is to use simulated data and the second is to use real data. For the first approach, the bubble noise shown in Fig. 4.12 is assumed to be the noise input and for the other, a second microphone is placed outside the mask cavity; more detail about this will be given later. As mentioned earlier, in order to cancel bubble noise, it is highly desirable to define a fixed filter so that both hardware, i.e. avoiding the use of a second channel, and software complexity can be reduced. Probably, using the simulated bubble noise input, the adaptive noise canceller may be used to achieve a digital filter with fixed coefficients.

In the application of the first approach, bubble noise signals extracted from recorded speech sequences are assumed to be stationary noise and applied to the noise input of the adaptive noise canceller, $n(k)$. A clear speech signal, $s(k)$, which was recorded while the diver was on the surface and wearing the *wet mask,* is utilised as illustrated in Fig.4.18. After that, the bubble noise signal is digitally added to the clear speech signal to simulate a noisy speech signal, $y(k)$ and it is assumed that there is no cross-talk between the bubble noise input and speech input. These two input signals are then applied to an adaptive noise canceller (ANC) with a filter length of 10 taps.

The LMS algorithm adapts the weights of the filter, in a manner described earlier in this chapter, and at the output of the filter an estimate of the bubble noise signal is obtained that is subtracted from the noisy speech signal. This process results in producing a clear speech signal. Under the assumption of strong correlation between the bubble noise and the noise signal added to speech, and the absence of crosstalk between the channels, the adaptive noise canceller performs satisfactorily. The aim of this simulation is to see whether the adaptive filter coefficients converge to a level or are time varying. If the noise canceller is stable, then we can use these coefficient values to represent a fixed filter. However, the current experiments show that the filter coefficients are time varying, hence the idea of using fixed filter is inapplicable.

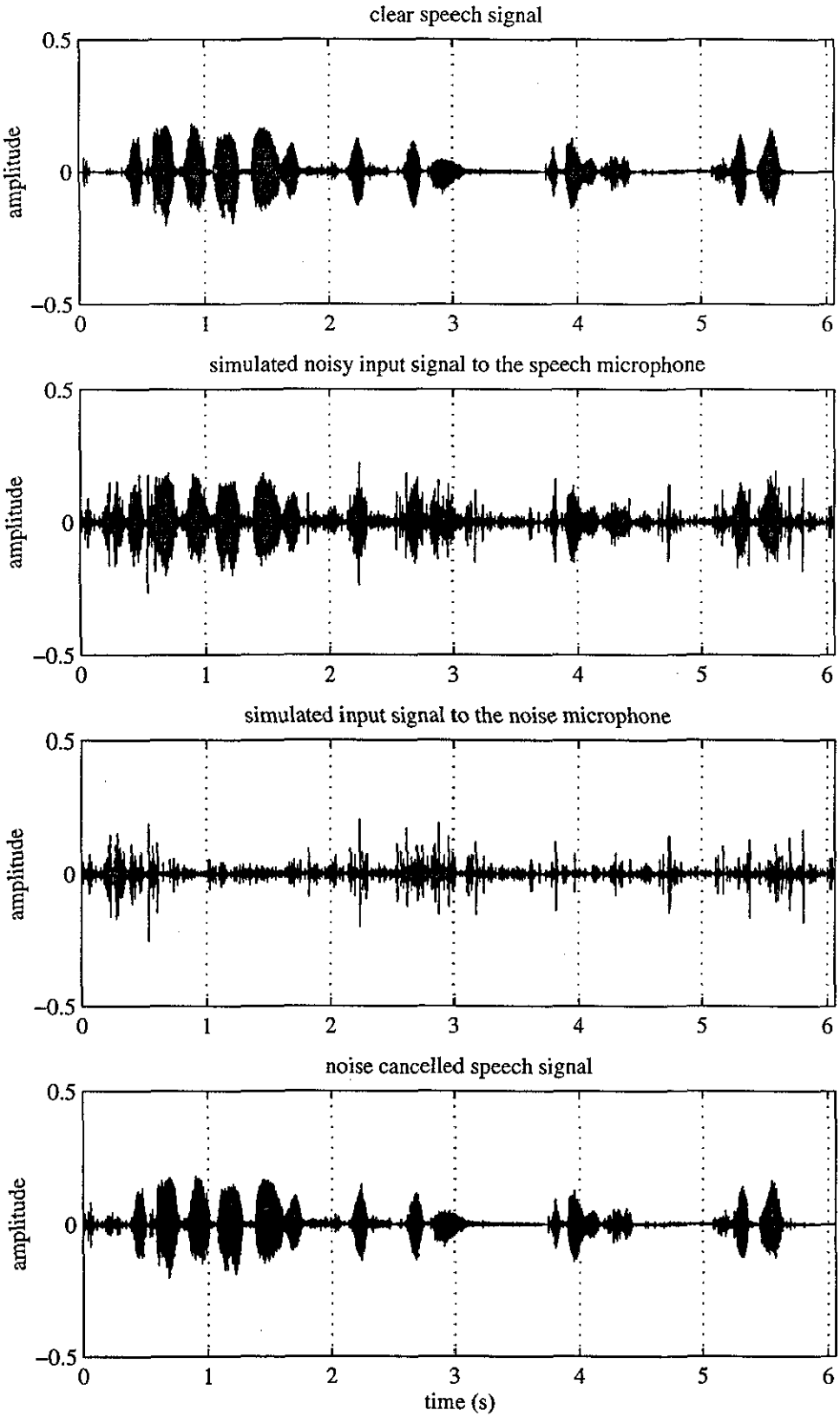


Figure 4.18 Adaptive noise cancellation performance on simulated data

Moreover, in real applications, it is highly difficult to isolate the noise source and the speech source, so there is crosstalk between them [96, 97]. In order to examine the performance of the adaptive noise canceller in a real environment, as a second approach mentioned above, a second microphone is utilised and placed outside the mask cavity in water to pick up the bubble noise. In principle, this microphone must be positioned far from the speech microphone inside the mask cavity so that the crosstalk effect is minimised [46]. Nevertheless, since the bubble noise is produced by escaping air from the demand valve which is only a few centimetres from the mouth the crosstalk effect is unavoidable. In order to see the performance of the noise canceller under practical conditions, two CF 2949 noise-cancelling microphones were employed. These were waterproofed by covering the sound ports and electrical connections with rubber. During a diving trial, signals from these two microphones were recorded with a multi-channel RACAL recorder, and waveforms are shown in Fig.4.19. When these signals are analysed in frequency domain, it can be seen that signals picked up by the noise microphone are attenuated in the frequency band 2000 - 4000 Hz and the overall attenuation of 3.55 dB is introduced on the signal energy picked up by the noise microphone.

There are two factors contributing to this result; one is the change of the frequency response of the microphone underwater. Although no experimental results are available for this, since the microphone is operating in water conduction rather than air, it is obvious that some sort of alteration in the frequency response should be expected. From Fig.4.19, it is noticeable that the microphone response decreases over the frequency range of 2000 Hz to 4000 Hz. The other possible explanation for this is the radiation loss in the underwater channel. Since the radiation loss increases with the increase the frequency of acoustic waves, as defined in Chapter 5, this phenomenon will contribute to the attenuation at the specified frequency band. Moreover, the mask, which acts as a barrier between two media, causes attenuation at high frequencies of the speech signal detected by the noise microphone underwater.

Performance of the ANC is tested with these two real input signals and no modification is introduced to the noise canceller to compensate for the crosstalk effect. When the processed speech signal is listened to, it is noticeable that there is no improvement in

speech signal quality introduced by the noise canceller. This may be explained by taking into account the condition for successful operation of the ANC as described in Section 4.5.3.3. It was stated that the noise component of the speech signal and the noise must be correlated for effective noise cancellation. In addition to this, the noise microphone must be acoustically isolated so that no speech signal is picked up. However, these two conditions are not provided. Due to a change in the medium, from water to air, noise becomes uncorrelated. Because there is cross-talk between the two channels, both microphones pick up the speech signal, and an estimation of the noise signal added to the speech by the ANC is not achieved.

However, the improvement in the noise cancellation in a real environment may be achieved by considering the crosstalk effect and modifying the ANC as recommended in [94]. In addition to this, the simplest solution would be to keep the bubble noise source far from the speech source by modifying the diving equipment, like in an *Aga mask* or with a rebreather system.

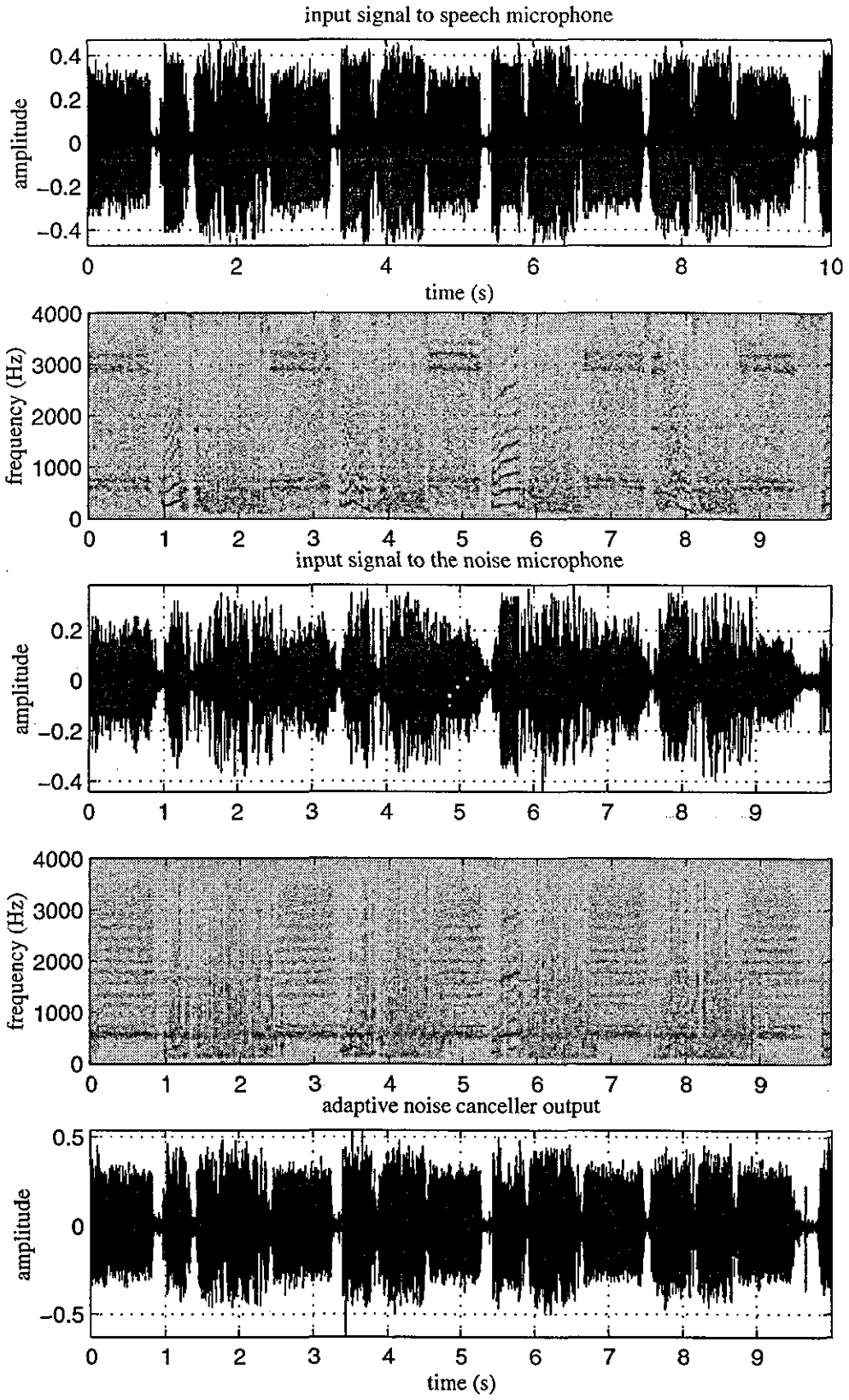


Figure 4.19 Real input signals to the adaptive noise canceller.

CHAPTER FIVE

DESIGN CONCEPTS OF A DIGITAL UNDERWATER ACOUSTIC VOICE COMMUNICATION SYSTEM

5.1 Introduction

In this chapter, we discuss properties of an underwater acoustic communication channel governing digital transmission of the quantized speech parameters extracted in Chapter 3. In the proposed system, for efficiency, some measures are defined against the limitations introduced by the underwater channel as well as the communication range factor. Moreover, the feasibility of commonly used modulation techniques for underwater telemetry systems is studied and implementation of the *digital pulse position modulation* (DPPM) method is considered.

5.2 Underwater Acoustic Voice Communication Channel

The underwater communication channel has proved to be an extremely difficult medium in which to achieve the high data rates required for many applications. High rate data transmission requires either wide bandwidths, limited for the underwater channel, or a noise free channel [98]. Therefore, unlike a radio communication channel, much less information can be conveyed acoustically from one point to the another. There are several characteristics of the channel that dominate the design of any underwater acoustic communication system which need to be considered in the voice communication system. These are path losses due to geometrical spreading and absorption, multipaths caused mainly by reflection on the surface and bottom boundaries, sound speed variation, amplitude, frequency and phase distortion, ambient noise and the self-noise of the system [99-102]. In the following sections, these factors are studied in detail, their effects on the performance of the system are discussed and solutions to improve the performance of the voice communication system are addressed.

5.2.1 Acoustic Attenuation

The voice communication system is intended to be used as a portable unit, but due to its power supply, a somewhat limited communication range will be achieved although long distances are desired. Nevertheless, because of the communication medium, the received acoustic signal is weakened as the distance between transmitter and receiver is increased. This is because of the attenuation introduced by *transmission loss* which is classified into *absorption loss* and *spreading loss*. Previous studies report that absorption loss, representing a conversion of acoustic energy into heat due to viscous friction, generally increases as the square of frequency, hence this factor determines the frequency band for the system [101]. Moreover, absorption loss decreases with increasing depth due to hydrostatic pressure. In sea water it is greater than for pure water below 100 kHz due to ionic relaxation of the magnesium sulphate molecules in sea water [100]. The absorption loss in sea water at 4°C and 1 atm, with the salinity of 0.035, is given in [103] as:

$$\alpha(f) = \frac{0.075f^2}{0.8^2 + f^2} + \frac{36f^2}{5000 + f^2} + 4.1x10^{-4} f^2 \quad \text{dB/km} \tag{5.1}$$

where f is frequency in kHz. When the carrier frequency for the communication system is decided, the absorption loss must be considered, because rapid attenuation of higher frequencies, as illustrated in Fig.5.1, impose a maximum frequency limit for a communication system.

If digital data transmission is required for a range of several hundreds meters similar to that of the analogue voice communication systems, either the transmission power must be increased or the carrier frequency must be decreased to accommodate for losses. Each solution possesses certain degree of limitations in considering a practical system. Increasing the power will decrease the operation-time of the system while decreasing the carrier frequency will reduce data transmission rate [104]. More detail about the latter issue will be discussed later in this chapter.

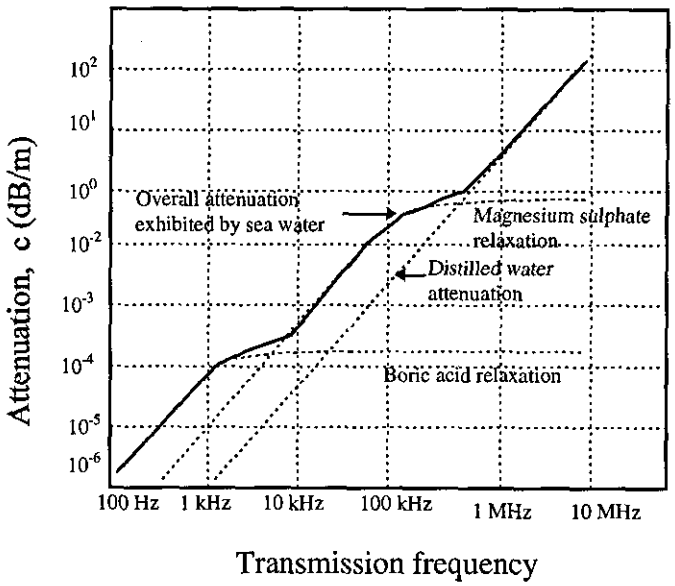


Figure 5.1 Attenuation in water with transmission frequency

In addition to absorption loss, transmission loss also caused by geometrical spreading of the acoustic signal, which introduces a strong dependence on the pressure level of the received waveforms [99-101]. This loss increases proportional to the distance, *R*, between the transmitter and the receiver. The overall transmission loss, *TL(f)*, for such a channel is expressed as:

$$TL(f) = 20\log R + \alpha(f)R \quad dB \tag{5.2}$$

However, transmission loss for a specified distance can be reduced by employing a directional hydrophone which focuses the acoustic signal and reduce the effects of attenuation due to spherical spreading. Implementation of such a solution will have a drawback in voice communications. If the direction of the transmit or the receiver transducer is deviated from line of sight the communication link may be interrupted and data lost. Therefore, directional radiation is not recommended for underwater acoustic voice communications where the direction and orientation of the source and the receiver are difficult to achieve.

5.2.2 Underwater noise

The level of background noise in an underwater communication system will influence the lower limit of detection of acoustic signals at a hydrophone. As in any other signal detection scheme, noise makes it difficult to determine if the desired signal is present or absent, and an understanding of the origins of the noise is essential. Indeed much work has been done in [17, 99, 105, 106] to evaluate noise sources. Noise in the sea is mainly produced by wave action, shipping, thermal agitation, rainfall, and sounds made by marine animals. It is also reported that wind-related noise is dominant in the mid-frequency band, i.e. from several hundred hertz to about 50 kHz, and decreases at the rate of -6 dB/octave at higher frequencies [17, 99]. Moreover, in the 10-20 kHz band, noise generated by rain and waves also dominates operation of communication systems [106]. In the high frequency band, 50 kHz to approximately 2.5 MHz, thermal noise originating in molecular motion in the sea becomes dominant and increases at a rate of 6 dB/octave as frequency increases [105]. The thermal noise as a function of frequency is defined as:

$$NSL(f) = -15 + 20\log f \quad dB\ re\ \mu Pa\ per\ Hz \tag{5.3}$$

where $NSL(f)$ is noise spectrum level and f is frequency in kHz. The overall noise spectral level in the sea is as shown in Fig.5.2 [101]

Ambient noise affects the received signal-to-noise ratio and largely influences the transmitter power. Since its level generally decreases in the high frequency range, most of the underwater communication systems are operated as close to the attenuation limit as possible. Moreover, ambient noise levels are very site-dependent, especially in shallow water, inshore environments and marine worksites which are inherently noisier environments [102]. The total noise level in an underwater channel with a receiver bandwidth of B is defined as:

$$NL(f) = NSL(f) + 10\log B \quad dB \tag{5.4}$$

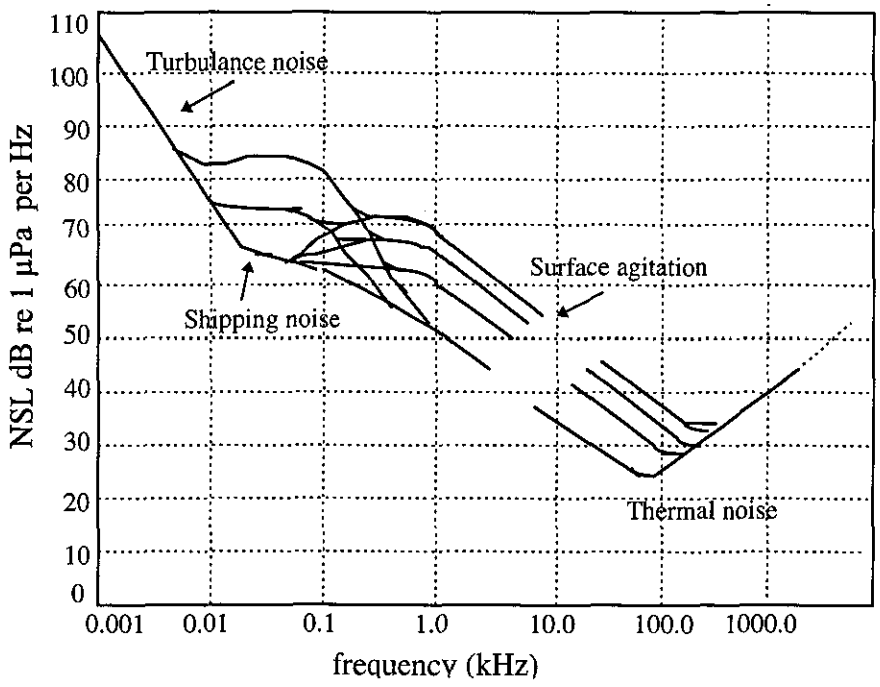


Figure 5.2 Underwater ambient noise spectrum level

5.2.3. Multipath Propagation

The most difficult problem encountered in underwater acoustic communication channels is considered to be the multipath propagation of limited bandwidth signals [107, 108] and it depends on the channel geometry, as displayed in Figure 5.3, i.e. location of transmitter and receiver, and also the frequency of the transmitted signal. In an underwater channel, the multipath propagation will most usually occur due to

(a) reverberation from the boundary of the communication medium and it is a complex function of the environmental as well as the geometrical parameters of the channel [109, 110].

(b) volume reverberation due to marine life and other objects in the channel

Moreover, during underwater voice communications, bubbles generated from the breathing apparatus of a SCUBA diver maybe also considered as a source of multipath propagation. As stated by [99] and [111], since air bubbles in water have a different density and compressibility than water and due to its resonant characteristics, it affects

the radiation of underwater sound. It is concluded that an oscillating bubble may radiate the incident sound wave as scattered sound in all directions and the influence of bubble clouds on acoustic attenuation increases with frequency. However, in this study, the effect of air bubbles is not investigated and it is assumed that it is insignificant.

Shallow channels, in which the system will normally be used are generally considered to be extremely hostile to high data rate communication systems, as they exhibit such phenomena as signal dispersion and phase fluctuation due to multipath. The multipath propagation causes severe degradation of the acoustic communication signals. The combination of high data rate and multipath propagation can generate intersymbol interference (ISI) [98], and an important figure of merit is the multipath spread in terms of symbol intervals. Typical multipath spreads in a horizontal underwater acoustic channel may be several tens or even a hundred symbol intervals for moderate to high data rates [112, 113]. Combatting the underwater multipath to achieve high data throughput is considered to be the most challenging task in designing an underwater acoustic communication system [113-117].

The design of modern data transmission systems has been accompanied by extensive research in communication engineering. In this section, an introduction of well established principles of wireless radio communications to underwater acoustic communications is presented. Recently developed sophisticated systems have achieved high data rates by allowing ISI in the received signal [115]. However, they employed channel equalizers [113, 114] or array processing [118] and directional arrays were carefully positioned in the channel to reduce the effect of ISI. Another approach used in array processing is adaptive beamforming, which has produced successful results in high rate transmission [110, 119, 120]. However, both methods require at least two hydrophones which are spatially separated by about half a wavelength ($\lambda/2$). Although these methods may be beneficial in communications applications for autonomous underwater vehicles, neither array processing nor adaptive beamforming is suitable for diver communications due to the increase in signal processing complexity as well as the physical size of the system. In the underwater voice communication system presented in this thesis, the transmitter and receiver are mobile, therefore a single omnidirectional

transducer-based design is prefered; this also reduces the software and hardware complexity. However, no ISI suppression method is employed in the prototype but in any future design of the system, appropriate techniques should be included.

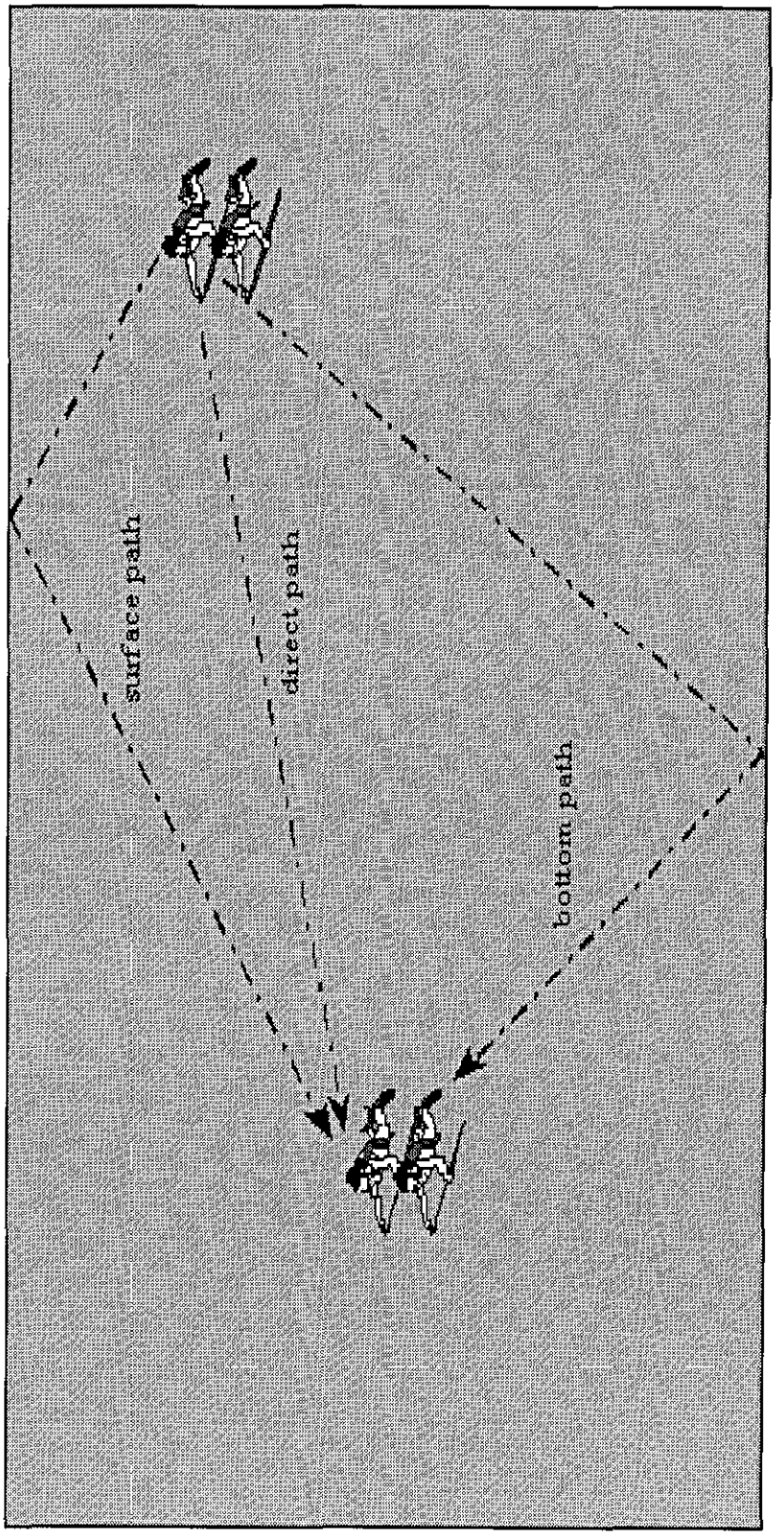


Figure 5.3 Multipath progation in underwater voice communication channel

5.3 Digital data transmission methods for underwater communications

Nowadays, sophisticated digital modulation methods are adopted for radio communication systems where high rate data transmission is achieved. Due to the channel properties as discussed in section 5.2, accomplishing the same data rate for underwater acoustic communication channel is impossible. Considering the frequency band limitations, an appropriate modulation method must be chosen to transmit speech parameters calculated in Chapter 3 for the underwater acoustic voice communication system. As shown in Figure 5.4 and listed in Table 5.1, there are several digital modulation techniques which have been frequently used for underwater telemetry. These are Amplitude Shift Keying (ASK), Frequency Shift Keying (FSK) and Phase Shift Keying (PSK) [122].

The early underwater acoustic telemetry systems employed amplitude and frequency shift keying modulation for data transmission. The ASK has performed well when the path is clean and reverberation is low, e.g. vertical radiation [123]. In a noisy channel the performance of ASK can be improved by introducing an error correction scheme [124] but this decreases the data rate. The difficulties with amplitude modulation in a reverberant environment is recognized and most of the underwater communication systems have used some form of FSK and PSK.

As mentioned in [125, 126], FSK modulation has traditionally been considered as the only alternative for channels exhibiting rapid phase variations, such as shallow water long- and medium-range channels. While incoherent detection eliminates the problem of carrier phase tracking, this technique is still prone to reverberation problems. FSK, which is more immune to the problems of multipath propagation, has been used for reverberant channels and performs well, for example, for transmission of biomedical signals from a SCUBA diver [127].

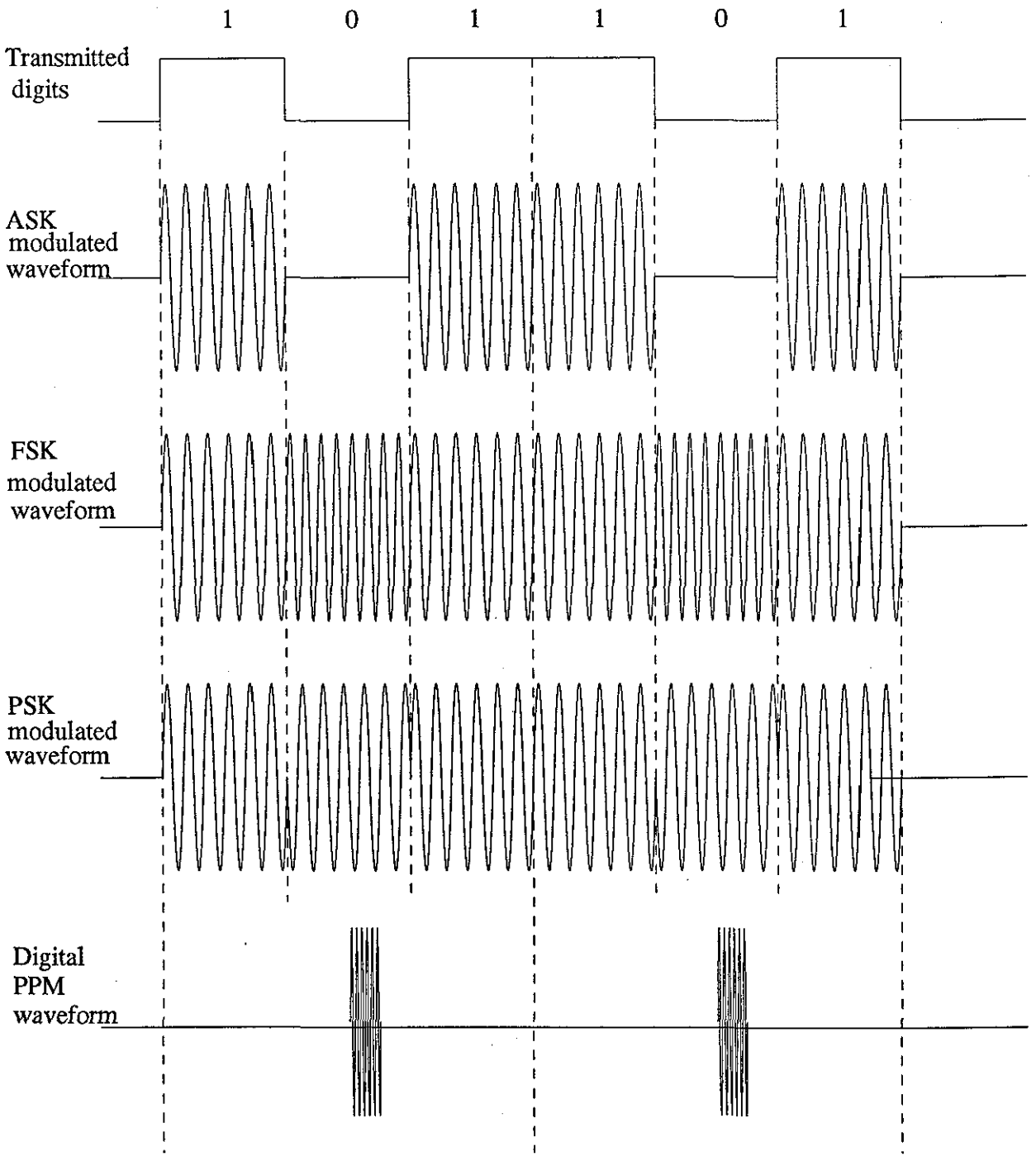


Figure 5.4 Digital modulation techniques for underwater communications.

Modulation	*Channel*	*Band*	*Data rate(kbit/s)*
ASK [123]	short/shallow	150 kHz	0.6
FSK [130]	vertical/horizantal	15 kHz	1.2
PSK [128]	medium	1.7 kHz	0.2
DPSK [118]	medium	50 kHz	10
QPSK [131]	shallow	15 kHz	5
8-slot DPPM [132]	test pool	70 kHz	2.4

Table 5.1 Digital modulation methods used for underwater communications.

Recent studies shows that PSK modulation achieving better SNR [98] and is a frequently applied technique for underwater communications[128, 129]. In this method, digital data is represented by a phase change. However, at the receiver, coherent detection is required so that the phase information of the input signal is recovered. Therefore its implementation is more complex than the other modulation techniques. Another shortcoming of PSK is its sensitivity to multipath propagation, which causes phase and amplitude changes on the signal. In order to overcome coherent detection complexity, *differential* PSK (DPSK) has been developed [98] and high rate data transmission, up to 20 kbit/s, has been achieved with a directional array [116, 118, 133]. In recent years, phase-coherent communications, 4-PSK and 8-PSK, previously not considered feasible [107, 108], have been demonstrated to be a viable way of achieving data transmission through underwater channels, at rates of 4.8 kbit/s [134] and 10 ksymbol/s [135]. Moreover, it is also reported that QPSK modulation has been employed for speech transmission at 6 kbit/s [136]

In order to suppress the ISI problem due to multipath propagation for the ASK and FSK modulation techniques, the existing noncoherent systems employ signal formatting with guard times. They are inserted between successive pulses to ensure that all the reverberation will disappear before each subsequent pulse is received [137]. The insertion of idle periods of time obviously results in a reduction of

available data throughput. However, the advantage of these methods is that they can be easily implemented in hardware and hence are relatively simple to design. Recently, for the DPSK and M-ary PSK modulation techniques, adaptive channel equalizers have been successfully implemented to overcome the multipath propagation effect [113, 114].

5.3.1 Digital Pulse Position Modulation

As an alternative to the previous modulation techniques, the pulse position modulation (PPM) method is also an effective technique for transmitting information. As stated in [138-141], the digital PPM has been shown to be a very effective modulation format for transmitting digital information in an optical channel. PPM is introduced for underwater data transmission in [142, 143] and implemented in [132]. The conclusion of these authors is that the PPM performance is limited by signal-to-noise ratio. The justification of considering this technique is because of its suitability for power-limited channels [144] and less complexity in implementation. In addition, since a burst of pulse is transmitted, detection of this signal is less affected by the multipath propagation. Moreover, bandwidth efficiency of the PPM, i.e. R_b/B where R_b is the bit rate and B is the bandwidth, must be also investigated in comparision with various digital modulation. As can be seen from Table 5.2 (it is assumed that a rectangular waveform modulates a 70 kHz carrier), by implementing the digital PPM, the bandwidth efficiency of the system decreases due to an increase in the required channel bandwidth [145]. With the receiver transducer bandwidth of 25 kHz, bit rate of 4.7 kbit/s for the 8-slot DPPM and 37.5 kbit/s for the 8-ary PSK can be achieved for through-water data transmission.

Since the envelope of the received signal may be used for data decoding, the binary ASK signal detection principle can be employed. Unlike for PSK there is no extra processing for phase detection and unlike FSK there is no need for bandpass filter banks. Hence, in terms of system complexity also, digital PPM is superior to the other techniques. However, there are several shortcomings of digital PPM. Since signal transmission and detection is as in ASK, the technique is sensitive to multipath propagation. The destructive effect of a multipath signal on the direct path signal will

therefore cause an error in data decoding. In addition to this, accurate synchronization of the transmitter and receiver is required since both utilize accurate timing for pulse position modulation and demodulation. More details of the synchronization issue is given in Chapter 7.

Modulation Techniques	*M*	*Bandwidth Efficiency*	*M*	*Bandwidth Efficiency*
M-ary ASK	2	0.5	8	1.5
M-ary PSK	2	0.5	8	1.5
M-ary FSK	2	0.25	8	0.1875
M-slot DPPM	2	0.25	8	0.1875

Table 5.2. Bandwidth efficiency comparison of digital modulation techniques.

5.4. Digital Pulse Position Modulation for Transmission of Speech Parameters

Because of the advantages of Digital PPM, this coding was chosen and implemented for underwater acoustic voice communications. Digital information is transmitted by dividing each data frame into M possible time slots and locating of a pulse in just one of these time slots per frame to indicate the data. For M-slot DPPM, the symbol interval of duration T_{symbol} seconds is sub-divided into M slots of equal duration T_{slot} seconds. A signal pulse is coded in only one of the M slots in a symbol interval. There are thus M different symbols, with the size of $k=log_2M$, each corresponding to one of the M possible slot positions in a symbol interval defined as:

$$T_{symbol} = \frac{log_2 M}{R_b} \tag{5.5}$$

The ratio of allowed band to the overall symbol duration defines a modulation index m defining the guard band interval, this avoids interframe interference (IFI) which may be caused by the band-limiting nature of the transducer. This is important because after transmitting a pulse, the hydrophone will continue ringing for several cycles causing pulse spreading at the source. Therefore, when the guard band interval is

defined, this property of the hydrophone must be considered so that during the transmission of successive data the hydrophone is in a quiescent state. The modulation index, m, is always selected to ensure that the symbol duration, T_{symbol}, consists of an integer number of slots, i.e. the result of (M/m) must be an integer. This is a practical requirement which simplifies the design of the PPM coder. The mathematical definition of a digital PPM pulse stream is given in [141] as:

$$x(t) = \sum_{n=-\infty}^{\infty} g(t - nT_{symbol} - t_n) \tag{5.6}$$

where $g(t)$ is the PPM pulse shape and t_n is the data coded into PPM. The slot duration T_{slot} and t_n are such that

$$T_{slot} = \frac{mT_{symbol}}{M} \tag{5.7}$$

and

$$0 \leq t_n \leq (M-1)T_{slot}$$

In the underwater voice communication application, the optimum slot number for a given bit rate is defined as a function of the bandwidth, B, of the transducer being employed for the transmission as

$$B \geq \frac{2MR_b}{m\log_2 M} \tag{5.8}$$

The digital PPM technique is implemented for the transmission of quantized speech parameters, as described in Chapter 3. Since the bit rate for the given parameters is 2.4 kbit/s, 8-slot digital PPM was decided on by considering power consumption which is 25 % of 2-DPPM. This is an important factor in the design of portable systems. Therefore, quantized speech parameters are separated into sub-groups consisting of 3 - bits, i.e. $k=3$, and a symbol interval is defined as $T_{symbol}=3/2400$ s, i.e. 1.25 ms. Eight

slots are required to represent a pulse position in each symbol interval and two slot intervals are used for the guard band, i.e. $m=0.8$. Thus, ten slots are used to transmit 3-bits of information from the speech parameters.

In the system implementation, $T_{symbol}= 1\ ms$ and hence T_{slot} = 100μs. For a 22.5 ms speech frame interval, the required bit rate is found to be 54 bits as defined in Chapter 3. This means that 18 symbols, i.e 54/3, must be transmitted for the speech frame interval, taking 18 ms, i.e. T_{frame}. The transmission of digital information in digital PPM format is illustrated in Fig. 5.5 for a symbol 100_2.

As presented in Fig. 5.5, three timing intervals are required for the digital PPM technique, T_{slot}, T_{symbol}, T_{frame} as well as timing interval for speech frame. During detection and decoding of the PPM signal by the receiver, the start of these times must be synchronized. Decoding of the digital PPM signal at the receiver requires an accurate estimation of pulse position. Because of multipath propagation and the receiver design, the received pulses tend to spread out, hence affecting the detection performance. However, measurement of the energy, E_{slot}, of the input signal, $x(t)$, in each slot interval gives useful information about the pulse position.

$$E_{slot} = \int_0^{T_{slot}} |x(t)|^2 dt \tag{5.9}$$

The maximum energy in the receiver will be detected during a slot interval representing the data slot in the transmitter, hence correct data decoding is achieved. During this process, the effect of multipath propagation must be taken into account. Depending on the constructive or destructive effect of the reverberant signal on the directly received signal, the energy of the signal during a slot interval may be altered, leading to a decoding error. Therefore, as previously mentioned, multipath suppression methods are required in order to improve the detection performance of the system.

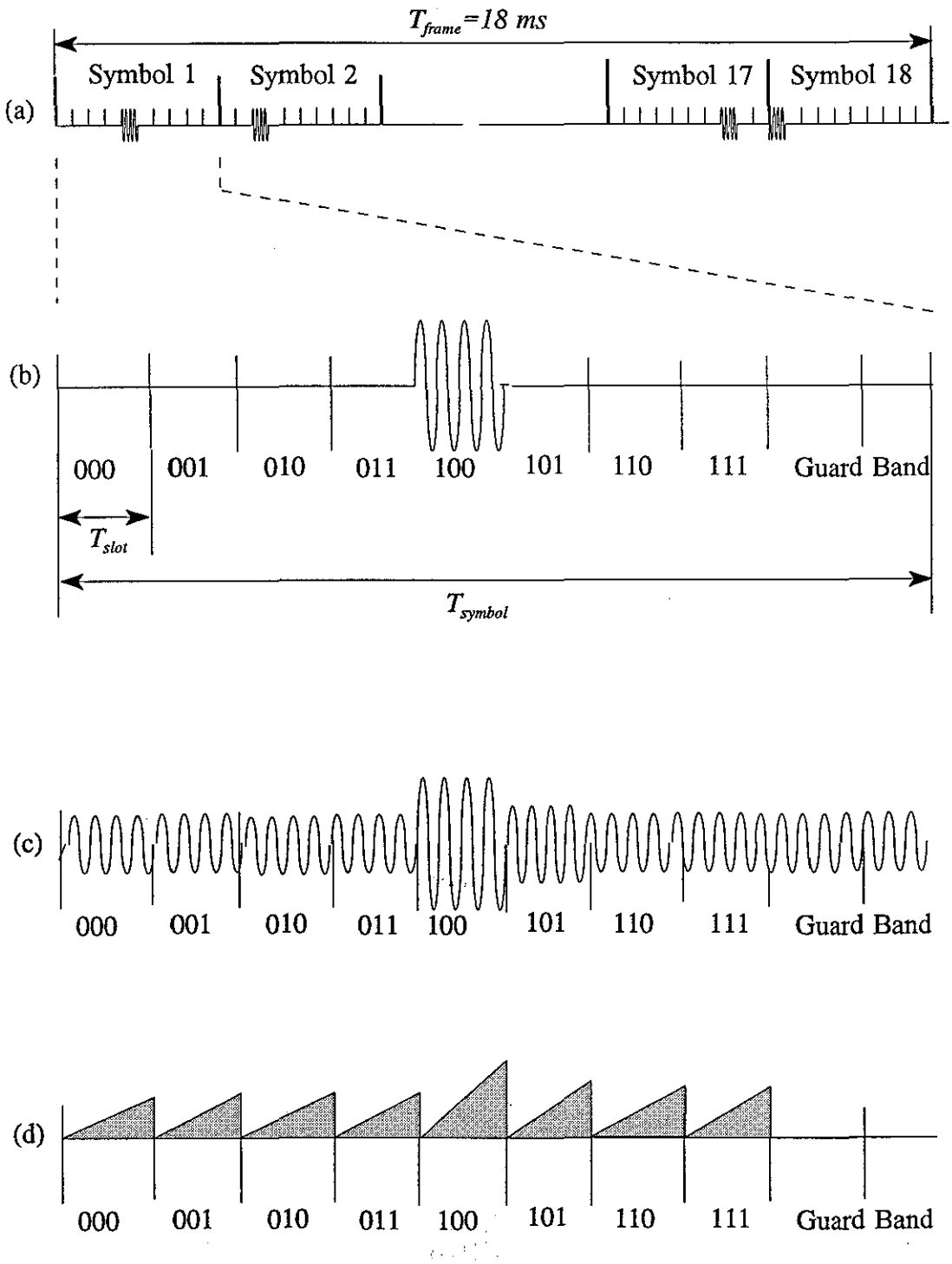


Figure 5.5 Transmission and detection principles for digital PPM.
(a) DPPM encoder output
(b) Transmission of encoded data in PPM format
(c) Received input signal to the decoder
(d) Digital decoder output

5.5 Choice of Modulation Frequency for Digital Voice Communications

In contrast with the strict legal allocation of frequencies for radio transmission, there is no regulatory body to govern the allocation of frequencies for underwater communications and telemetry.

