IMPLEMENTATION OF DIGITAL TECHNOLOGY FOR UNDERWATER VOICE COMMUNICATIONS

B. Woodward(1), H. Sari(1)

(1) Department of Electronic and Electrical Engineering,Loughborough University of Technology, Leicestershire, LE11 3TU, United Kingdom

1. INTRODUCTION

Modern telecommunications systems are now very sophisticated but underwater voice communication systems are still comparatively archaic. But there are 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 into a band-limited microphone in a mask or helmet, which distorts the speech signal. 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 noise from breathing, free-flowing air and bubbles. Even in comparatively quiet rebreather systems, breathing noise is still present. Non-linear microphone distortion and indistinct articulation compound these problems. In through-water acoustic systems, the noise sources include wave action, sounds of biological origin, shipping and rain. 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 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 are virtually unchanged from the "wire-less" or "through-water" communication systems introduced in the 1950s for divers using Self Contained Underwater Breathing Apparatus (SCUBA)[2,3].

Nowadays, there are many commercially available through-water diver communication 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]. In later versions, the baseband spectrum was shifted to a higher band by modulating a carrier frequency with the speech signal. Improved systems employed amplitude modulation (AM) [4]. This has the disadvantage that the carrier is transmitted continuously 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. The commonest diver communication systems now use the single sideband (SSB) modulation [5,6]. This counters many of the shortcomings of AM transmission because during periods when no speech is being transmitted, SSB dissipates no power and has only half the bandwidth of AM systems and Double Sideband Suppressed Carrier (DSBSC) systems. 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]. A major factor governing the choice of frequency is the effect of ambient noise, which is a particular problem 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 commercial 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. 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 a practical limit to the number of carrier frequencies that can be accommodated in the bandwidth of a transducer. With a digital system using an 8-bit synchronization code there are effectively 256 possible channels available for a single carrier frequency.

Although there have been rapid advances in mobile telephone technology, the adoption of such technology for diver communications has been totally neglected. Studies are now being done 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.

2. ELECTRONIC DESIGN

A block diagram of the prototype system is shown in Fig.2. It 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 Analogue Interface Circuit (AIC). The speech signals are band pass filtered with a bandwidth of 3.2 kHz, sampled at a rate of 8 kHz, then quantized to 14-bit 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. For memory, the system has four 8k x 8 bit EPROMs for the program and four 8k x 8 bit SRAMs for data.

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 1 output of the DSP, and a matching transformer. This timer is for generating a square wave carrier frequency of 70kHz, defined by the omnidirectional transducer used. The same transducer is used for both transmission and reception by isolating the transmitter section from the receiver section with an analogue switch controlled by the XF1 input/output pin of the DSP. The receiver section consists of a preamplifier, an 8th-order Butterworth band pass filter centred at 70 kHz, an envelope detector and an 8-bit MAX153 analogue-to-digital convertor, which sends digital samples to the DSP for the necessary processing. One of the important requirements of the system is to establish a communication link 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 less versatile than making a normal telephone call and would need to include a magnet-operated keyboard to allow the diver to "dial" underwater. The outputs of the 4 x 4 keypad encoder (MM74C922N) are directly connected to the least significant four data bits of the DSP, as shown in Fig. 2.

Since the system is initially set to the receive mode, the output of the receiver's envelope detector is continuously sampled and quantized by the ADC. This signal is then processed by the DSP in order to decode the transmitted data. Whenever the system recognises its ID number, it decodes the received data and the AIC generates speech. If one diver wants to talk to another, the appropriate number must be dialled. This results in the generation of an interrupt request signal to the DSP and the system is switched to transmit mode. After that the encoded data is transmitted.

3. SPEECH CODING METHODS

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. 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 minimum sampling rate needed is at least 6 kHz. The corresponding sampling period is therefore 167 μs, so whatever modulation technique chosen this is the maximum time allowed to transmit the encoded information.