In principle, any frequency can be chosen, although it is advisable to avoid the recommended emergency frequencies for manned submersible beacons (10 kHz), hazard markers (13 kHz) and diving bells (37.5 kHz) and also the carrier frequencies of communications systems used by NATO (8.0875 kHz), commercial submarines (25 kHz), the Royal Navy (40.2 kHz) and the NATO Navies (42 kHz) [23]. Fig. 5.6 illustrates the frequency bands in which various systems are located. Frequencies of the order of 40-42 kHz have been commonly used for analogue speech communication systems between divers and between divers and the surface to give a communication range of up to 1 km depending on the transmitter's output power [20].

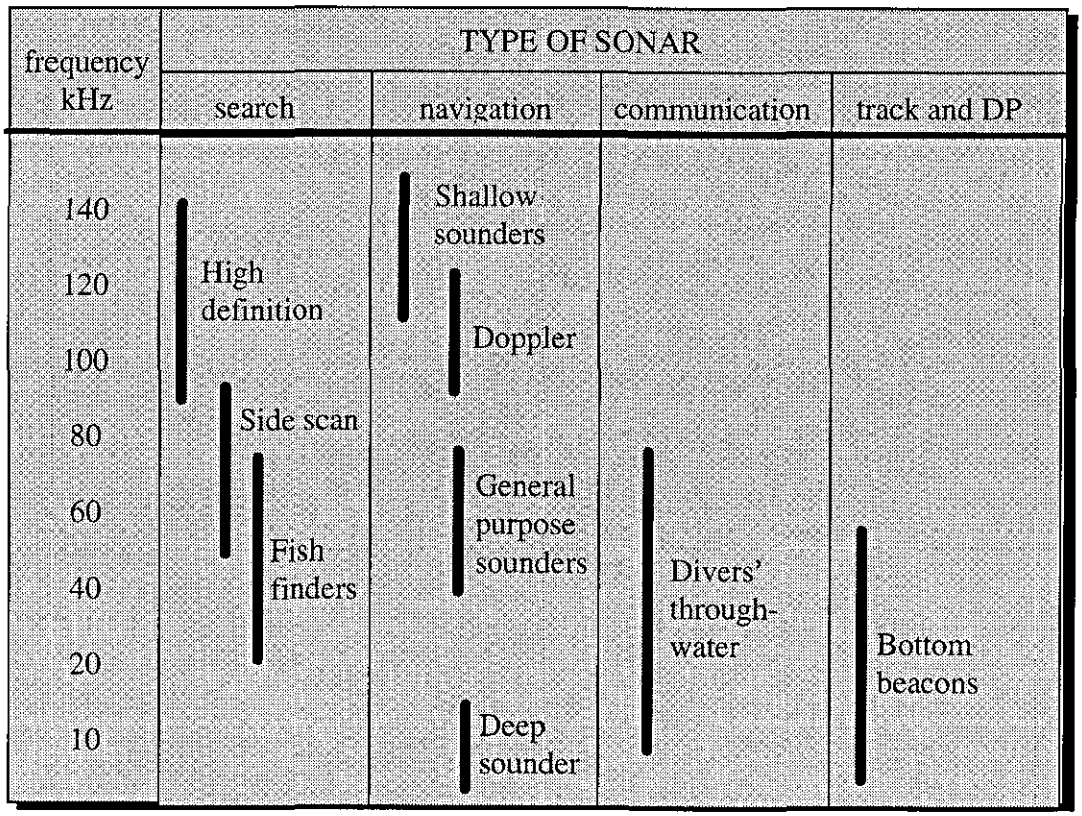


Figure 5.6 Frequency distribution in underwater acoustic signals

For the digital underwater voice communication system it will be useful to set a frequency band so that transmission of digitized speech at the rate of approximately 16 kbit/s will be achieved for a range of several hundred metres. However, at this point it is unavoidable to consider the influence of the various speech coding methods. As discussed in previous chapters for satellite communication systems, there is tendency towards the application of low bit rate speech coding and improving the decoded speech signal quality.

Nowadays good communication quality speech is achieved for speech encoding techniques using bit rates of 8 kbit/s or above. In Table 5.3 the estimated minimum carrier frequencies for different speech coding methods and different modulation techniques are listed when the Q factor of the hydrophone is approximately equal to 5. From Fig. 5.6 and Table 5.3 (frequency entries are calculated as $f = Q \times R_b$), it may be seen that for the digital underwater voice communications, a frequency band over 150 kHz may be utilized if one of the low bit rate speech coding standards, e.g. VSELP, CELP and LPC10, is implemented. Although this frequency band limits the communication range, it can still be beneficial for local area communications for the sake of speech quality. For the system designed and implemented in this research, a 70 kHz hydrophone was chosen, since transmission of high quality speech at a data rate of 13 kbit/s, i.e. RPE-LTP, can be accomplished although LPC10 at 2.4 kbit/s is employed.

speech encoding	*bit rate kbit/s*	*ASK* f_0*(kHz)*	*FSK* f_0-f_1*(kHz)*	*PSK* f_0*(kHz)*	*QPSK* f_0*(kHz)*	*DPPM* f_0*(kHz)*
PCM	64	320	320 - 448	320	160	1066.6
ADPCM	16	80	80 - 112	80	40	266.6
LD-CELP	16	80	80 - 112	80	40	266.6
RPE-LTP	13	65	65 - 91	65	32.5	216.66
VSELP	8	40	40 - 56	40	20	133.33
CELP	4.8	24	24 - 33.6	24	12	80
LPC10	2.4	12	12 - 16.8	12	6	40

Table 5.3. Recommendation of carrier frequencies for digital underwater voice communications for different speech encoding and modulation methods (8-slot DPPM is assumed).

5.6 The Sonar Equation and Estimation of the Received Signal

In order to process the acoustic signal at the system's receiver, its level must be known. As mentioned earlier in this chapter, there are many criteria limiting the design and performance of the system and the sonar equation provides a variety of statements for the efficiency of the system. Since one-way transmission exists for underwater voice communications, the passive sonar equation [99, 100] is used in the form of:

$$SL + DI_T = TL + NL + DT - DI_R \quad dB \tag{5.10}$$

where SL is the source level of the transmitter, DI_T is the directivity index of the transmitter, TL is the transmission loss because of spreading and absorption, NL noise level of the channel, DT is the detection threshold at the receiver and DI_R is the receiver directivity index.

For underwater voice communications, spherical radiation is recommended since the directions of the transmitter and the receiver are unknown in the channel, therefore two omnidirectional hydrophones with a *70 kHz* resonance frequency (D1/170, Universal Sonar Ltd., U.K.) are employed. Each therefore has zero directivity index. The power amplifier is designed to apply approximately *20* $V_{p\text{-}p}$ to the transmitter hydrophone and the communication range of the system is assumed to be 100 m. The source level related to acoustic power output, P_a, is defined as:

$$SL = 170.8 + 10\log P_a \quad dB\ re\ 1\mu Pa\ @\ 1m \tag{5.11}$$

where the acoustic power, P_a, is the product of the electrical power, P_e, and transduction efficiency ($\eta = P_a/P_e$) is assumed as 0.8. Eq. 5.11 can be re-written in term of electrical power as:

$$SL = 170.8 + 10\log P_e + 10\log\eta \quad dB\ re\ 1\mu Pa\ @\ 1m \tag{5.12}$$

where $P_e=V_t^2 / R_{eq}$ (V_t is the rms voltage applied to the transmitter hydrophone during transmission and R_{eq} is its equivalent resistance, which was measured to be 276.6 Ω). The source level is calculated as 162.4 *dB re 1 μPa @ 1m* with the given parameters.

Since the transmitted signal level is decreased by the channel due to spreading and attenuation, the input signal at the receiver should be at an acceptable level for correct detection and therefore transmission loss must be computed. The transmission loss caused by spherical radiation and attenuation at the operating frequency of the system, is calculated for a range of 100 m. Assuming an attenuation coefficient, $\alpha(f)$, of 22 dB/km [103], the transmission loss is found to be 42.2 dB from Eq. 5.1 and Eq.5.2.

Before considering the noise level at the receiver and the detection threshold, the receiver level can be defined as:

$$RL = SL - TL \tag{5.13}$$

which is calculated as 120.2 dB. From this result, finding the input signal level at the terminals of the hydrophone is possible providing the hydrophone specifications are known. The receiving sensitivity of the hydrophone, R_s, at 70 kHz is found as -201 dB re 1V μPa^{-1} in Chapter 6. Since the detected voltage level at the receiver is the receiving sensitivity times receiver level, i.e.

$$20\log V_r = RL + R_s \tag{5.14}$$

this suggests that an output voltage level *91.20 μV rms* is expected at the output terminal of the hydrophone. The next step is to calculate the noise level at 70 kHz when the wave action at the surface is the dominant source. For this purpose the noise spectrum level is estimated as 30 *dB re μPa/Hz* from [99]. Then the total noise level is found as *72 dB* from Eq. 5.4, as the bandwidth of the hydrophone is equivalent to *18 kHz*

$$RL - NL \geq DT \tag{5.15}$$

The predicted power level at the farthest receiver was calculated earlier as 120.2 *dB*, and the noise level, as *72 dB*. For a signal to be detected this has to be significantly greater than the detection threshold. The detection threshold, DT, which is equivalent to the SNR at the receiver, plays an important role in this definition; in Eq. 5.15 and it is assumed to be 25 dB for the communication channel. More details on detection threshold can be found in [99, 100]. From these results, it can be seen that the receiver signal-to-noise ratio is greater than the detection threshold level. This suggests that a system with the source level assumed above is capable of communicating at longer distances.

5.7 Short Range Underwater Channel Data Transmission

In order to determine the response of a shallow and short range underwater communication channel, several experiments were conducted in the tank room in Department of Electronic and Electrical Engineering at Loughborough University. The tank *(9m long* x *5m wide* x *2m deep)* is filled with fresh water and the sound speed is assumed as 1500 ms^{-1} in this particular channel. Two omnidirectional hydrophones having a resonant frequency of 70 kHz were used for transmission and reception and positioned at 0.8 m deep and 1.5m distance from each other at the centre of the tank, as illustrated in Fig. 5.7.

In order to simulate a transmission rate of 50 bit/s, 100μs pulse with a 20ms pulse period as shown in Fig.5.8(a) is used to gate a signal generated at the resonant frequency of the hydrophone. For the given channel geometry and location of transmitting and receiving hydrophones, a surface scattered signal arrives at the receiver approximately 400 μs and a bottom scattered signal approximately 600μs after the direct path signal. It may also be observed that the direct path signal magnitude is greater than the multipath signal, hence decoding the signal may be easily done by employing a simple threshold detector. The amplitude difference between the direct path signal and reverberant signal is simply due to the long pulse

period; the multipath signal magnitude is negligible when the direct path signal is received. However, this kind of arrangement results in low rate data transmission. When the pulse period is decreased as shown in Fig. 5.8(b), aiming to increase transmission rate, the reverberant signal seriously influences signal detection, causing difficulty in distinguishing the direct path signal from multipath signal. Several other experiments were conducted in the tank to find the maximum transmission rate and it was observed that a 5ms pulse period, i.e. a transmission rate of 200 bit/s, was found to be satisfactory for this specific channel. Therefore, for the specified channel geometry, the transmission rate can be increased up to 1 kbit/s by utilizing the time delay between the arrival of the direct path and multipath signal as suggested in [146]. Experimental results for speech data transmission in the tank requiring much higher transmission rates are given in Chapter 7.

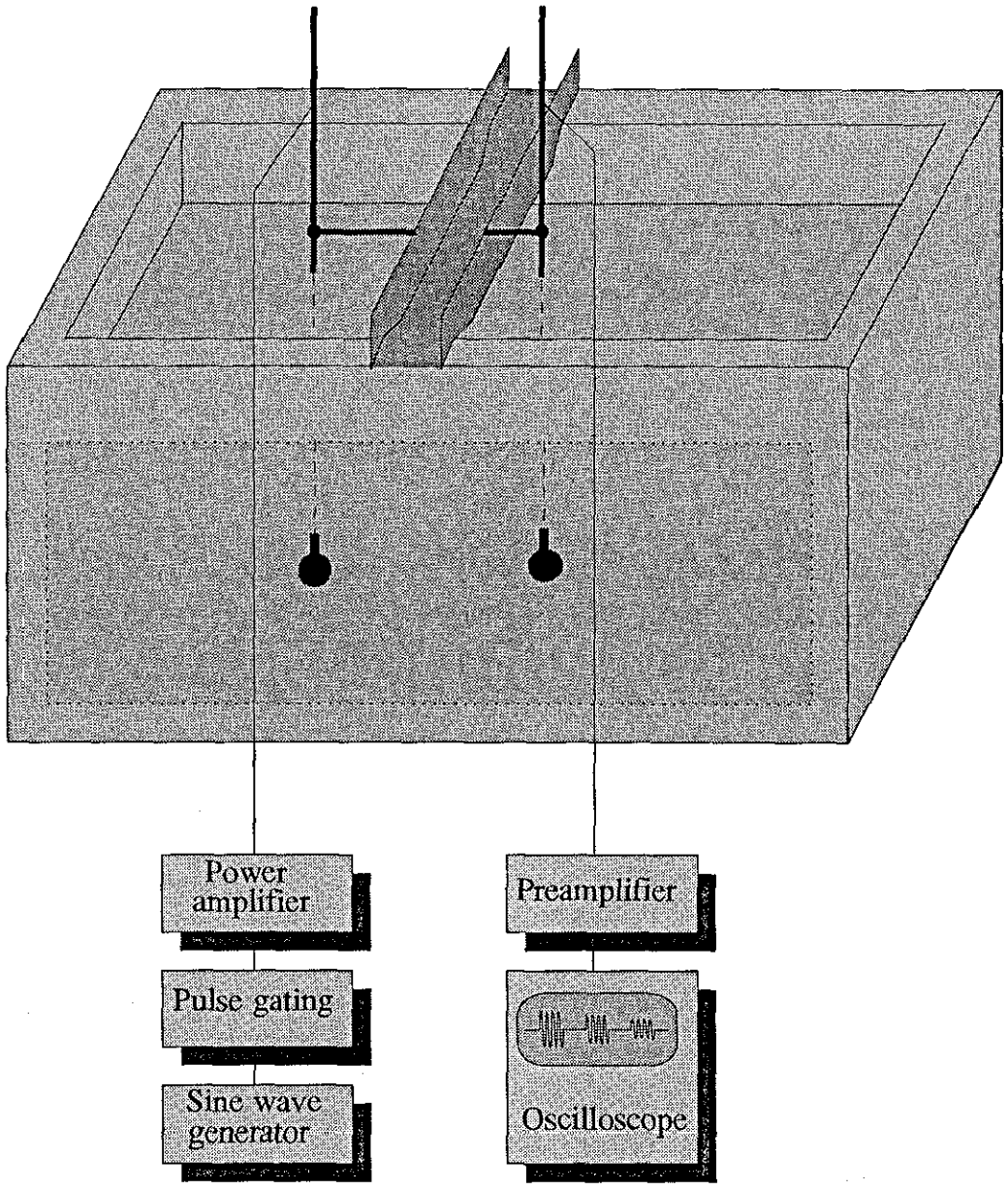
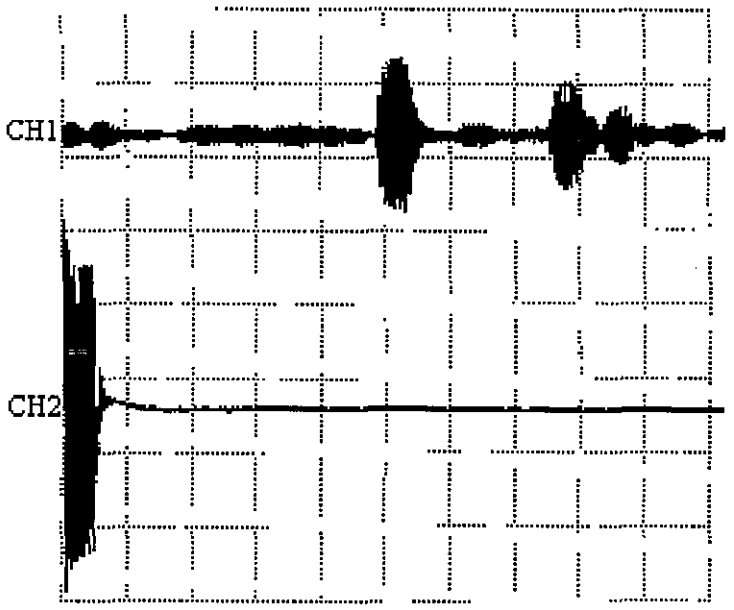
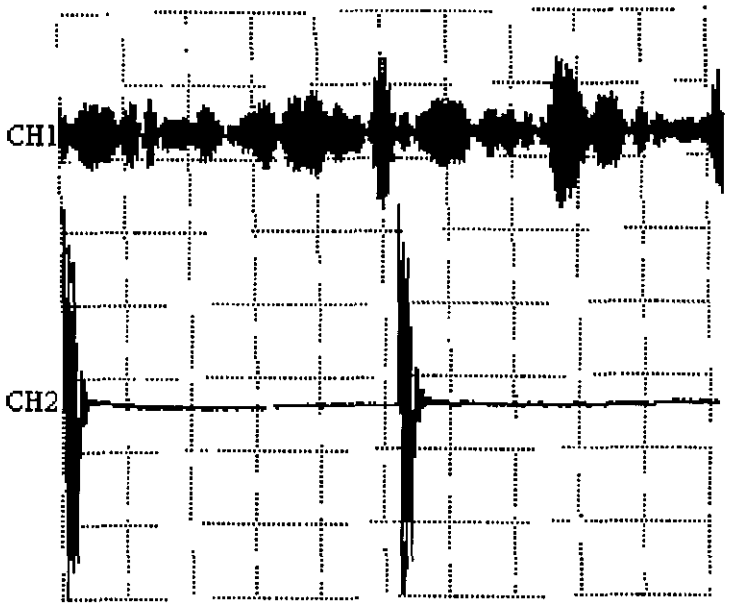


Figure 5.7 Experimental set up in the tank room to investigate data transmission



(a)



(b)

Figure 5.8 An illustration of data transmission in the tank. (a) Data rate of 20 bps (b)Data rate of 833 bps. (Time/Div=200 μs, Volt/Div=1 V for CH1 and 5 V for CH2).

CHAPTER SIX

DSP BASED UNDERWATER ACOUSTIC VOICE COMMUNICATION SYSTEMS

6.1 Introduction

As described in Chapter 2, most commercially available underwater acoustic voice communication systems employ analogue technology from the 1950s. Today's digital technology is widely used in many communication applications, but the same technology has not been adopted for underwater voice communications. In this chapter, an implementation of such technology is presented.

To achieve real-time, digital underwater acoustic voice communications, the speech signal must be processed, as described in Chapter 3. The processed speech samples must also be transmitted and received by implementing a suitable digital modulation and demodulation technique (refer to Chapter 5). However, these require fast and powerful processors; because of their processing power, available microprocessors are inappropriate for such applications. Therefore, the use of a digital signal processor (DSP) is essential to implement complex signal processing computations. This follows the trend of the last decade, which has witnessed a vast advance of DSP applications in real-time communication systems such as mobile telephones and in household products, such as televisions and hi-fi systems. Due to their sophisticated structure and small size, they minimize the number of components needed in a design, thus allowing the realization of a very small stand-alone, portable system. Moreover, because of the programmable structure of the system, modifications in speech coding, modulation and demodulation algorithms are feasible. The system designed here operates in two different modes:

In transmission mode operation, it must be able to:

(a) digitise analogue speech signal

(b) compress (analyse) speech samples

(c) encode compressed speech parameters

(d) implement digital PPM for the transmission of speech parameters

In receive mode operation, it must be capable of:

(a) receiving the digital PPM signals and decoding speech parameters,

(b) synthesising speech signals

(c) producing analogue speech signals.

Apart from the above functions, in the design of the system, private communication facility between selected units and providing an emergency communication channel are considered, however experimental results are not available at present. This chapter is dedicated to the hardware and software design methodology of a digital underwater acoustic voice communication system utilizing most of the above functions.

6.2 System Architecture

The system structure is shown in Fig.6.1. It may be described as both a transmitter and a receiver with associated algorithms. The transmitter section consists of a keypad unit, a speech pre-processing circuit (preamplifier, anti-aliasing filters and analogue-to-digital converter), a power amplifier and a transmitting transducer. The receiver section includes the same transducer as in the transmitter section as well as a preamplifier, bandpass filter, envelope detector and analogue-to-digital converter. At the speech output end, a post-processing circuit, which includes a digital-to-analogue converter, lowpass filter and amplifier, is also provided. In the design of the system, a TMS320C31 Texas Instrument DSP [146] was chosen because of its high performance in signal processing applications. This is the core of the system and controls operation of the transmitter and the receiver sections and it executes speech

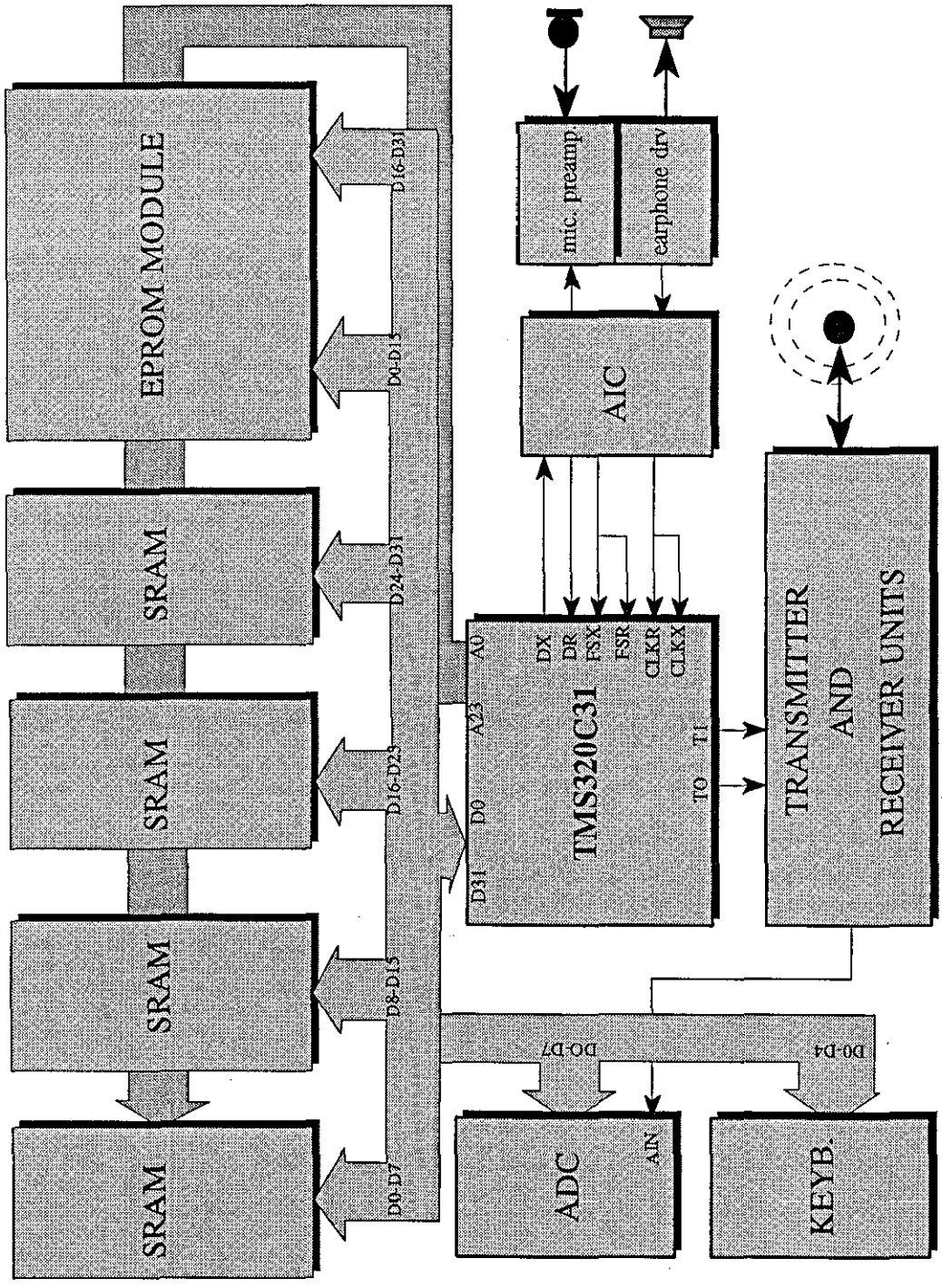


Figure 6.1 Block diagram of the digital underwater acoustic voice communication system

processing algorithms (details can be found in Chapter 3). The analogue processing and digitising of speech signals is done by an Analogue Interface Circuit (AIC) [147] that performs both in the transmitter and the receiver. In the construction of the system, a modular based approach is considered. This provides the advantage of easy and cost effective hardware modifications, although the physical size of the system is increased.

This approach is acceptable for the primary design stage of the system. Once its operation meets the designer's demands, then the secondary design phase, i.e. miniaturisation of the system, can be considered. This can be achieved by re-engineering the system with surface mount components.

6.3 Digital Signal Processor Structure

In the previous studies, it has been suggested that the use of fast and powerful processors is essential to implement speech encoding and decoding, digital modulation and demodulation of encoded speech data (discussed in Chapters 3 and 5) in real-time. Apart form accomplishing signal processing, the processor must also maintain complete control of the system and perform communication tasks with peripheral units in the system. In this prototype, depicted in Fig.6.1, a 32-bit TMS320C31 (33 MHz) DSP is utilised. It is a third generation Texas Instrument (TI) DSP, which is capable of executing 33.3 million floating point operations per second (MFLOPS), and it is widely used where complex mathematical and fast signal processing applications are required.

The TMS320C31 DSP, whose architecture is shown in Fig. 6.2, accommodates key design requirements of small size and +5V single supply operation. The other features are two general purpose 32-bit programmable timers, TCLK0 and TCLK1, capable of generating clock signals from 970 μHz to 8.325 MHz and one serial port supporting a maximum data transmission rate of 4.1625 Mbit/s. Allocated registers for the timers and the serial port can be programmed to perform a wide variety of tasks, and the timers are effective in the generation of accurately timed, high-speed pulse sequences

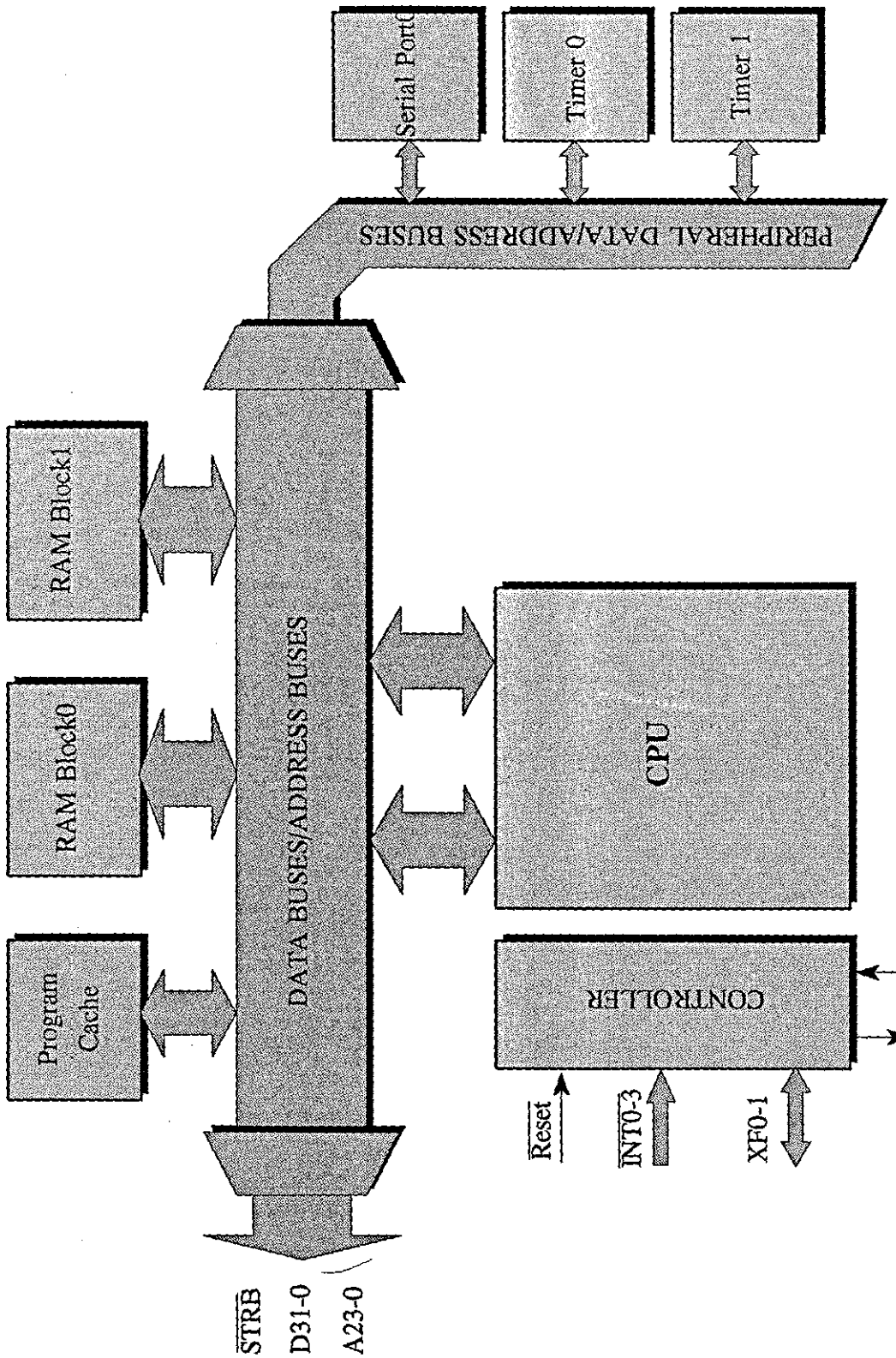


Figure 6.2 Block diagram of the TMS320C31 DSP

for the different pulse characteristics. Both the timers and the serial port utilise internal interrupts and their functions will be discussed later. The DSP also incorporates two general purpose input/output ports, XF0 and XF1, which can be operated under software control. Besides the above features, the DSP has four external interrupt inputs, INT0 to INT3, allowing efficient communications with the external peripheral units.

Moreover, the DSP features constitute *2kWord* internal Static RAM and *64 Word* program cache memory, supporting maximum efficiency in program execution time. The on-chip SRAM can be used either to execute programs from slow speed off-chip memory devices or as temporary memory to store speech samples and relevant data. The 24-bit address bus of the DSP enables expansion of the off-chip memory to a maximum of *16.77 Mword.* In the current design, due to the limited size of the internal SRAM and the exclusion of internal EPROM, off-chip memory devices are incorporated, SRAMs as the data memory, EPROMs as the program memory. Therefore , four KM6865-20ns *8 x 8k* SRAMs and four NMC2764-150ns *8 x 8k* EPROMs are configured to form *32 x 8k* as separate modules and they are interfaced to the DSP as given in Appendix A.

The EPROM memory allows the system to perform the required speech processing (analysis and synthesis), and to execute the transmission and receiving algorithms except for the time consuming computationally complex algorithms. During the system initialization, the speed critical algorithms in EPROMs, are transferred to the internal SRAM of the DSP and executed there.

6.4 Memory Organization and Access to Peripheral Units

For the successful operation of a system incorporating a microprocessor or a DSP, the unique requirement is to provide communications within the system's peripherals. This can only be achieved by allocating each peripheral to a specific address. The same process is valid for the voice communication system presented in this chapter. An address map of the system is organized as shown in Table 6.1. Access to the peripherals is administered by the DSP and by an address decoder. A 74HCT138

octal-to-decimal decoder, as shown in Fig.6.3, is utilized to ensure that only one external peripheral is selected at a time. The most significant three bits of the address bus, A23-A21, are employed to define the memory map. They are applied to the decoder's input in order to access eight possible peripherals. In this application, since the system incorporates the EPROM, the SRAM, the receiver ADC and the keypad unit as the external peripherals, only four outputs of the decoder are employed. They provide $\overline{CS}$ (chip select) signals for the peripherals. Moreover, the decoder is disabled to eliminate any possible bus conflict when internal peripherals are accessed. This is controlled by the $\overline{STRB}$ output of the DSP which is active only when the addresses of an external peripheral is loaded into the address bus.

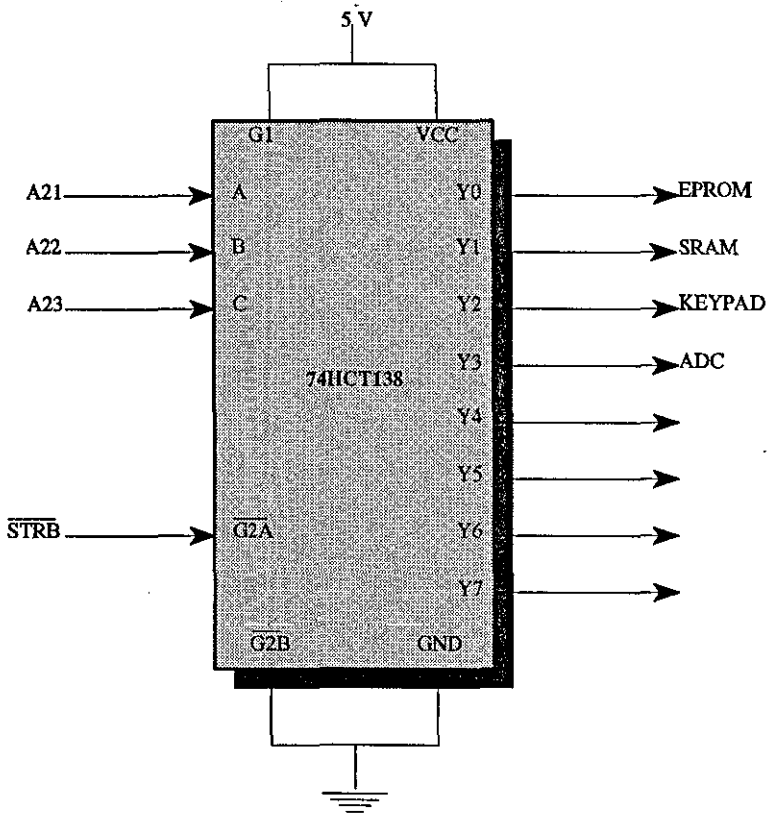


Figure 6.3 Address decoder for the system.

Address	*Usage*
000000_h $00003F_h$	Interrupt Locations (*EPROM*)
000040_h $1FFFFF_h$	EPROM
200000_h $3FFFFF_h$	External RAM
400000_h $5FFFFF_h$	Keypad
600000_h $7FFFFF_h$	Receiver ADC
808020_h $80802F_h$	Timer 0 Registers
808030_h $80803F_h$	Timer 1 Registers
808040_h $80804F_h$	Serial Port Registers
809800_h $809FFF_h$	Internal RAM
$80A000_h$ $FFFFFF_h$	Not Used

Table 6.1 Memory allocation of the system

6.5 Interrupt Management and Operational Control

Although four external peripherals are incorporated in the system, their number can always be increased if a modification in the design is required. Communications with these peripherals is essential for the DSP to perform its operations. Some of the peripheral units, such as the EPROM and SRAM are continuously accessed under the entire control of the DSP. Some only instruct the DSP through its interrupt inputs to initiate communication. The last one enhances the system's operational speed since

the DSP does not need continuously to access these units. The only drawback is that a limited number of peripheral units is allowed to utilize this feature of the processor. This is determined by the number of interrupts of the processor. The TMS320C31 DSP supports multiple external and internal interrupts, as illustrated in Table. 6.2. These are controlled by appropriately setting the internal *Status Register* (ST), *Interrupt Enable Flags* (IE) and *Interrupt Flags* (IF) of the DSP. During the initialization stage of the registers, appropriate interrupts are enabled by configuring the IE flags. While the system is operating, an interrupt request maybe received either from an internal or an external peripheral. A corresponding IF bit presents an access to the specified interrupt subroutine. In this system, each interrupt input associates an interrupt subroutine in the software. An efficient partitioning of the interrupts and the management of priorities are important because they will improve the operational performance of the system. While the external interrupts are allocated for the external peripherals, internal ones are for the internal peripherals. The external peripherals (the receiver ADC, the keypad encoder, the transmitting/receiving mode selection network) and the internal peripherals (the timers, the serial port) produce interrupt requests to the CPU. As soon as an interrupt request is received from a peripheral, the corresponding interrupt subroutine is accessed and communication with the peripheral is established. The other interrupts are ignored. This may cause serious problems if data from the peripheral are essential. Therefore, prioritisation of the interrupts must be defined.

Although the priority of the interrupts is set by the DSP, as illustrated in Table 6.2, this can be modified in software. The priorities are organized according to the operation mode of the system and the most important interrupt request is given the highest priority For example, in the transmission mode operation, the interrupt produced by **Timer 1** (TINT1) is the most significant due to accuracy requirement in timing in DPPM. Therefore, the highest interrupt processing priority is given to TINT1. During the receiving mode operation, the internal interrupt (TINT0) produced by **Timer 0**, operating at the sampling clock of the receiver ADC, has the highest priority. TINT0 is used in the demodulation of DPPM signals and given the highest priority since accuracy in timing measurement is important. External interrupt **INT0** is also utilized during the receive mode operation for reading samples from the ADC. More details

about these interrupts and their operations are presented in Chapter 7. In the design, **INT1** and **INT2** external interrupts of the DSP are employed to provide communication with the keypad unit and to select the operation mode of the system, respectively. The external interrupts are assigned to low interrupt priorities since they do not significantly affect the performance of the system.

initial priority	*Interrupt*	*Usage*
1	INT0	Receiver ADC (data ready)
2	INT1	Keypad (number dialed)
3	INT2	Mode selection (press-to-talk switch)
4	INT3	Not used (left for future use)
5	XINT0	Not used
6	RINT0	AIC (speech sample received)
7	XINT1	Reserved
8	RINT1	Reserved
9	TINT0	Demodulation of PPM signal (in Rx mode)
10	TINT1	Modulation(PPM transmission, in Tx mode)

Table 6.2 DSP interrupts and their utilization

6.6 Speech Signal Conditioning and Digitization

Speech signal conditioning and digital or analogue conversion are achieved by utilizing an application specific speech codec (TLC320AC01 Analogue Interface Circuit (AIC) of Texas Instrument [147]) in the hardware. The AIC has a synchronous, serial, digital interface designed for ease of connection to many DSP chips. Its configuration with the TMS320C31 DSP is illustrated in Fig.6.4.

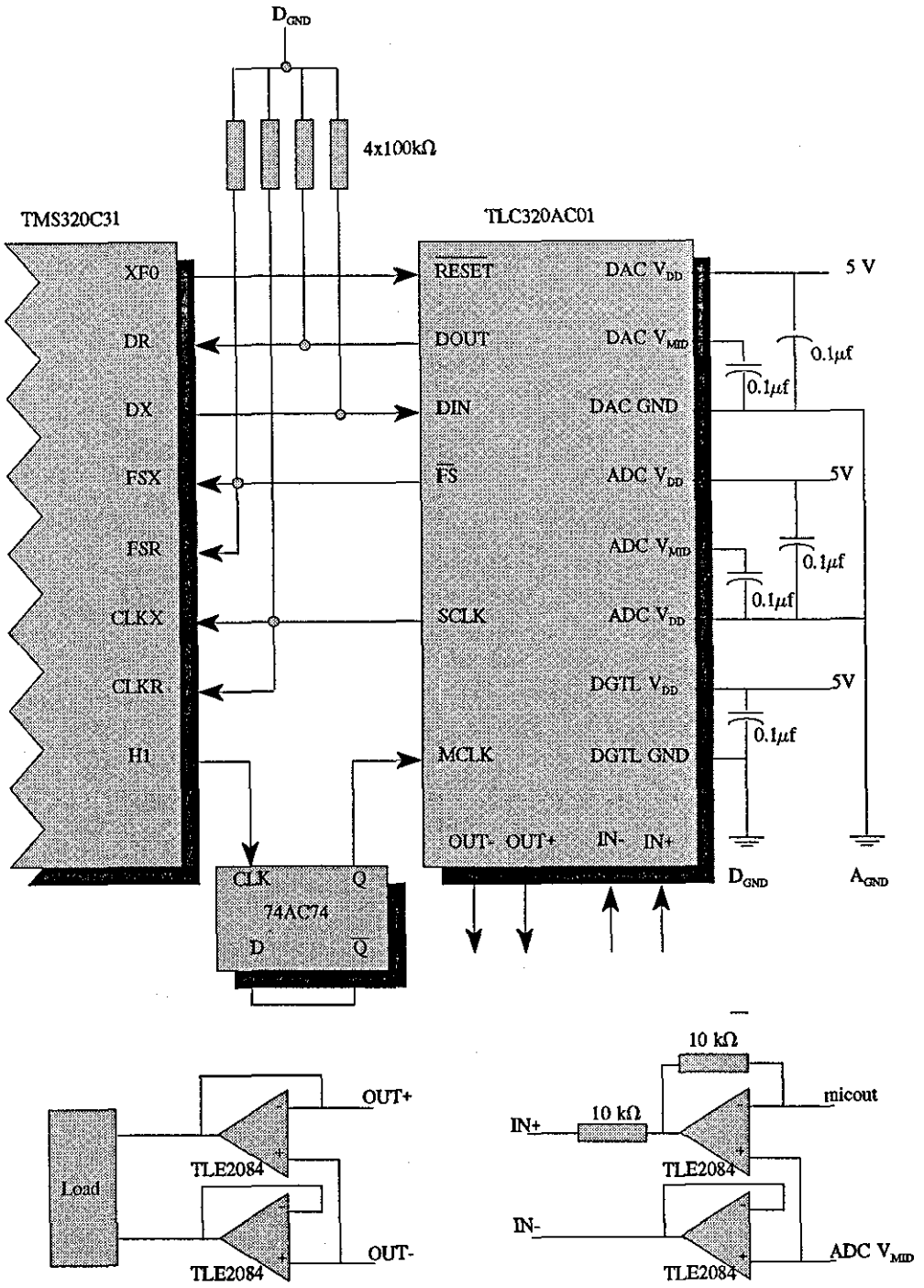


Figure 6.4 Configuration of the TLC320AC01 AIC with the DSP

Register	Initialized Data
Register 1 (A Register)	$(011E)_h$
Register 2 (B Register)	$(0212)_h$
Register 3 (A' Register)	$(0300)_h$
Register 4 (Amplifier and Gain Register)	$(0419)_h$
Register 5 (Analogue Conf. Register)	$(0501)_h$
Register 6 (Digital Conf. Register)	$(0600)_h$
Register 7 (Frame-Sync Delay Register)	$(0700)_h$
Register 8 (Frame-Sync No. Register)	$(0801)_h$

Table 6.3 Internal registers of the AIC and their initialization.