In this research, speech data is sampled at a rate of f_c = 8 kHz. The coded information representing each sample must therefore be transmitted every 125 μs. 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 techniques are Pulse Code Modulation (PCM) and Adaptive Differential Pulse Code Modulation (ADPCM)[12]. PCM has a data rate of 64 kbit/s and ADPCM is normally encoded with 5, 4, 3 or 2 bits/sample, i.e. at rates of 40, 32, 24 or 16 kbit/s. Because of the limitation of the transducer, with a resonance frequency of 70 kHz, only 16 kbit/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. 4. 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 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 methods 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 kbit/s and Residual Excited Linear Prediction Coding (RELP) [14] with a data rate of 13.2 kbit/s, which are used in digital mobile telephones.

The coding method adopted here is 'conventional' linear predictive coding (LPC) which provides an accurate representation of the relevant speech parameters that can reduce transmission rate of 2.4 kbit/s [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 (i) the amplitude, (ii) a voiced / unvoiced decision and pitch period for voiced sounds, and (iii) a set of so-called linear prediction coefficients. These are computed by the digital signal processor.

In the linear prediction representation, the speech production system is modeled as 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 modelled 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, i.e. 180 sampling periods of 125 μs, which is sufficiently long to enable the pitch period to be determined. 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. 5 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 N = 180.

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 low-pass filter with a cut-off 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. In order to estimate the voice characteristics, energy and zero-crossing measurements are employed. There are several methods of estimating the pitch period, including the Average Magnitude Difference Function (AMDF) method [16], which is implemented here because it requires less computational time for real-time operation. 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, then transmitted [17,18].

4. THROUGH-WATER TRANSMISSION

Although digital modulation methods are adopted in mobile communication systems, 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 metres is required [19]. For the present system, implementation is achieved by Digital Pulse Position Modulation (DPPM), as shown in Fig. 6. This encodes M bits of PCM quantized speech 'in a frame of duration T_f by transmitting a single pulse in one of the n slots within the frame, where $n = 2^M$, which is therefore capable of accommodating all the M-bit permutations [20]. A guard band whose length is defined by the modulation index m is also included in each frame to avoid inter-frame interference (IFI). The slot duration T_s is defined as T_f/n. Here, the quantized speech parameters are separated into sub-groups consisting of 3-bits, i.e. M = 3, giving a data rate of 2.4 kbit/s and a frame of $T_f = 3/2400$ s , i.e. 1.25ms. Therefore, 8 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 $T_s = T_f/(n+2)$, i.e. 125 μs, is needed, corresponding to a slot frequency of 8 kHz.

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 analogue-to-digital convertor. Since the envelope of the received signal can alter at a rate of 8 kHz, it is sampled at 32 kHz, i.e. 4 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 found. The corresponding code, here 100_2 in Fig. 6, represents the LPC parameters transmitted and is used to synthesise the speech signal.

5. SPEECH ENHANCEMENT

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 [21].

Cancelling the breathing noise has a significant effect on the clarity of communications. Also, the power consumption of the system is reduced if this noise is not transmitted. 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 pre-set threshold, the signal frame is considered to be breathing noise and no data is transmitted.

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, the AGA 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, since the microphone was placed on the opposite side from the regulator and the bubble noise was only slightly audible.

To cancel the bubble noise, several algorithms were tested using speech recorded during diving trials.

The first of these was the spectral subtraction method [22], 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 a Fast Fourier Transform (FFT) and assuming stationarity 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 [22]. Finally, two spectra are subtracted; by employing an inverse FFT and using the phase information, the speech signal in the time domain can be reconstructed.

The second method attempted for bubble noise suppression was Adaptive Noise Cancellation [23,24]. This method needs two inputs, one from the microphone placed inside the mask for the speech signal, the other from a sample of the bubble noise stored in memory. 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 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 coefficients. The result of this simulation illustrated that the adaptive noise cancellation method was superior to the spectral subtraction method in terms of speech quality.

6. CONCLUSIONS

This paper describes 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 ranges of typically a few hundred metres are to be achieved; this is because the acoustic range is reduced with increasing carrier frequency, which here is 70 kHz. This means that the high data rates used in mobile telephones (32 or 64 kbit/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 kbit/s using digital pulse position modulation. The complete process of transmission and synchronous reception is controlled by a digital signal processor, which for each diver's unit can 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.

The main reason for improving on presently available 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 of kilometres away. Such a jump ahead from the outdated present-day technology dictates the replacement of single sideband analogue systems with digital technology for encoding and decoding the speech and for minimising the noise. As with the 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.

UNDERWATER VOICE COMMUNICATIONS

7. REFERENCES

[1] L E Virr, 'The role of electricity in subsea intervention', *IEE Proc.*,part A, 131 p547 (1987)

[2] R J Hicks, and L E Virr, 'Underwater communication - a review', *Proc. Int. Conf. Divetech '81: The Way Ahead in Diving Technology*, London, 24-26 Nov (1981).

[3] J Maloney, 'Diver communications in the police service', *Proc. Int. Conf. Divetech '81: The Way Ahead in Diving Technology*, London, 24-26 Nov (1981).

[4] H O Berktay, B. Gazey & C A Teer, 'Underwater communication past, present and future', *J. Sound Vib.*, 7 p62 (1968).

[5] A Clark, 'Diver communications - the case for single sideband', *Underwater Systems Design*, p16 (1989).

[6] B Woodward, 'Underwater telephony: past, present and future' *Colloque De Physique C2*, p591 (1990).

[7] J K Sear, 'Standardisation of sonar communications', *Proc. Int. Conf. Divetech '81: The Way Ahead in Diving Technology*, London, 24-26 Nov (1981).

[8] B Woodward & H. Sari, 'Underwater voice communications using digital techniques' *J. de Physique IV, Colloque C5*, 4 p469 (1994).

[9] A Goalic, J Labat, J Trubuil, S Saoudi, & D Rioualen, 'Toward a digital acoustic underwater phone', *Proc. OCEANS'94*, III, p489 (1994).

[10] C T Morrow & A J Brouns, 'Acoustic impedance calibrator for mask and microphone measurements', *J. Audio Eng. Soc.*, 18 p519 (1970).

[11] C T Morrow & A J Brouns, 'Speech communication in diving masks. I. Acoustics of microphones and mask cavities', *J. Acoust. Soc. Am.*, 50, p1 (1971).

[12] L M Lafuente, 'Adaptive differential pulse code modulation for low bit rate transmission of speech signals', *Electrical Comm.*, 58 p225 (1983).

[13] National Communications System 'Details to assist in implementation of Federal Standard 1016 CELP', *Technical Bulletin* 92-1 (1992).

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

[15] J Makhoul, 'Linear prediction: a tutorial review', *Proc. IEEE*, 63 p561 (1975).

[16] M J Ross, H L Shaffer, A Cohen, R Freudberg & H J Manley, 'Average magnitude difference function pitch extractor', *IEEE Trans. Acoust., Speech, Signal Proc.*, ASSP-22 p353 (1974).

[17] A H Gray, R M Gray & J D Markel, 'Comparison of optimal quantization of speech reflection coefficients', *IEEE Trans. Acoust., Speech, Signal Proc.*, ASSP-25 p9 (1977).

[18] T E Tremain, 'The government standard linear predictive coding algorithm:LPC10', *Speech Technology*, 1, p40 (1982).

[19] A Falahati, S C Bateman & B Woodward, 'Underwater acoustic channel models for 4800bps QPSK signals' *IEEE J. Ocean. Eng.*, OE-16 p12 (1991).

[20] S Riter & P A Boatrigth, 'Design considerations for a pulse position modulation underwater acoustic communication system', *Digest IEEE Conf. Eng. Ocean Environment*, p21 (1970).

[21] D J Meares, 'Broadcast quality speech from diving helmets', *J. Audio Eng. Soc.*, 37, p929 (1989).