The AIC comprises software selectable variable gain amplifiers providing common mode rejection ratio (CMRR) of typically 55 dB, anti- aliasing filter (a low pass and a high pass) and reconstruction filter built in switched-capacitor technology. Eight internal registers, listed in Table 6.3, are supplied to control these features of the AIC. In addition, a 14-bit Analogue-to-Digital Converter (ADC) and Digital-to-Analogue Converter (DAC) are included. A serial port which enables the AIC to communicate with the DSP is also built within the component. In this study, since the design of a portable system is aimed, the choice of the surface-mount TLC320AC01 in single supply operation provides an optimum solution as a speech signal conditioning device.

The input speech signal to the AIC is supplied by a CF2949 electret-condenser close range speech microphone having an in-built preamplifier circuitry (receiving sensitivity of -65 dB re V/μbar) as shown in Fig.6.5. The microphone is connected to the AIC in single-ended input configuration where all single-ended inputs are referenced to the internal reference voltage (ADC V_{MID}) rather than ground, presented in Fig.6.4. This allows single supply operation.

With such a configuration, a maximum input signal to the AIC of *3* $V_{p\text{-}p}$ is allowable. However, this does not utilize the full scale of the ADC, which is *6* $V_{p\text{-}p}$ so the input speech signal must be amplified, either internally or externally. As mentioned earlier,

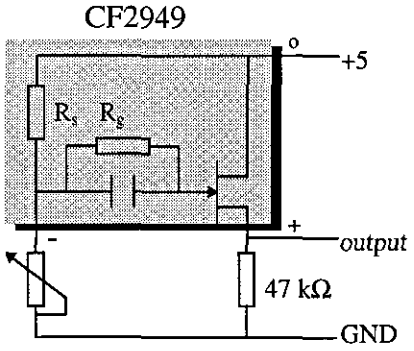


Figure 6.5 Microphone connection and preamplifier circuitry

the AIC incorporates an input amplifier with software selectable gain of 0, 6, or 12 dB as well as an attenuator (squelch mode, 60 dB attenuation). A 6 dB gain is adequate to use full scale conversion and this is selected by setting the contents of the *amplifier gain select register* (Register 4) to $(0419)_h$. The amplified speech signal is then passed through the anti-aliasing filters to the ADC. In the following, the bandwidth settings of the filters and the sampling frequency for the ADC are studied. The required sampling rate for the elimination of the aliasing effect is dependent on the quantization level and bandwidth of the speech signal of interest. The anti-aliasing filter should attenuate the levels of the frequencies in the stop band to less than the *rms* quantization noise level for the ADC [148] so that they are not detected by the ADC.

Since the internal filters are constructed using switched-capacitor technology, they require a timing clock (FCLK) to determine the filters' cut-off frequencies. Two internal registers of the AIC (*A Register* and *B Register*) are used to define the filters' bandwidth and the sampling frequency. The FCLK input to the AIC is derived from the input master clock (*MCLK=8.325 MHz*) of the DSP and from the content of the A register as [147]:

$$FCLK = \frac{MCLK}{(A\ Register\ Value).2} \tag{6.1}$$

The content of the A Register is defined by the lowpass filter cut-off frequency (f_{LP}) which varies with the type of application. The frequency performance of the lowpass switched capacitor filter is proportional to the **FCLK** of the AIC. In practice, this is

usually chosen to be considerably higher than the highest signal frequency component to be filtered [149]. For the hardware presented here, this relation is defined in [150] as:

$$f_{LP} = \frac{FCLK}{40} = \frac{MCLK}{40.(A\ Register\ Value).2} \tag{6.2}$$

Since f_{LP} is chosen as *3.4 kHz* for speech communications, *A Register Value* can be calculated as 30_d, $(1E_h)$. Hence FCLK is found to be *138.75 kHz*. The *B Register* is used for sampling frequency definition:

$$f_s = \frac{MCLK}{(A\ Register\ Value)(B\ Register\ Value).2} \tag{6.3}$$

The ratio of the sampling rate to the anti-aliasing filter corner frequency is set by the *B Register Value*. To satisfy Nyquist's sampling theorem and to prevent the aliasing effect on the digitized speech signal, the sampling frequency must be at least twice the f_{LP} frequency, as defined:

$$\frac{f_s}{f_{LP}} \geq 2 \quad , \ i.e.\ (B\ Register\ Value) \leq\ 20 \tag{6.4}$$

For voice communication systems, an *8 kHz* sampling frequency is found adequate and implemented in many speech processing applications. After considering the above definitions, the *B Register Value* is computed as 18_d (12_h). Since a switched-capacitor high pass filter is also included in the signal path, its corner frequency, f_{HP}, is given by:

$$f_{HP} = \frac{f_s}{200} = \frac{MCLK}{200.(A\ Register\ Value)(B\ Register\ Value).2} \tag{6.5}$$

After setting the register values and defining the sampling frequency, it can be said that the speech signal spectrum occupies the frequency band of 40 Hz to 3400 Hz. To

provide this bandwidth, both lowpass and highpass filters are included in the signal path and this is done by setting the content of the *analogue configuration register* of the AIC to (0501_h). Then the filtered speech signal is applied to the ADC operating at a sampling frequency of 8 kHz with *14-bit* resolution. For the ADC, the quantization step size is given by:

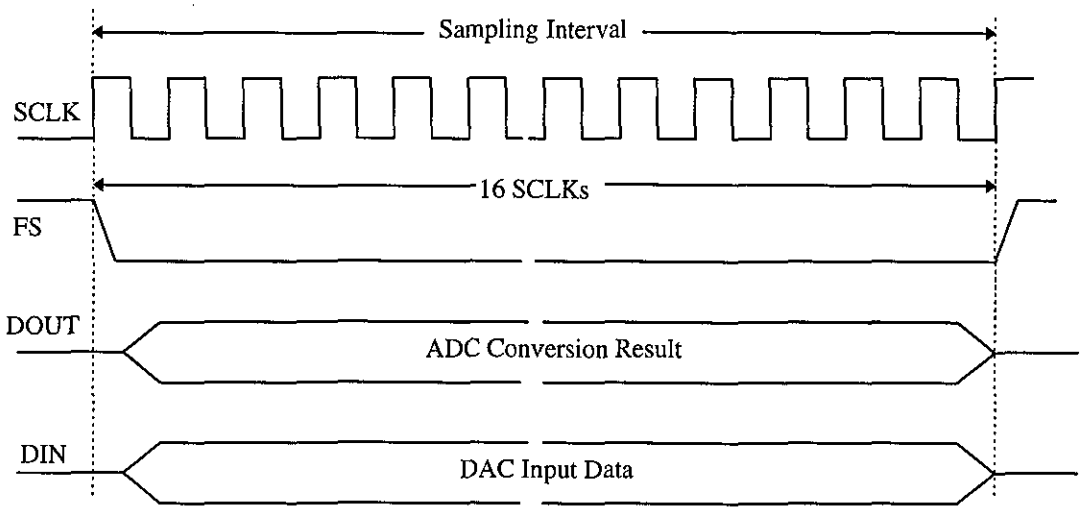
$$\Delta V = \frac{V_{FS}}{2^{14}} \tag{6.6}$$

where V_{FS} is the dynamic range of the signal, i.e. $6\ V_{p\text{-}p}$, hence the ΔV is found to be 366.23μV. The maximum quantization error is defined as $\pm\frac{\Delta V}{2}$. This error is assumed to be random and uniformly distributed. For such a case, the quantization noise power is given by $\frac{\Delta V^2}{12}$ [148]. The signal-to-quantisation noise power ratio (SQNR) can be used theoretically to define the recommended stopband attenuation level of the bandpass filter. The SQNR figure for the 14-bit AIC is found to be 86.02dB.

The ADC converts the analogue speech signal value at the sampling time into digital in 2s-complement format and it forms the data in 16-bit (most significant 14 bits are the sample value, the rest are control data for the AIC). These 16-bit digital words are clocked out through **DOUT** of the serial port as shown in Fig.6.6a. At the same time, quantized speech samples, processed in the DSP, are transmitted to the AIC through **DIN** in 16 bits format. The digital speech samples are converted to analogue voltage by the DAC and then passed through a *(sinx)/x* correction circuit [151] and a smoothing filter. At the audio outputs of the AIC, an analogue speech signal is produced.

In the analogue output signal path of the AIC, an amplifier with software selectable gain is also included. The gain factor can be adjusted by setting appropriate bits of the *analogue configuration register* as it was in the input signal path. However, in this

application, the speech signal level at the output port is found satisfactory, therefore, no signal amplification is introduced.



(a)

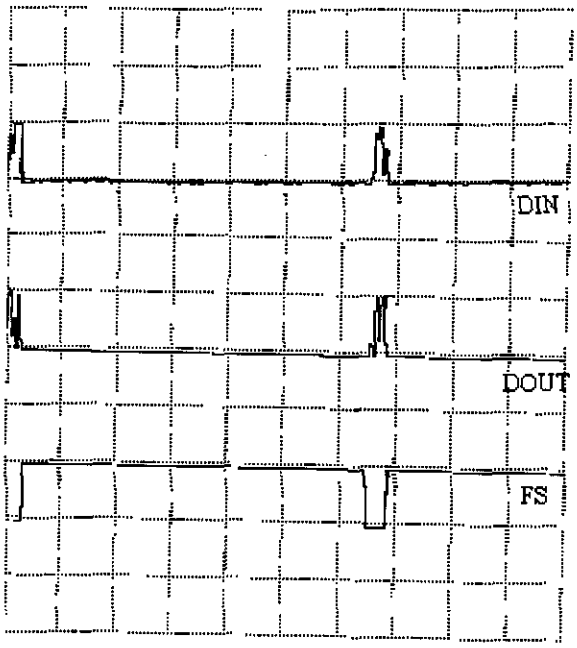


Figure 6.6 AIC timing (a) detailed illustration (b) measured waveforms (20μs/Div, 5V/Div)

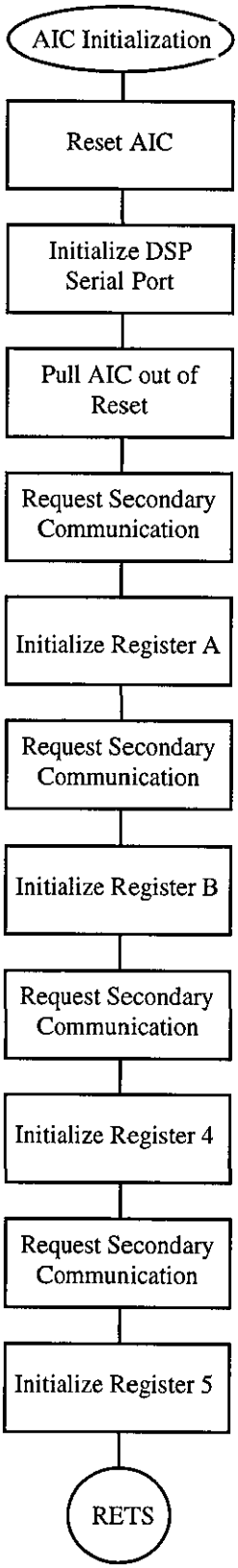


Figure 6.7 Initialization of AIC

6.7 Keypad Unit

When designing the digital underwater acoustic voice communication system, it was aimed to achieve a close approximation of a mobile telephone and to provide a private communication facility for the user. Therefore, a 4x4 keypad unit is included and configured as shown in Fig.6.8. Since the keypad will be subjected to full ambient pressure in water, it must be waterproofed and the keys must spring back after being pressed. For this facility, a keypad encoder IC, a Motorola NM74C922, is utilized. When a key is pressed, INT1 of the DSP is triggered to execute an associated interrupt subroutine. Nevertheless, due to time limitation, the software design has not been fully completed. Therefore, further study is recommended for its future application.

In the design of the keypad unit, it is assumed that each system can be coded with a personal ID number like in personal telephones. This number is expected to be dialed by the user, who requests communications, to establish a voice communication link. For successful operation of the system, a private communication protocol is essential. A possible design, as illustrated in Fig.6.9, can be utilized to allow only one pair of users to communicate in a band limited channel.

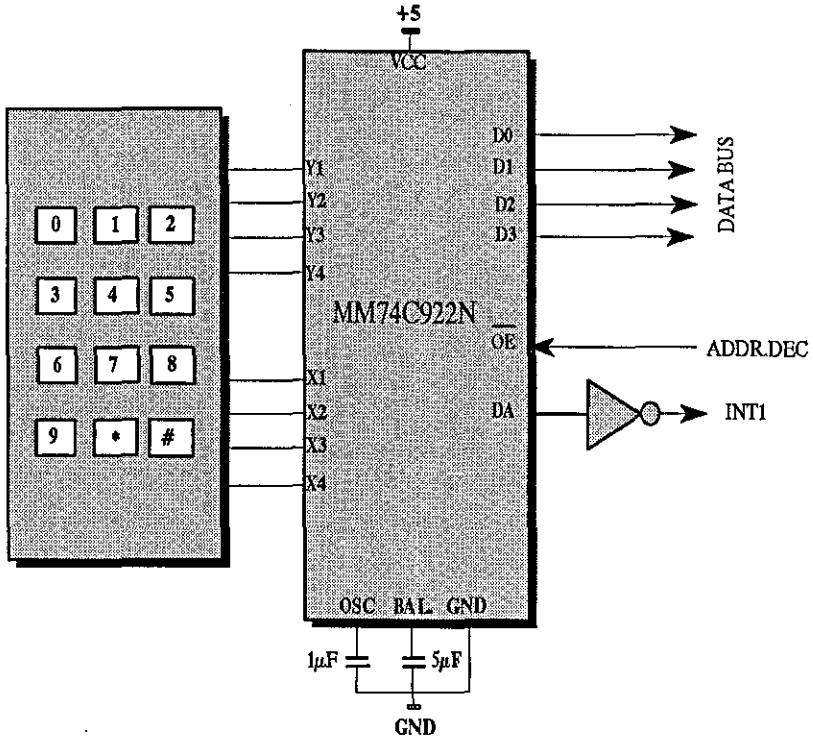


Figure 6.8 Design of keypad unit

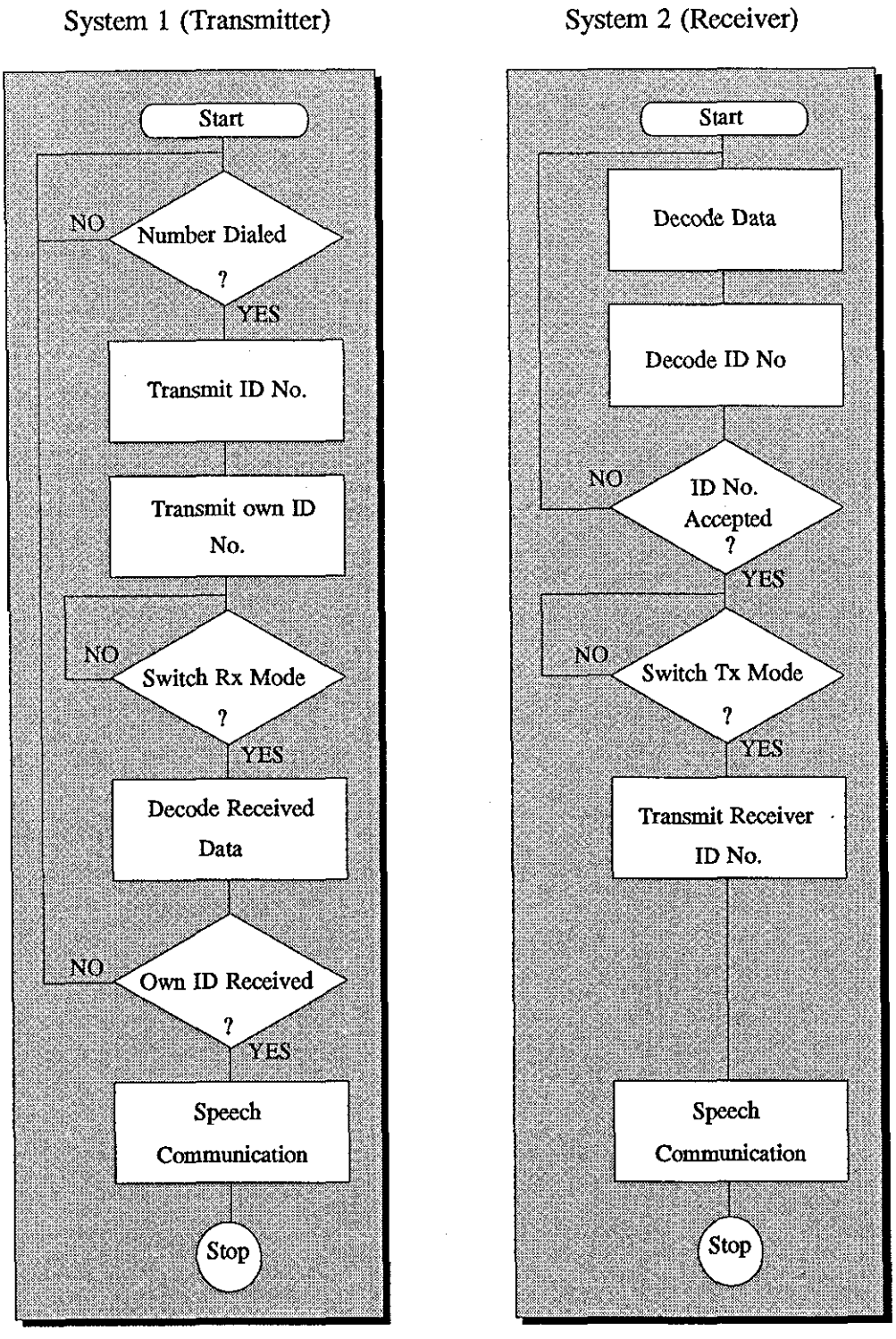


Figure 6.9 A possible private communication protocol recommended for the keypad unit.

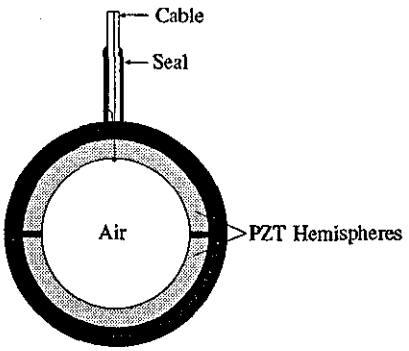
6.8 Transmit/Receive Transducer

The underwater voice communication system described here is based on transmission and reception of acoustic pulses through water. It requires an appropriate transducer which can convert electrical energy into mechanical energy or produce an electrical signal when excited by a pressure wave. Piezo-electric transducers [152] are commonly used for this type of application because of their robustness, cost and variety of shapes and sizes. Natural occurring crystals, such as quartz, exhibit the piezo-electic effect, i.e. an electrical signal is generated when the crystal is compressed and mechanical distortion is produced with the application of voltage. Commonly used man-made electrostrictive transducers, such as lead zirconate titanate (PZT) ceramics, exhibit the same effect. A single ball-shaped transducer (D-170 Universal Sonar Ltd., U.K), as shown in Fig. 6.10a, operating at its resonance frequency and radiating spherically is employed for the current system as both projector and hydrophone. The transducer can be modelled as a series tuned LCR circuit as shown in Fig.6.10b [101] where the L_m and C_m are compliance of the element. The resistive component in the LCR circuit has been split into two components as loss resistance, R_L, and radiation resistance, R_r. Since the drive elements are ceramic acting as a dielectric material with electrodes, the capacitor C_o, called static capacitance, is present in parallel with the LCR circuit. We can define the overall voltage transfer function for the given electrical model as:

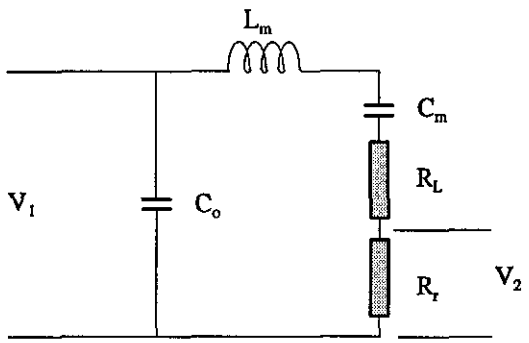
$$|H(f)| = \frac{\omega_o C R_r}{\sqrt{\frac{1}{Q^2} + \left(\frac{\omega}{\omega_o} - \frac{\omega_o}{\omega}\right)^2}} \tag{6.7}$$

where $\omega_o = \frac{1}{\sqrt{LC}}$ and $Q = \frac{1}{\omega_o C(R_L + R_r)}$

The electrical characteristics of the omnidirectional hydrophone are measured by using HP-4192A Hewlett-Packard impedance analyser and these are listed in Table 6.4. The susceptance and conductance plot of the hydrophone is given in Fig.6.11. These measured parameters are utilized in the design of a power amplifier circuitry.



(a)



(b)

Figure 6.10 Illustration of a ball-shaped hydrophone and its electrical representation.

Resonant frequency (f_o)	69.8 kHz
Q factor	4.1
Low cut-off frequency	61 kHz
High cut-off frequency	78 kHz
Susceptance in water (B_o)	3.117×10^{-3} S
Conductance in water (G_o)	3.615×10^{-3} S
Equivalent resistance (R_s)	276.6 Ω
Static capacitance (C_o)	7.107 pF
Bandwidth :	18 kHz

Table 6.4 Properties of the hydrophone at resonance.

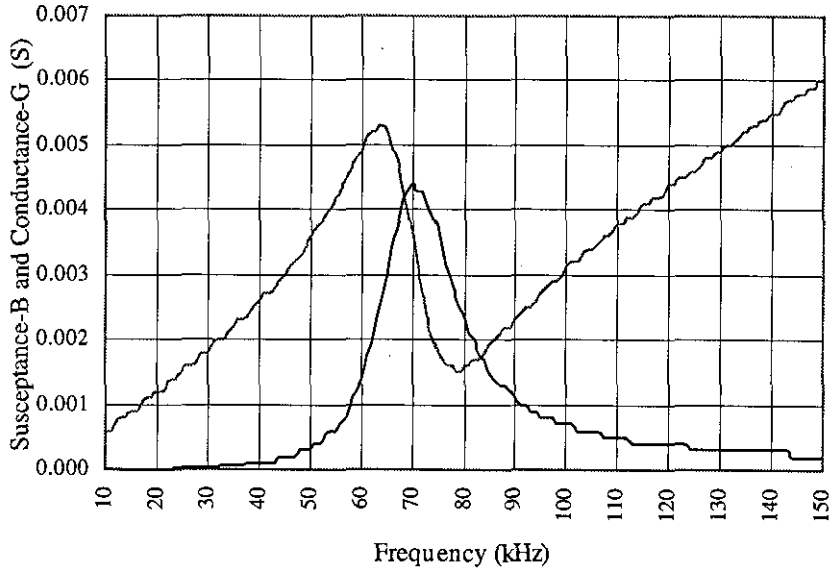


Figure 6.11 Conductance and Susceptance measurements of the hydrophone

6.9 Transmit and Receive Mode Selection

The system is designed to operate in half-duplex mode. Its operation mode, as a receiver or as a transmitter, is controlled by the user. There are different methods available to regulate the operation mode. Voice-activated switches and press-to-talk switches (**PTT**) are amongst the ones widely implemented in the analogue systems. A PTT switch is utilized because it provides the user with a total control over the

communication process. This also reduces the hardware and the software complexity. The PTT switch circuitry is configured as illustrated in Fig.6.12 and it is used to control an analogue switch (**MAXIM HI3-0201HS** [153]). When the PTT switch is pressed (logic high output), transmission mode (**Tx** mode) is selected. When it is released (logic low output), receive mode (**Rx** mode) is selected.

During transmission, the solid-state MAXIM HI3-0201HS switch at high impedance isolates the receiver amplifier although it has the limitation of a maximum permitted voltage level of 30 V_{p-p} across its terminals. The solid-state changeover network is an essential part of such a design [101] since it is inappropriate to connect the receiver amplifier directly to the transmit/receive transducer while the system is transmitting. As shown in Fig.6.12, parallel diodes, BAT85, in series with the secondary winding of the transformer are considered a part of the switching network. They present a low impedance connection between the transformer and the transducer during transmission. On reception, the voltage level across the transducer terminals will be small; as a consequence, the diodes will present a relatively high impedance, effectively isolating the transmission circuitry from the transducer. Moreover, the analogue switch in the ON state produces a low impedance connection between the transducer and the receiver amplifier.

So far a hardware design for the operation mode selection has been described. However, as emphasized in the previous chapters, these operation modes execute different algorithms. Therefore, the DSP must be informed about which operation mode is activated by the user and which algorithms are required. This is done by configuring the output of the **PTT** switch circuitry to provide a control signal to the INT2 of the DSP.

6.10 Power Amplifier Design

The primary requirement of the transmitter is to efficiently deliver adequate power to the hydrophone at the desired frequency (70 kHz) and to achieve a desired communication range (approximately 100 m). Therefore, a power amplifier is required.

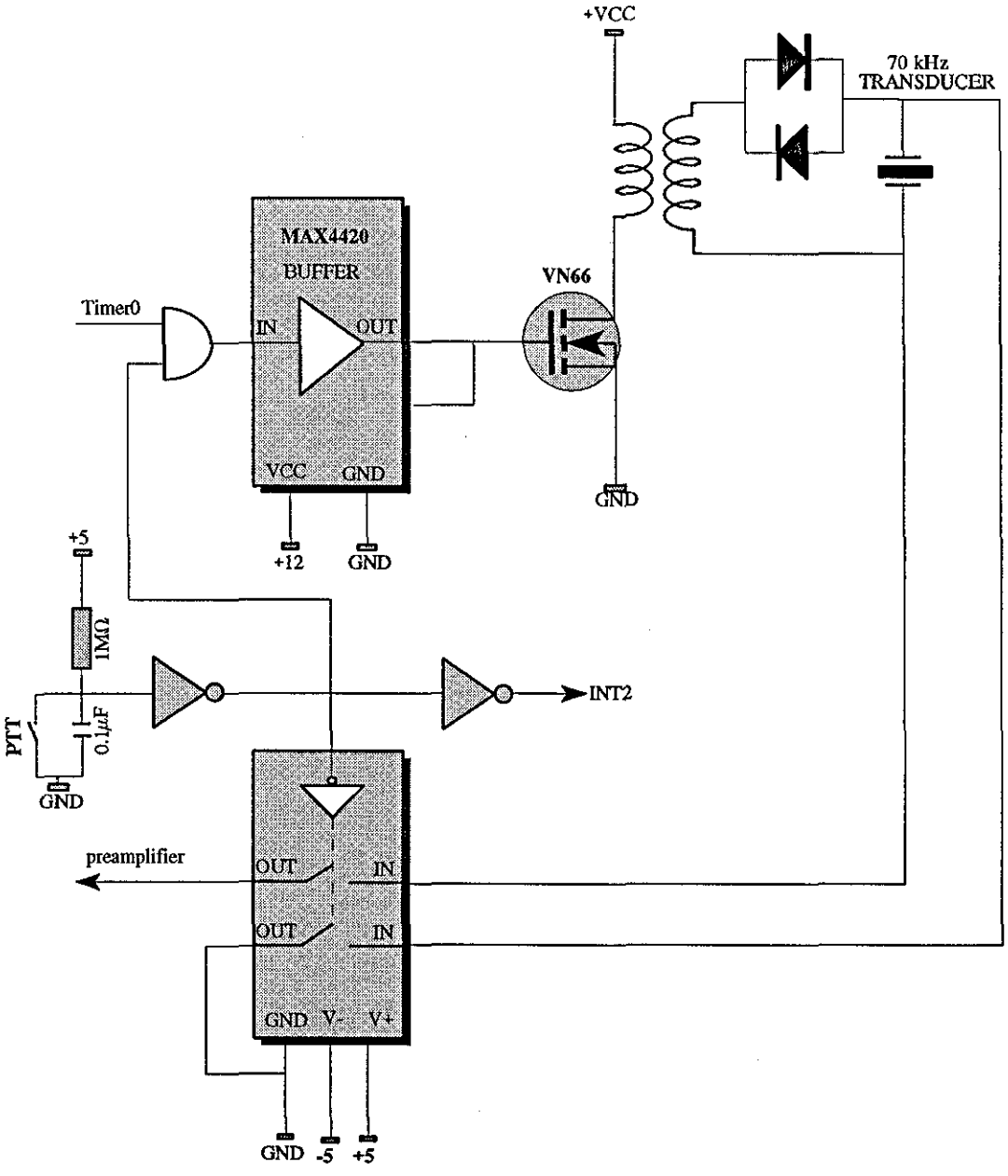


Figure 6.12 Illustration of communication mode selection and power amplifier circuitry

There are various types of power amplifiers which are identified by their classes of operations such as A, B, C, ... etc, and their operations are well defined in literature [154]. Classes A and B are linear amplifiers which produce a magnified replica of the input signal and typically used in AM and SSB systems. Class A amplification is an inefficient form of signal amplification because current taken by the active device from the power supply is independent of the input signal amplitude and it remains constant even when the input signal is removed. The maximum efficiency for this is 25% but, in practical systems, it is even less. The class B amplifiers use two matched devices in push-pull. A great advantage of this class is that the combined current of the two devices taken from the supply is proportional to the amplitude of the input signal and the current is very low in the absence of the input signal. Theoretically, the efficiency of this amplifier is about 78% and is implemented in mobile telephones [155]. However, to improve efficiency in transmission where linear amplification is not required, class C, D or E amplifiers are utilized. Class C amplifiers can be used only for the amplification of constant-amplitude sinusoidal signals. Hence this class amplifier is more often applied for FM transmission [154], providing a high efficiency of around 90%. Specifically, class D and E amplifiers are suitable to operate in switched mode. The increased efficiency of these amplifiers results from techniques which reduce power dissipation. In switch mode amplifiers, this degree of efficiency is accomplished by employing active devices as switches rather than current sources as in classes A, B and C. In class D operation, a pair of active devices are used with a tuned output circuit. The active devices act as a two-pole switch, which is regularly switched at a high frequency between the off state or on state at all times. The efficiency is increased with this amplifier of about 90%. However, the class D amplifier, which uses of a pair of drivers, requires the input signal magnitude at least to be equal to the supply voltage. On the other hand, the class E amplifier, which employs a single active component as a switch and provides approximately 90% efficiency in power transmission, has become attractive in recent years. A detailed study of class E amplifier can be found in [156-159].

Before choosing a class of amplifier for underwater acoustic transmission, it is important to have *a priori* knowledge of the waveform which will be transmitted. As an example, for the analogue underwater voice communication systems implementing

AM or SSB techniques, the classes A and B are suitable since the modulated carrier signal is required to be amplified and transmitted. However, for underwater digital data transmission, such as telemetry and digital voice communications, generating square waveforms at desired frequencies is a relatively simple task. Therefore, it will be convenient to select an amplifier which allows this type of waveform as driving signal. This criterion is only valid for switched mode amplifiers which provide high efficiency at the range of 80-90% in spite of their simple hardware [154].

6.10.1 Implementation of Power Amplifier

The features of a suitable power amplifier are low output impedance, high efficiency and adequate current output capacity. After the above discussion, it is suggested that the class E amplifier is the most suitable one for the transmission of DPPM signals since it utilises a single switching device and it provides high efficiency. In its implementation, *a priori* knowledge of the operating frequency and communication range are required. Moreover, hydrophone parameters are also essential in the design of the power amplifier.

The amplifier is designed with VMOS power FET, VN66AFD. This allows a maximum 3 A of drain current in switching mode operation and it is fast enough (4ns) to be utilized in high speed applications. It presents a high OFF drain-source resistance (1 MΩ) and a low ON resistance (3Ω). The E class power amplifier is configured as illustrated in Fig.6.12. A step-up transformer, which also acts as a tuning device, is included. The power amplifier is required to supply high current to the primary of the transformer. The transformer must be carefully designed and efficiently used as an impedance matching for maximum power transmission. In high frequency applications, pot-core type ferrite is recommended [101]. For the current design, this is a Mullard ferrite core, RM6/160, with an inductance factor of A_L=*160 nH/turn*. In the successful construction of the transformer, primary and secondary winding inductances and turn ratios must be defined.

Since the hydrophone is connected to the secondary winding of the transformer, at resonance, only the static capacitance (C_o=7.10 pF) is effective in parallel with the load resistance of the hydrophone. If the power amplifier drives purely resistive load, i.e. necessary condition to improve efficiency, this capacitance must be tuned out at

the resonance frequency of the hydrophone [152]. This is achieved with the secondary inductance of the transformer and its value is defined as

$$L_s = \frac{1}{4\pi^2 f_o^2 C_o} \tag{6.8}$$

Substituting the values, the secondary winding impedance was found to be Ls=731 μH. From the inductance factor of the core and secondary inductance, the secondary turns (N_s) is calculated as 67 turns from [160]:

$$N_s = \sqrt{\frac{L_s}{A_L}} \tag{6.9}$$

where the turn ratio is chosen to perform impedance matching between the load network and output resistance of the amplifier. As emphasized before, the maximum power transmission at resonance can be achieved if the load impedance of the hydrophone is equal to the source resistance. This is done by using the transformer as an impedance matching network. The source resistance R_p is the ON resistance of the switching VMOS (VN66AF), specified as 3 Ω. Since the secondary turns, source and load resistance values are known, it is possible at this stage to calculate the primary turns [161]:

$$\frac{R_p}{R_s} = \left(\frac{N_p}{N_s}\right)^2 \tag{6.10}$$

where R_s is secondary load resistance and it is calculated from the measured conductance of the hydrophone as:

$$R_s = \frac{1}{G_o} = 276.6\Omega \tag{6.11}$$

From the Eq.6.10, N_p is calculated as 7 turns. However, in the practical system the number of turns in the primary winding is slightly adjusted to improve the performance of the amplifier as recommended in [137]. In testing the amplifier, it was observed that the hydrophone was not driven by the first harmonic of the input signal and some sort of distortion occurred. It is concluded in [101] that the secondary winding inductance should not be used to tune the static capacitance of the hydrophone if high efficiency voltage step-up is required. Therefore a separate tuning inductance must be provided. However, experimental results showed that the power amplifier performed satisfactorily within the design criteria.

Switching ON and OFF of the VMOS is done by a MAX4420, which is a MOSFET driver designed to translate TTL/CMOS inputs to high-voltage/high-current outputs. Its low 1.5 Ω output impedance and 6A peak current output allow to rapidly switch high-capacitance power MOSFETs. In this application, a 74AC00 TTL device provides the input signal to the driver IC. The input waveform, in digital PPM format, is generated by **Timer 1** of the DSP. Further discussion about this will be given in Chapter 7.

6.11 Receiver Implementation

This section is dedicated to the design of the receiver hardware and its associated software. Since the transmitted signal is a bandpass signal, the frequency response and amplification factor of the receiver must be carefully investigated for optimum detection. Its design criteria significantly affect the performance of the over all system. A single transducer is used for both transmission and reception. It is connected to the receiver circuitry via the analogue switch as discussed before. The design of the receiver is considered in three different sections: (i) preamplifier circuitry and bandpass filter (ii) envelope detection and lowpass filtering (iii) analogue-to-digital conversion and demodulation of the digital PPM signal. Before further studying these points, it is important to present the receiving response of the hydrophone which dominates the design of amplifiers and filters.

6.11.1 Receiver Sensitivity Measurement and Bandwidth Definition

Although several properties of the transducer (in normal use, a hydrophone) are described in Section 6.8, the definition of its receiving sensitivity and receiving bandwidth are also important parameters that need to be mentioned because they govern the design of the receiver hardware. The receiving sensitivity of the hydrophone was measured in the department's tank room. Details of this experiment are described in Appendix B, and the result is presented in Fig.6.13. It is evident from this illustration that the receiving sensitivity of the omnidirectional hydrophone at 70 kHz resonance frequency is -201 dB re 1V/μPa. From this measurement, the -3dB bandwidth of the hydrophone is also found to be approximately 25 kHz (extending from 60 kHz to 85 kHz). This suggests that the hydrophone permits the detection of 25 kbit/s data rate of transmission. Nevertheless, this is limited by the transmission bandwidth of the hydrophone, i.e. 18 kHz.

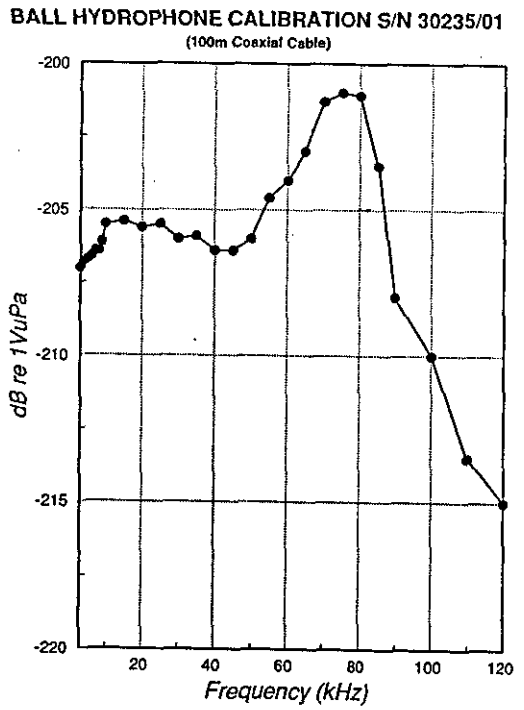


Figure 6.13 Receiver response of the hydrophone employed for the system.

6.12.2 Preamplifier Design for Underwater Acoustic Transducer

Acoustic waveforms detected by the hydrophone must be amplified since the voltage level across the terminals of the transducer will be low in magnitude, especially when the distance between the transmitter and the receiver increases. From the sonar equation studied in Chapter 5, the expected rms voltage level at the terminals of the hydrophone is found to vary from 91.20μV to 11.92 mV while the communication range is decreased from 100 m to 1 m. Such input signal levels are not high enough to be processed. Therefore, an amplifier is required to improve the accuracy in signal detection and a preamplifier circuitry was designed, as illustrated in Fig.6.14. It is important for the first stage receiver amplifier to be designed with low noise components. Although the transducer will be selective at around the resonance frequencies, it may respond to a number of unwanted frequencies because of the existence of the spurious resonance [100, 101]. It is necessary to make the first stage low-noise amplifier broad band selective at the transmit frequency. If this is not done, out of band ambient noise saturates the amplifier, hence limiting the detection of wanted signal.

The preamplifier is designed by employing a JFET op-amp (TLE2074). Its low-noise performance ($11.6\ nV/\sqrt{Hz}$ at 10 kHz), coupled with a 28 V/μs minimum slew rate, results in low distortion in the received signal. The preamplifier is configured as a highpass preamplifier with the cut-off frequency of 10 kHz. The passband gain factor of 20 dB is also applied. Since the amplified signal includes out of band frequency spectrum, signal-to-noise power ratio decreases. These may introduce errors in the decoded data. Therefore, a bandpass filter (BPF) with the bandwidth of the transmitter must be designed.

6.12.3 Receiver Bandpass Filter Design and DPPM Demodulation

Making the preamplifier broad band selective does not provide sufficient bandpass filtering to eliminate out of band ambient noise and preamplifier noise. To achieve a high signal-to-noise ratio (SNR) at the receiver, the received signal must be filtered.