[22] J A Lim, & A V Oppenheim, 'Enhancement and bandwidth compression of noisy speech', *Proc. IEEE*, 67 12, p1586 (1979).

[23] B Widrow, J R Glover, M J McCool, J Kaunitz, S C Williams, H R Hearn, J R Zeidler, E Dong & R C Goodlin, 'Adaptive noise cancelling: Principles and applications', *Proc. IEEE*, 63 12, p1692 (1975).

[24] J Dunlop, M J Al-Kindi, L E Virr & H M S Nelson, 'Application of adaptive noise cancelling to diver voice communications', *Proc. Int. Conf. ASSP*, p1708 (1987).

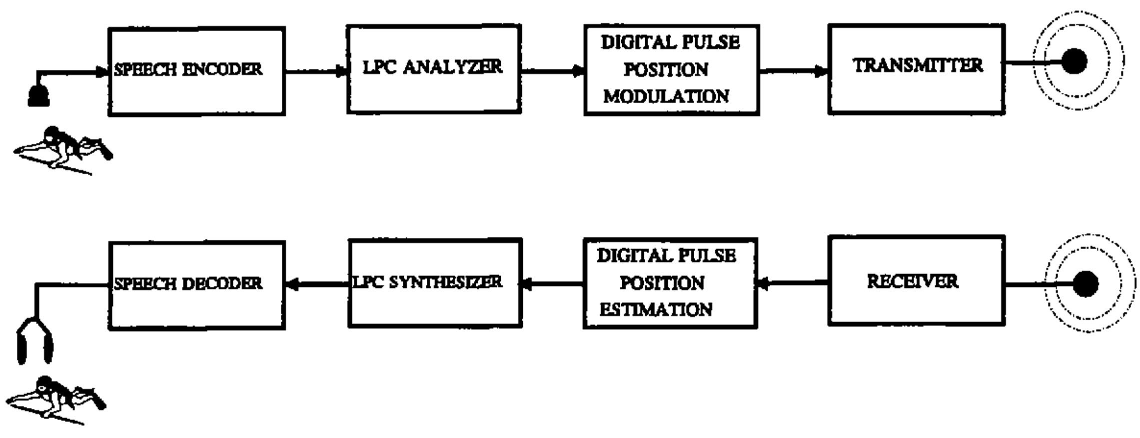

Fig.1. Principle of digital underwater acoustic voice communication system

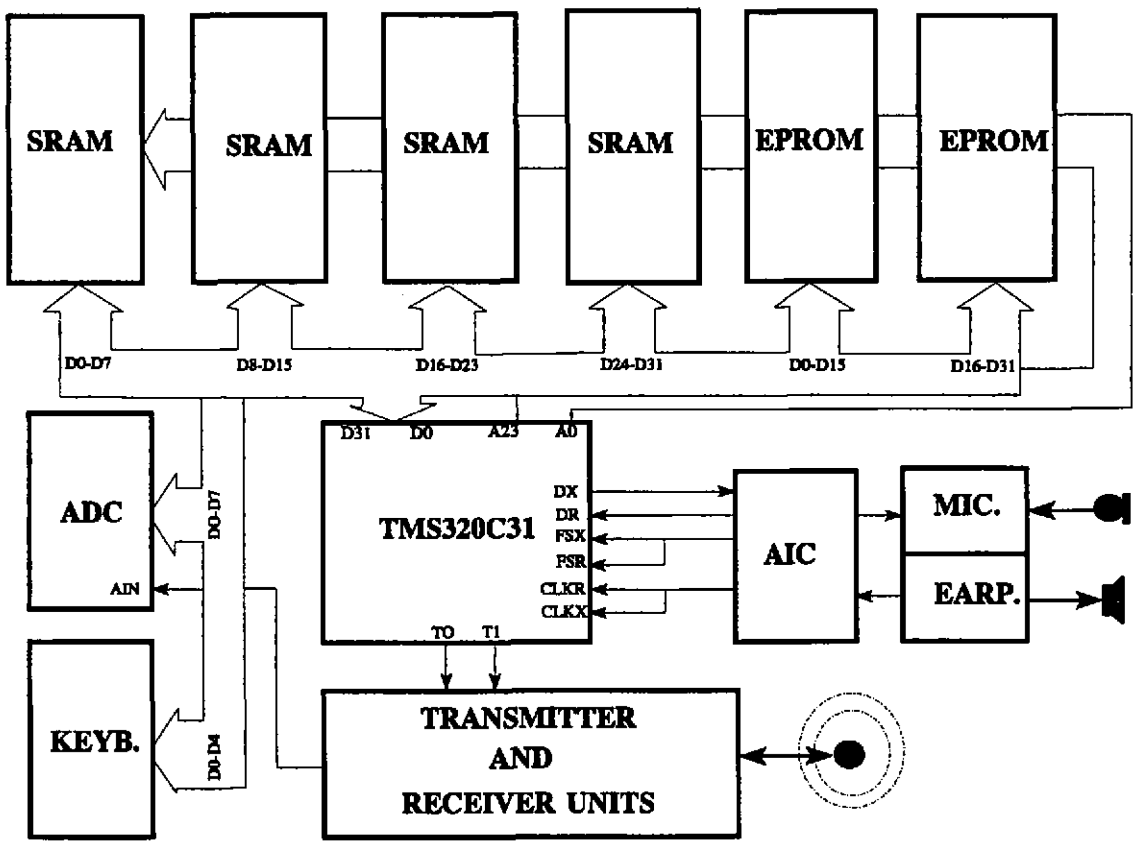

Fig.2. TMS320C31 DSP based underwater acoustic voice communication system

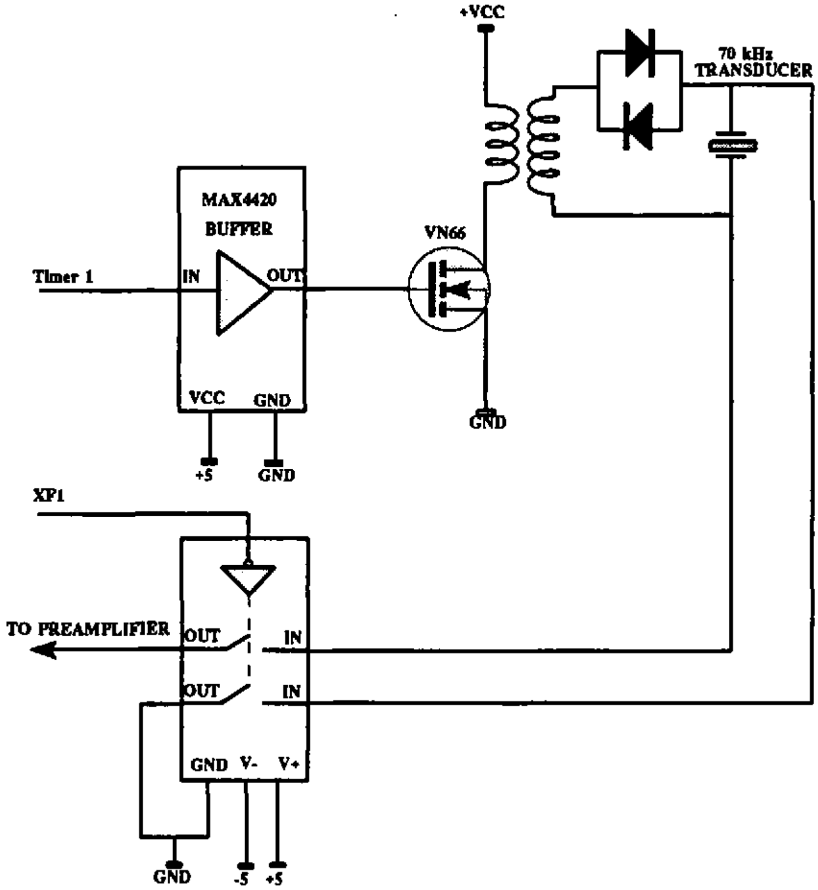

Fig.3. Transmitter section and its control.

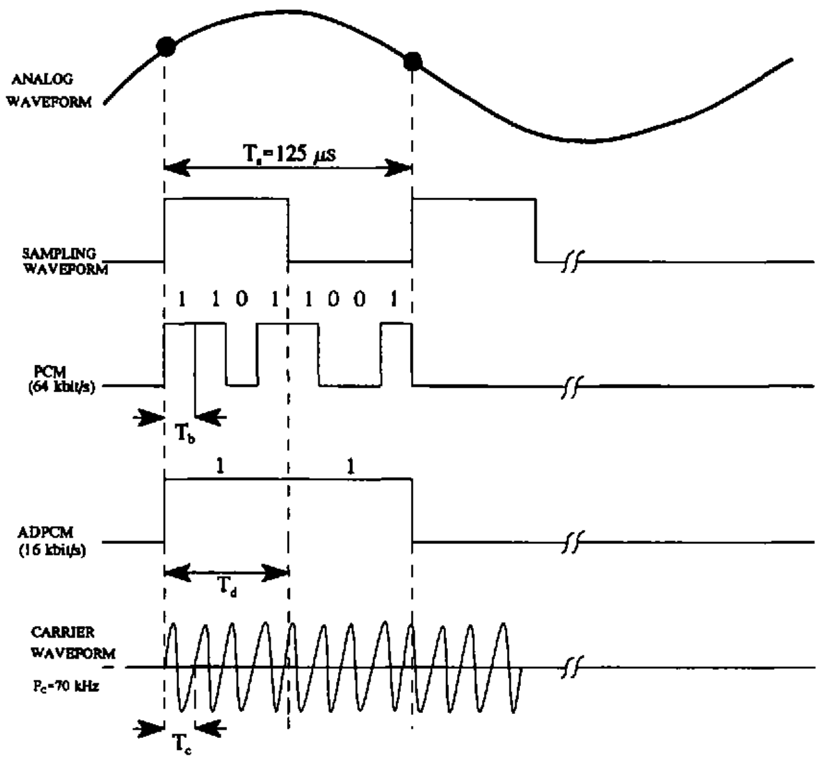

Fig.4. Transmission of PCM and ADPCM encoded speech samples

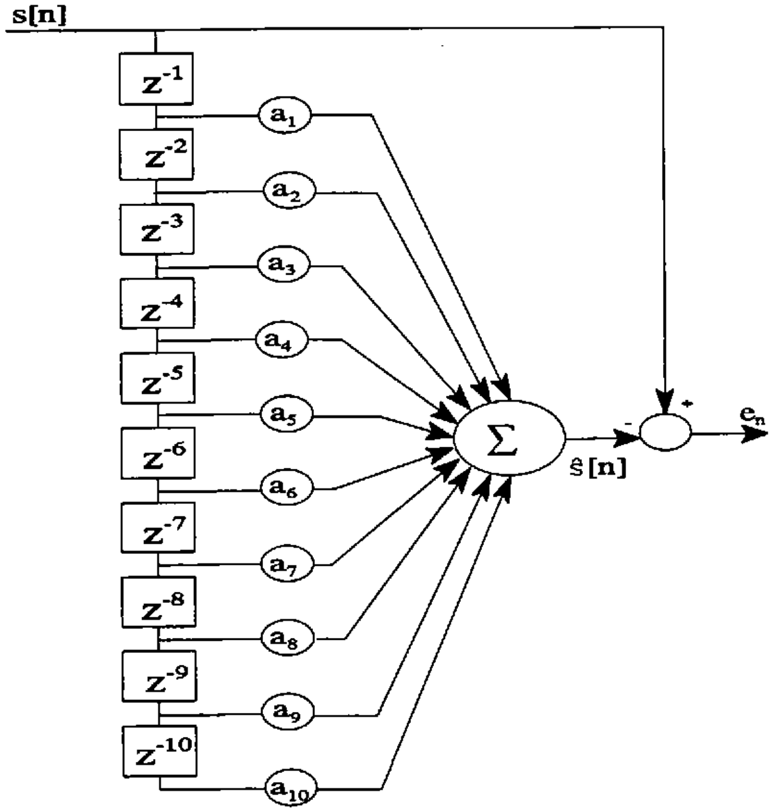