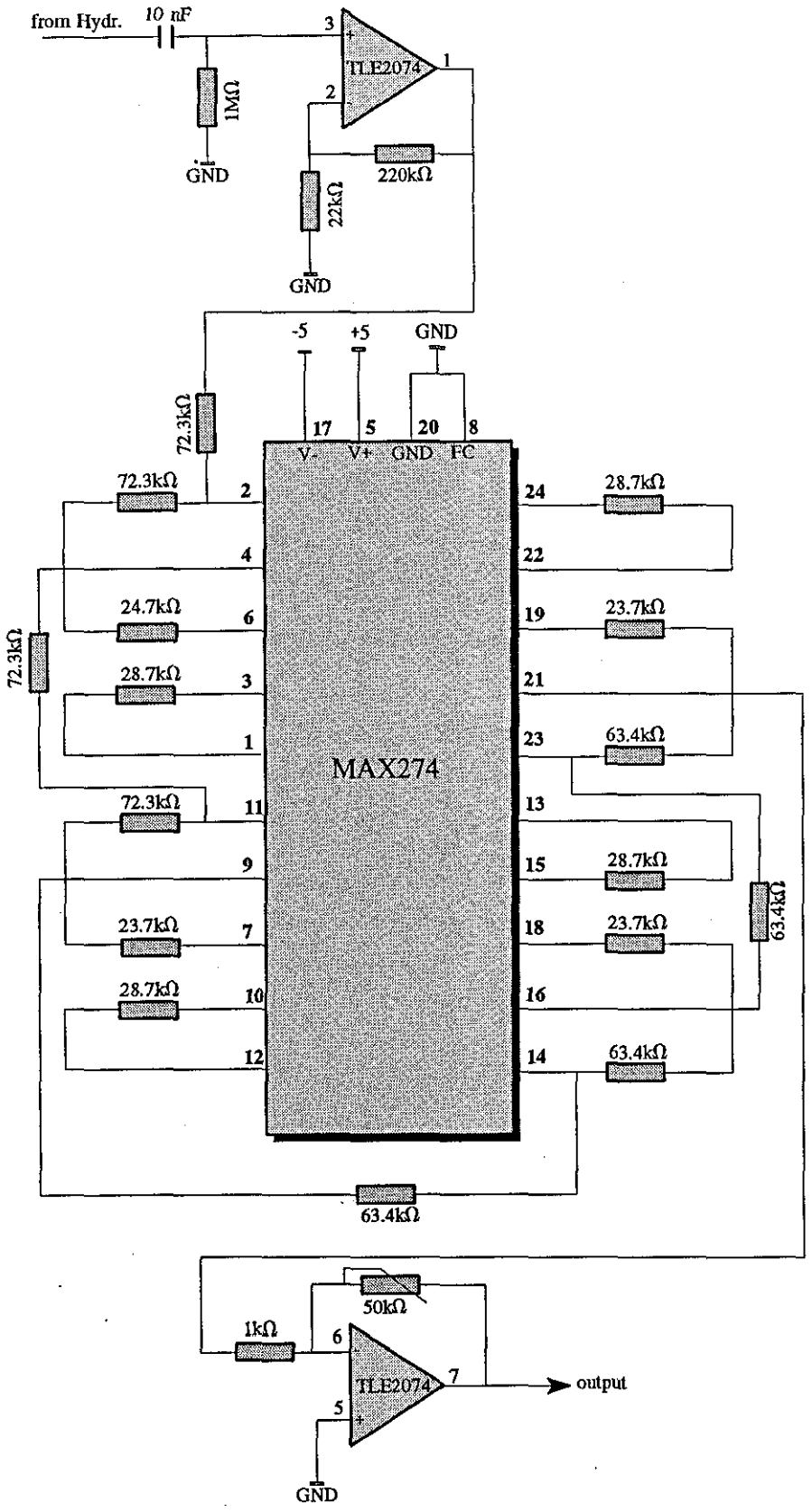
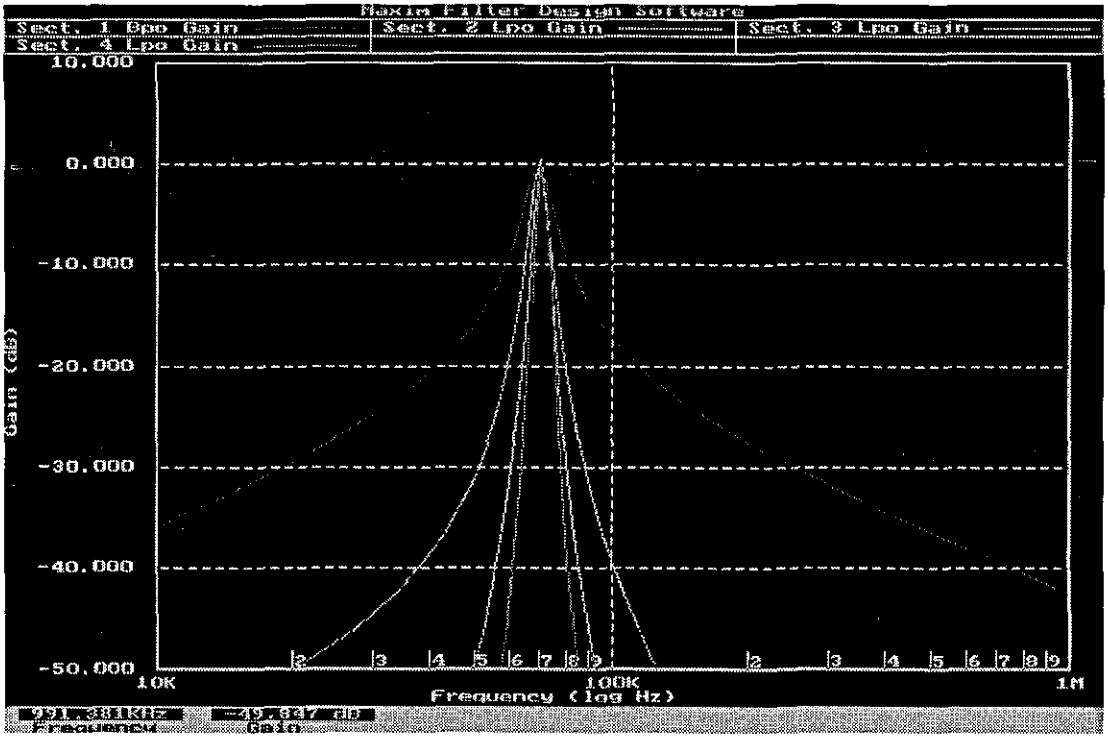


Figure 6.14 Receiver amplifiers and bandpass filter implementation.

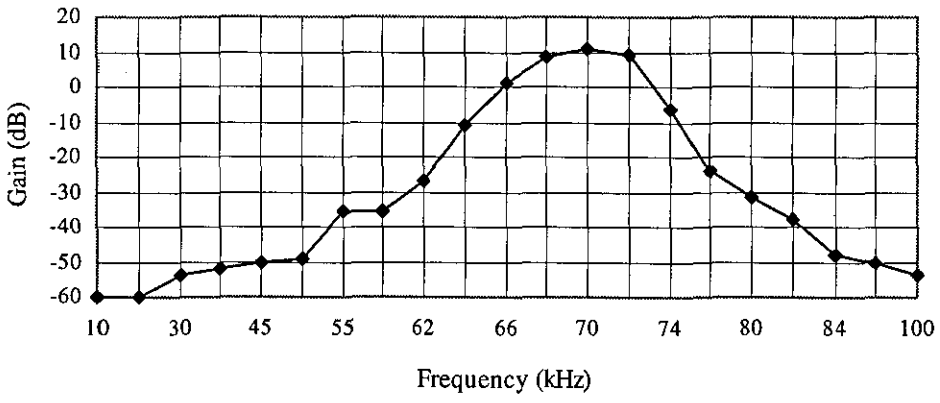
Since a single carrier frequency is used to convey information for underwater acoustic communications, a narrow bandpass filter is suitable. The main specification of the filter is to provide filtering within the -3dB frequencies of the hydrophone. The filter used in the system is an 8th order active filter (MAX274). It is a single chip IC consisting of four independent cascadable 2nd order sections which filter frequencies up to 150 kHz.

Before considering the component specifications, the design criteria must be outlined. Because the centre frequency of the filter is dominated by the hydrophone resonance frequency, it is selected as 70 kHz. Due to the hydrophone bandwidth, the bandwidth of the filter is decided to be 18 kHz. The filter structure is based on a Butterworth filter, as illustrated in Fig.6.14, and component values are calculated by using a special simulation software provided by the manufacturer (Maxim Integrated Products, Inc., CA). In the implementation of the filter, high precision resistors are used. If they are not properly chosen both the magnitude and frequency response of the filter will diverge from the simulated result. This is illustrated in Fig.6.15. Although unity gain is considered in the design stage, due to inaccuracy in the resistance values it is observed that 10.8 dB of gain is introduced by the bandpass filter at the resonance frequency. A second stage variable gain amplifier is also supplied for the receiver. For the system, when a short range communication is taking place, signal magnitude may exceed the maximum allowable level, because of the amplification (the amplifier saturates). Although out of band noise signals are attenuated, because of such amplification, the noise signal is also amplified. This results in a detection error. Moreover, the gain factor is also related to the digitization process of the signal (i.e V_{REF}) which will be defined later.

Demodulation of the DPPM signal is based on the envelope detection rather than on the frequency or phase measurement of the signal. The envelope of the signal is extracted by passing it through a high precision full wave rectifier, as illustrated in Fig 6.16. Then the rectified signal is lowpass filtered to smooth the waveform and to provide a baseband signal. The bandwidth of this filter is selected as 10 kHz and is defined by the baseband signal. Then the signal is digitized using an analogue-to-digital converter studied in the following section.



(a)



(b)

Figure 6.15 Bandpass filter response, (a) simulated design (b) actual design

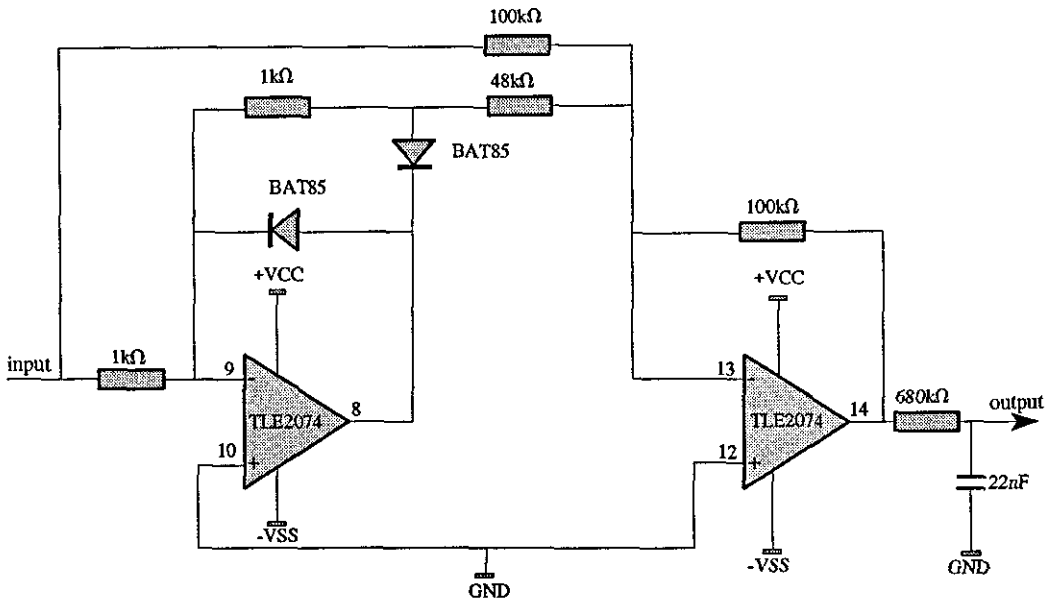


Figure 6.16 Baseband detection circuitry

Before concluding this section, it is important to emphasize the bandwidth definition of the receiver circuitry. The transmitted short pulses are distorted by the Q factor of the hydrophone as well as the multipath distortion introduced by the underwater communication channel. For optimum detection of such signals, the receiver bandwidth (controlled by the hydrophone, the preamplifier and the BPF) must be equivalent to the essential bandwidth of the DPPM signal (this is 20 kHz and more details can be found in Chapters 5 and 7).

However, the bandpass filter presented here was not implemented according to this design criteria and bandwidth of the filter is set as 4 kHz for the purpose explained above. This has affected decoding of the DPPM signal since the fast rising and decaying edge of the baseband signal is essential. Therefore, the first stage of the BPF is utilized for the system throughout the experiments, but in the future design of the receiver circuitry, the above phenomena are recommended to be considered.

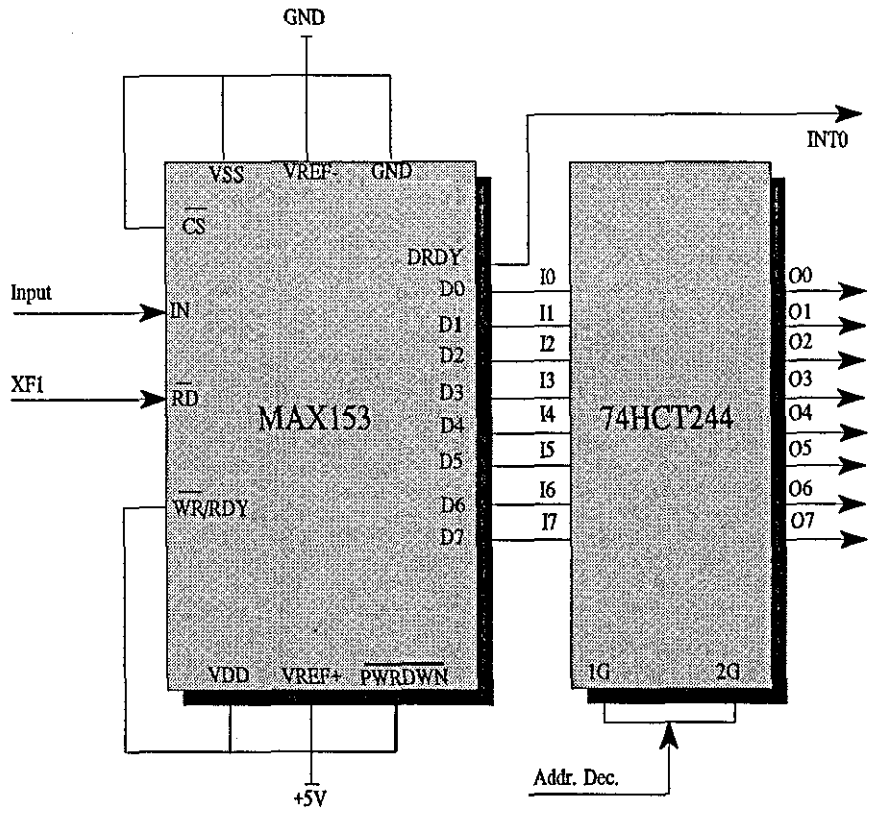
6.13 Digitization of Received DPPM Signal

In order to accurately decode the DPPM baseband signal, it is necessary to apply digital signal processing techniques. Therefore, the signal must be digitized for further analysis. Before choosing an A/D converter, the sampling frequency and quantization rate must be defined by considering input signal properties. Since the input signal bandwidth is 10 kHz, the sampling frequency must be at least twice this, as discussed previously. However, this is selected as 40 kHz to improve the performance of DPPM decoding. Although an 8-bit quantization rate is arbitrarily chosen, its effect with the V_{REF} (5V) on the decoding process is worthwhile to investigate. As noted before, estimated signal level at the hydrophone's terminals and the receiver gain factor contribute to the decoding of the input signal. At the closest communication range, the input signal magnitude (11.92 mV at 1 m distance) to the ADC must not exceed the V_{REF} to avoid overloading. Therefore, the overall receiver gain is selected as 52.6 dB. For the communication range of 100 m, the expected signal level, after amplification, will be 38 mV. Since the preamplifier and the BPF gains are defined as 20 dB and 10.8 dB respectively, the second stage amplifier gain is, therefore, adjusted to be 21.8 dB.

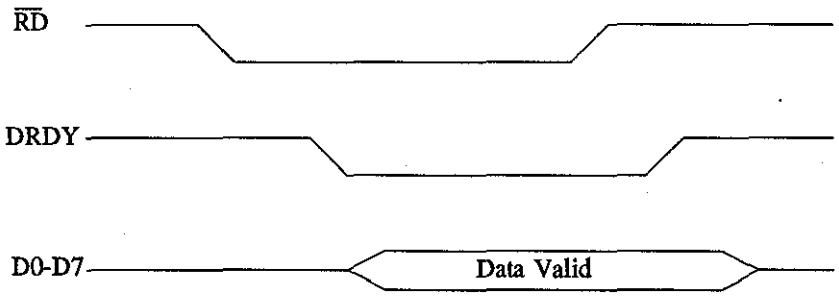
For the digitization of the baseband signal, an 8-bit MAX153 ADC with 725 ns access time is employed. To define the smallest voltage step, ΔV, for the ADC, Eq.6.6 is used and this is found to be 19.53 mV which results in 34 dB SNR (noise signal level is calculated from Eq.5.4 and 5.14). This result suggests that at the distance of 100 m or above the receiver performance is limited by the quantization rate of the ADC for the approved specifications of the system. This can be improved by either replacing the ADC with a higher quantization rate ADC or increasing the emitted acoustic power.

Configuration of the MAX153 ADC with a three-state 74HCT244 buffer IC is illustrated in Fig.6.17a. Control signals for these components are provided by the TMS320C31. As the sampling frequency, the **Timer 0** interrupt of the DSP is utilized. XF1 output of the DSP is engaged to initiate the conversion by triggering the $\overline{RD}$ input of the ADC while the $\overline{INT0}$ input detects completion of the conversion. To

operate the ADC as specified above, an appropriate control program was written, as presented in Fig.6.18. (analysis of the digital data is discussed in Chapter 7). As with many A/D converters, when a digitized sample is read by the DSP, the data bus of the MAX153 enters a high impedance state. This occurs after the $\overline{RD}$ line becomes inactive, as illustrated in Fig.6.17b. With this type of device, it is common that the data output buffers require a substantial amount of time to reach a full high-impedance state. This is given as 80 ns for the MAX153. Since the internal clock of the TMS320C31 is operating at 16.65 MHz, any ADC used for this purpose must disable the outputs not later than 65 ns after the deselection of that particular device. This is important because the DSP may begin to use the data bus for consecutive operations. If this timing requirement is not provided, bus conflicts between the TMS320C31 and MAX153 will occur. This will cause failure of the system due to damaged data bus. Therefore, buffers are required to isolate the A/D converter output from the DSP. For this design, the 74HCT244 buffer, introducing an 18 ns delay, is used. The state of the buffer is controlled by the address decoder. The buffer is enabled while digital data is read and disabled after reading process is completed.



(a)



(b)

Figure 6.17 Configuration of the ADC and timing signals

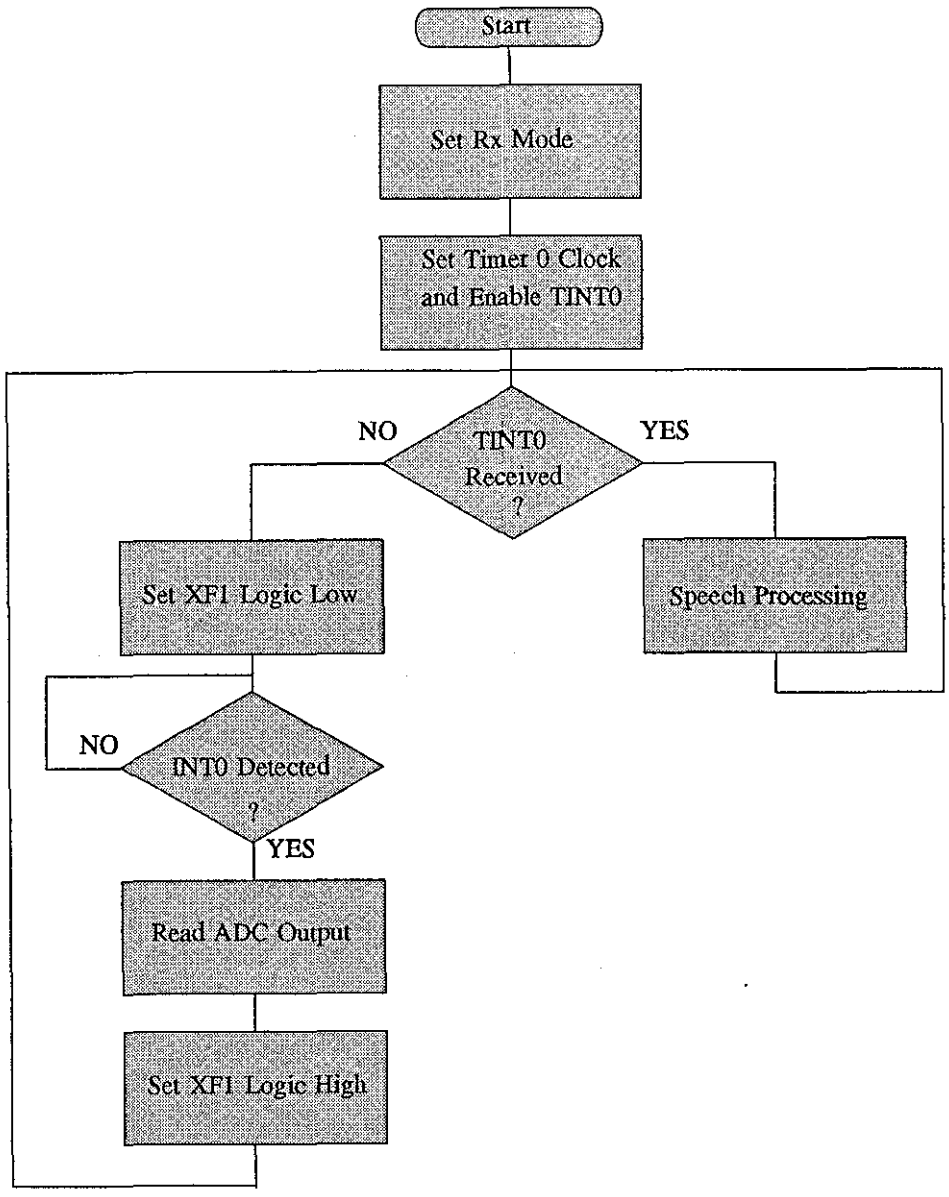


Figure 6.18 Software control of the ADC

CHAPTER SEVEN

ACOUSTIC TRANSMISSION AND DETECTION OF SPEECH PARAMETERS

7.1 Introduction

In this chapter, the transmission and detection methods for quantized speech parameters for the digital underwater voice communication system will be discussed. As illustrated in Fig.7.1, the system includes a source encoder and a PPM encoder in the transmitter section and a PPM decoder and source decoder in the receiver section. The source encoder, as described previously, implements the speech compression algorithm and extracts appropriate speech parameters to be transmitted. The PPM encoded speech parameters are acoustically transmitted through an underwater communication channel by using the Digital PPM method. Due to the nature of the communication channel the received signal is susceptible to a variety of disturbances previously described in detail in Chapter 5. This suggests that the quality of the decoded speech signal at the receiver is highly dependent on the performance of the signal detection scheme.

The system utilises noncoherent-based signal detection, reducing the amount of software as well as the complexity of the required hardware. The PPM and source decoders resemble the corresponding encoders in the transmitter.

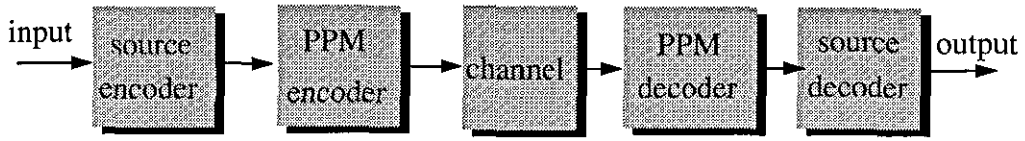


Fig 7.1. Block diagram of an underwater communication system

7.2 Encoding of Speech Parameters for Transmission

Since an 8-slot Digital PPM method was selected for digital data transmission through an underwater acoustic channel, the quantized speech parameters, described in Chapter 3, must be encoded accordingly so that transmission can be achieved. The PPM encoder accepts as input 54 bits representing the speech parameters and implements a block coding method in which these parameters are transmitted in 3-bit symbols. The results of the encoding process are appropriately stored in two 32-bit memory locations.

In this system, to achieve the block coding method with less algorithm complexity, it is necessary to consider the size of the quantized speech parameters being stored in each DPPM memory location. Therefore, before deciding how these parameters must be classified we introduce a simple parameter selection criterion.

Firstly, the overall length of the block of quantized speech parameters for each DPPM location must not exceed the size of memory, i.e. 32 bits. Secondly, the product of the number of bits in a location and n^{-1} ,where n is the symbol size which for this application is 3 bits, must be an integer. By taking into account these points, it was decided to classify the parameters in two groups, as illustrated in Table 7.1. One group includes reflection coefficients ***k1*** to ***k6*** and ***k10***, a total size of 30 bits; the other group consists of ***k7*** to ***k9***, ***gain***, ***pitch period*** and ***voiced/unvoiced decision*** bit, amounting to 24 bits. Although these groups are arranged as described, they can be organized in several different ways by providing the criteria discussed in this section.

After defining the groups, the next step is the block coding of these parameters. In the present system, block coding of a mixture of rows and column is implemented for software convenience, as shown in Table 7.1. Therefore, equally quantized speech parameters in each group, i.e. ***k1***, ***k2***, ***k3***, ***k4*** and ***k5***, ***k6***, ***k7*** and ***k8***, are separated into sub-groups and a column block coding method is employed to form symbols. In this method, each symbol is composed of one bit for different parameters if the sizes of columns are equal or greater than the symbol size. Otherwise, this may be changed, i.e.

a symbol is composed of one bit from one parameter, two bits from another. For the rest of the parameters, row block coding is implemented.

The block coding methods described here do not include an error correction scheme. A random error occurring during transmission will alter the decoded speech parameters at the receiver. Since we transmit quantized speech parameters as a packet

k1	*b4*	*b3*	*b2*	*b1*	*b0*
k2	*b4*	*b3*	*b2*	*b1*	*b0*
k3	*b4*	*b3*	*b2*	*b1*	*b0*
k4	*b4*	*b3*	*b2*	*b1*	*b0*

k5	*b3*	*b2*	*b1*	*b0*
k6	*b3*	*b2*	*b1*	*b0*

k10	*b1*	*b0*

k7	*b3*	*b2*	*b1*	*b0*
k8	*b3*	*b2*	*b1*	*b0*

k9	*b2*	*b1*	*b0*

gain	*b5*	*b4*	*b3*	*b2*	*b1*	*b0*

pitch	*b5*	*b4*	*b3*	*b2*	*b1*	*b0*

v/uv	*b0*

Table 7.1 Grouping the quantized speech parameters

of 54 bits, i.e. 18 symbols, it will be worthwhile to investigate the effect of a random error in a symbol to the overall data packet. The 18 symbols are built up in the way described above. Eight of them comprise bits from different speech parameters while the other eight consist of two bits from the same parameters. The rest of the symbols include all the bits from the same parameter. A burst of pulses is transmitted for each symbol and it is assumed that an error is detected in one of the symbol. This error will affect at most two speech parameters depending in which slot the error is detected. In

addition to this, one more point needed to be addressed is whether the error is significant or not. Since a single bit from each parameter will be affected, depending on its significance in the parameter, the error may be ignored. However, if the error is detected at most the significant bits of a parameter, then this will affect the synthesized speech signal quality.

After completing the grouping process, each symbol is stored in the appropriate DPPM memory locations as illustrated in Table 7.2, where the top row indicates the number of bits used in each location, the middle row shows the output bit stream of the PPM encoder and the bottom row illustrates the number of symbols used for transmission.

31	30	29	28	27	26	25	24	23	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0
x	x	$k10_0$	$k10_1$	$k6_3$	$k5_3$	$k5_2$	$k5_1$	$k5_0$	$k3_2$	$k3_1$	$k3_0$	$k4_3$	$k4_2$	$k4_1$	$k4_0$	$k2_3$	$k2_2$	$k2_1$	$k2_0$	$k1_3$	$k1_2$	$k1_1$	$k1_0$	$k8_3$	$k8_2$	$k8_1$	$k8_0$	$k7_3$	$k7_2$	$k7_1$	$k7_0$	$k6_4$
		10			9			8			7			6			5			4			3			2			1		

31	30	29	28	27	26	25	24	23	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0
x	x	x	x	x	x	x	x	p_{19}	p_{18}	p_{17}	p_{16}	p_{15}	p_{14}	p_{13}	p_{12}	p_{11}	p_{10}	p_9	p_8	p_7	p_6	$k9_2$	$k9_1$	$k9_0$	$k8_4$	$k8_3$	$k8_2$	$k8_1$	$k8_0$	$k7_4$	$k7_3$	$k7_2$
								18			17			16			15			14			13			12			11		

Table 7.2. Allocation of block encoded speech parameters into appropriate memory.

Once the DPPM memory locations are filled with data, the system is ready for transmission. However, since the system computes the speech parameters and transmits simultaneously, there is a time difference between the parameters' transmission time (i.e. 19 ms) and the parameters' computation time (i.e. approximately 10 ms). This means that the system must be set in a waiting mode until the transmission of previous parameters is completed.

7.3 Clock Generation for Digital PPM

As suggested in Chapter 3 , digital PPM implementation requires two clock generators for information to be carried by the position of successive pulses. One is operating at the slot frequency, 10 kHz, the other is at the carrier frequency determined by the hydrophone resonant frequency of 70 kHz. Two timers of the DSP (T0 and T1 timer outputs) are utilized for this purpose, as shown in Table 7.3. Since the system is in transmission mode, the T0 and T1 *timer_period_registers* of the TMS320C31 which specify the timers' signaling frequency, are assigned as slot clock and carrier clock generators at 808028h and 808038h addresses, respectively. They must be appropriately initialized to operate at the necessary frequencies. Thus, as described in [146], the *period_register* values are extracted from

$$f_{wanted} = \frac{f_{timer\,clock}}{2.\,period\ register} \tag{7.1}$$

where f_{wanted} is the frequency value at which the timer will operate, f_{timer_clock} is an internal clock source with a frequency of $f_{HI}/2$, i.e. 8.3325 MHz. f_{HI} is the clock output of the DSP and half of the system clock 33.33 MHz. After evaluating the Eq.7.1, T0 and T1 timer *period_registers* are set to (01A0)h and (003B)h, respectively. Operation of the timers can be controlled by altering the contents of the timers' *global_control_registers* at 808020h and 808030h locations for T0 and T1. Since the system is in transmit mode the timer outputs are utilized to control some external circuitry or generate internal interrupt signal. Hence these registers are initialized to (03C1)h when they are needed as defined in [146].

Timer 0 registers	**Address**	**Tx mode**	**Rx mode**
T0 global_control_register	808020h	03C1h	03C1h
T0 period_register	808028h	01A0h	0068h
Timer 1 registers	**Address**	**Tx mode**	**Rx mode**
T1 global_control_register	808030h	03C1h	03C1h
T1 period_register	808038h	003Bh/0000h	X

Table 7.3. Initialization of timer registers for transmit and receive mode operations.

T0 timer of the TMS320C31 is employed for both transmission and receiving operation of the system. During the receive mode, its *timer_period_register*'s content is modified since T0 is utilized to generate the sampling clock for the A/D converter in the receiver circuitry, i.e. $f_s = 4.f_{in}$ where f_{in} is 10 kHz slot frequency. Therefore the new value for 40 kHz is simply calculated again from Eq.7.1 as (0068)h and stored into an appropriate memory location. The importance of the sampling frequency value will be discussed in the following sections.

7.4 Digital PPM Transmission of Speech Parameters

The transmission of speech signal parameters for a speech frame is manipulated as illustrated in Fig.7.2a. This transmission is accomplished by sending a synchronization frame and 18 symbols. Since no transmission is wanted during breathing noise intervals or before computing speech parameters, operation of the transmitter circuitry must be controlled by employing an internal supervisory control.

As the system is in transmit mode, the *slot_counter*, DPPM *symbol_counter* and *ready_to_transmit_flag* are initialized appropriately as shown in Fig.7.3a after the *first* block coding process is completed and DPPM memory locations are filled with data. Because we intend to transmit data and do signal processing simultaneously for real time operation, interrupt controlled execution is adopted. The T0 timer interrupt, TINT0, is utilized for this purpose. The rising edge of the T0 clock generates an interrupt request then the DPPM_TX_ISR interrupt subroutine is accessed for transmission of the SYNC signal and digital data as shown in Fig.7.2b. In the following section we will discuss how the transmission process is performed.

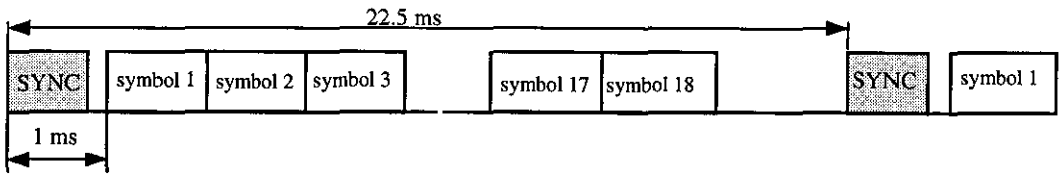


Fig.7.2 (a). Digital PPM transmission of a frame of speech signal parameters

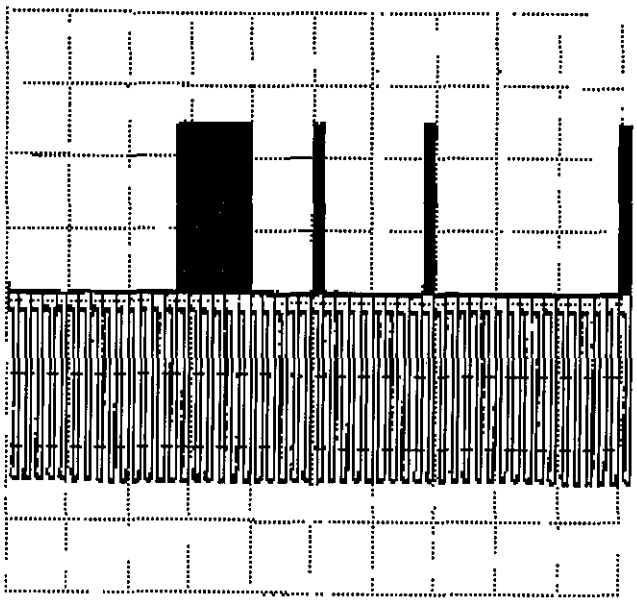


Fig.7.2 (b) Transmission of SYNC signal and (601) octal data in DPPM format (Volt/Div=2V and and Time/Div=0.5 ms for the CH1 and CH2).

7.4.1 Transmission of Synchronization Signal

Symbol synchronization is required in every digital communication system which transmits information synchronously. By the nature of digital PPM, signal synchronization is more difficult than for the more familiar digital Pulse Amplitude Modulation (PAM) systems. This is because the PPM format encodes the information by the temporal slot position of the pulse in the symbol interval. However, in order for the system to work correctly, the transmitter and receiver must be perfectly synchronized. The question of how best to achieve synchronization of digital PPM has received some attention in the literature [138,140].

The synchronization issue is considered in three aspects, speech frame synchronization, DPPM symbol synchronization and slot synchronization, which all depend on each other. In this section we will only consider the first one. As shown in Fig.7.2, 18 symbols are used for transmission of a 22.5 ms speech signal frame. When the receiver unit is operated or if the communication is inadvertently suspended, the receiver has to be synchronized to the beginning of a speech frame in order to correctly decode the received DPPM slots.

The loss of synchronization reduces the efficiency in the data detection process, because the inaccurate symbol synchronization directly reduces the probability of making correct decisions. Moreover, when a loss of synchronization occurs, it may sometimes lead to successive errors before it is regained; these successive errors affect the overall performance of the system. Although the symbols are decoded properly at the receiver, they may not represent the encoded speech parameters. Therefore, in order to synchronize the transmitter and the receiver at the beginning of each speech parameter transmission interval a SYNC pulse is transmitted, as illustrated in Fig.7.2.

From Fig. 7.2, it may be noted that the frequency of the SYNC signal is 44.444 Hz, i.e. *1/22.5ms.* To distinguish the SYNC signal from the DPPM data signal its duration must be carefully decided. After data transmission, the received signal will be detected as spread out due to the response of the hydrophone and multipath propagation in the underwater acoustic communication channel. The DPPM data signal may occupy

several successive slot intervals. This must be considered in the definition of the SYNC signal duration to distinguish from the data slots. By taking into account all these factors, the duty cycle of the SYNC signal is defined as $t_{duty}=n.T_{slot}$, where $1<n<2^k$, k is the block size, i.e. 3, and n is the number of slots, selected as 6 for this application.

After the system is set to the transmission mode, the *slot_counter* is reset, the TINT0 interrupt is enabled, at the rising edge of the T0 clock, and DPPM_TX_ISR is accessed. Every time this subroutine is called, the *slot_counter* is increased. As the SYNC pulse is being transmitted, the T1 *timer_control_register* is set to (03C1)h, as explained before to generate a waveform at the resonant frequency of the hydrophone, i.e. 70 kHz, so that the hydrophone emits acoustic pulses at this frequency. This emission is continued for $6.T_{slot}$. After that the T1 *timer_control_register* is set to (0000)h, indicating no acoustic emission.

7.4.2 Transmission of Block Coded Speech Parameters

Once the system completes the transmission of the SYNC signal, the data transmission mode is selected by initializing the *data_tx_rdy* flag and clearing the contents of the *slot_counter* and DPPM *symbol_counter* registers as illustrated in Fig.7.3c. As described before, the T0 timer of the TMS320C31 is used again for slot timing. When its associated interrupt request is detected, the DPPM_TX_ISR interrupt subroutine is called. As soon as the subroutine is accessed, i.e. the first slot interval of the first symbol is being processed, the content of appropriate DPPM memory location in Table 7.2 is shifted to the left by 3 bits in order to form a symbol to be transmitted. The resulting octal number is stored in a buffer and kept constant during each symbol transmission interval. Every time this interrupt subroutine is accessed (i.e. 10 times for each symbol), the contents of the *slot_counter* are compared to the number in the buffer and the *slot_counter* is increased. If these two numbers are found to be equal then the 70 kHz clock signal is generated for that slot interval by setting the T1 timer control register to (03C1)h, the system emits acoustic pulses and the DPPM data is transmitted, as illustrated in Fig.7.2a. As soon as the consecutive interrupt request is received this timer is disabled. Since a symbol interval is divided into 8 data and 2

guard slots, the *slot_counter* must be equal to the total slot number in the symbol duration. Once this condition is provided, the *slot_counter* is reset, thereby acknowledging that data transmission is completed.

For a 22.5 ms speech frame represented by 54-bit quantized speech parameters, 18 symbols are required to be transmitted. Therefore after transmitting each symbol, the DPPM *symbol_counter* is increased. This procedure is repeated until all the data transmission is completed. The total transmission time for a 22.5 ms speech signal frame takes 19 ms for the quantized speech parameters with 100 μs for T_{slot}. However, with careful study of the slot interval, this can be decreased. Since data is transmitted as a short burst consisting of several cycles at the resonant frequency of the hydrophone, i.e. f_0=*70* kHz, the transmitting response of the hydrophone plays an important role in the definition of T_{slot} and must be considered. The Q-factor of the hydrophone is also a significant factor [162]. As described in Chapter 5, a slot interval is defined as:

$$T_{slot}=k/f_o \tag{7.2}$$

where k is the number of cycles. During transmission, the hydrophone must be excited for at least for Q cycles, i.e. $Q \leq k$, for efficient acoustic power transmission and it can be shown that

$$Q=\frac{f_0}{f_{BW}} \tag{7.3}$$

where f_{BW} is the bandwidth of the hydrophone. The hydrophone parameters were previously extracted as f_0 = *70 kHz* and f_{BW} = *17.1 kHz*, hence Q = *4.1*. This suggests an acceptable minimum slot interval should be 58.57 μs. Assuming this slot interval, then the data transmission rate in 8-slot DPPM format will be increased to 5.122 kbit/s. In addition, this will correspond to 17.1 kbit/s for the ASK and PSK modulation methods. Considering a noise-free channel, these data rates can be successfully achieved. However, since an underwater communication channel introduces a multipath problem, the latter modulation methods with this slot interval are susceptible to errors.

If this optimized slot interval is used, transmission of a symbol will take 585.7 μs. Since 18 symbols per speech frame and the SYNC signal are required to be transmitted during a 22.5 ms speech interval, with the aid of the optimized slot time, 11.13 ms will be used for data transmission. Since *full-dublex* operation as in mobile telephone technology is desired for the system, this can be achieved by utilising two hydrophones at diferent resonance frequencies; one for the transmission the other is for the receiving. However this will increase the complexity and the cost of the system. *Two-way* communications in real-time is possible to implement. This can be done in two ways: One is by replacing the transmit/receive hydrophone with a high frequency one, i.e. less time for the transmission of the speech parameters, and the other is by increasing the speech signal frame interval, which is 22.5 ms for the current design. Adopting both methods will obviously introduce some alteration in the hardware and software design of the system; however, the first one is easier since only the receiver filter is required to be modified.

During 22.5 ms of speech signal, transmitting and receiving and also decoding processes must be completed. With the speed of the TMS320C31 DSP, the system has enough signal processing power to accomplish all this processing for *two-way* operation. However, the availability of the 70 kHz transducer and concerns due to limitations in communication range and multipath propagation in the communication channel dictated the adoption of a *half-duplex* operation mode. Due to the modular design of the software and hardware, the system can be modified to incorporate *two-way* operation by using a higher frequency transducer. Use of such a transducer permits implementation of Time Division Multiplexing (TDM), i.e. 11.25 ms for the transmission and 11.25 ms for the receiving intervals.

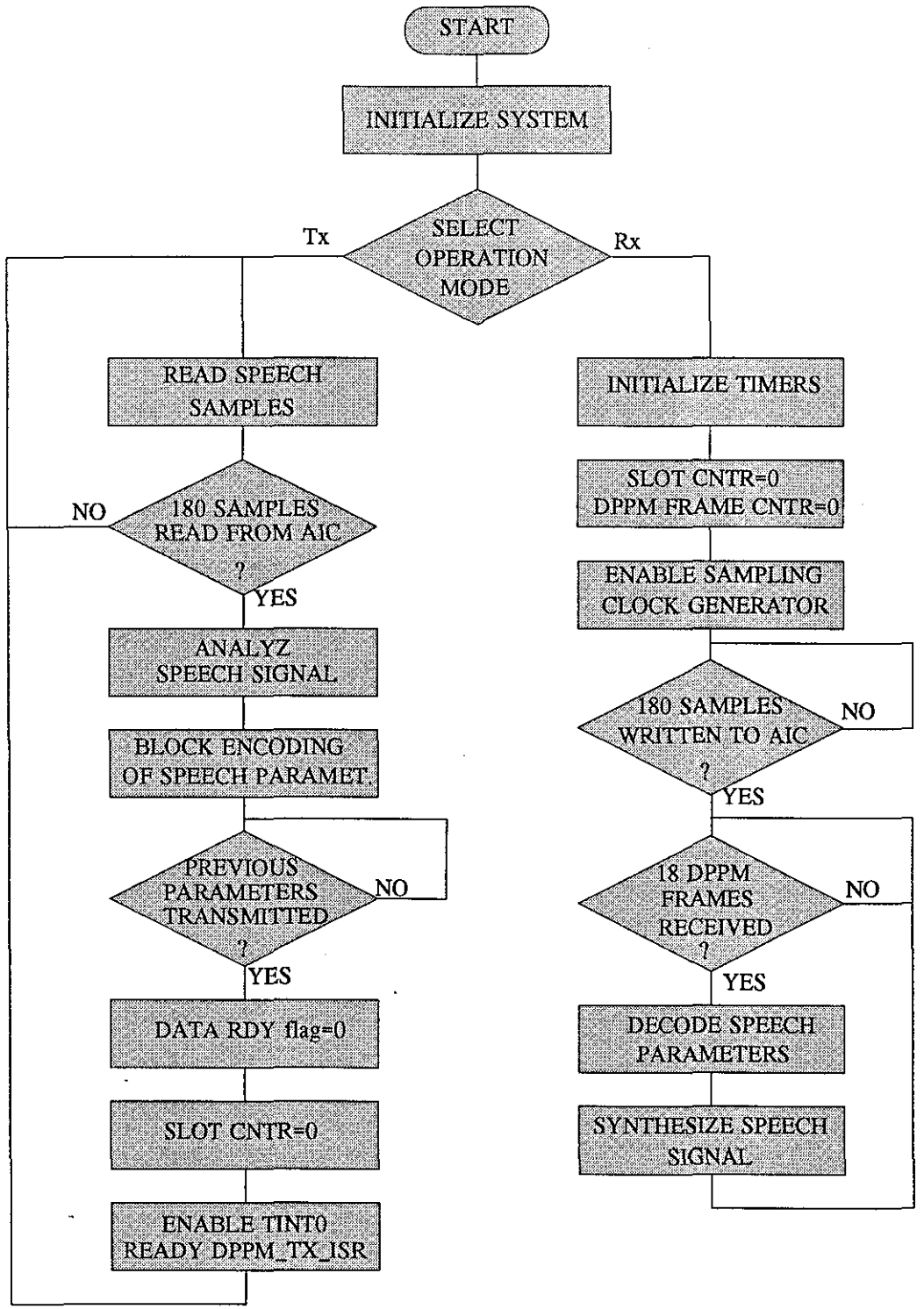


Fig.7.3 (a) Algorithmic illustration of voice communication system

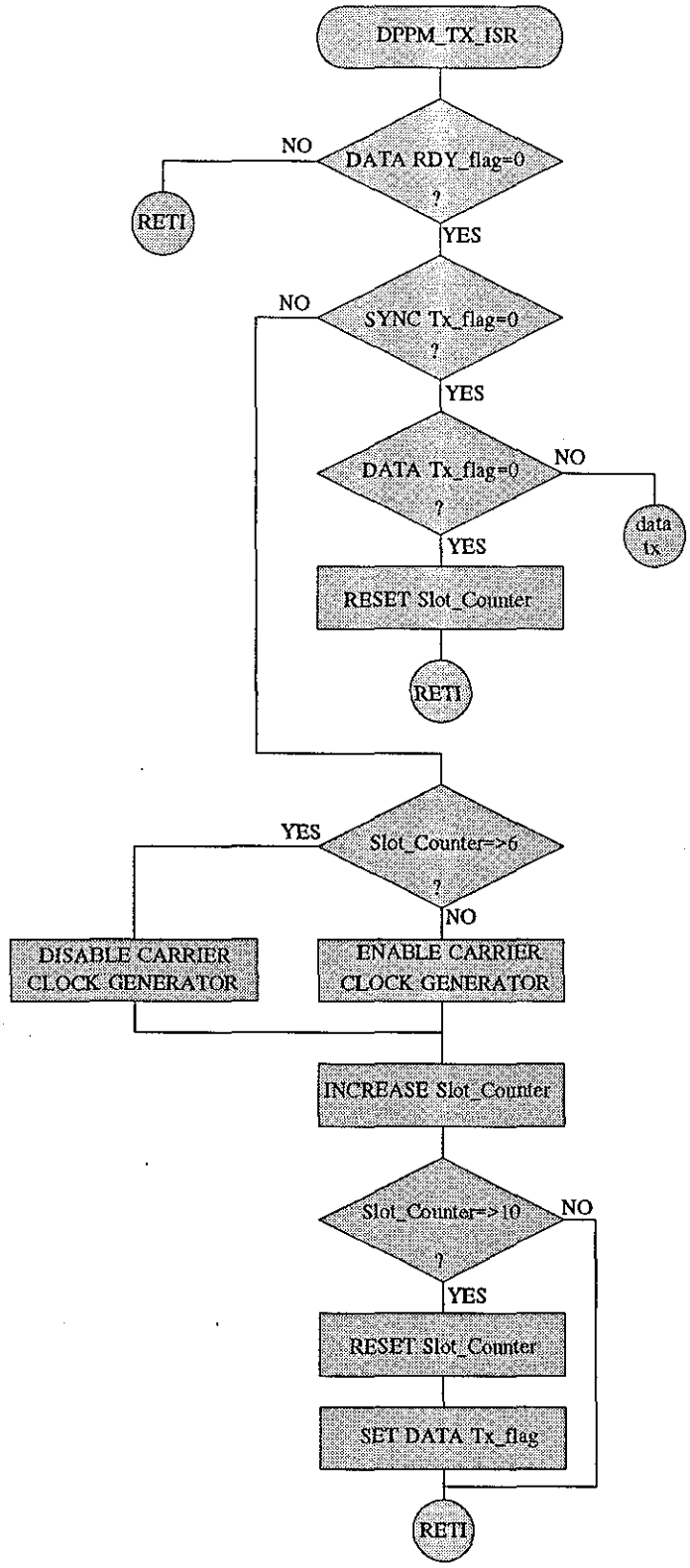


Fig. 7.3 (b) Operation of internal timer interrupt and SYNC signal generation

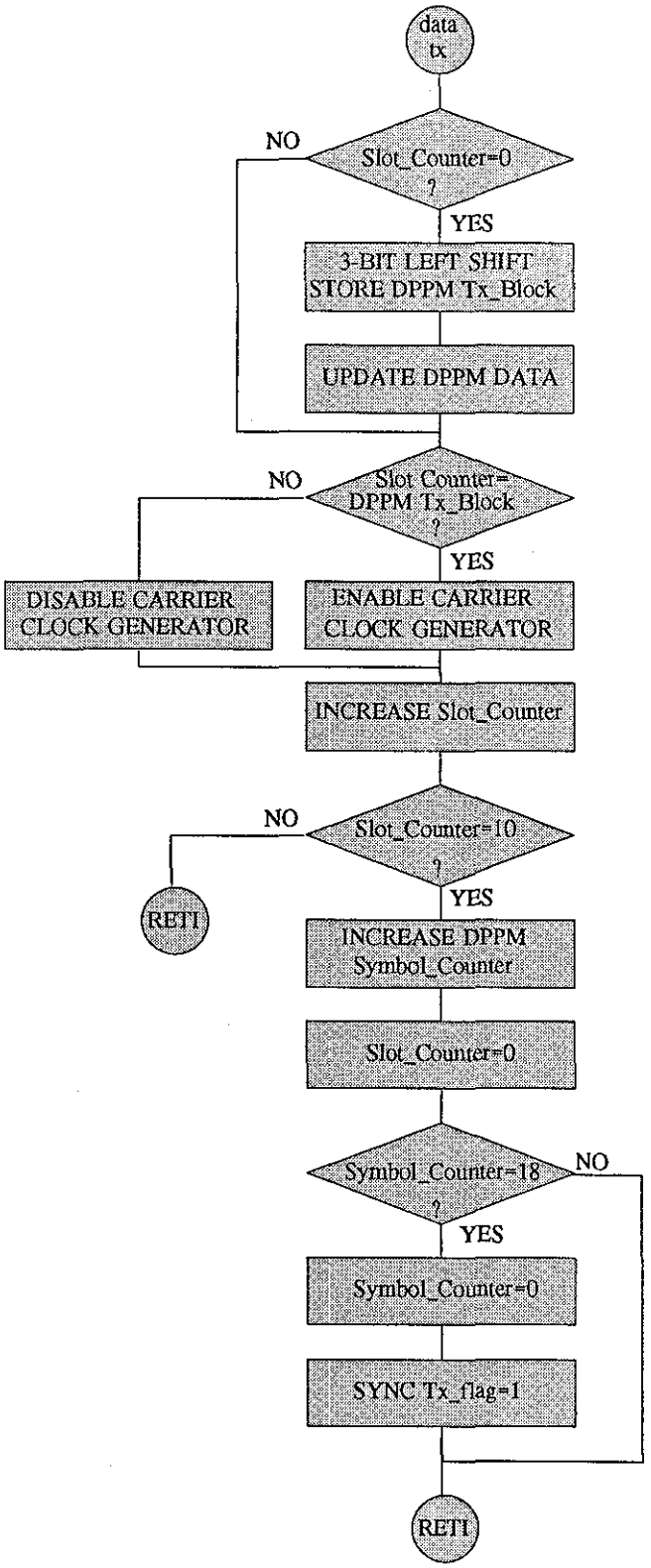


Fig.7.3 (c) Digital PPM of speech signal parameters

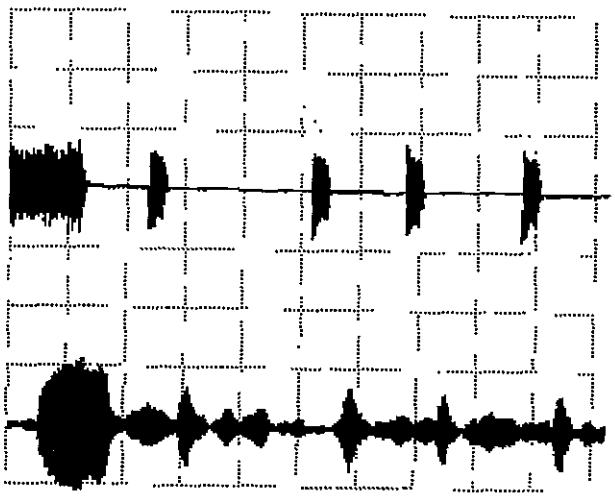
7.5 Detection and Decoding of Digital PPM Signal

As studied in Chapter 6, acoustic pulses transmitted in DPPM format through an underwater communication channel are detected by the omnidirectional hydrophone at the receiver. The input signal is applied to the bandpass filter, an example of the output waveform of being shown in Fig.7.4a. Since the DPPM decoding algorithm processes the baseband signal, an incoherent detection method is utilised. Therefore, the full wave rectifier and the low pass filter are designed to be utilised to obtain the baseband signal which is a pulse signal with an approximate pulse duration of T_{slot} (100 μs), as depicted in Fig.7.4b. The actual data decoding process is conducted after obtaining the baseband signal. Since the system is set to receiver mode operation as illustrated in Fig.7.3a, SYNC signal detection, pulse position detection and the speech synthesizing algorithms are operated. In the following section, details of this process will be discussed.

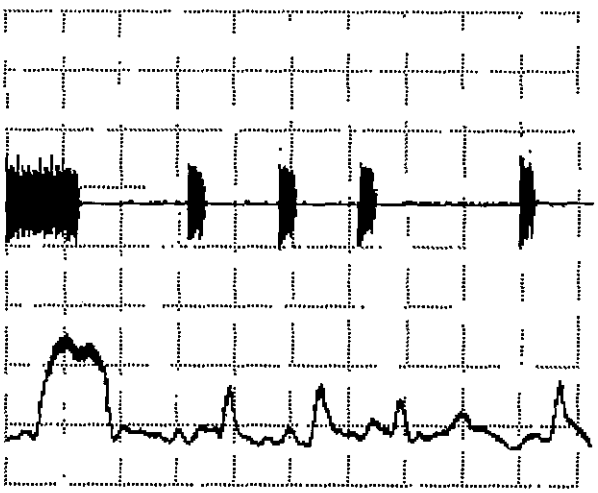
In DPPM signal detection, as emphasized during the hardware design, the limited bandwidth of the receiver affects the performance of the system by introducing a finite rise time to the received baseband signal. The slow baseband pulse build up may result in an error in the DPPM data decoding because the necessary timing accuracy in the pulse position may not be achieved. In order to process the baseband input signal shown in Fig.7.4b, as described in Chapter 6, the MAX153 ADC is utilized with the sampling frequency of 40 kHz and an 8-bit quantization rate. Once the system is set to receiver mode operation, the **Timer 0** of the DSP is initialized to operate at the sampling frequency of 40 kHz as described earlier in Section 7.3. Access to the ADC is controlled by the **TINT0** internal interrupt of the DSP and DPPM_RX_ISR interrupt subroutine, as presented in Fig.7.5, is executed to detect the SYNC signal and decode DPPM signal as described in the following sections.

7.6 DPPM Baseband Signal Consideration and Synchronization

As studied earlier in this chapter, the transmitted pulse chain for a speech frame comprises of DPPM data as well as the SYNC signal (this may be addressed as speech



(a)



(b)

Figure 7.4 Transmitted and received DPPM waveforms (a) output of bandpass filter (b) output of the envelope detector (CH1:10V/Div; CH2:2V/Div; Time/Div=0.5ms).

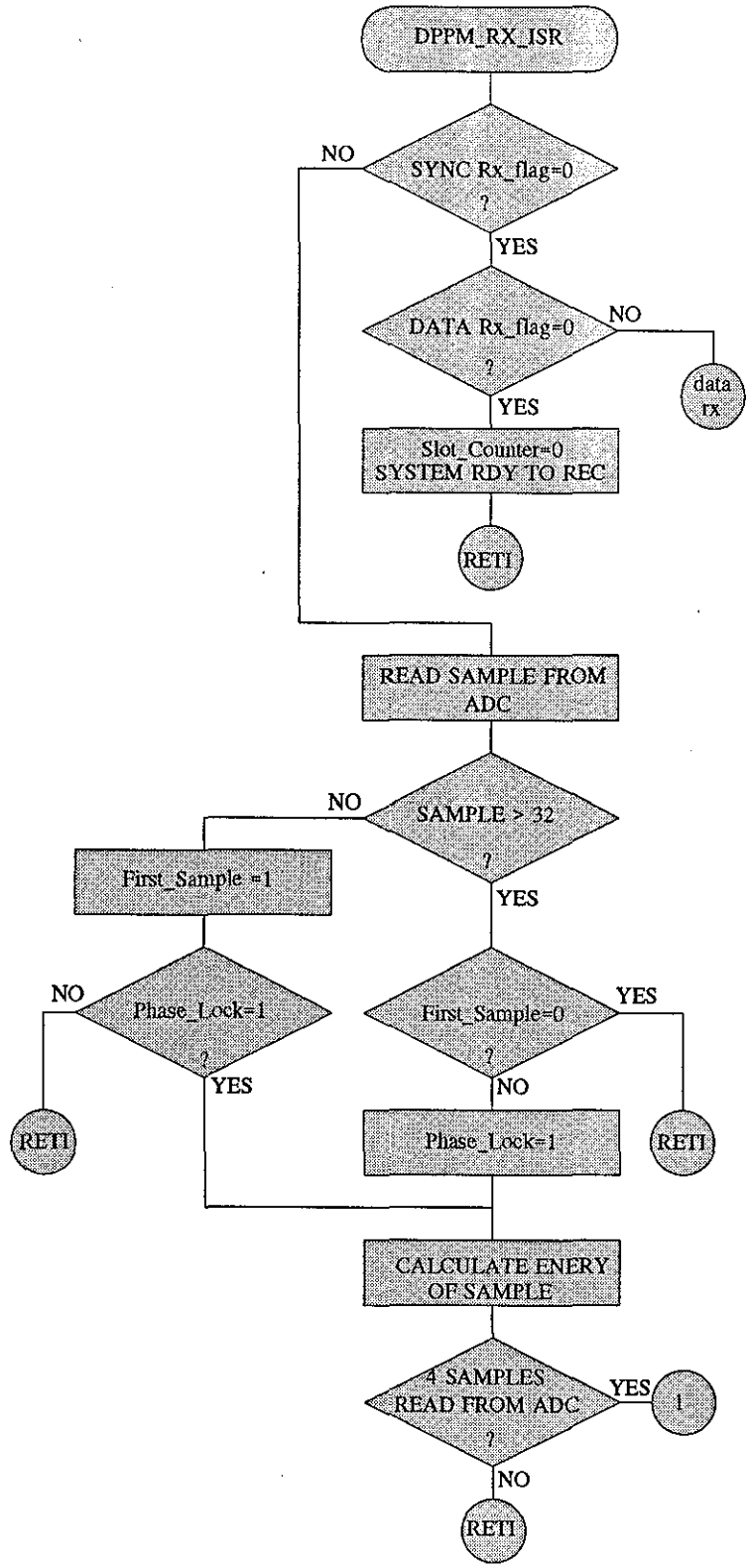


Figure 7.5 SYNC signal detection and DPPM decoding

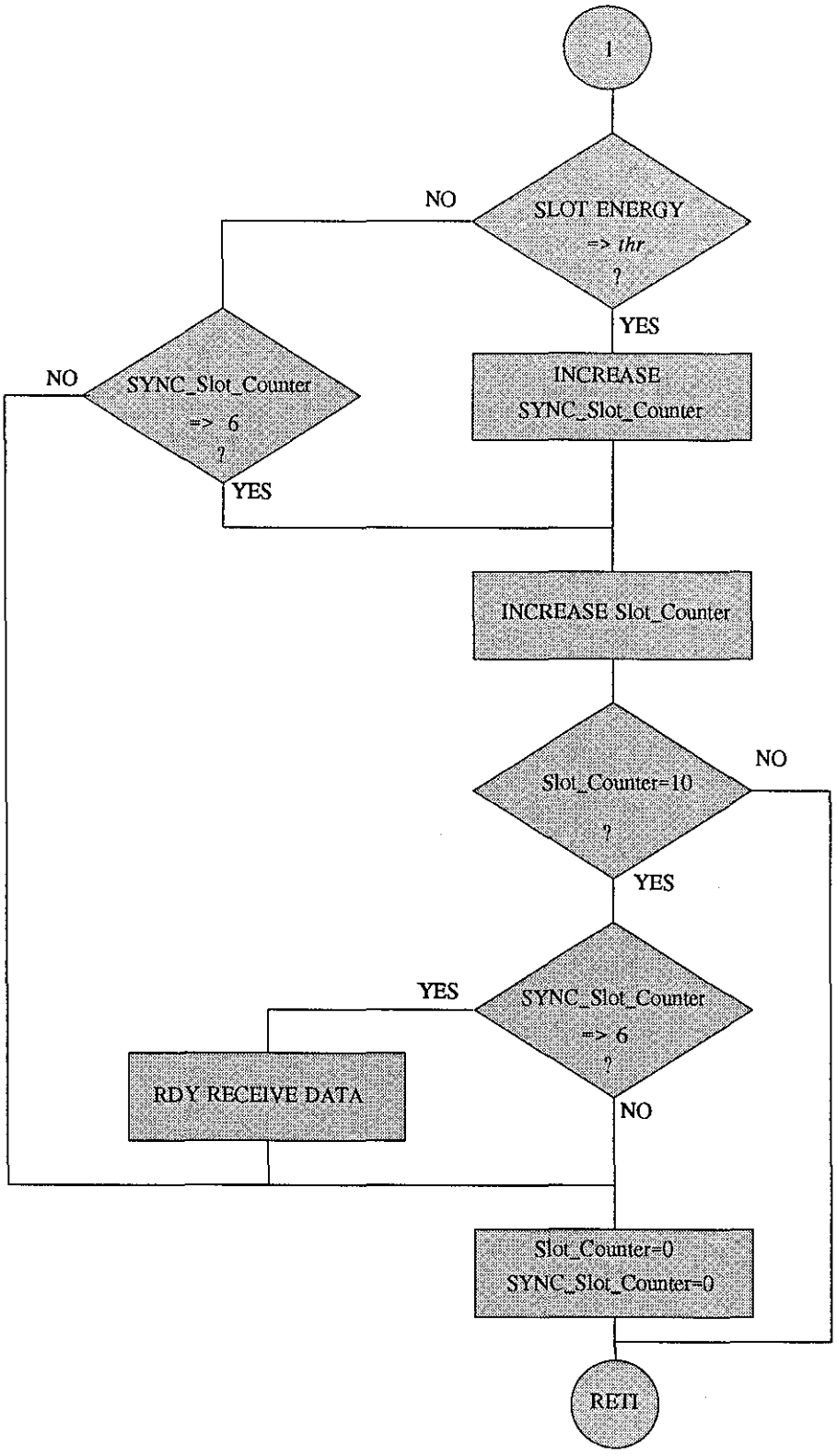


Figure 7.5 SYNC signal detection and DPPM decoding

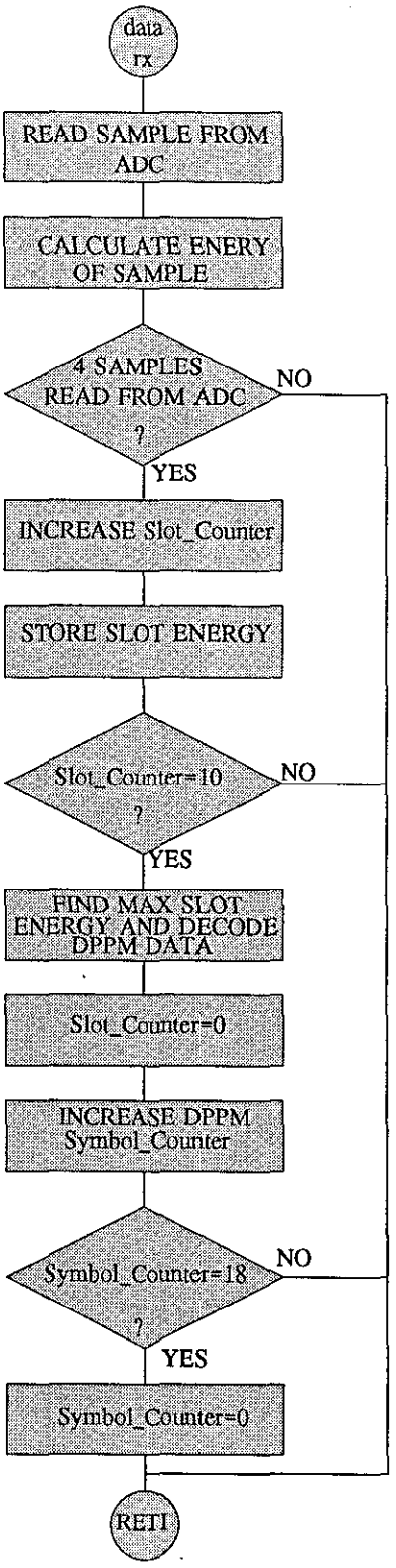


Figure 7.5 SYNC signal detection and DPPM decoding

frame synchronization signal). With the data transmission protocol employed in the system, the frame synchronization between the transmitter and the receiver is updated every 22.5 ms. This is necessary because, at the receiver, the pulse position may be significantly diverged from its original location as a result of a timing error during demodulation process. Moreover, in some cases, the communication link may be intermittently intercepted and need to be resumed. Under these circumstances, the transmitter and the receiver must be synchronized and this is achieved by transmitting the SYNC signal as previously defined. Detection of this signal is also an important aspect of the receiver and it is entirely based on the DPPM baseband signal. In other words, the slot synchronization must be established before seeking the SYNC signal. Once the above condition is provided, the rising edge of the SYNC baseband signal can be detected. In the following sections, presence of a signal in the communication channel, either a DPPM signal or the SYNC signal, is examined and processing of this signal is performed.

When the receiver is operated, it is likely that the transmitter and the receiver are not synchronized. Therefore, a difficulty will arise to make a decision about the input signal, i.e. either the SYNC signal or a DPPM baseband signal or multipath signal or noise signal. The presence of a DPPM baseband signal must be verified and the slot synchronization must be established. In the system, this is achieved by introducing a threshold-based detection method.

If the first input sample magnitude is above the threshold level, this could be a sample from a DPPM pulse or SYNC signal or even from a multipath signal. Moreover, it may not be the closest sample to the rising edge of the input baseband signal. Such a case is considered as late synchronization of the baseband signal, as illustrated in Fig.7.6. If the first sample is from the DPPM pulse, the system does not perform any decoding process, since the speech frame synchronization is not established. Lets assume that the first sample is taken from the SYNC signal. If this is not the closest sample to the rising edge and the system is allowed to continue the DPPM baseband signal decoding, it is likely that a high error rate will be experienced. This is simply because of the accurate timing relation between the SYNC signal and the DPPM signal. Therefore, it is important for the decoder that the first sample magnitude must be lower than the threshold level so

that slot synchronization is established. This phenomenon is called as early synchronization of the baseband signal as performed in Fig.7.6. The source for this sample is assumed to be either from noise or a multipath signal. The receiver continuously samples the input baseband signal and compares its magnitude with the fixed threshold value. As soon as a sample whose magnitude is higher than that of the threshold level is detected, the slot synchronization is verified and the system is enabled to search the SYNC signal.

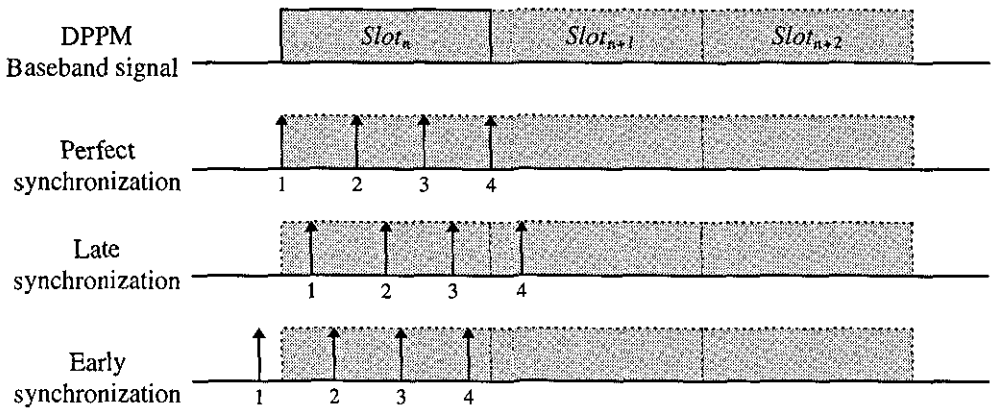


Figure 7.6 DPPM baseband signal sampling and synchronization

7.7 Threshold Definition and Slot Synchronization

In the system, threshold based signal detection is implemented. Although this simplifies the design of the receiver section, it reduces the performance of the system because of the fixed threshold level. In the controlled test environment, the multipath signal magnitude and ambient noise magnitude play an important role. For the system, the threshold level is assumed to be greater than both of these signal magnitudes. However, such an arrangement will obviously limit the effective range of the system because of a reduction of input signal magnitudes with increased separation of two units. If the threshold level is inappropriately selected, the receiver may also attempt to synchronize to the multipath signal as if it is the direct path signal. Depending on its effect, either constructive or destructive, the DPPM baseband signal detection performance is improved or decreased. For the current system, a 0.6 V fixed threshold level (i.e. equivalent to *32d* when digitized) was chosen after the system was tested in the

department's water tank. In this situation, the maximum communication range of the system is approximately 10 m.

The system is designed for a communication range of 100 m. To meet these requirements, a fixed threshold comparison method must be abandoned and an optimum slot synchronization method must be introduced. The slot synchronization issue in optical communications is studied in detail [138-141] and the threshold level is defined by considering the noise at the receiver. Although this kind of threshold setting is desirable for underwater communications, it is restricted by the multipath propagation, as studied later. Another method of achieving synchronization, for both speech frame and slot, is the transmission of a pilot tone as the SYNC signal at a different frequency from the carrier signal [163]. However, this will require modification in the system design, hence it is not considered here. The underwater communication system is designed to support threshold-based signal detection and synchronization. In the following, we propose a method which implements a variable threshold setting and this must be considered in any future development of the system.

A threshold level setting algorithm must be developed and executed as soon as the system is set to receiver mode operation. Samples of the input signal for the duration of several symbol intervals are processed. The number of symbol intervals is arbitrarily decided and this is limited by the available memory size. However, it is better if a speech frame interval is utilized since a SYNC signal is also included. From these samples, the maximum and minimum sample magnitudes representing the DPPM signal and the noise are extracted. Then, a SYNC reference signal consisting of 24 samples is generated in software to implement digital synchronization and its magnitude is set equal to that of the sample with maximum magnitude. Then the sum of the errors, $e[n]$, between the simulated SYNC signal, $a[k]$, and input samples, $x[n]$, can be computed from:

$$e[n] = \sum_{k=0}^{23} a[k] - x[n+k] \qquad n = 0,\ 1,\ 2,\ldots N-k \tag{7.4}$$

where k is the number of samples in the SYNC signal, n is the number of samples in the analysis frame and N is equal to 900 if a speech frame interval is used. The minimum sum of the errors will occur when two signals are synchronized and its index represents

the position of the rising edge of the SYNC signal. Once the synchronization process is completed, the mean value of the SYNC signal, μ_{sync}, can be calculated and used to define the threshold level. However, selecting the mean value as a threshold level itself will result in a SYNC detection error, therefore the mean value of the adjacent samples to the SYNC signal, $\mu_{adjacent}$, may be used for the fine tuning. By considering these, the threshold level can be defined as:

$$thr = \mu_{adjacent} + \frac{\mu_{sync} - \mu_{adjacent}}{2} \tag{7.5}$$

However, if any sample of the SYNC signal is lower than the threshold level, the *threshold level (thr)* must be readjusted. In this case, the division component of the Eq.7.5 can be multiplied by a factor of g, i.e. $0 < g \leq 1$. For a reliable threshold definition, the maximum sum of the errors, i.e. two signals are dissimilar to each other, must also be utilized. This is essential, especially when the minimum and maximum of the sum of the errors, i.e. e_{min} and e_{max}, become close to each other. This suggests that there is no transmission in the channel and the analysed signal is a noise signal. Therefore, the condition of $e_{max} \geq 2\ e_{min}$ must be provided, where the multiplication factor is arbitrarily selected, then *thr* must be defined as in Eq.7.5. Once the initial threshold setting is achieved, it can be updated by considering the DPPM signal magnitude for subsequent speech frames.

7.8 SYNC Signal Detection

When the slot synchronization is achieved with a fixed threshold level as studied before, the speech frame synchronization signal, i.e. SYNC signal, is searched for. In this section, detection of the SYNC signal is presented. As described before, the baseband signal is sampled at 40 kHz, i.e. 4 samples per slot, and quantized at 8-bit resolution. Then the energy of each slot, *E[slot],* is calculated as:

$$E[slot] = \sum_{k=1}^{4} x[k]^2 \qquad 0 \leq slot \leq 7 \tag{7.6}$$

where k is the sample number, $x[k]$ is the digitized baseband signal. Six of the ten slot intervals are reserved for the transmission of the SYNC signal, as shown in Fig.7.2a. At the receiver, it is expected that the energies of these six slots are higher than that of the non-pulse intervals. Moreover, because of the fixed threshold level, i.e. 0.6 V in analogue and *32d* when digitized, it is possible to define the expected minimum slot energy value. By using Eq.7.6, this is calculated as *4096d* which represents 39 μJoule in the analogue baseband signal. After the slot synchronization is completed as described in Section 7.6, the energy of each successive slot is calculated and compared to the predefined energy level as performed in Fig.7.5. When six successive slot energies are found to be equal or greater than the threshold energy level as presented in Eq.7.7, it is said that the SYNC signal is detected and speech frame synchronization is established.

$$if \begin{cases} E[0] \geq thr \\ \quad \cdot \\ \quad \cdot \\ E[5] \geq thr \\ E[6] \leq thr \\ E[7] \leq thr \end{cases} \quad SYNC\ signal\ is\ present \tag{7.7}$$

If the speech frame synchronization is not achieved, the DPPM decoding process is not implemented, thus its detection is absolutely necessary. Moreover, because the rising edge of the SYNC signal is taken as a reference in the demodulation of the DPPM signal, the sharper the edge and the higher the sampling frequency the more accurate the pulse position detection is achieved.

7.9 DPPM Baseband Signal Demodulation

Since the DPPM demodulator works on the position estimation principle, a timing error significantly increases bit error rate. Therefore, accuracy in the detection of the SYNC signal is important. Once this is achieved, as described before, the DPPM slot data decoding can be easily implemented and this process is studied in the following pages.

During the experiments conducted in the department's tank room, the transmitter, unit 1, and the receiver, unit 2, were closely positioned so that the magnitude of the multipath signal is minimized although it is more significant in the SYNC signal detection. The receiver demodulates the DPPM baseband signal as it was in the SYNC signal, without considering the threshold level. The threshold level is only utilized during the synchronization process and here we assume that the transmitter and the receiver are synchronized. Baseband signal energy for each slot interval, i.e. 8-slot intervals, is calculated by means of Eq.7.6 and their results are stored in eight memory locations, as shown in Table 7.4, to be decoded during the guard interval (see Fig.7.2a).

Index	*0*	*1*	*2*	*3*	*4*	*5*	*6*	*7*
Slot Energy	E[0]	E[1]	E[2]	E[3]	E[4]	E[5]	E[6]	E[7]

Table 7.4 Slot energy values

The decoding of the DPPM data is done by comparing the slot energy values from Table7.4. The index of the maximum energy, which indicates the possible pulse position, is accepted as the transmitted digital data. The same processing is done for the subsequent 17 symbol intervals. While the DPPM decoding is in process, the decoded digital data are stored in appropriate memory locations in a similar way to that shown in Table 7.2. After that they are further decoded to form the quantized speech parameters in a similar way to those presented in Table 7.1. By implementing such a decoding process, the transmitted speech parameters are successfully recovered and utilized in the speech synthesizing algorithm in Chapter 3.

During the DPPM decoding scheme, the multipath signal energy is assumed to be relatively small compared with the direct path signal energy. Providing that perfect synchronization is achieved, the decoded speech parameters should be free of error. However, if the above conditions are not fulfilled, an error in the pulse position detection will occur. When speech synthesizing is conducted, because of this error and depending on their significance, speech quality can be decreased.

7.10 Significance of Multipath Propagation in DPPM Signal Detection

In order to eliminate the effect of multipath propagation during system testing, the transducers of the two units were positioned 1m apart at 1m depth. In this arrangement, the direct path signal magnitude was always much greater than the multipath signal as, illustrated in Fig. 7.7, and the threshold level was set to a value greater than the multipath signal magnitude. It is necessary to mention here that threshold based signal detection is only used for the SYNC signal recognition and it is an essential part of the system to operate accordingly.

The multipath propagation effect is considered in two aspects as:

-synchronization and SYNC signal detection

-DPPM data demodulation

The multipath propagation may significantly affect the synchronization process of the baseband signal and the SYNC signal detection. If it appears just in the front of the SYNC signal and its magnitude is greater than the threshold level, as illustrated in Fig 7.8, the transmitter and the receiver will be synchronized to the multipath signal. Then the succeding demodulated digital data are going to be error corrupted.

The multipath signal may appear in any of the slot intervals and it may have a constructive or destructive effect on the direct path DPPM signal, as shown in Fig.7.8. If it is constructive, the energy of the data slot is increased. This clearly improves the detection of the pulse position. However, if the multipath signal is destructive, i.e. energy of the data slot is lower than other slot energies, then the decoded data will be error corrupted. In the system design, no precaution is taken to overcome this limitation.

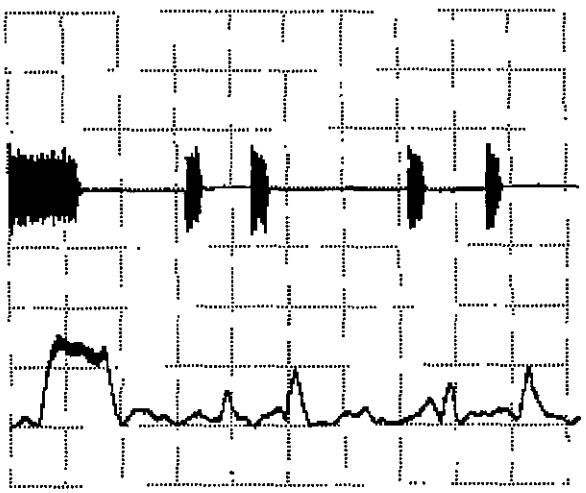


Figure 7.7. Transmitted and received DPPM signal

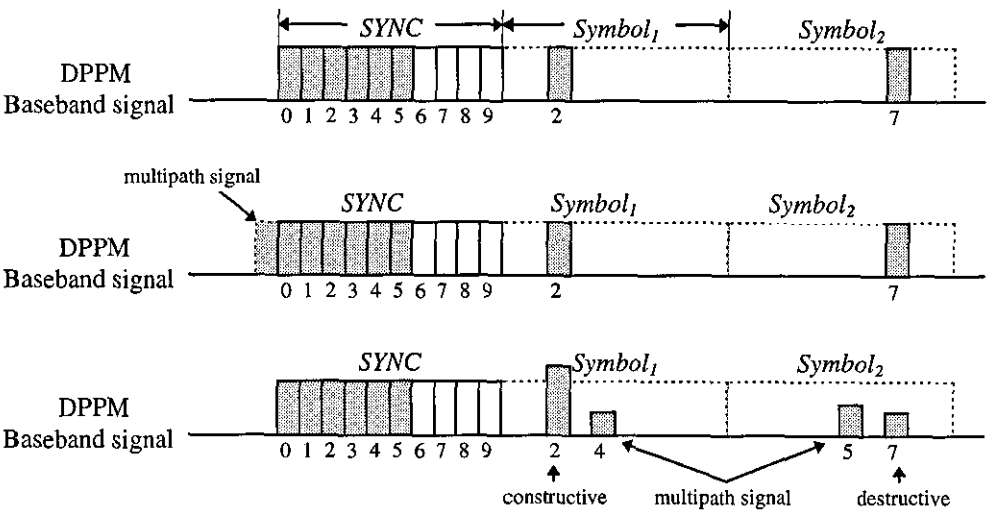
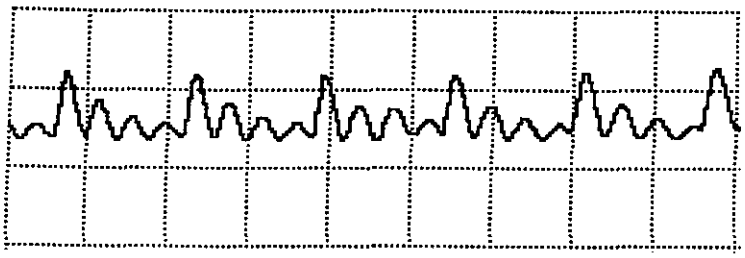


Figure 7.8 Possible multipath effect in DPPM signal detection

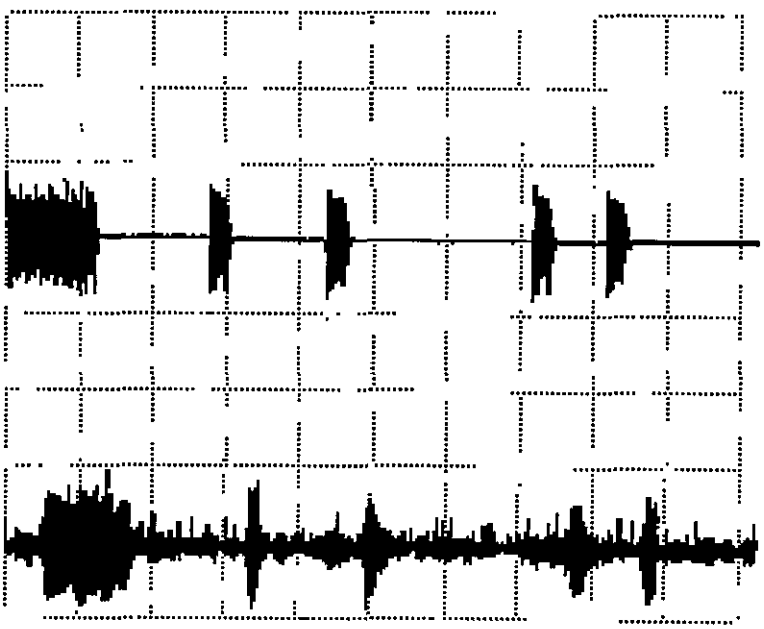
7.11 Experimental Results and Discussions

The digital underwater acoustic voice communication system is designed as described in Chapter 6. Speech encoding and decoding algorithms are implemented as studied in Chapter 3, and modulation and demodulation of speech data are achieved as defined in Chapter 5. When the two systems are tested off-line, where almost perfect square wave baseband signal is obtained and multipath signal and channel noise are absent, the problems discussed above are not encountered. This confirms that both systems operate accordingly with the design criteria. During field testing, one of the two units was configured as a transmitter, the other as a receiver, and their transducers were positioned approximately 1 m apart in the department's tank. Dry testing, where only the hydrophones are submerged, was applied since wet testing was planned only after circuit miniaturisation.

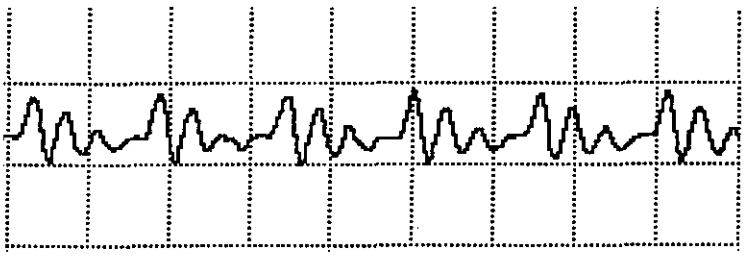
The encoded speech parameters were successfully transmitted by unit 1 through water and decoded at the receiver, i.e. unit 2, then the speech signals were synthesized as illustrated in Fig.7.9. As expected, the speech quality at the receiver was synthetic. With the current speech coding algorithm, achieving a good quality speech signal is a difficult task and requires modification of the algorithms. During the testing of these two systems, we observed that there were occasional interrupts in the synthesized speech signal and these interrupts resulted in the generation of noise for a period of 22.5 ms at the output of the AIC (Analogue Interface Circuit). The interrupts were caused by the SYNC signal detection error. In the software design, no precautions were taken to correct such an error. The system is designed to execute the synthesizing algorithm although the SYNC signal is not detected appropriately. During the synthesizing of speech signals, the contents of the speech parameters' memory (similar to Table 7.2) are modified. When there is an error in the SYNC signal detection, the synthesizing algorithm uses these incorrect data and a speech frame is lost. This is the reason why interrupts in the speech signal synthesis is experienced and noise is obtained. Several solutions can be recommended to eliminate such a problem; one is to use previous speech parameters, the other is to generate a comfort noise and one of these may be considered in a future design of the system.



(a)



(b)



(c)

Figure 7.9 Speech communications and implementations of DPPM transmission and detection (a) input speech signal, (b) transmission of encoded speech parameters, (c) synthesized speech signal.

For the voice communication system presented in this thesis, no bit error rate measurements are performed due to the hardware limitations. However, such an analysis is essential to illustrate the performance of the system and efficiency of the modulation and demodulation tasks. In order to calculate the bit error rate, use of a computer is inevitable, but this will require an interface circuit and modification of the software. The only possible method of measuring the error figure, or pointing out its significance, is to compare the synthesized speech signals at the transmitter, unit 1, and receiver, unit 2. Therefore, the transmitter is configured to analyse and synthesize the speech signal.