Fig.5. Inverse all-pole filter for LPC coefficients estimation

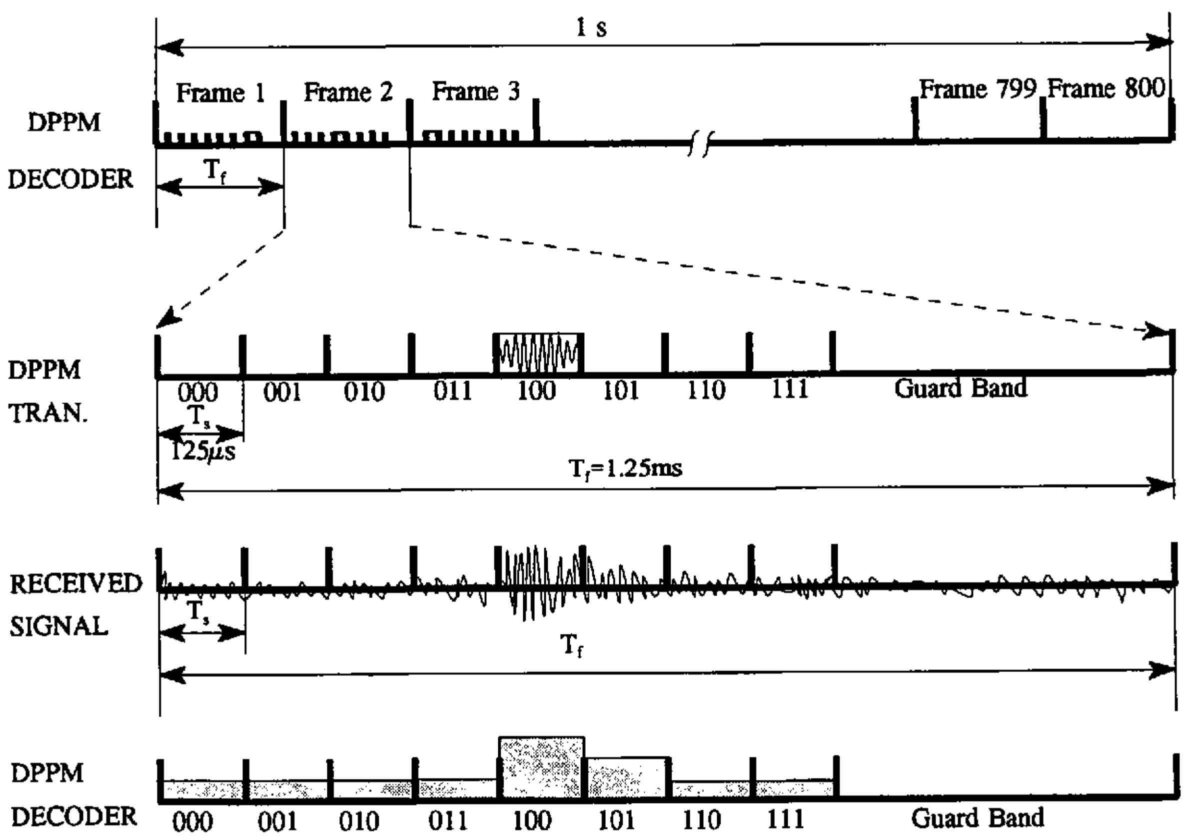

Fig.6. Digital PPM signal format for transmit and receive operation