Since the synthesized speech signal at the transmitter represents error free transmission and at the receiver, it represents possible error corrupted transmission, a subjective test may be utilized to make a conclusion about the significance bit error. In this study, we only present the waveform similarities between the synthesized speech signal at the transmitter and at the receiver, as shown in Fig.7.10. We can conclude that the two waveforms are identical, i.e. no error or insignificant error is detected. Moreover, a listening test was conducted on five adult subjects and it was concluded that the speech quality of the two systems was the same. However, in the literature, a theoretical study of the probability of a bit error in the DPPM transmission is given in detail [164]. It is assumed that perfect synchronization is established and a rectangular envelope of the signal is detected. Under these conditions, the probability of bit error, *PBE*, and the signal-to-noise ratio, *SNR*, are given as:

$$PBE = \left(\frac{M}{2}\right)\frac{1}{\sqrt{2\pi\frac{E_s}{N_o}}}e^{-\frac{E_s}{2N_o}} \tag{7.8}$$

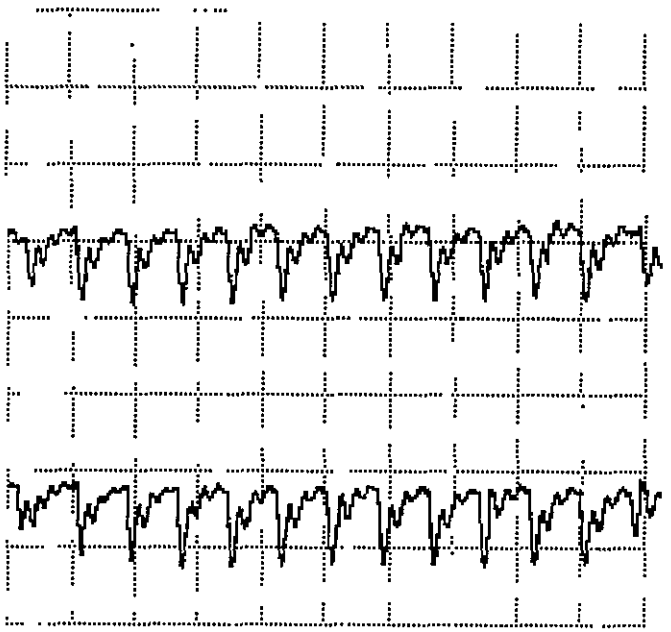
$$SNR = 10\log_{10}\frac{E_s}{N_o} \tag{7.9}$$

where M is number of the DPPM data slots, E_s and N_o are signal energy and noise and multipath signal energy respectively. The numerical results of the above equations are

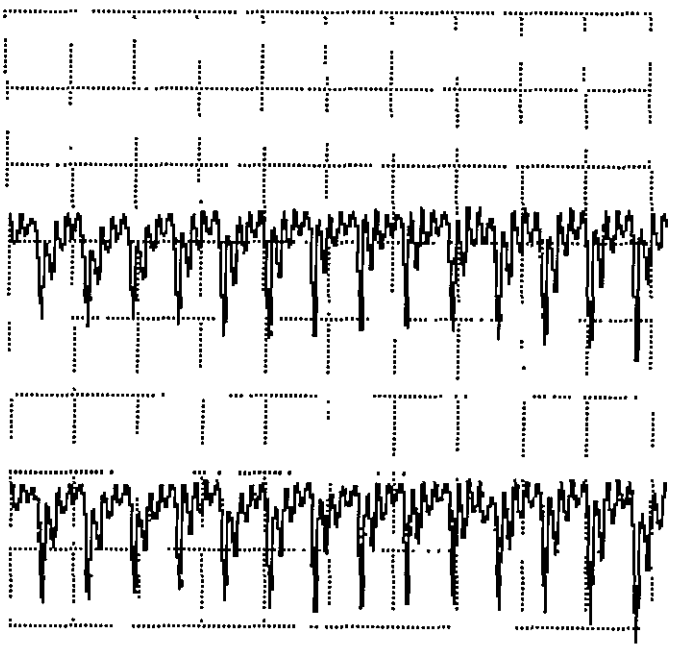
listed in Table 7.5. As expected, when the SNR is increased the DPPM transmission performance is improved.

SNR,dB	0	4	6	8	10	12
PBE	0.967	0.284	0.1092	0.0279	0.0034	0.00014

Table 7.5 Probability of bit error of 8-slot DPPM transmission.

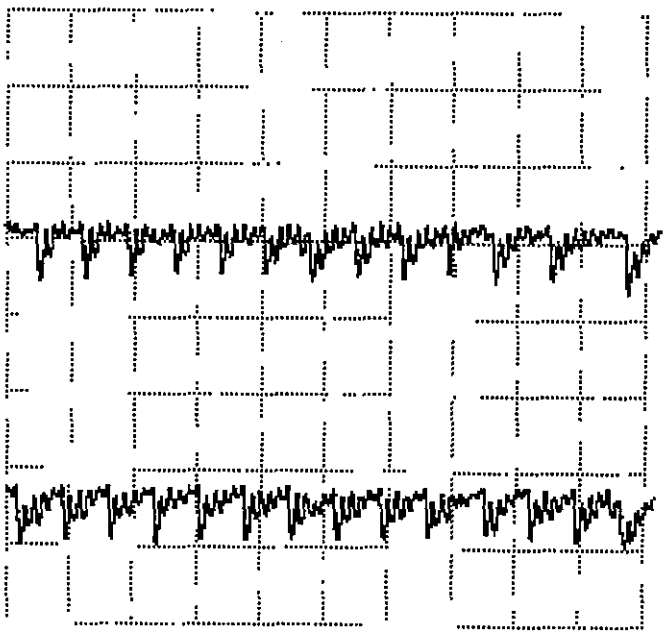


(a)



(b)

Figure 7.10 Analysed and Synthesized speech signals at the transmitter and at the receiver (cont.).



(c)

Figure 7.10 Analysed and Synthesized speech signals at the transmitter and at the receiver.

CHAPTER EIGHT

CONCLUSIONS AND RECOMMENDATIONS

8.1 Conclusions

The main aim of the thesis has been to design and develop a real-time digital underwater acoustic voice communication system suitable for SCUBA divers and for communication between underwater vehicles. To improve the speech signal quality, several noise-cancelling techniques were considered and simulation studies were carried out. From the foregoing work and the experimental results obtained, the following conclusions may be drawn:

1. A new digital underwater voice communication system based on a single-chip TMS320C31 DSP has been designed to operate as a transmitter and receiver. Its design topology is considered in two sections as hardware and software. The design allows implementation of complex digital speech coding techniques and digital modulation methods for transmission and reception of acoustic signals. It also allows the transmission of any desired signals or data, such as ECG signals and breathing rate data from a diver, with minor alterations in the hardware and software.

2. A multi-channel interrupt-based operation has been developed. This allows the DSP to execute multi-task operations without introducing errors in the modulation and demodulation or distortion in the speech signals. Three of the four external interrupt inputs are employed to select the operation mode, detect the keypad encoder output and read data from the receiver ADC. The internal interrupts of the serial port and the **Timer0** are utilised for reading and writing speech samples to the AIC and modulation of speech parameters respectively.

3. In the system, the speech signal is sampled at 8 kHz with a resolution of 14 bits (i.e. the required bit rate is 112 kbit/s). Transmission of the quantized speech samples demands a high channel bandwidth. This is limited for the underwater acoustic channel although such a transmission rate can be achieved by employing a hydrophone with a high resonance frequency and wide bandwidth. However, using such a hydrophone with the given specifications decreases the communication range. In order to decrease the data rate, the speech signal is compressed by implementing the LPC algorithm in real-time with the DSP. The calculated LPC parameters are quantized at a rate of 2.4 kbit/s. Because LPC is a lossy compression technique, the quality of the reconstructed speech is degraded.

4. The 8-slot DPPM technique is successfully implemented for the transmission of LPC parameters through an underwater channel at the symbol rate of 800 baud. Decoding of the received acoustic signal is accurately done although errors due to loss of the synchronisation signal are occasionally detected. This results in loss of a speech frame and it introduces distortion of the decoded speech signal.

Although detection errors can occur at the receiver, a bit error rate measurements have not been performed for the present system. Accurate measurement of the error rate would require an interface circuitry to a computer, but this is restricted by the hardware. However, an approximate figure for this can be taken by writing a separate assembly program and controlling one of the timers. The computed error rate can be used to define the frequency of a timer and the bit error rate can be extracted from the frequency measurement.

5. Two systems were built and configured to work as a transmitter and receiver. Operation and performance of these systems was tested in the department's tank which is built to simulate a shallow and multipath-dominant channel. During the experiments, one of the systems was selected as a transmitter while the other operating as a receiver. The experimental results illustrated that both systems performed satisfactorily in such a communication channel.

6. Since breathing and bubble noises also limit the intelligibility of the speech signals, several simulation studies were carried out to cancel these noises and to enhance speech quality. Breathing noise is successfully cancelled by measuring the zero crossing rate and energy of each speech frame. To cancel bubble noise two different methods were examined, spectral subtraction and an adaptive noise cancelling. Simulation results from these studies suggested that the adaptive noise-cancelling method can be used. Moreover, it is also concluded that the type of microphone and diver mask used for these simulations determines speech signal quality. Hence, an AGA-type mask incorporating a noise cancelling microphone was shown to improve the speech quality.

In summary, the digital design concept presented in this thesis provides a means of achieving a digital underwater acoustic voice communication system. This is confirmed by the experimental results taken throughout study. Since the system now operates satisfactorily, its design must be reconsidered to miniaturise it by using appropriate components as in mobile telephone technology.

8.2 Recommendations

The digital underwater acoustic voice communication system presented in this thesis is a prototype. The system has hardware and software related limitations and it is suggested that they must be taken into consideration in future designs. In this section, these limitations are introduced and improvement methods are discussed.

An important requirement for an underwater voice communication system is that it must be worn with a diving mask, i.e. its physical size must be as small as possible, and it must consume low power. Moreover, since the system is being used in an underwater environment it should not be damaged by the high ambient pressure. To achieve these requirements, an ergonomic design of the system is necessary and suitable packaging must be investigated.

In the system, external EPROM and SRAM units occupy a considerable amount of physical space and their elimination will greatly contribute to the miniaturisation process

of the system. Recently, many DSPs have been manufactured with improved features and speed. Therefore, in future applications, it is important to employ a DSP with on-chip ROM and RAM. *TMS320C53 - TMS320C57* DSPs are recommended since they incorporate internal *4 kWord RAM* and *16 kWord ROM* and two serial ports, etc. In addition, they can be easily interfaced to a TLC320AC01 AIC. If low power consumption is required, a *TMS320C546* DSP is suggested since a 3 V power supply is adequate to operate it. After the design stage is successfully completed, an Application Specific DSP incorporating memory units and an AIC as well as general purpose A/D and D/A converters can be manufactured. Since these ICs can be fabricated as surface mount devices, miniaturisation of the system becomes achievable.

In the design of the digital underwater acoustic voice communication system, mobile telephone technology is taken as a model and the same features are aimed to be included. Therefore a key pad unit is incorporated in the system in order to enable divers to communicate privately if necessary. However, associated algorithms for the key pad control and communication protocol have not been completed. The key pad is considered as an essential part of the communication system and it should be fully functioning in future designs. Because the key pad unit will be subjected to the ambient pressure, it should be waterproofed and constructed so as not to be affected by pressure variations.

To select the operation mode of the system a press-to-talk switch is included. The fact that the switch occupies one hand of the operator during communications is an obvious limitation. In order to overcome it, as used by some of the analogue systems, a voice activated switch must be included to replace the press-to-talk switch.

The speech signal from a diver is corrupted by breathing and bubble noise. In this thesis, simulation studies were carried out to cancel these noises. However, especially for the bubble noise cancellation, further study is required to improve speech quality. The noise-cancelling algorithms are excluded from the present system but they must be included in future designs.

Linear Prediction Coding of the speech signal causes speech quality reduction while providing the advantage of a 2.4 kbit/s transmission rate. To improve the quality, several speech coding algorithms are available and these are widely used in mobile telephone technology and multimedia applications. Amongst them, the CELP speech coding method is particularly suitable for underwater acoustic voice communications. Recently, this coding method has been made available in VLSI packages. For example, the Qualcomm CELP voice coder Q4401 (supplied by Qualcomm Inc. USA) can be utilised. This is capable of implementing a variable rate coding from 800 bit/s to 9.6 kbit/s and it can be interfaced to any DSP through its serial communication port. However, using such a component would require modifications in the hardware and software in order to enable the DSP to communicate with the Q4401.

In this design, implementation of channel coders is not considered because they would increase the data rate in spite of reducing the error rate in decoded LPC speech parameters. Therefore, in future designs, a suitable error correction coding technique can be implemented to improve the system's performance.

The Digital PPM method has been successfully applied to underwater acoustic data transmission. However, if the encoded speech data rate is increased by using a different speech coding method (i.e. higher channel bandwidth is required and this may not be supported by the hydrophone) the DPPM may not perform sufficiently well. Therefore, the M-ary PSK modulation method, which provides efficient bandwidth utilisation, can be implemented even though it requires modification to the hardware and software. In such an application, the transmitter section must constitute a D/A converter to generate a carrier waveform at appropriate phases for data modulation. However, in the receiver side, detection of the phase modulated acoustic signal necessitates an A/D converter capable of operating at frequencies above twice of the Nyquist frequency. An 8-bit MAX153 ADC at a maximum sampling frequency of 1.5 MHz can be used for this purpose since it is already included in the present system.

The synchronous communication principle is utilised to demodulate DPPM signals in the present system. Due to multipath propagation and fixed threshold level setting, the synchronisation signal is occasionally undetected. However, this limitation can be improved by introducing a correlation detector although it will increase the software complexity. Missing the synchronisation signal results in the loss of a speech frame, hence introducing noise in the reconstructed speech signal. However, no precaution has been taken to overcome the effect of synchronisation signal loss. The noise signal introduced due to synchronisation signal loss can be eliminated by switching off the speech synthesising algorithm and producing "comfort noise". A better solution would be to use the last correctly detected speech parameters for speech synthesising. This improvement can be achieved by including a synchronisation signal-checking algorithm and modifying the speech synthesising algorithm.

The multipath propagation phenomenon of underwater communication channels introduces serious limitations in signal decoding and it is recommended to implement a suitable multipath elimination method. This will be essential for the DPPM signal since the demodulation process utilises threshold-based detection of the direct path signal. If the M-ary PSK modulation technique is implemented, a channel equaliser becomes an essential part of the system since multipath propagation may introduce significant distortion on the received signal.

Last but not least, the digital underwater acoustic voice communication system may be used for biotelemetry system to transmit biomedical signals such as the ECG from a diver to a remote receiver. This can be achieved by utilising appropriate sensors and signal conditioning units within the system. This operation mode can be activated while voice communication is not taking place.

Appendix A

DSP Configurations and Memory Units

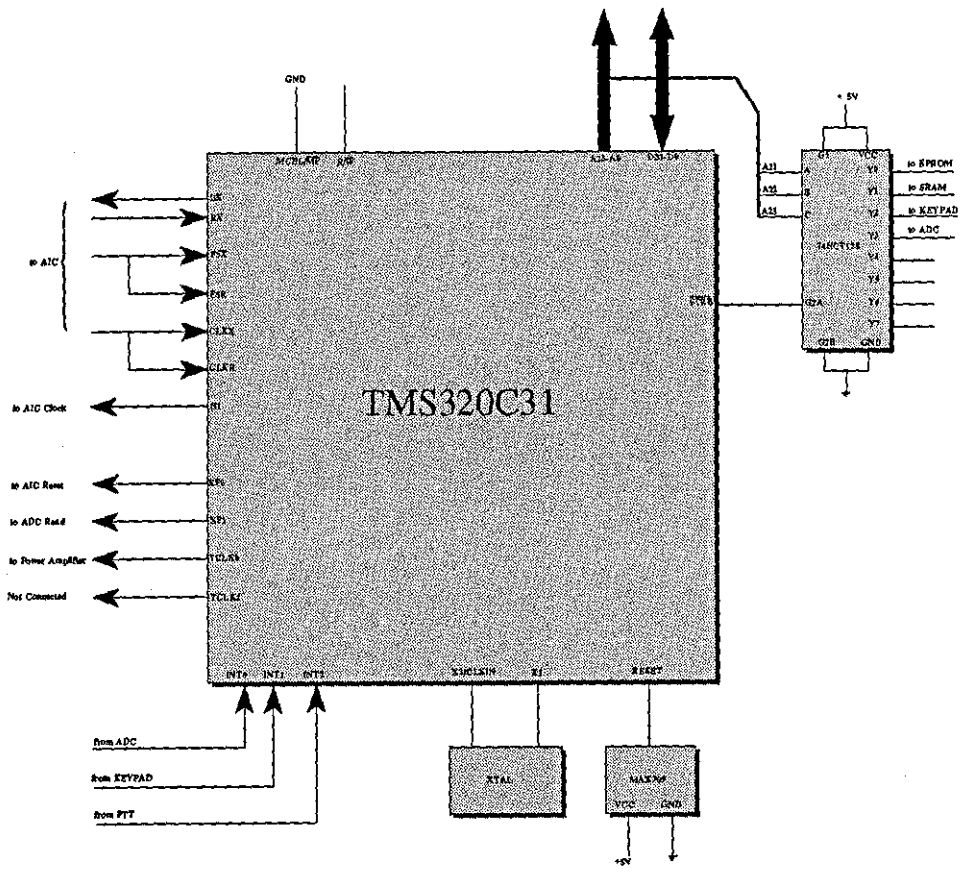


Figure A.1 Configuration of the TMS320C31 DSP with the memory units (cont.)

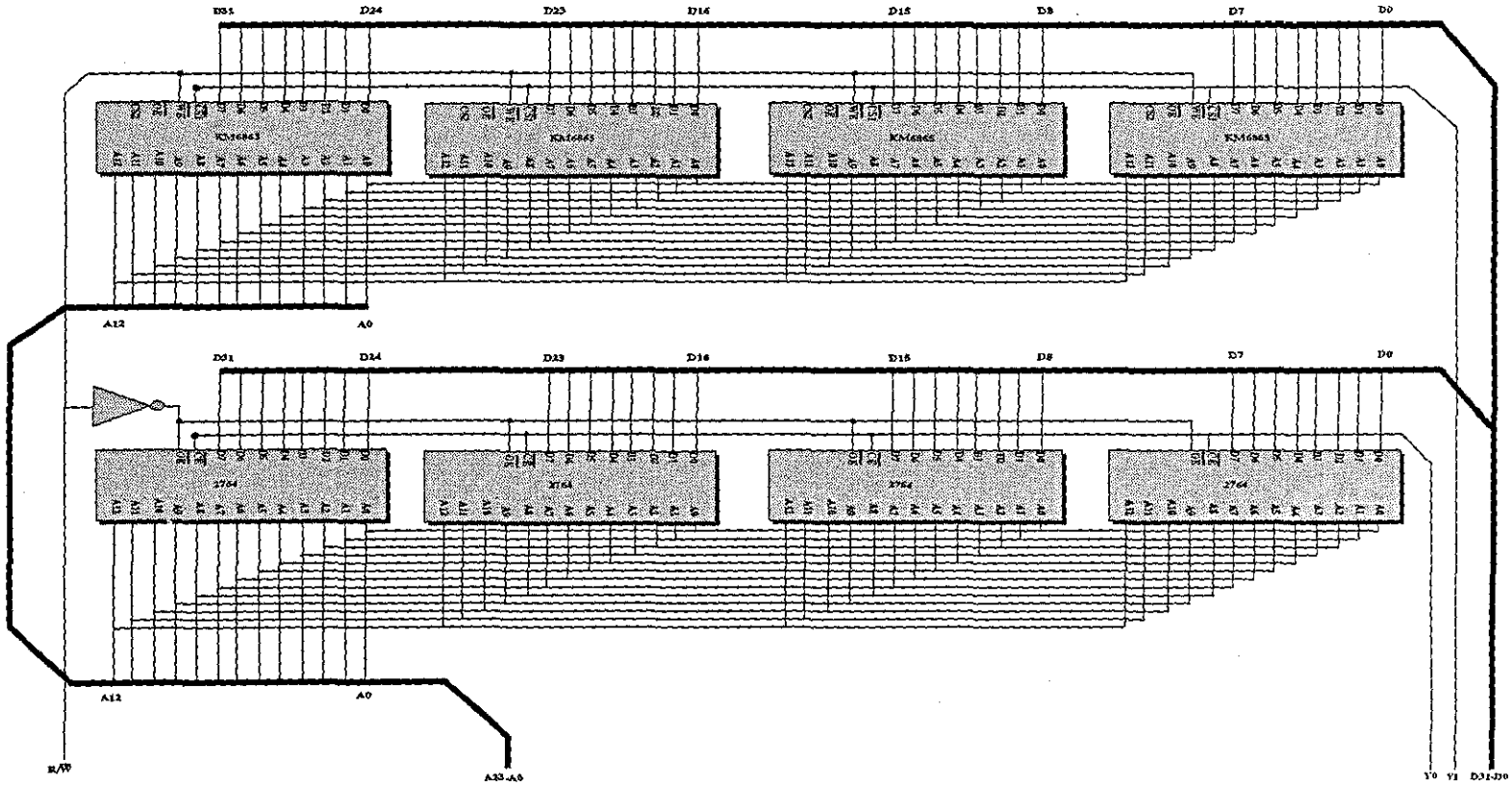


Figure A.1 Configuration of the TMS320C31 DSP with the memory units.

Appendix B

Receiving Sensitivity Measurement of a Hydrophone

The electrical properties of the hydrophone employed in the system is described in Chapter 6. However, the receiving sensitivity of the hydrophone is also an important parameter for the receiver design. Measurement of this parameter is based on a comparison calibration method. This is basically applying the same free-field pressure, produced by a projector, to the hydrophone under test and a reference hydrophone and then comparing their electrical output voltages.

This measurement is done in the department's tank and the transducers are configured as illustrated in Fig. B.1. During this experiment, a power amplifier (EIN Model 1140 LA, 9kHz-250 kHz bandwidth and maximum output power of 2000 W), a frequency generator (Farnel Sine-Square Oscillator LF1, 10 Hz - 1 MHz) and a gating system)Bruel & Kjaer, Type 4440 Gating System) are utilised to drive the projector (30 $V_{p\text{-}p}$ fixed voltage across its terminal and variable frequency input). Then the output voltages for the reference hydrophone, e_r, and for the test hydrophone, e_x, are measured. Since the free-field voltage sensitivity of the reference hydrophone, M_r, is known in advance, then the sensitivity of the test hydrophone, M_x, can be found from [152,162]

$$M_x = \frac{M_r e_x}{e_r} \tag{B.1}$$

or in dB

$$20\log M_x = 20\log M_r + 20\log e_x - 20\log e_r \tag{B.2}$$

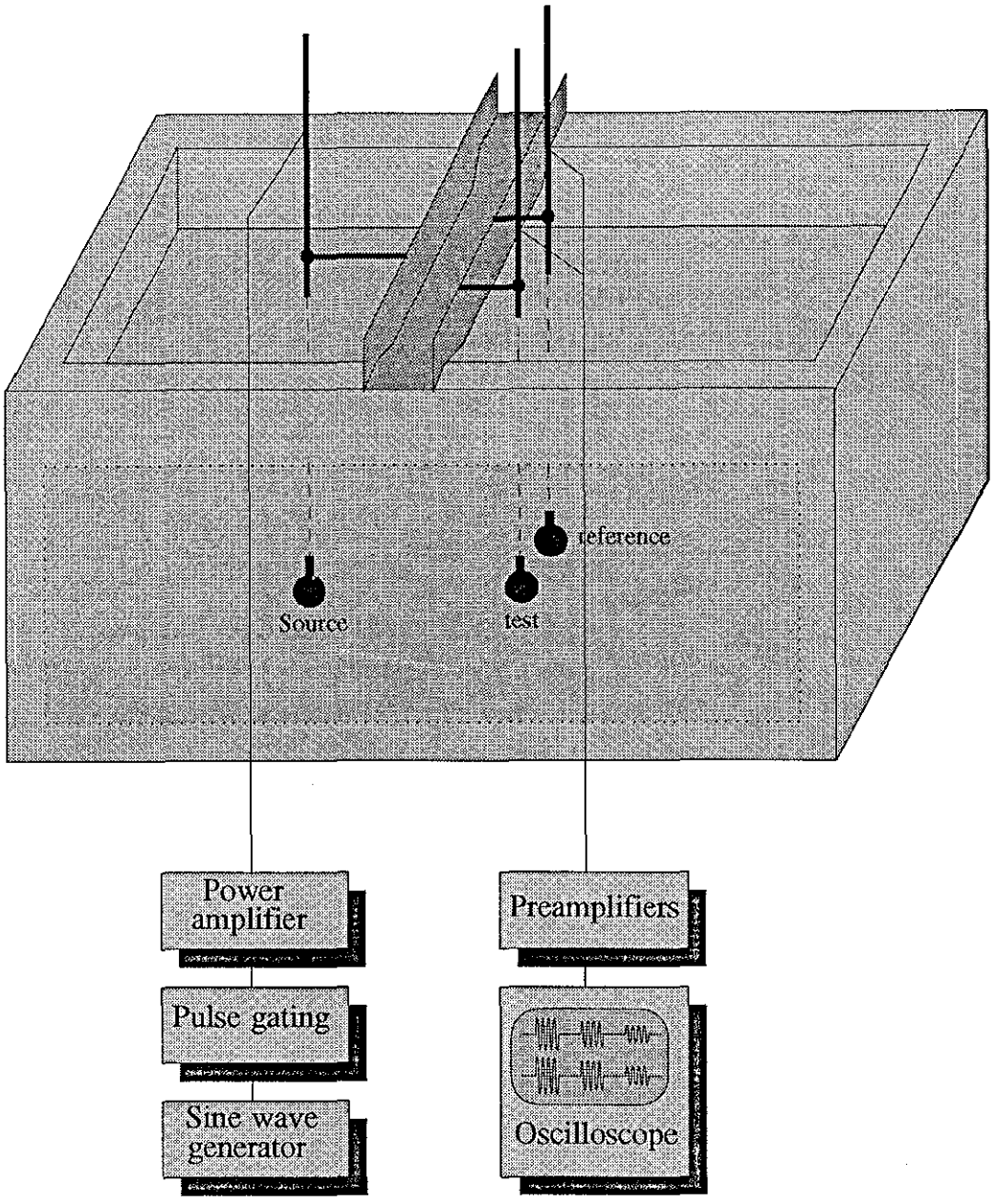


Figure B.1 Hydrophone calibration set-up.

REFERENCES

[1] Wong, W.T.K., Mack, R.M., Cheetham, B.M.G., and Sun, X.Q. "Low rate speech coding for telecommunications," *BT Technol. J.* Vol.14, No.1, pp.28-43 (1996).

[2] Barret, P.A., Voelcker, R.M., and Lewis, A.V., "Speech transmission over mobile radio channels," *BT Tech. J.*, Vol.14, No.1, pp.45-55 (1996).

[3] Woodward, B. "Underwater Telephony: Past, Present and Future," *Colloque De Physique*, Colloque C2, No.2, pp. C2_591-C2_594(1990).

[4] Baume, D., Godden, D. and Hipwell, J. " Improving diver communications," *Underwater Systems Design*, pp.21-23 (December 1979/January 1980).

[5] McIntosh, W., "Underwater communications: An oil company viewpoint," *J. Soc. Underwater Technology*, Vol.11, No.3, pp.14-26 (1985).

[6] Hicks, R.J. and Virr, L.E. "Underwater communications-a review," *Int. Conf. Divetech'81: The Way Ahead in Diving Technology*, Society for Underwater Technology, London,(1981).

[7] Mackie, R.D.L., and Smith, N. "Diver communications and helium speech unscrambling," *Underwater System Design*, pp.19-21 (June/July 1983).

[8] Bud, A.M.G., Holmes, J., White, I., Dtaham, R.J. and Kramer, J. " Improvements in diver communications" *Report OTI 87 500*, Dept. Energy, H.M. Stationary Office (1986).

[9] Moore, R.K., "Radio communication in the sea," *IEEE Spectrum*, pp.42-51 (1967).

[10] Virr, L.E., "Role of electricity in subsea intervention," *IEE Proc.*, Vol.134, Pt. A, No.6, pp.547-576 (1987).

[11] Burt, E.G.C., and Rigby, L., "Electromagnetic through-water communications," *Underwater Technology*, pp.14-26 (Autumn, 1985).

[12] Anderson, V.C., "Acoustic communication is better than none," *IEEE Spectrum*, pp.63-68 (1970).

[13] Berktay, H.O., Gazey, B. and Teer C.A. "Underwater communication past, present and future," *J. Sound Vib.* Vol.7, Part 1, pp.62- 70 (1968).

[14] Hollien, H., Coleman, R.F., and Rothman, H.B. "Evaluation of Diver communication systems by a diver-to-diver technique," *IEEE Trans. Communication Tech.*, Vol. COM-19, No.4, pp.403-409 (1971).

[15] Woodward, B. " Model 230 wireless diver communication system", Internal Report, Loughborough University of Technology (1988).

[16] Webb, H.J., and Webb, J.R., "An underwater audio communicator," *IEEE Trans. Audio Electroacouctic*, Vol.AU-14, No.3, pp.127-135 (1968).

[17] Wenz, G.M. "Acoustic ambient noise in the ocean: spectra and sources," *J. Acoust. Soc. Am.*, Vol.32, No.12, pp.1936-1956 (1962).

[18] Lathi, B.P., *Modern Digital and Analog Communication Systems*, Holt, Rinehart and Winston, Inc., USA (1989).

[19] Mulcahy, M., "A through-water diver communication system," *Sea Technology*, Vol.20, No.8, pp.27-29 (1979)

[20] Gazey, B.K. and Morris, J.C., "An underwater acoustic telephone for free-swimming divers," *Electronic Eng.*, pp.364-368 (1964).

[21] Peck, M.J. " Wireless underwater communications past, present, and future," *Sea Technology*, pp.61-65 (1992).

[22] Overfield, T. " Modern Through-Water Diver Communications," *Underwater Systems Design*, pp.8-13 (March/April 1988).

[23] Clark, A. " Diver communications- The case for Single Sideband," *Underwater Systems Design*, pp.16-18 (January 1989).

[24] Sear, J.K. "Standardisation of Sonar Communications," *Int. Conf. Divetech'81: The Way Ahead in Diving Technology*, Society for Underwater Technology, London,(1981).

[25] Maloney, J. "Diver communication in police service," *Int. Conf. Divetech'81: The Way Ahead in Diving Technology*, Society for Underwater Technology, London,(1981).

[26] Jayant, S.N. and Noll, P., *Digital Coding of Waveforms-Principles and Applications to Speech and Video*, Prentice-Hall, New Jersey (1984).

[27] Konduz, A.M., *Digital Speech Coding for Low Bit Rate Communication Systems*, John Wiley and Sons, New York (1995).

[28] Benvenuto, N., Bertocci, G., Daumer, W.R. and Sparrell, D.K. " The 32-kb/s ADPCM coding standard," *AT&T Tech. J.*, Vol.65, pp.12-19(1986).

[29] Lafuente, L.M. "Adaptive differential pulse code modulation for low bit rate transmission of speech signals," *Electrical Commun.*, Vol.58, No.2, pp.225-229. (1983).

[30] Koo, B. and Gibson, J.D. " Experimental comparison of All-pole, all-zero and pole-zero predictors for ADPCM speech coding," *IEEE Trans. Commun.*, Vol.Com-34, No.3, pp.285-290 (1986).

[31] Bonnet, M., Macchi, O. and Saidane,M.J. " Theoretical analysis of the ADPCM CCITT algorithm," *IEEE Trans. Commun.*, Vol.38, No.6, pp.847-858 (1990).

[32] Sherif, M.H., Bowker, D.O., Bertocci, G., Orford, B.A., and Mariano, G.A. "Overview and performance of CCITT/ANSI embedded ADPCM algorithms," *IEEE Trans. Commun.*, Vol.41, No.2, pp.391-399 (1993).

[33] Spanias, S.A. "Speech coding: A tutorial review," *Proc. IEEE*, Vol.82, No. 10, pp.1541-1582 (1994).

[34] Woodward, B. and Sari, H. "Underwater voice communications using digital techniques," *J. de Physique IV, Colloque C5*, Vol.4, pp. 469-472 (1994).

[35] Jayant, S.N. " Coding speech at low bit rates," *IEEE Spectrum*, pp.58-63 (1986).

[36] Carmody, J. and Rothweiler, J. "Speech coding at 800 and 400 bps," *Electrical Commun.* pp.260-265 (1986).

[37] Tremain, T.E. " The government standard linear predictive coding," *Speech Technology*, Vol.1, pp 40-49 (1982).

[38] "Details to assist in implementation of Federal Standard 1016 CELP," National Communications System, Tech. Bulletin, 92-1 (1992).

[39] Kroon, P., Deprettere, E., and Sluyeter, R.J. "Regular-Pulse excitation- A novel approach to effective and efficient multipulse coding of speech," *IEEE Trans. Acoustic, Speech, Signal Proc.*, Vol. ASSP-34, No.5, pp (1986).

[40] Rowden, C, *Speech Processing*, McGraw-Hill, Cambridge (1990).

[41] Owens, F.J, *Signal Processing of Speech,* Macmillan, England, (1993).

[42] Samuel, D.S. and Ruth, A.D, *Signal Processing Algorithms*, Prentice-Hall, New Jersey, (1989).

[43] Markel, J.D. and Gray, A.H, *Linear Prediction of Speech*, Springer-Verlag, Berlin (1982).

[44] Makhoul, J. "Linear Prediction: A tutorial review," *Proc. IEEE*, Vol.63, No.4, pp.561-580 (1975).

[45] Rabiner, L.R. and Schafer, R.W, *Digital Processing of Speech Signals*, Prentice-Hall, New Jersey (1978).

[46] Deller, Jr.J.R., Proakis, J.G., and Hansen, J.H.L, *Discrete-Time Processing of Speech Signals*, Macmillan Publishing Comp., New York (1993).

[47] Saito, S. and Nakata, K, *Fundamentals of Speech Signal Processing*, Academic Press, Tokyo (1985).

[48] Flanagan, J.L., Ishizaka, K., and Shipley, K.L. " Synthesis of speech from a dynamic model of the vocal tract," *The Bell System Tech. J.*, pp.485-506 (1975).

[49] Krubsack, D.A. and Niederjohn, R.J. " An autocorrelation pitch detector and voicing decision with confidence measures developed for noise-corrupted speech," *IEEE Trans. Signal Proc.*, Vol.39, No.2, pp.319-329 (1991).

[50] Tsakalos, N. and Zigouris, E. " An investigation of failures and comparison of correlation measurement pitch trackers and pre-processing filters," *Int. J. Electronics*, Vol.75, No.2, pp.269-283 (1993).

[51] Gold, B. and Rabiner, L.R. " Parallel processing techniques for estimating pitch periods of speech in the time domain," *J. Acoust. Soc. Am.*, Vol.46, pp.442-448 (1969).

[52] Sukkar, R.A., Locicero, J.L. and Picone, J.W. " Design and implementation of a robust pitch detector based on a parallel processing technique," *IEEE J. Select. Areas Commun.*, Vol.6., No.2., pp.441-450(1988).

[53] Markel, J.D. "The SIFT algorithm for fundamental frequency estimation," *IEEE Trans. Audio Electroacoust.*, Vol.AU-20, pp.367-377 (1972).

[54] Rabiner,L.R. "On the use of autocorrelation Analysis for pitch detection," *IEEE Trans. Acoust., Speech, and Sig. Proc.*, Vol.ASSP-25, No.1, pp.24-33 (1977)

[55] Ross, M.J. " Average magnitude difference function pitch extractor," *IEEE Trans. Acoust., Speech, Sig. Proc.*, Vol.ASSP-22, pp.353-362 (1974).

[56] Tsakalos, N., and Zigouris, E. " Threshold based magnitude difference function pitch determination algorithm," *Int. J. Electronics*, Vol.71, No.1, pp.13-28 (1991).

[57] Medan, Y., Yair, E., Chazan, D. " Super resolution pitch determination of speech signals," *IEEE Trans. Signal Proc.*, Vol.39, No.1, pp.40-48 (1991).

[58] Campell, P.J., and Tremain, E.T. "Voiced/unvoiced classification of speech with applications to the U.S. Government LPC-10E algorithm," *Int. Conf. Acoustic, Speech and Signal Proc.*, pp.473-476 (1986).

[59] Ifeachor, C.E., and Jervis, B.W., *Digital Signal Processing- A Practical Approach*, Addison-Wesley, Wokingham (1993).

[60] Buzo, A., Gray, A.H., Gray, M.R. and Markel, J.D. "Speech Coding Based upon Vector Quantization," *IEEE Trans. Acoustic, Speech, Sig. Proc.*,Vol.ASSP-28, pp.562-574 (1980).

[61] Gray, A.H, Gray, R.M., and Markel, J.D. "Comparison of optimal quantization of speech reflection coefficients," *IEEE Trans. Acoustic, Speech, Sig. Proc.*,Vol.ASSP-25, No.1, pp.9-21 (1977).

[62] Itakura, F. "Line spectrum representation of linear predictive coefficients of speech signals," *J. Acoust. Soc. Am.*, Vol.57, pp.535(A) (1975).

[63] Lepschy, A., Mian, G.A., and Viaro, U. "A note on line spectral frequencies," *IEEE Trans. Acoustic, Speech, Sig. Proc.*, Vol.36, No.8, pp.1355-1357 (1988).

[64] "Details to assist in implementation of Federal Standard 1016 CELP," National Communications System, Tech. Bulletin, 92-1 (1992).

[65] Mendel, L.L., Hamill, B.W., Crepeau, L.J. and Fallon, E. " Speech intelligibility assessment in a helium environment," *J. Acoust. Soc. Am.*, Vol. 97, No. 1, pp. 628-636 (1995).

[66] Fant, G. and Sonesson, B. "Speech at high ambient air pressure," in *Speech Transmission Lab. Quart. Prog. and Status Report*, STL-QPSR No. 2/1964, pp.9-21 (1964).

[67] Fant, G. and Lindqvist, J. " II. Studies related to diver's speech. A: Pressure and gas mixture effects on diver's speech," in *Speech Transmission Lab. Quart. Prog. and Status Report* STL-QPSR No. 2/1968, pp.7-17 (1968).

[68] Paul, S., *Sport Diving: BS-AC Diving Manual*, Butler and Tanner, London (1993).

[69] Crummett, W.P., and Western, A.B., *University Physics - Models and Applications*, Wm.C. Brown, New York (1994).

[70] Richards, M.A. "Helium speech enhancement using the short-time fourier transform," *IEEE Trans. on Acous. Speech and Signal. Process.*, Vol.ASSP-30, No.6, pp.841-853(1982).

[71] Jack, M.A. and Duncan, G. "The helium speech effect and electronic techniques for enhancing intelligibility in a helium-oxygen environment," *The Radio and Electronic Engineer*, Vol.52, No.5, pp.211-223 (1982).

[72] Mackie, R.D.L., and Smith, N. "Diver communications and helium speech unscrambling," *Underwater System Design*, pp.19-21 (June/July 1983).

[73] Beet, S.W. and Goodyear, C.C. "Helium speech processor using linear prediction," *Electronics Letters*, Vol.19, No.11, pp.408-410 (1983).

[74] Duncan, G. and Jack, M.A. "Residually excited LPC processor for enhancing helium speech intelligibility," *Electronics Letters*, Vol.19, No.18, pp.710-711 (1983).

[75] Morrow, C.T. "Reaction of small enclosures on the human voice. Part I. Specifications required for satisfactory intelligibility," *J. Acoust. Soc. Am.*, Vol.19, No.4, pp.645-652 (1948).

[76] Morrow, C.T. "Reaction of small enclosures on the human voice. Part II. Analyses of vowels," *J. Acoust. Soc. Am.*, Vol.20, No.4, pp.487-497(1948).

[77] Bond, Z.S., Moore, T.J., and Gable, B. "Acoustic-phonetic characteristics of speech produced in noise and while wearing an oxygen mask," *J. Acoust. Soc. Am.*, Vol.85, No.2, pp.908-912 (1989).

[78] Morrow, C.T., Brouns, A.J. "Acoustic impedance calibrator for mask and microphone measurements," *J. Audio Eng. Soc.*,Vol.18,No.5, pp.519-523(1970).

[79] Morrow, C.T., Brouns, A.J. "Speech communication in diving masks. I. Acoustics of microphones and mask cavities," *J. Acoust. Soc. Am.*, Vol.50, No.1, Part 1, pp.1-9 (1971).

[80] Morrow, C.T., Brouns, A.J. "Speech communication in diving masks. I. Psychoacoustic and supplementary tests," *J. Acoust. Soc. Am.*, Vol.50, No.1, Part 1, pp.10-22 (1971).

[81] Meares, D.J. "Broadcast-quality speech from diving helmets," *J. Audio Eng. Soc.*, Vol.37, No.11, pp.927-933 (1989).

[82] Brandt, J.F. and Hollien, H. " Underwater hearing thresholds in man as a function of water depth," *J. Acoust. Soc. Am.*, Vol.46, No.4, pp.893-894 (1969).

[83] Brandt, J.F. and Hollien, H. " Underwater hearing thresholds in man," *J. Acoust. Soc. Am.*, Vol.42, No.5, pp.966-971 (1967).

[84] Hollien, H., and Brandt, J.F. "Effect of air bubbles in the external auditory meatus underwater hearing thresholds," *J. Acoust. Soc. Am.*, Vol.46, No.2, pp.384-387 (1969).

[85] Mathers, C.D. and Baird, M.D.M. "A simple means of improving the quality of speech from a diving helmet," *Report BBC RD 1989/16*, Engineering Division(1989).

[86] Lim, J.S., *Speech Enhancement*, Prentice-Hall, New Jersey (1983).

[87] Lim, J.S. "Evaluation of a correlation subtraction method for enhancing speech degraded by additive white noise," *IEEE Trans. Acous. Speech and Signal Process.*, Vol. ASSP-26, No.5, pp.471 -472 (1978).

[88] Boll, S.F. "Suppression of acoustic noise in speech using spectral subtraction" *IEEE Trans. on Acous. Speech and Signal. Process.*, Vol. ASSP-27, No.2, pp.113-120(1979).

[89] Crozier, P.M., Cheetham, B.M.G., Holt, C., and Munday, E. "Speech enhancement employing spectral subtraction and linear predictive analysis," *Electronics Letters*, Vol.29, No.12, pp. 1094-1095 (1993).

[90] Lim, J.S., and Oppenheim, A. V. "Enhancement and bandwidth compression of noisy speech," *Proc. IEEE*, Vol.67, No.12, pp.1586 - 1604 (1979).

[91] Proakis, J.G., Rader, C.M., Ling, F., and Nikias, C., *Advanced Digital Signal Processing, Macmillan, New York (1992).*

[92] Widrow, B., and Stearns, S.D., *Adaptive Signal Processing*, Prentice-Hall, New Jersey (1985).

[93] Widrow, B., Glover, J.R., McCool, J.M., Kaunitz, J., Williams, C.S., Hearn, R.H., Zeidler, J.R., Dong, E., and Goodlin, R.C., "Adaptive noise cancelling: Principles and Applications," *Proc. IEEE*, Vol.63, No.12, pp.1692-1716 (1975).

[94] Dunlop, J., Al-Kindi, M.J., Virr, L.E., and Nelson, H.M.S. "Application of adaptive noise cancelling to diver voice communications," *Proc. Int. Conf. Acous. Speech and Signal Proc.*, pp.1708-1711 (1987).

[95] Pritzker, Z., and Feuer, A. "Variable length stochastic gradient algorithm," *IEEE Trans. on Signal Process.*, Vol.39 No. 4, pp.997 - 1001 (1991).

[96] Harrison, W.A., Lim, J.S., and Singer, E. "A new application of adaptive noise cancellation," *IEEE Trans. Acous. Speech and Signal Process.*, Vol. ASSP-34, No. 1, pp.21 -27 (1986).

[97] Mirchandani, G., Zinser, R.L., and Evans, J.B. " A new adaptive noise cancellation scheme in the presence of crosstalk," *IEEE Trans. Circuits Syst: II-Analog and Digital Signal Process.*, Vol.39, No.10, pp.681-694 (1992).

[98] Proakis, J.G., *Digital Communications*, McGraw-Hill, New York (1995).

[99] Urick, R.J. *Principles of Underwater Sound*, McGraw-Hill, New York (1975).

[100] Burdick, W.S. *Underwater Acoustic System Analysis*, Prentice-Hall, New Jersey (1984).

[101] Coates, R.F.W. *Underwater Acoustic Systems*, Macmillan, London (1990).

[102] Knight, W.C., Pridham, R.G., and Kay, S.M., "Digital signal processing for Sonar," *Proc. IEEE*, Vol.69, No.11, pp.1451-1506 (1981).

[103] Fisher, F.H. and Simmon, V.P. " Sound absorption in sea water," *J. Acoust. Soc. Am.*, Vol.62, No.3, pp.558-564 (1977).

[104] Haykin, S., *Communication Systems*, John Wiley and Sons, New York (1994).

[105] Meller, R.H. "Thermal noise limit in the detection of underwater acoustic signals," *J. Acoust. Soc. Am.*, Vol.24, pp.478 (1952).

[106] Nystuen, J.A., McGlothin, C.C. and Cook, M.S. "The underwater sound generated by heavily rainfall," *J. Acoust. Soc. Am.*, Vol.93, No.6, pp.3169-3177 (1993).

[107] Catipovic, J.A. "Performance limitations in underwater acoustic telemetry," *IEEE J. Oceanic Eng.*, Vol.OE-15, No.3, pp.205-216 (1990).

[108] Baggeroer, A.B. " Acoustic telemetry- An overwiev," *IEEE J. Oceanic Eng.*, Vol.OE-9, No.4, pp.229-234 (1984).

[109] Jourdain, D., "Acoustical propagation analysis in shallow water," *Proc. 2nd European Conf. Underwater Acoustics*, Vol.II, pp.781-786 (1994).

[110] Owen, R.H., Smith, B.V. and Coates, R.F.W. "An experimental study of rough surface scattering and its effects on communication coherence," in *Proc. Oceans'94*, pp.III.483-III.488 (1994).

[111] Herwig,H. and Nutzel, B. "The influence of bubbles on acoustic propagation and scattering," *Underwater Acoustic Data Processing*, Kluwer Academic, Boston, pp.105-111 (1989).

[112] Geller, B., Brossier, J.M., and Capellano, V., "Equalizer for high data rate transmission in underwater communications," in *Proc. Oceans'94*, pp.I.302-I.306 (1994).

[113] Geller, B., Capellano, V., Brossier, J.M., Essebbar, A. and Jourdain, G. "Equalizer for video rate transmission in multipath underwater communications," *IEEE J.Oceanic Eng*, Vol.21, No. 2 (1996).

[114] Proakis, J.G. "Adaptive equalization techniques for acoustic telemetry channels," *IEEE J. Oceanic Eng.*, Vol.OE-16, No.1, pp.21-31 (1991).

[115] Stojanovic, M. "Recent advances in high-speed underwater acoustic communications," *IEEE J.Oceanic Eng*, Vol.21, No. 2, pp. 125-136 (1996).

[116] Tarbit, P.S.D, Howe, G.S., Hinton, O.R., Adams, A.E. and Sharif, B.S. "Development of a real-time adaptive equalizer for a high-rate underwater acoustic data communications link," in *Proc. Oceans'94*, pp.I.307-I.312 (1994).

[117] Bessios, A.G. and Caimi, F.M. "Multipath compensation for underwater acoustic communication," in *Proc. Oceans'94*, pp.I.317-I.322(1994).

[118] Galvin, R. and Coates, R.F.W. "Analysis and performance of an underwater acoustic communications system and comparison with a stochastic model," in *Proc. Oceans'94*, pp.III.478-III.482 (1994).

[119] Billon, D. and Quellec, B. "Performance of high data rate acoustic underwater communication systems using adaptive beamforming and equalizing," in *Proc. Oceans'94*, pp.III.507-III.512 (1994).

[120] Henderson, G.B., Tweedy, A., Howe, G.S., Hinton, O., and Adams, A.E., "Investigation of adaptive beamformer performance and experimental verification of applications in high data rate digital underwater communications," in *Proc. Oceans'94*, pp.I.296-I.301 (1994).

[121] Labat, J. "Real time underwater communications," in *Proc. Oceans'94*, pp.III.501-III.506 (1994).

[122] Brock, D.C., Bateman, S.C. and Woodward, B. " Underwater acoustic transmission of low-rate digital data," *Ultrasonics*, Vol.24, pp.183-188 (1986).

[123] Andrews, R.S. and Turner, L.F. "On the performance of underwater data transmission system using amplitude shift keying techniques," *IEEE Trans. on Sonic and Ultrasonics*, Vol.SU-23, pp.64-71 (1976).

[124] Dawoud, M.M., Halawani, T.U. and Abdul-jauwad, S.H. "Experimental realisation of ASK underwater digital acoustic communications system using error correcting codes," *Int. J. Electronics*, Vol.72, No.2, pp.183-196 (1992).

[125] El-Hennawey, M.S. and Shehadah, W.H. "Non-coherent FSK receiver for underwater communications," *Int. J. Electronics*, Vol.79, No.3, pp.265-280 (1995).

[126] Catipovic, J., Baggeroer, A.B., Von Der Heydt, K. and Koelsch, D. "Design and performance analysis of a digital acoustic telemetry system for the short-range underwater channel," *IEEE J.Oceanic Eng*, Vol. OE-9, pp.252-252 (1984).

[127] Woodward, B. and Bateman, S.C. "Diver monitoring by ultrasonic digital data telemetry," *Med. Eng. Phys.* Vol. 16, pp.278-286 (1994).

[128] Thompson, D., Neasham, J., Sharif, B.S., Hinton, O.R., Adams, A.E., "Performance of coherent PSK receivers using adaptive combining and beamforming for long range acoustic telemetry," *3rd European Conf. Underwater Acoustics*, pp.747-752 (1996).

[129] Plaisant, A., "Long range acoustic communications," *3rd European Conf. Underwater Acoustics*, pp.759-764 (1996).

[130] Merriam, S., and Porta, D., "DSP-based acoustic telemetry modems," *Sea Technol.*, May 1993.

[131] Johnson, M., Herold, D., and Catipovic, J., "The design and performance of a compact underwater acoustic network node," in *Proc. Oceans'94*, pp.III.467-III.471 (1994).

[132] Woodward, B. and Sari, H. "Digital underwater acoustic voice communications," *IEEE J.Oceanic Eng*, Vol.21, No. 2, pp. 181-192 (1996).

[133] Howe, G.S., Tarbit, P.S.D, Hinton, O.R., Sharif, B.S. and Adams, A.E. "Sub-sea acoustic remote communications utilising an adaptive receiving beamformer for multipath suppression," in *Proc. Oceans'94*, pp.I.313-I.316 (1994).

[134] Falahati, A., Bateman, S.C. and Woodward, B. " Underwater acoustic channel models for 4800bps QPSK signals," *IEEE J. Oceanic Eng.*, Vol.OE-16, No.1, pp.12-20 (1991).

[135] Stojanovic, M., Catipovic, J.A. and Proakis, J.G. "Phase-coherent digital communication for underwater acoustic channels," *IEEE J. Oceanic Eng.*, Vol.19, No.1, pp.100-111 (1994).

[136] Goalic, A., Labat, J., Trubuil, J., Saoudi, S. and Riouaten, D. "Toward a digital acoustic underwater phone," in *Proc. Oceans'94*, pp.III.489-III.494(1994).

[137] Habib Istepanian, R.Sh., *Use of Microcontrollers for Diver Monitoring by Underwater Acoustic Biotelemetry in Multipath Environments*, Ph.D. Thesis, Loughborough University (1994).

[138] Ling, G. and Cagliardi, R.M. " Slot synchronization in optical PPM communications," *IEEE Trans. Commun.*, Vol.COM-34, No.12, pp.1202-1208 (1986).

[139] Garrett, I. "Pulse-Position modulation for transmission over optical fibers with direct or heterodyne detection," *IEEE Trans. Commun.*, Vol.COM-31, No.4, pp.518-527 (1983).

[140] Georgehiades, C.N. "Optimum joint slot and symbol synchronization for optical PPM channel," *IEEE Trans. Commun.*, Vol.COM-35, No.6, pp.518-527 (1987).

[141] Elmirghani,J.M.H. and Cryan, R.A. " Analytic and numeric modelling of optical fibre PPM slot and frame spectral properties with application to timing extraction," *IEE Proc.-Commun.*, Vol.141, No.6, pp.379-389 (1994)

[142] Riter, S. and Boatrigth, P.A. "Design considerations for a pulse position modulation underwater acoustic communication system", *Digest IEEE Conf. Eng. Oceanic Environment*, pp.21-24 (1970).

[143] Riter, S. " Pulse position modulation communications via the underwater acoustic communication channel," IEEE SWIEECO Rec. 22nd Southwestern Conf. & Exhib. pp.453-457 (1970).

[144] Kwon, M.H. and Birdsall, G.T. "Channel Capacity in bits per joule," *IEEE J. Oceanic Eng.*, Vol.OE-11, No.1, pp.97-99 (1986).

[145] Proakis, J.G., and Salehi, M., *Communication Systems Engineering*, Prentice-Hall, New Jersey (1994).

[146] *TMS320C3x User's Guide*, Texas Instruments (1992).

[147] *TLC320AC01 Analogue Interface Circuit - Data Manual,* Texas Instruments (1993).

[148] Ifeachor, C.E., and Jervis, B.W., *Digital Signal Processing A Practical Approach*, Addison-Wesley, Wokingham, England (1993).

[149] Huelsman, L.P., *Active and Passive Analog Filter Design*, McGraw-Hill, New York (1993).

[150] *Designing with the TLC320AC01 Analogue Interface for DSPs,* Texas Instruments (1994).

[151] Oppenheim, A.V., and Schafer R.W., *Discrete-Time Signal Processing*, Prentice-Hall, New Jersey (1989)

[152] Stansfield, D., *Underwater Electroacoustic Transducers - A Handbook for Users and Designers*, Bath University Press (1990).

[153] *New Releases Data Book - Volume II,* Maxim (1993).

[154] Krauss, H.L., Bostain, C.W., and Raab, F.H., *Solid State Radio Engineering*, John Wiley & Sons, New York (1980).

[155] *Telecommunications Applications with the TMS320C5x DSPs*, Texas Instruments (1994).

[156] Sokal, N.O. and Sokal, A.D., "Class E -A new class of high-efficiency tuned single-ended switching power amplifiers," *IEEE J. Solid-State Circuits*, Vol. Sc-10, No.3, pp.168-176 (1975)

[157] Raab, F.H., "Idealized operation of the Class E tuned power amplifier," *IEEE Trans. Circuits and Systems*, Vol.Cas-24, No.12, pp.725-735 (1977).

[158] Herman, K.J. and Zulinski, R.E., "The infeasibility of a zero-current switching Class E amplifier," *IEEE Trans. Circuits and Systems*, Vol.37, No.1, pp.152-154 (1990).

[159] Li, C.H. and Yam, Y.O., "Maximum frequency and optimum performance of Class E power amplifiers," *IEE Proc. -Circuits Devices Syst.*, Vol.141, No.3, pp.174-184 (1993).

[160] *Soft Ferrites Data Handbook*, Philips (1993)

[161] Kosow, I.L., *Circuit Analysis*, John Wilesy and Sons, New York (1988).

[162] Bobber, R.J., *Underwater Electroacoustic Measurements*, Peninsula Pub., California (1988).

[163] Morgera, S.D., "Digital filtering and prediction for communications systems time synchronization," *IEEE J. Oceanic Eng.*, Vol. OE-7, No.3, pp.110-119 (1982).

[164] Cagliardi, R.M., and Karp, S., *Optical Communications*, New York, Wiley (1995).

AUTHOR'S PUBLICATIONS

[1] Woodward, B. and Sari, H. "Digital underwater acoustic voice communications," *IEEE J. Oceanic Eng.*, Vol.21, No. 2, pp. 181-192 (1996).

[2] Woodward, B. and Sari, H. "Underwater speech" *Acustica, Acta Acustica,* Vol.82, pp 233 (1996).

[3] Woodward, B. and Sari, H. "Implementation of digital technology for underwater voice communications," *Proceedings of the Institute of Acoustics: Sonar Signal Processing*, pp.119-128 (1995)

[4] Woodward, B. and Sari, H. "Digital communications for divers," *2nd Underwater Science Symposium* (1995).

[5] Woodward, B. and Sari, H. "Underwater voice communications using digital techniques," *J. de Physique IV,* Colloque C5, Vol.4, pp. 469-472 (1994).

IEEE JOURNAL OF OCEANIC ENGINEERING, VOL. 21, NO. 2, APRIL 1996

Digital Underwater Acoustic Voice Communications

Bryan Woodward, *Member, IEEE,* and Hayri Sari, *Student Member, IEEE*

Abstract—**This paper describes the design of an underwater acoustic diver communication system controlled by a digital signal processor. The speech signal transmission rate is compressed by using linear predictive coding (LPC) and the extracted parameters are transmitted through the water to a synchronized receiver by employing digital pulse position modulation (DPPM). The pulse position in each time frame is estimated by an energy detection and decision algorithm which enables the received LPC parameters to be recovered and used to synthesize the speech signal.**

I. Introduction

MODERN telecommunications systems are now very sophisticated, especially since the advent of mobile telephones, but most systems available for underwater voice communication are still comparatively archaic. As with any other communication system, all that is required in principle is a method of conveying a meaningful message from one person to another. The essential constituents are a source (a voice signal encoded into electrical signals from a microphone), a channel (a wire link carrying electrical signals or the water itself carrying acoustic signals), and a destination (the decoded voice signal received by a headphone).

But there are particular problems associated with an underwater system that are not shared with its conventional counterpart. One of the main problems is the poor quality of the source signal caused by the diver speaking, often unclearly, into a band-limited microphone placed in a small resonant cavity. This affects the frequency spectrum of the speech, and therefore distorts the signal from the microphone. The signal is further degraded by the introduction of noise at the source, in the transmission channel and at the receiver [1].

At the source there is breathing noise, noise from free-flowing air (in helmets and band masks) and bubble noise. Even in comparatively quiet rebreather systems, breathing noise is still present. Nonlinear microphone distortion and indistinct articulation compound these problems. In directly-wired systems there is noise from crackle due to the ingress of water into connectors and umbilical cables, and electrical noise from extraneous sources like motors, generators, and switches. In through-water acoustic systems, the noise sources include wave action, sounds of biological origin, shipping, rain, and many other mechanisms. There is also multipath interference which causes delayed versions of the transmitted signal to reach the receiver and blur the original message. At the receiver, there is often audible noise from generators, gas flow, clanging metal, telephones, and conversation.

The quality and reliability of an underwater communications system are important to diver safety. In the event of an emergency underwater, a diver's life may depend on fast and effective action at the surface, so good communication between a pair of divers or between the divers and the surface is essential. In most commercial underwater work, a communication link is a statutory requirement because it allows coordination with surface activities, monitoring and control of decompression stops, warnings of dangers and calls to assist other divers. Most presently available systems, both hard-wired and through-water acoustic, are virtually unchanged from those developed three decades ago.

Voice communication systems have been used underwater since 1905. The early versions had a wire link between the diver and the surface supervisor and used direct transmission of the voice signal without modulation, which rendered them susceptible to pick-up from audible sources. One advantage was that the diver's microphone and speaker worked in the comparatively large, dry volume of a traditional "hard hat." The voice signal would therefore have been better than that from a modern band mask which has a very small air volume in front of the diver's mouth. Like the diving equipment of the time, these systems were simple and unsophisticated, and were eventually replaced by systems that used some form of modulation of a carrier frequency well above the audio range [2]. In spite of great technological advances, these systems are still used because of their proven reliability. For many commercial, police, navy, and scientific divers, the wire link can be conveniently included in the umbilical line supplying the breathing gas [3].

The two basic types of hard-wired communications systems have two-wire and four-wire configurations. With the two-wire system, each diver is connected to the surface unit by two wires, which allow one-way communication, referred to as *half-duplex* operation. The supervisor's system is usually on receive mode, allowing him to monitor the divers' speech and breathing noise, which is a useful safety feature. To talk to them he uses a press-to-talk (PTT) switch which reverses the direction of communication. This means that while he is speaking he cannot hear the divers, an obvious potential hazard. An alternative is to use a voice operated switch, which allows transmission automatically when a diver speaks. With a four-wire system, each diver is connected to the surface by four wires in a "round-robin" configuration which allows two-way communication, referred to as *full-duplex* operation.

Manuscript received March 15, 1995. This work was supported by the Turkish Ministry of Education.

The authors are with the Department of Electronic and Electrical Engineering, Loughborough University of Technology, Leicestershire LE11 3TU, U.K.

Publisher Item Identifier S 0364-9059(96)03416-4.

Because of the obvious encumbrance of an umbilical line, especially in areas where it is difficult for a surface supervisor to operate, "wire-less" or "through-water" communication systems were introduced in the 1950s; these were more suitable for divers using Self Contained Underwater Breathing Apparatus (SCUBA).

Nowadays, there are many commercially available through-water diver communication systems. Like the hard-wired systems, the earliest versions used baseband methods in which signals were transmitted without modulation. But these systems were strongly affected by ambient acoustic noise which is dominant below 6 kHz [1]. Moreover, since baseband methods require more power at low frequencies, they are not suitable for mobile diver communications. Therefore, in the design of later systems, the baseband spectrum was shifted to a higher band by modulating a carrier frequency with the speech signal.

Improved versions of the early systems employed amplitude modulation (AM) [4]. This has inherent shortcomings because the carrier is transmitted, which requires two-thirds of the total power transmitted, with the other third required for the upper and lower sidebands, each of which carries the actual speech signal. This means that the carrier is transmitted continuously, even during periods when no speech signal is being sent, so power is wasted without any benefit.

The commonest diver communication systems now use the more convenient single sideband (SSB) modulation [5], [6]. This effectively counters many of the shortcomings of AM transmission because during periods when no speech is being transmitted, SSB dissipates no power. The net result for SSB transmission is significantly longer battery life. Also, the technique has the advantage of having half the bandwidth of earlier conventional AM systems and double sideband suppressed carrier (DSBSC) systems. When the SSB modulation technique is applied to underwater communications, experience has shown that the quality and intelligibility of the received speech are superior to that of the AM technique. Typically, the speech is transmitted as a band-limited signal that modulates a carrier frequency centred on the resonant frequency of an omnidirectional electrostrictive transducer. There is no legislation on the use of frequency bands underwater, although certain frequencies have been adopted unofficially for communications [7]. The lower the frequency, the greater the range of transmission achievable, since range is inversely proportional to the square of frequency.

Another factor that governs the choice of frequency is the effect of ambient noise, which is noticeable at low frequencies, particularly in the audible range. While some systems use frequencies as low as 8 kHz and as high as 70 kHz, the most common frequency used in commercially available systems is 42 kHz [7]. This is high enough to avoid excessive extraneous noise and low enough to give a range of about 1 km, depending on the power output of the transmitter.

For both DSBSC and SSB modulation, the multipath effect and channel fading can produce distortion of the received signals unless precautions are taken to avoid it. Also, there are limitations to any analog system which a digital system can overcome. One of these is to have a private communication link between divers or between a diver and the surface, so that there is no unwanted cross-talk with any other divers in the same area. Although this is possible in principle with DSBSC and SSB, there is a practical limit to the number of carrier frequencies that can be accommodated in the bandwidth of a transducer. By contrast, with a digital system using an 8-b synchronization code (described later), there are effectively 256 possible channels available for a single carrier frequency.

It is surprising that although there has been considerable progress in mobile telephone technology, nothing substantially new for diver communications has been achieved for almost three decades. The only important use of digital technology seems to have been in the unscrambling of oxy-helium speech. This is necessary because the speed of sound in helium, being nearly three times higher than in air, causes the speech to become so garbled that it is often referred to as the "Donald Duck" effect. The principle of the device is to write samples of speech into memory and subsequently to read them back at a slower rate to render the speech intelligible. There are, in fact, extensive studies continuing on the application of digital techniques to speech coding and their implementation for diver voice communication [8], and more complex speech coding methods have also been simulated for this purpose [9]. The rest of this paper presents the design of a digital acoustic voice communications system for divers, with a digital signal processor (DSP) as its central feature as shown in Fig. 1. Its principle of operation and the choice of carrier frequency, speech coding, and modulation method are now given in detail.

II. System Design

The main consideration in the design of a digital underwater communications system is to provide a diver with a comparable level of communications capability as is provided by a digital mobile telephone. The availability of surface mount devices and digital signal processors allows the design of a compact system which can be worn with conventional diving equipment. A block diagram of our prototype system is shown in Fig. 2.

The system consists of a speech signal conditioning section, digital signal processor and memory, transmitter and receiver unit and keypad encoder. For reliable operation, the first essential is to select a suitable microphone to pick up the speech signals in the harsh, high pressure environment of a mask [10]. From studies of different type of microphones in diving masks [11] it has been found that noise-cancelling microphones provide intelligibility independent of the cavity dimensions. Therefore in this work a pressure compensated, noise cancelling microphone is used; this is placed inside a mask and connected to the input of a TLC320AC01 analog interface circuit (AIC). The speech signals are bandpass filtered with a bandwidth of 3.2 kHz, sampled at a rate of 8 kHz, then quantized to 14-b resolution. Digital data representing each speech sample are transferred to the DSP via its serial port.

Since complex algorithms are required for speech signal processing for underwater voice communications, the type of DSP used here is the TMS320C31, which is configured in a microprocessor mode. For memory, the system has two 128 k x 16 b electrically programmable read only memories

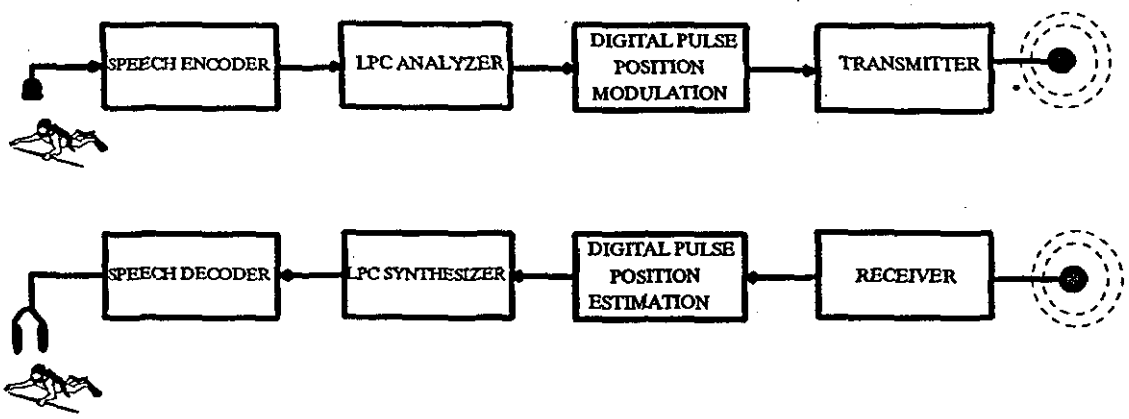


Fig. 1. Principle of digital underwater acoustic voice communication system.

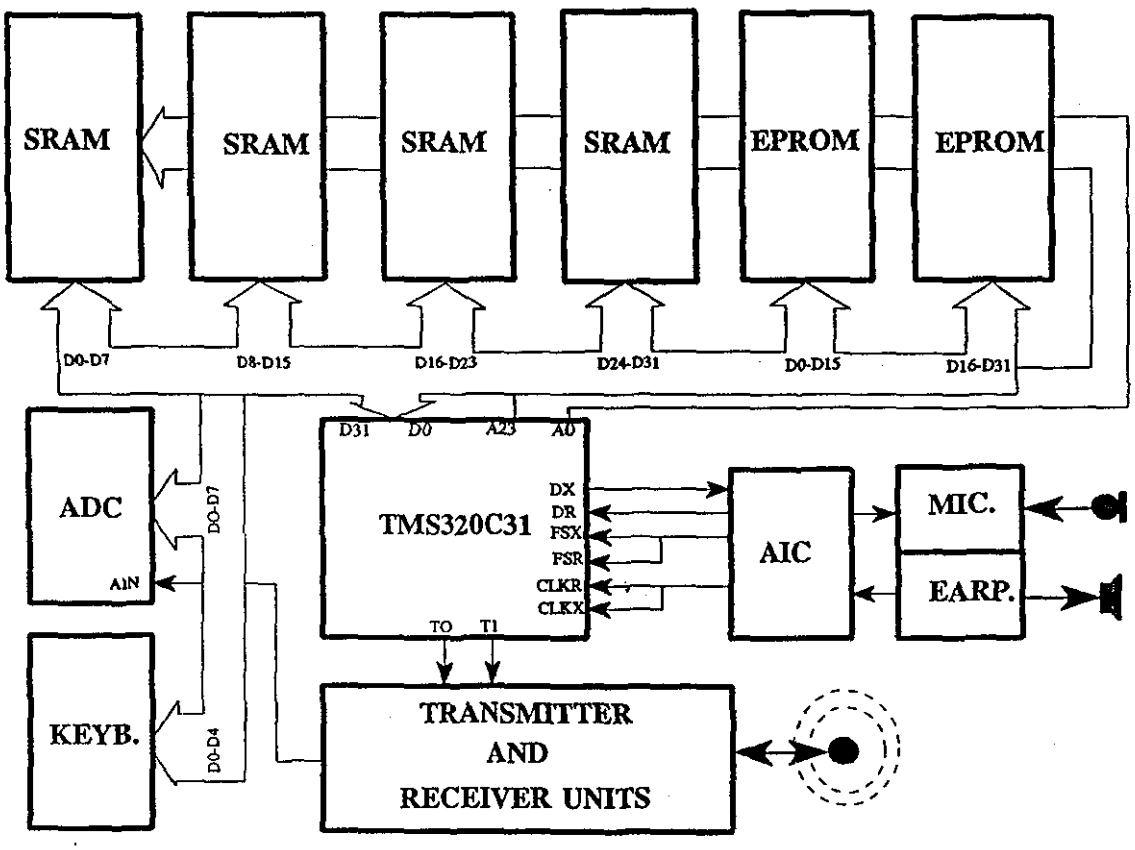


Fig. 2. A TMS320C31 DSP-based underwater acoustic voice communication system.

(EPROM) for the program and four 8 k × 8 b static random access memories (SRAM) for data. The signal processing algorithms are run from the SRAM because of the slow speed of EPROM's.

Another important feature of the system is the design of the transmitter section, shown in Fig. 3, which consists of a power MOSFET operated in a switching mode by the timer outputs of the DSP and an impedance matching transformer. Timer $T0$ is engaged for generating the carrier frequency, defined by the type of transducer, while Timer $T1$ selects the transmit or receive mode (Tx or Rx). An omnidirectional transducer with a 70-kHz resonance frequency is employed, so $T0$ is set to generate a square waveform output at this frequency. Since the same transducer is used for both transmission and reception, isolating the transmitter section from the receiver section is important, therefore an analog switch with a low on-resistance is employed.

The receiver section consists of a preamplifier, an 8th-order Butterworth bandpass filter centred at 70 kHz, an envelope detector and an 8-b MAX166 analog-to-digital convertor, which sends digital samples to the DSP for the necessary signal processing.

One of the important features of a digital underwater voice communication system is to establish a communication link

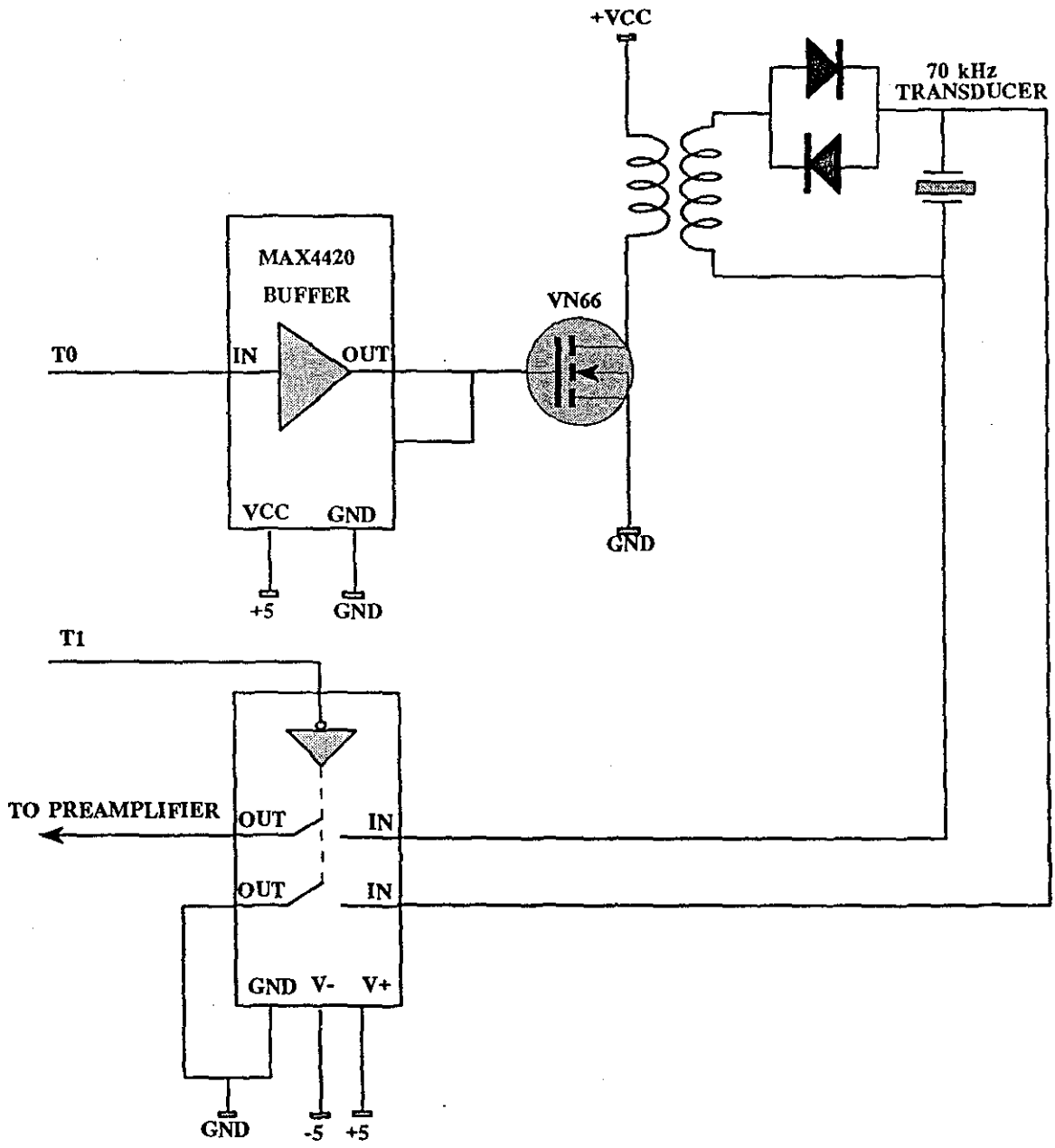


Fig. 3. Transmitter section and its control.

with another diver, using a specific identification code (ID number). Therefore, a keypad is included in the prototype for laboratory tests. In normal diving procedure, the ID of each diver's system would be encoded prior to the dive; this is, of course, less versatile than making a normal telephone call and would need to include a further feature to allow the diver to "dial" underwater. The outputs of the 4 × 4 keypad encoder (MM74C922N) are directly connected to the least significant four data bits of the DSP, as shown in Fig. 4. Whenever a key is pressed, an interrupt request is generated and detected by the DSP, then the encoder outputs are read and the ID number of a specific receiver is transmitted.

When an acoustic signal arrives at the transducer, the system is switched automatically to the receive mode, and the output of the receiver's envelope detector is continuously monitored by the DSP. When the unit recognizes its ID number, the DSP reads the AIC output and processes the input speech signal. If the diver wants to talk to another diver, the appropriate ID number is encoded the DSP then switches to the transmit mode and reads the output of the encoder. After that the digital data is transmitted, as described in the following sections.

III. Speech Coding

Advances in technology allow the implementation of complex speech processing algorithms in real time. In principle digital encoding of speech can be achieved by any of the well-known modulation techniques. These include pulse width modulation (PWM), pulse position modulation (PPM), pulse code modulation (PCM), and delta modulation (DM), each

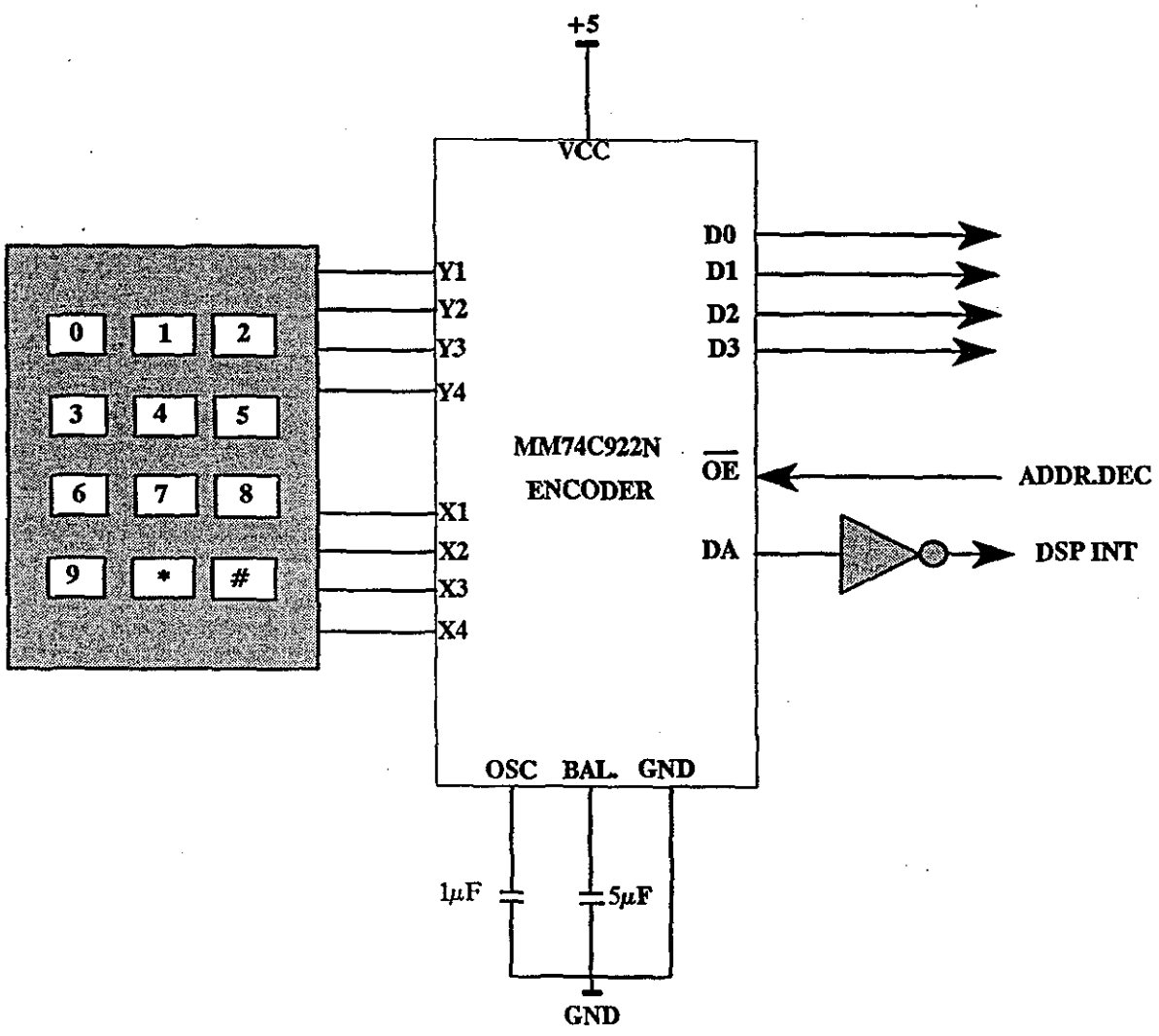


Fig. 4. Keypad unit for diver telephony.

of which has certain advantages and disadvantages. These are perfectly feasible for hard-wired systems, although they do not seem to have been adopted in any presently available product. For through-water acoustic systems, there are severe problems that need to be addressed in addition to the phenomenon of multipath interference. One of these is the low sampling rate achievable. With a speech bandwidth of about 3 kHz, the Nyquist sampling theorem dictates that the minimum sampling rate needed for recovery of a digitized signal back to its original analog form is at least twice this, i.e., 6 kHz. The corresponding sampling period is therefore 167 μs, so whatever modulation technique is chosen this is the maximum time allowed to transmit the encoded information. This is a typical duration for a single pulse transmitted from an electrostrictive transducer. In view of the complex nature of speech signals, long trains of bits are implied with the techniques mentioned above and therefore the use of some other kind of speech coding is called for.

In our research, speech data is sampled at a rate of $f_s = 8$ kHz. The coded information representing each sample must therefore be transmitted during the corresponding sampling period of $T_s = 125$ μs, as shown in Fig. 5. There are several waveform quantization and encoding techniques for speech signals which conform to standards laid down by the Consultative Committee for Telephone and Telegraph (CCITT). The most widely used ones are pulse code modulation (PCM) and adaptive differential pulse code modulation (ADPCM) [9]. PCM has a data rate of 64 kb/s and ADPCM is normally encoded with 5, 4, 3, or 2 bits/sample, i.e., at rates of 40, 32, 24, or 16 kb/s. Because of the limitation of the transducer, with a resonance frequency of 70 kHz, only 16 kb/s is worth considering for speech transmission through water. This may be appreciated by studying the relationships between the sampling period ($T_s = 125$ μs), the bit period for PCM ($T_b = 15.625$ μs), the bit period for ADPCM ($T_d = 62.5$ μs) and the carrier waveform period ($T_c = 14.286$ μs) in Fig. 5. A necessary condition for the bit period is that it should be about Q times greater than the carrier waveform period, where $Q = 7$ for the transducer used here.

It is evident from these considerations that conventional PCM is not suitable for a digital voice communications system other than if a very high carrier frequency is used for a very short range [8]. ADPCM is also not suitable because the bit period T_d is less than $7T_c$. The new speech coding method

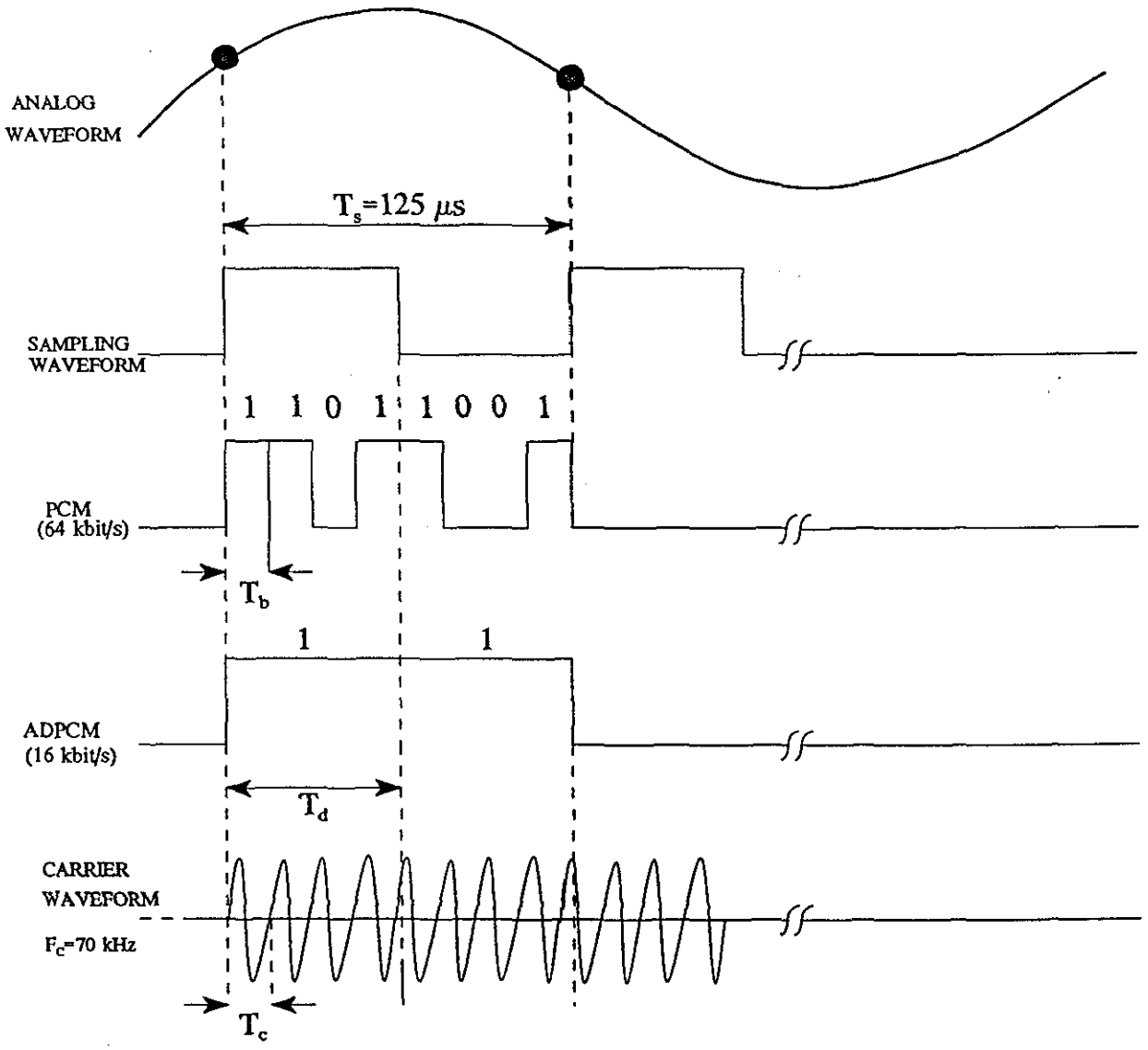


Fig. 5. Transmission of PCM and ADPCM encoded speech samples.

available also have bit rates that are too high for a diver system. Among these are code excited linear predictive coding (CELP) [13] with a data rate of 4.8 kb/s and residual excited linear prediction coding (RELP) [14] with a data rate of 13.2 kb/s, which are used in digital mobile telephones. A coding method requiring a lower data rate must therefore be used.

The solution adopted here is to use "conventional" linear predictive coding (LPC) which provides an accurate representation of the relevant speech parameters that can reduce transmission rates to 2.4 kb/s, including error correction codes [15]. This reduction is at the expense of a slight quality impairment of the reproduced speech. In an LPC voice coder, efficient speech synthesis is achieved by transmitting frames of the speech waveform as a set of *parameters* which are 1) the amplitude, 2) a voiced/unvoiced decision and pitch period for voiced sounds, and 3) a set of so-called linear prediction coefficients. These are computed by the digital signal processor in this system, although commercial vocoders working at 2.4-k baud are available.

In the linear prediction representation, the speech production is modeled as the output of an all-pole filter $H(z)$, given by

$$H(z) = \frac{G}{1 + \sum_{k=1}^{p} a[k] z^{-k}} \tag{1}$$

where p is the number of modeled poles, selected here to be 10, G is the gain of the filter, and the parameters $a[k]$ are the coefficients characterizing the filter. Processing the speech signal is done every 22.5 ms, or 180 sampling periods of 125 μs. This is sufficiently long to enable the pitch period to be determined, as described later. An estimate of the speech signal amplitude is made by a linear combination of p values of $a[k]$, as represented by the inverse all-pole filter shown in Fig. 6 and defined as

$$\hat{s}[n] = \sum_{k=1}^{p} a[k]\ s[n-k]. \tag{2}$$

The error between the actual amplitude of $s[n]$ and the predicted value $\hat{s}[n]$ is minimized over N samples, where

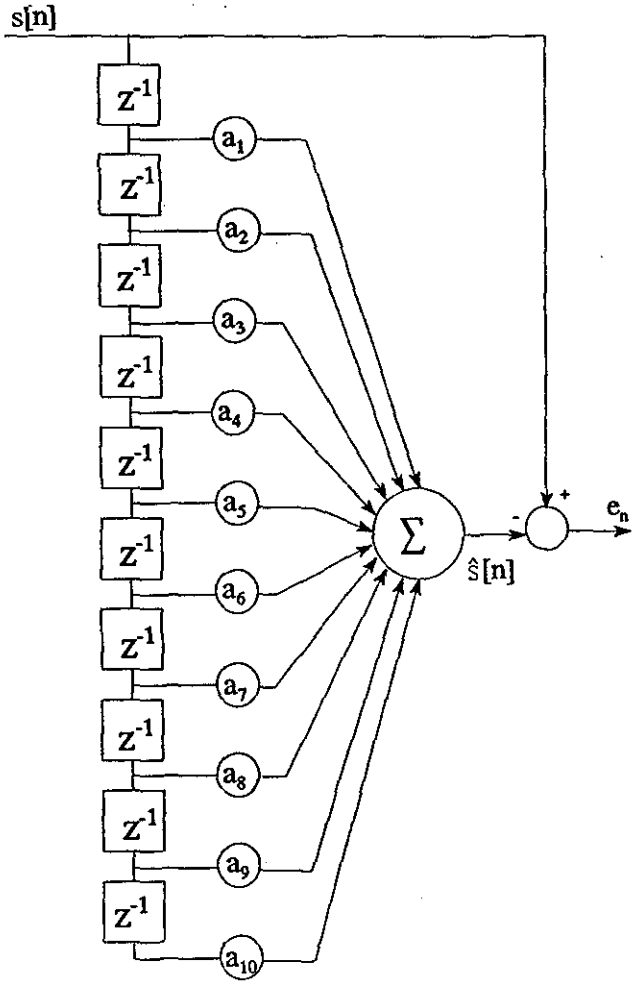


Fig. 6. Inverse all-pole filter for LPC coefficients estimation.

$N = 180$, which results in

$$\sum_{k=1}^{p} a[k]\ R[i-k] = -R[k] \quad 1 \le i \le p \tag{3}$$

where

$$R[k] = \sum_{n=0}^{N-1} s[n]\ s[n+k] \tag{4}$$

is the autocorrelation function of the signal $s[n]$. The coefficients $R[i-k]$ form what is often known as an autocorrelation matrix, given by

$$R[i-k] = \sum_{k=1}^{p} s[n]s[n+i-k]. \tag{5}$$

The values of $a[k]$ are then found by applying the so-called Levinson–Durbin LPC algorithm [15]. Since these coefficients are sensitive to quantization noise, more stable reflection coefficients are extracted and quantized for transmission [16].

Another important parameter required for transmission is the pitch period for voiced speech frames. Pitch information is estimated by operating on speech which is filtered by a lowpass filter with a cutoff frequency of 800 Hz. A second-order inverse filter is used to enhance the pitch estimator for input signals whose frequency content is below 300 Hz.

TABLE I
BANDWIDTH EFFICIENCY COMPARISON OF DIGITAL MODULATION TECHNIQUES

	M	bandwidth efficiency	M	bandwidth efficiency
M-ary ASK	2	0.5	8	1.5
M-ary PSK	2	0.5	8	1.5
M-ary FSK	2	1	8	0.75
M-slot DPPM	2	0.5	8	1.5

There are several methods of estimating the pitch period, including the average magnitude difference function (AMDF) method [17], which is implemented here because it requires less computational time for real-time operation. The AMDF is given by

$$\text{AMDF}[k] = \frac{1}{N}\sum_{n=0}^{N-1} |x[n] - x[n-k]| \quad 20 \le k \le 156. \tag{6}$$

The minimum value of AMDF gives the pitch period. The pitch frequency is given in terms of the sampling frequency as $f_p = f_s/k$, so for these values of k, the corresponding pitch frequencies range from a maximum of 400 Hz to a minimum of 51 Hz. A voiced/unvoiced decision is made from the result of the zero-crossing rate and the energy in each frame. These parameters are quantized and encoded as given in [18] then transmitted. The major advantage of the LPC method is that the speech transmission rate is decreased to 2.4 kb/s.

IV. TRANSMISSION OF SPEECH PARAMETERS

Today sophisticated digital modulation methods are adopted in mobile communication systems. But they use much higher carrier frequencies and sampling rates than are appropriate for a through-water channel if a reasonable acoustic range of the order of hundreds of meters is required.

Digital modulation techniques, such as amplitude shift keying (ASK), frequency shift keying (FSK), and phase shift keying (PSK) can be employed for underwater acoustic data transmission [19]. As an alternative, pulse position modulation (PPM) is also an effective technique [20]. The analogue PPM method is suited to serial data transmission; the principle is to transmit a short pulse that is delayed with respect to the sampling instant by an amount that is a linear function of the signal amplitude. For the present system, implementation is achieved by digital pulse position modulation (DPPM), as shown in Fig. 7.

The justification for using this technique is because of its superior bandwidth efficiency, i.e., R_b/B where R_b is the bit rate and B is the bandwidth, and because it is less complex to implement. Table I shows a comparison of the bandwidth efficiency for various digital modulation techniques.

By implementing the DPPM technique the bandwidth efficiency of the system is increased by $\log_2 M$ compared to binary ASK and PSK and gives as high a bandwidth efficiency as M-ary ASK and M-ary PSK. Since the envelope of the received signal is used for data decoding, the binary ASK signal detection principle is employed. Unlike for PSK there is no extra hardware for phase detection, and unlike FSK there

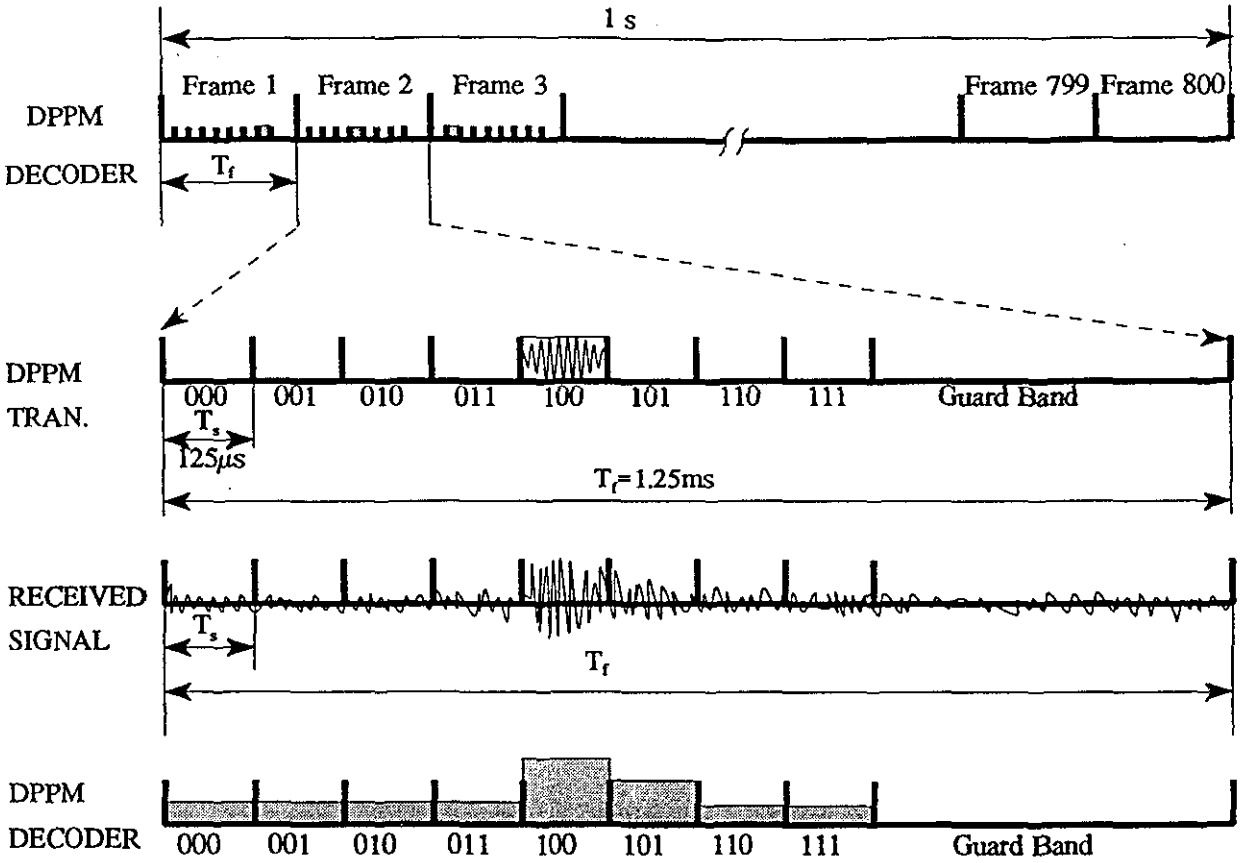


Fig. 7. Digital PPM signal format for transmit and receive operation.

is no need for two bandpass filters. Hence, in terms of system complexity also, DPPM is superior to the other techniques. However, synchronization in reception is a significant factor in DPPM system performance and therefore the PLL algorithm is included for synchronization frame identification.

With DPPM, m bits of linear quantized speech parameters are encoded in each frame by transmitting a single pulse in one of the M slots within the frame, where $M = 2^m$, which is therefore capable of accommodating all the m-bit permutations. A guard band is also included in each frame to avoid interframe interference (IFI). The mathematical definition of DPPM is given in [21] as

$$x(t) = \sum_{l=-\infty}^{\infty} g(t - lT_f - t_l). \tag{7}$$

Here, $g(t)$ is the DPPM pulse shape, T_f is the frame duration, and t_l is the data coded into DPPM and $0 \le t_l \le (m-1)T_s$, where the slot duration T_s is defined in terms of the modulation index as

$$\text{modulation index} = \frac{MT_s}{T_f}. \tag{8}$$

The quantized speech parameters are separated into subgroups consisting of 3-b, i.e., $m = 3$, giving a data rate of 2.4 kb/s and a frame of $T_f = 3/2400$ s, i.e., 1.25 ms. Therefore, eight slots are required to represent a pulse position in each frame and two slot intervals are used for the guard band. Thus a slot time of $T_s = T_f/(M+2)$, i.e., 125 μs, is needed, corresponding to a slot frequency of 8 kHz.

The DPPM principles are illustrated in Fig. 7 for data represented by 100_2. Decoding the received data requires an accurate estimation of the pulse position, therefore a synchronization scheme has to be applied. During synchronization, the transmitter sends a code taking a full frame interval (1.25 ms) to enable the receiver clock to be synchronized with the transmitter clock.

Detection of the DPPM signals is achieved by a filter, envelope detector and an analog-to-digital convertor. Since the envelope of the received signal can alter at a rate of 8 kHz, i.e., the slot frequency, it is sampled at 32 kHz, i.e., four samples per slot. The energy in each slot is then computed by the DSP and the result is stored in memory. This is done for all the slot intervals until the position of maximum energy is established. The corresponding code, here 100_2 for illustration purposes, is taken as the LPC parameters transmitted and used to synthesize the speech signal.

Many acoustic channel characteristics could lead to spurious data detection. These include transmission loss, ambient noise, temporal spreading due to multipath propagation, and Doppler spreading due to source/receiver motion. They are well described in literature [26] and they dominate the design of the receiver section of the system.

For the purpose of illustrating the effect of transmission loss on the received signal level, consider a communication

Fig. 8. EXO-26 full-face mask.

range of 100 m, which would be far greater than the actual range in many diving scenarios. In our system, the following specifications apply: projector resistance, 276 Ω; hydrophone sensitivity, −201 dB re 1 V/μPa. For an applied projector voltage of 20 V_{pp}, the received signal level, RL, is found from (9) below to be 121 μV

$$\mathrm{RL} = 170.8 + 10\log_{10} P_e - 20\log_{10} r - 0.022r. \quad (9)$$

The received signal is amplified by 26 dB, which is determined by the minimum voltage level detected by the analog-to-digital converter (see Fig. 2), and used for data decoding as described above. Because the amplifier gain is selected for a range of 100 m, the transmitted data will not be detectable at greater ranges if the calculation above holds. Ambient noise generally controls the signal-to-noise ratio (SNR) but by employing a 4-kHz bandpass filter centred on 70 kHz this noise is considered insignificant.

The most important factor influencing the system performance is the multipath problem. This point is studied in detail elsewhere [26]. In our design, the detection principle is based on energy computations. The energy of each slot is computed and the one with maximum energy is accepted as the correct data slot. In theory data errors may occur if the multipath signal energy is equal to or greater than the direct path signal energy, but in a practical situation with a diver moving about this is only likely to occur briefly and intermittently.

V. Speech Enhancement

Transmission of clear speech signals is essential for reliable communication between divers. Therefore, any kind of noise that would degrade the speech intelligibility must be cancelled or at least minimized if possible. In diver communications, there are two types of noise that affect speech quality. One is the breathing noise produced by air flow through the regulator during inhalation in the case of a conventional SCUBA, or by free-flowing air in the case of a helmet or band mask. The other is bubble noise which is generated during exhalation and during speech periods by escaping air from the regulator. Although the bubble noise occurs outside the mask cavity, the microphone inside the mask picks up this noise as well as the breathing noise. Several methods of noise cancellation have been applied elsewhere, especially to cancel the breathing noise [22], but these are all done by analog circuitry and need some modification before they can be incorporated into the present system.

First, we concentrate on cancelling the breathing noise because it has a significant effect on the clarity of communications. Also, the power consumption of the system is reduced if this noise is not transmitted. Our experiments have shown that breathing noise has white noise characteristics and higher signal magnitudes than either the speech signal or the bubble noise. These features of the noise are extracted every 22.5 ms (180 signal frames) and the energy magnitude and zero-crossing rates are measured. When the input signal is higher than some preset threshold, the signal frame is considered to be breathing noise and no data is transmitted. In this way the breathing noise is cancelled.

Bubble noise also degrades the speech quality, but its significance depends on the type of mask worn by the diver. In our experiments with three different full-face masks, the EXO-26, shown in Fig. 8, the AGA, shown in Fig. 9, and the AQUA LUNG, it was clear that the position of the microphone in the mask and the type of regulator used had a significant effect on speech quality. The closer the microphone was to the regulator, the more serious was the bubble noise. In this respect, the AGA mask yielded much better speech quality performance, since the microphone was placed on the opposite side from the regulator and the bubble noise was only slightly audible. Moreover, the bubble noise was suppressed when high amplitude speech signals were produced, i.e., for voiced sounds, and was hardly noticeable.

In order to cancel the bubble noise, thus enhancing the speech quality, several algorithms were tested using speech recorded during diving trials. The first of these was the spectral subtraction method [23], based on an estimation of the short-term spectral magnitude of the speech signal and bubble noise. From the recorded speech signal, the bubble noise amplitude spectral density is estimated by employing fast Fourier transform (FFT) analysis; the condition of stationarity is assumed during speech generation. Then the speech-plus-bubble noise amplitude spectral density for each frame is computed. In order to estimate the speech signal itself, it is also necessary to estimate its phase spectrum [23]. Finally, the bubble noise amplitude spectral density is subtracted from the speech-plus-bubble amplitude spectral density. The result should be a noise-free speech signal spectrum; by employing inverse FFT analysis and using the phase information, the speech signal in the time domain can be reconstructed. This method tends to enhance speech quality, but at times when the speech spectral amplitude is lower than the reference bubble noise spectral amplitude, it results in discontinuities in the

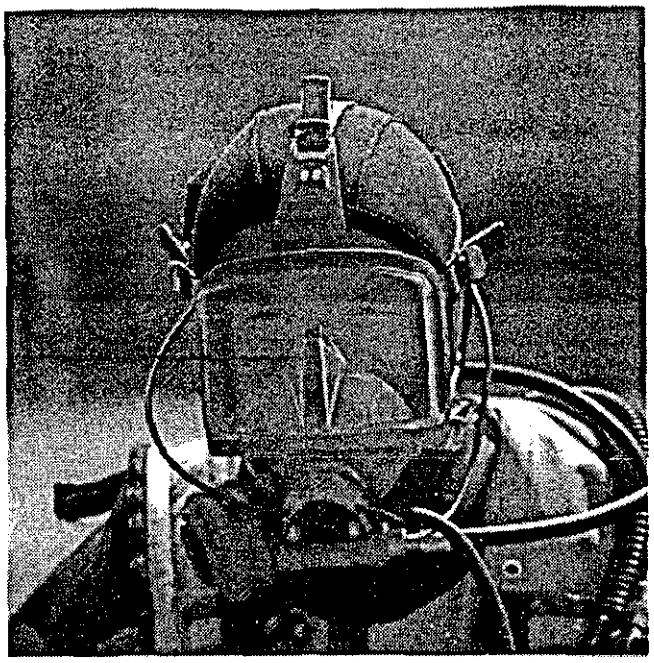


Fig. 9. AGA full-face mask.

speech signal. Moreover, since the reference spectrum is an estimate of bubble noise, it does not characterize the bubble noise, which can therefore be audible.

The second method attempted for bubble noise suppression was adaptive noise cancellation [24], [25]. This method needs two microphones, one placed inside the mask for the speech signal, the other placed outside the mask to pick up the bubble noise. This type of noise canceller consists of an adaptive filter that produces an estimate of the bubble noise, which is then subtracted from the speech signal. The overall output of the noise canceller is used to control the coefficients of the adaptive filter. In our experiments, bubble noise was added to clear speech signals in order to generate the input signal to the microphone. The same bubble noise was applied to be the input to a 20-tap adaptive filter. The least mean square (LMS) method was then applied to estimate the filter taps. The result of this simulation illustrated that the adaptive noise cancellation method was superior to the spectral subtraction method in terms of speech quality. More details of speech quality improvement will be presented in a future publication.

VI. Results

The system described here is still being developed but the results of initial tests show that it is capable of encoding and decoding signals according to the design principles. To test the system, a frame of 180 speech samples, shown in Fig. 10(a), was analyzed by a host computer to extract the vocal tract filter coefficients, pitch period, gain values and voiced/unvoiced frame decision data. These parameters were then quantized and stored in the EPROM's of the prototype system. Identical processing was done by the system and the two computed sets of parameters were found to be the same, thus verifying correct encoding.

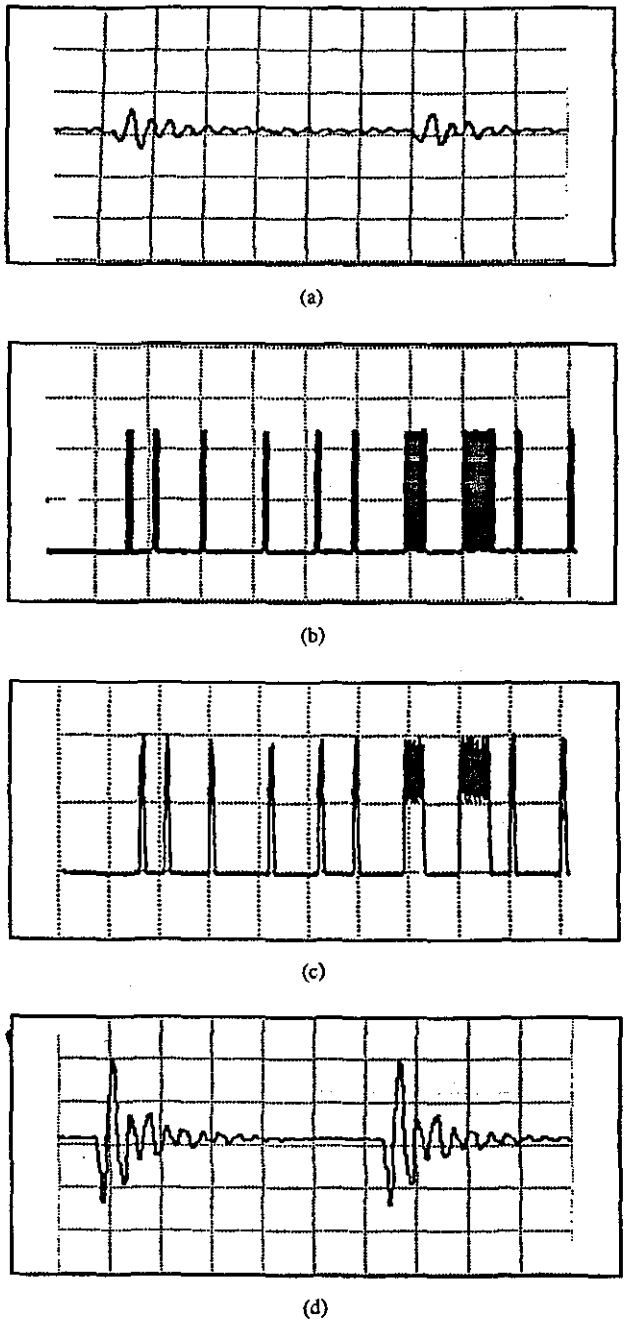


Fig. 10. Waveforms of (a) voiced input speech frame, (b) DPPM pulses representing speech parameters, (c) received DPPM pulses, and (d) synthesized speech waveform.

The computed parameters were transmitted in DPPM format, as shown in Fig. 10(b). In Fig. 10(b) the widest pulse is the synchronization pulse, occupying six slots, the second widest pulse signals the end of transmission, and the narrow pulses represent data slots. All of these pulses, after modulating a 70-kHz carrier signal, were transmitted to the receiver of a second system through a channel having a low pass filter response. From the received DPPM pulses, shown in Fig. 10(c), the speech parameters were obtained. These were then used to synthesise the speech signal, as shown in Fig. 10(d), which was the output of the second system's AIC It is clear that the transmitted and received codes shown hav the same form.

TABLE II
EXPECTED BIT ERROR RATE OF EIGHT-SLOT DPPM TRANSMISSION

SNR,dB	0	4	6	8	10	12
BER	0.967	0.284	0.1092	0.0279	0.0034	0.00014

During these tests no errors were detected due to the nonexistence of the multipath problem. For this type of channel the bit-error rate (BER) and the SNR are [27]

$$\mathrm{BER} \cong \left(\frac{M}{2}\right) \frac{1}{\sqrt{2\pi S_p/N_p}} e^{\frac{-S_p/N_p}{2}} \tag{10}$$

$$\mathrm{SNR} = 10 \log_{10} \frac{S_p}{N_p} \tag{11}$$

where S_p and N_p are signal and noise power levels, respectively.

Table II shows the dependence of BER upon the SNR. Clearly, it is important for the system to have as large a SNR as possible.

VII. CONCLUSION

This paper has described the design of an underwater acoustic voice communications system based on a digital signal processor (DSP). A comparatively low transmission rate has to be used if moderate ranges of typically a few hundred meters are to be achieved; this is because the acoustic range is reduced with increasing carrier frequency, which in this research is 70 kHz. This means that the high data rates used in mobile telephones (32 or 64 kb/s) cannot be adopted for diver communications except for very short range applications, when much higher carrier frequencies can be used, for example 1 MHz. To meet this limitation, the linear predictive coding method has been implemented and speech parameters are transmitted at 2.4 kb/s using digital pulse position modulation. The complete process of transmission and synchronous reception is controlled by a TMS320C31 DSP, which for each diver's unit can at present be switched from transmit to receive mode by a press-to-talk switch; in future, this function will be carried out by a voice-operated switch. The prototype system described here is currently being appraised in laboratory tests and during open water diving trials.

The main reason for improving on presently available underwater voice communication systems is to enhance the quality and reliability of communications and, for some scenarios, to enable direct interfacing with conventional telecommunications networks. This would make it possible, for example, for a diver engaged in a scientific project to be in direct contact with an expert hundreds or even thousands of kilometers away. Such a jump ahead from the rather dated present-day technology dictates the replacement of single sideband analogue systems and the adoption of digital technology for encoding and decoding the speech, and for minimizing the noise. As with the impressively small mobile telephones now available, there is consequently scope for miniaturising circuits as well as producing ergonomically acceptable designs, especially for the mass market of sports divers. An obvious feature is to mount the entire system on the diver's face mask to obviate the need for long wires that get snagged. This and other innovations will at last bring this long-neglected aspect of underwater technology up to date.

REFERENCES

[1] L. E. Virr, "The role of electricity in subsea intervention," *IEE Proc.*, 1987, pt. A, vol. 131, pp. 547–576.

[2] R. J. Hicks and L. E. Virr, "Underwater communication—A review," presented at Int. Conf. Divetech '81: The Way Ahead in Diving Technology, Workshop B: Aids to Underwater Operations, London, U.K., Nov. 1981.

[3] J. Maloney, "Diver communications in the police service," presented at Int. Conf. Divetech '81: The Way Ahead in Diving Technology, Workshop B: Aids to Underwater Operations, London, U.K., Nov. 1981.

[4] H. O. Berktay, B. Gazey, and C. A. Teer, "Underwater communication past, present, and future," *J. Sound Vib.*, vol. 7, pp. 62–70, 1968.

[5] A. Clark, "Diver communications—The case for single sideband," *Underwater Syst. Design*, pp. 16–18, 1989.

[6] B. Woodward, "Underwater telephony: Past, present, and future," *Colloque De Physique* C2, pp. 591–594, 1990.

[7] J. K. Sear, "Standardization of sonar communications," presented at Int. Conf. Divetech '81: The Way Ahead in Diving Technology, Workshop B: Aids to Underwater Operations, London, U.K., Nov. 1981.

[8] B. Woodward and H. Sari, "Underwater voice communications using digital techniques," *J. de Physique IV*, Colloque C5, vol. 4, pp. 469–472, 1994.

[9] A. Goalic *et al.*, "Toward a digital acoustic underwater phone," in *Proc. OCEANS'94*, 1994, vol. III, pp. 489–494.

[10] C. T. Morrow and A. J. Brouns, "Acoustic impedance calibrator for mask and microphone measurements," *J. Audio Eng. Soc.*, vol. 18, pp. 519–523, 1970.

[11] ——, "Speech communication in diving masks—I: Acoustics of microphones and mask cavities," *J. Acoust. Soc. Am.*, vol. 50, pp. 1–9, 1971.

[12] L. M. Lafuente, "Adaptive differential pulse code modulation for low bit rate transmission of speech signals," *Electrical Comm.*, vol. 58, pp. 225–229, 1983.

[13] "Details to assist in implementation of federal standard 1016 CELP," Natl. Commun. Syst., Tech. Bull. 92-1, 1992.

[14] F. J. Owens, *Signal Processing of Speech*. Basingstoke, U.K.: Macmillan, 1993.

[15] J. Makhoul, "Linear prediction: A tutorial review," in *Proc. IEEE*, vol. 63, pp. 561–580, 1975.

[16] A. H. Gray, R. M. Gray, and J. D. Markel, "Comparison of optimal quantization of speech reflection coefficients," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-25, pp. 9–21, 1977.

[17] M. J. Ross *et al.*, "Average magnitude difference function pitch extractor," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-22, pp. 353–362, 1974.

[18] T. E. Tremain, "The government standard linear predictive coding algorithm: LPC10," *Speech Technol.*, vol. 1, pp. 40–49, 1982.

[19] A. Falahati, S. C. Bateman, and B. Woodward, "Underwater acoustic channel models for 4800 bps QPSK signals," *IEEE J. Ocean. Eng.*, vol. 16, pp. 12–18, 1991.

[20] S. Riter and P. A. Boatrigth, "Design considerations for a pulse position modulation underwater acoustic communication system," in *Dig. IEEE Conf. Eng. Ocean Environment*, 1970, pp. 21–24.

[21] J. M. H. Elmirghani and R. A. Cryan, "Analytic and numeric modeling of optical fiber PPM slot and frame spectral properties with application to timing extraction," in *IEE Proc. Commun.*, 1994, vol. 141, pp. 379–389.

[22] D. J. Meares, "Broadcast quality speech from diving helmets," *J. Audio Eng. Soc.*, vol. 37, pp. 929–933, 1989.

[23] J. A. Lim and A. V. Oppenheim, "Enhancement and bandwidth compression of noisy speech," *Proc. IEEE*, vol. 67, no. 12, pp. 1586–1604, 1979.

[24] B. Widrow *et al.*, "Adaptive noise cancelling: Principles and applications," *Proc. IEEE*, vol. 63, no. 12, pp. 1692–1716, Dec. 1975.

[25] J. Dunlop, M. J. Al-Kindi, L. E. Virr, and H. M. S. Nelson, "Application of adaptive noise cancelling to diver voice communications," in *Proc. Int. Conf. ASSP*, 1987, pp. 1708–1711.

[26] R. J. Urick, *Principles of Underwater Sound for Engineers*. New York: McGraw-Hill, 1965.

[27] R. M. Cagliardi and S. Karp, *Optical Communications*. New York: Wiley, 1995.

Bryan Woodward (M'95) was born in Gloucestershire, U.K., in 1941. He received the B.Sc., M.Sc., and Ph.D. degrees in physics from the University of London in 1964, 1966, and 1968, respectively, and the Diploma of Membership of Imperial College (D.I.C.), London, in 1966.

During 1964–1968 he did postgraduate research at Imperial College of Science and Technology (University of London), on noise detection in liquid-cooled nuclear reactors by acoustic waveguides. During 1969–1971 he was a Research Associate at Guy's Hospital Medical School, London, investigating the medical hazards of ultrasound, then during 1972–1975 he was a Research Scientist with the Australian Atomic Energy Commission, Sydney, studying the use of acoustic emission for structural integrity monitoring. Since 1975, except for a period of study leave, he has been at Loughborough University of Technology, Leicestershire, U.K., where he is presently Reader in Underwater Acoustics. During 1982–1983 he was Visiting Professor at Laboratoire de Mecanique et d'Acoustique, CNRS, Marseille, France. His main research interests are in underwater communications, telemetry and navigation, parametric sonar, correlation sonar, bioacoustics, transducers and diver instrumentation.

Dr. Woodward is a Chartered Engineer, European Engineer, Fellow of the Institution of Electrical Engineers, Fellow of the Institute of Acoustics, and Fellow of the Royal Geographical Society. He is also a Chief Examiner for the Engineering Council in the United Kingdom.

Hayri Sari (S'96) was born in Samsun, Turkey, in 1965. He received the B.Sc. degree in electronic engineering from Marmara University, Turkey, in 1988 and the M.Sc. degree in instrument design and application from the University of Manchester Institute of Science and Technology in England in 1992. He is now preparing a thesis for the Ph.D. degree from Loughborough University of Technology, Leicestershire, U. K. His research interests include speech signal processing, underwater communications, and DSP implementation into digital data transmission systems.

JOURNAL DE PHYSIQUE IV
Colloque C5, supplément au Journal de Physique III, Volume 4, mai 1994

Underwater voice communications using digital techniques

B. WOODWARD and H. SARI

Department of Electronic and Electrical Engineering, Loughborough University of Technology, Leicestershire LE11 3TU, U.K.

Abstract: A digital signal processor (DSP) based voice communication system for divers with a 2400 bit/s transmission rate is described. Each speech signal frame is analysed using linear predictive coding (LPC) and the essential *parameters* such as amplitude, pitch period and digital filter coefficients are extracted. These parameters are transmitted through a water channel at a rate of 2400 bit/s by employing digital pulse position modulation (DPPM).

1. INTRODUCTION

The role of communications in diving is important not only from the point of view of information exchange and operational efficiency, but also to achieve greater safety. Since 1905, divers have been using voice communication systems, the early versions having a direct wire link between a *hard hat* diver on the bottom and a supervisor at the surface. Like the diving equipment of the time, these systems were simple and unsophisticated, and they were replaced in the 1960s by through-water acoustic systems which were more suitable for divers using Self Contained Underwater Breathing Apparatus (SCUBA).

There are several commercially available analogue *wireless* diver communication systems [1-3]. Some early systems used baseband communication methods in which acoustic signals were transmitted without modulation. However, these systems were affected by ambient acoustic noise which is dominant below 6 kHz [4]. In most subsequent versions this problem has been obviated by shifting the spectrum of the baseband signals to a higher band and transmitting by using a modulation method such as amplitude modulation (AM), double sideband suppressed carrier (DSBSB), single sideband (SSB) or frequency modulation (FM). In view of its lower power consumption, SSB has been the most commonly employed technique [5]. When used underwater with acoustic signals for diver-to-diver or diver-to-surface communication, SSB has been shown to provide all the proven advantages of its counterpart for radio transmission. Nearly all the analogue systems that are presently on the market are virtually identical in their technical specifications and have not progressed beyond the technology of three decades ago. It is all too clear that the rapid development in digital signal processor technology and the advent of mobile communication systems has been completely ignored by the manufacturers of divers' voice communication systems.

2. FEASIBILITY OF DIGITAL SPEECH CODING FOR UNDERWATER COMMUNICATIONS

In this research project speech data is transmitted by an acoustic transducer with a resonance frequency of 70 kHz, therefore a suitable kind of speech coding method has to be applied. After sampling the speech signal at a rate of 8 kHz, all the coded information representing this sample must be transmitted during the corresponding sampling period of 125 μs. There are several *waveform quantization* and encoding techniques for speech signals that have been widely used in telecommunications since the 1950s. In particular, we consider adaptive differential pulse code modulation (ADPCM).

In 1990, the Consultative Committee for International Telephone and Telegraph (CCITT) approved Recommendation G.727 for embedded ADPCM [6]. This specifies that the ADPCM algorithms can operate with 5, 4, 3 or 2 bits per sample, i.e. at rates of 40, 32, 24, and 16 kbit/s. Because of the limitation of the projector, with a resonant frequency of f_0=70 kHz, only 16 kbit/s, or 4 cycles per bit, may be used for speech transmission. Therefore, the maximum time allowed to transmit each bit is less than 62,5μs. In a noiseless channel, this rate is acceptable. But during transmission through water, the acoustic signal is seriously distorted by multipath propagation, which makes signal recovery very difficult. A method requiring a lower transmission rate must therefore be employed.

One possibility is to use one of the so-called *parametric* coding methods, which are recommended for low bit rate coding of speech signals [7, 8]. These methods reduce the transmission rate to 400 bit/s by separating the excitation component of the speech from the spectral envelope component. The excitation is then characterized as either a pulse train for voiced sounds or noise for unvoiced sounds. The spectral envelope can be characterized by the parameters of a digital filter having the same transfer characteristics as the vocal tract.

Linear predictive coding (LPC) provides an accurate representation of the relevant speech parameters that can reduce transmission rates at the expense of a reduction in the quality of the reproduced speech [9]. In an LPC voice coder, efficient speech synthesis is achieved by transmitting frames of the speech waveform as a set of *parameters*, which are (i) the amplitude, (ii) a voiced/unvoiced decision and pitch period for voiced sounds, and (iii) the filter coefficients.

In the linear prediction representation, the speech signal is modelled as the output of an all-pole filter H(z).

$$H(z) = \frac{G}{1 + \sum_{k=1}^{p} a[k]\, z^{-k}} \qquad 1$$

where p is the number of modelled poles, G is the gain of the filter and the a[k]s are the coefficients characterizing the filter. Generation of the synthetic speech sequence requires a knowledge of the pitch period, the linear prediction coefficients and the power of the waveform in each speech frame. These parameters are computed by means of a TMS320C30 digital signal processor (DSP) [10]. The advantage of the LPC method is that the speech transmission rate is decreased to 2400 bit/s, including error correction codes. In comparing the *waveform* and *parametric* coding methods, as can be seen from their transmission rates, the LPC method is superior to ADPCM and is therefore preferred for underwater acoustic transmission.

3. PULSE POSITION MODULATION FOR UNDERWATER SPEECH TRANSMISSION

Digital modulation techniques, such as amplitude shift keying (ASK) [11], frequency shift keying (FSK) [12] and phase shift keying (PSK) [13] can be employed for underwater acoustic data transmission. As an alternative, pulse position modulation (PPM) is also an effective technique. The analogue PPM method is suited to serial data transmission systems [14]; its mode of operation is to sample the analogue signal and to transmit a short pulse that is delayed with respect to the sampling instant by an amount that is a linear function of the signal amplitude. The drawback is that during transmission a pulse can be seriously distorted by multipath propagation. Therefore estimating its relative position at the receiver is difficult, especially for high data rates.

Since the encoded speech signal is digital in nature, a digital pulse position modulation (DPPM) system is proposed. Because the speech parameters are digitally coded with a transmission rate of 2400 bit/s each 1-second of data is divided into frames consisting of 3 bits, i.e. 800 frames per second as shown in Fig.1. A frame, with a period of T_f = 3/2400 s (1.25 ms), is further divided into 10 slots, each of width T_s = 125 μs; these comprise 8 data slots which represent the eight possible 3-bit combinations of data and two guard band slots. A single PPM pulse is transmitted in each frame, allocated to one of the slots according to the digital information. The demodulation of such signals therefore requires two clock signals, one at the frame frequency of 800 Hz and the other at the slot frequency of 8 kHz.

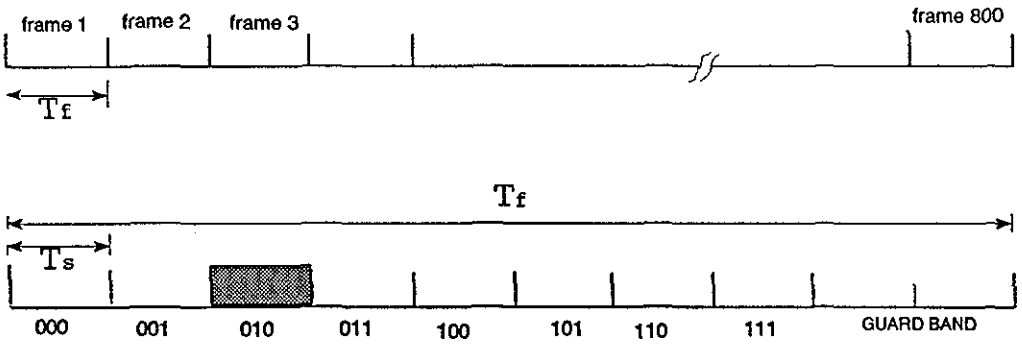


Fig.1. Digital PPM signal format for 8 data slots per sub-frame

Different synchronization schemes have been suggested in the literature for digital PPM transmission [15]. The system described here operates asynchronously, since there is no phase synchronism between the transmitter of one diver's system and the receiver of another. But in each separate system the transmitter and receiver are controlled by the same clock, which is simply a multiple of the slot frequency. Detection of the digital pulse position signal is achieved with the system shown in Fig.2.

Since a digital pulse representing the position of the data modulates a 70 kHz sinusoidal waveform, a narrow bandpass filter is included in the receiver. The output of this filter is then passed through an envelope detector. Consider the situation when one diver starts speaking. The filter output of another diver's receiver first detects eight synchronization data pulses, i.e. 1111 1111_2, which trigger a 4-bit counter running 8 kHz. The outputs of the counter control the integrators. When it counts to 1000_2 the outputs of the eight integrators are compared to a threshold level set just above a level representing the amplitude of the unwanted multipath signal. During the pulse transmission interval the output of a particular integrator will be high and this becomes the input to a 8-to-3 decoder. The 3-bit decoder output is the transmitted parameter that is fed to the DSP-based LPC synthesizer.

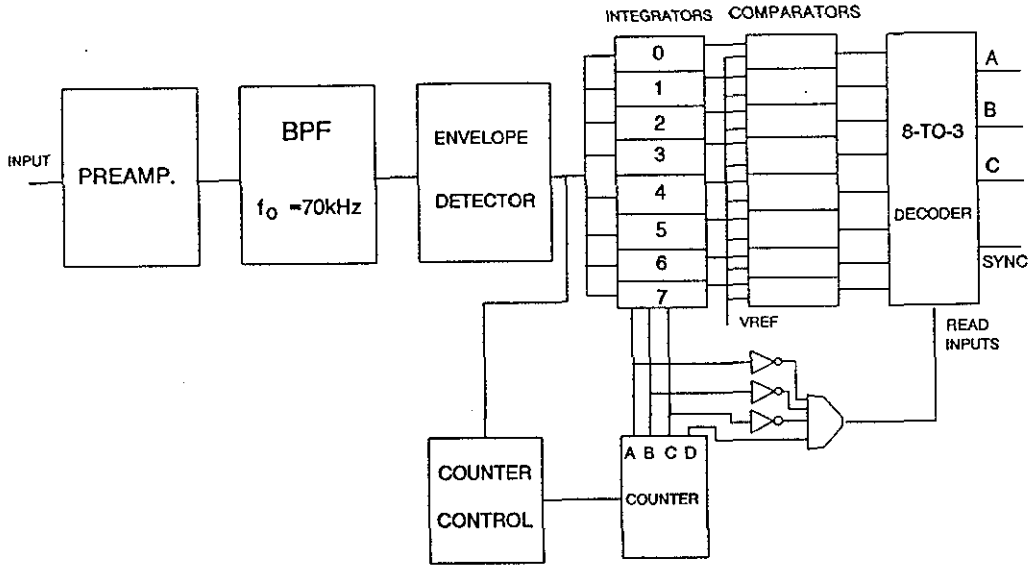


Fig.2 Analogue system for digital pulse position detection.

REFERENCES

[1] Overfield T., *Underwater Systems Design*, (March/April 1988) pp.8-13.

[2] Clark A., *Underwater Systems Design*, (January 1989) pp. 16-18.

[3] Peck M.J., *Sea Technology*, (1992) pp. 61-65.

[4] Virr L.E., *IEE Proc* **6** (1987) pp. 547-576.

[5] Woodward B., "Underwater Telephony: Past, Present and Future," Colloque De Physique **2** (1990) pp. C2_591-C2_594.

[6] Sherif M.H., Bowker D.O., Bertocci G., Orford B.A. and Mariano G.A., *IEEE Trans. Comm.*, **2** (1993) pp. 391-399.

[7] Rabiner L.R. and Schafer R.W., Digital Processing of Speech Signals (Prentice-Hall, New Jersey 1978).

[8] Deller Jr. J.R., Proakis J.G. and Hansen J.H.L., Discrete-Time Processing of Speech Signals (MacMillan Publishing Comp., New York 1993).

[9] Makhoul J., *Proc IEEE* **63** (1975) pp.561-580.

[10] Tremain T.E., *Speech Technology* **1** (1982), pp. 40-49.

[11] Dawoud M.M., Halawani T.U. and Abdul-jauwad S.H. *Int. J. Electronics* **72** (1992) pp. 183-196.

[12] Brock D.C., Bateman S.C. and Woodward B., *Ultrasonics*, **24** (1986) pp. 183-188.

[13] Falahati A., Bateman S.C. and Woodward B., *IEEE J. Ocean. Eng.*, **OE-16** (1991) pp. 12-18.

[14] Riter S. and Boatrigth P.A., "Design considerations for a pulse position modulation underwater acoustic communication system," Digest IEEE Conf. Engineering in the Ocean Environment, Panama City, Fla., September 1970 pp. 21-24.

[15] Ling G. and Cagliardi R.M., *IEEE Trans. Comm.* **COM-34** (1986) pp. 1202-1208.