

A Real-time Underwater Speech Communication System

Jiang Jingsai, Ma Xiaofeng, Li Ye,
Shandong Computer Science Center
Shandong Provincial Key Laboratory of computer
Network
Jinan, China
jiangjs@keylab.net

Fan Yanhong, Hao Qiuyun
Shandong Computer Science Center
Shandong Provincial Key Laboratory of computer
Network
Jinan, China

Abstract—With frequency hopping (FH) communication system and ultra-low-rate speech coding technology, the paper designed a real-time underwater speech communication system, for which the hardware solution is also given. In the system, the part of source coding is a 300bps vocoder based on enhanced mixed excitation linear prediction (MELPE). In order to overcome the multi-path interference in underwater acoustic communication, FH technology is used because of its low probability of intercept as well as anti-interference. The test results showed that the system had good performance and could be used for underwater communication.

Keywords—Speech signal processing, MELPE, BPVC, LSF, FH, 4FSK

I. INTRODUCTION

Digital speech communication has been used widely in many fields of communication in recent years, but the underwater speech communication is still in its initial stage. With the development of modern military as well as the improvement of utilizing the resource of the sea, the research on underwater communication is getting more and more important and has been extensively studied in recent years [1]. Real-time underwater speech communication is of great significance in civilian and military. However, due to the limited bandwidth, strong multi-path interference and fast time-varying unfavorable factors in the underwater channel, real-time underwater communication is some difficult. Nevertheless, with the development of digital speech signal processing, vocoder with ultra-low-bit rate can be designed [2-6]. In view of this, the paper combines ultra-low-rate speech coding technology and frequency hopping (FH) communication to design an acoustic speech communication system.

II. SYSTEM ARCHITECTURE

The whole underwater speech communication system consists of four parts: input/output module, vocoder module, modem module and transmitter/receiver module. Input/output module is composed of microphones and speakers, which are mainly used to acquire and play back the speech signal. Vocoder module and modem module are the core modules which carry out speech coding and signal modulation. The transmitter and receiver modules are implemented by acoustic transducers. The algorithm of core

modules is implemented on high-performance DSP chip. TI's high-speed processor TMS320VC5509A is adopted in the system. The whole block diagram of the system is shown in Figure 1.

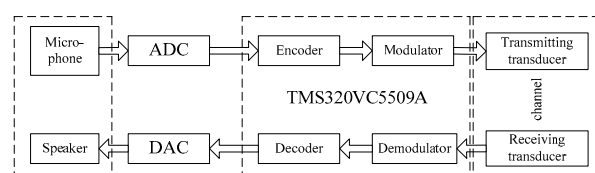


Figure 1. System Block Diagram

III. DESIGN OF 300BPS VOCODER

The enhanced mixed excitation linear prediction (MELPE) is an effective model for low bit rate speech coding. Based on this model, algorithms whose rates are 2400bps, 1200bps as well as 600bps has been proposed. In the MELPE model, band-pass voicing strengths coefficients (BPVC), line spectral frequency (LSF) parameters, pitch, short-term energy and fourier amplitudes of the residual are extracted in each speech frame. For our algorithm, the model extracts all parameters except the fourier amplitudes per frame (Frame length is 22.5ms, 180 sampling points, 8 kHz sampling frequency). The fourier amplitudes of the residual are usually set to 1 in the decoder part. The principle of MELPE encoder and decoder are shown in Figure 2 and Figure 3.

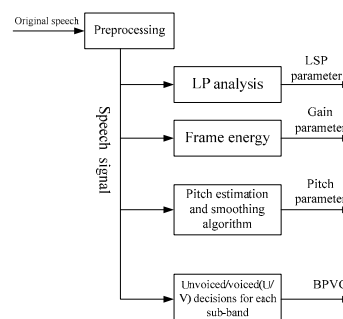


Figure 2. MELPE encoder

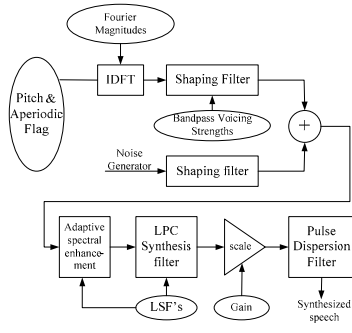


Figure 3. MELPE decoder

In our 300bps MELPE vocoder, the super-frame consists of 8 speech frames. Parameters are quantized in each super-frame. Each super-frame lasts 180 milliseconds and is quantized with 54 bits. The bit allocation is shown in Table 1.

TABLE I. 300BPS VOCODER BIT ALLOCATION SCHEME

Parameters	Bits
BPVC	5
pitch	10
gain parameter	8
LSF	31
total	54

In the algorithm, single-stage inter-frame prediction vector quantization is introduced into the gain and pitch quantization, while multi-stage vector quantization with both inter-frame and inter-stage prediction is used in the LSF quantization. The number of quantization bits for each stage is 10bits, 8bits, 7bits, 6bits respectively. The 300bps vocoder designed in this paper is of great intelligibility. DRT test can get up to 80% above.

IV. IMPLEMENTATION OF MODEM SYSTEM

In the underwater acoustic communication system, limited channel bandwidth and signal multi-path transmission can cause serious inter-symbol interference. Therefore, FH-4FSK technology which has strong anti-interference capability is introduced. In the transmitter, the carrier signal is modulated by the output of the encoder. Then rearrange the order of frequency according to the order of frequency hopping sequence. The signal rearranged is amplified through the power amplifier and is sent to the transmitting transducer for transmitting. In the receiver, the transducer receives the weak signal from water and sends it to the preamplifier circuit to amplify. Then demodulation system catches the frame synchronization from the received signal and de-hops the hopping sequence. The de-hopped signal is processed by FFT algorithm and demodulator obtains the original useful information. Finally, this useful information is sent to the decoder to decode.

V. HOPPING SEQUENCE

In the frequency hopping system, hopping sequence reflects the rule of the carrier frequency for both sides of the communication. There are several commonly frequency hopping sequences which are based on pseudo-random sequence such as m sequence, M sequence, RS code, etc. Among them, m sequence is the most widely used one in the frequency-hopping communication. So the m sequence is adopted in this system.

The m sequence is the short for the longest linear feedback shift register. The m sequence is generated by linear feedback shift register. The block diagram of the n-linear shift register is shown in Figure 4.

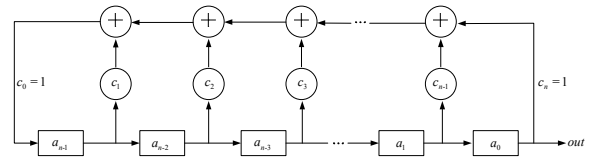


Figure 4. n-linear shift register block diagram

Linear feedback logical expression can be described as follows.

$$a_n = c_1 a_{n-1} \oplus c_2 a_{n-2} \oplus \dots \oplus c_n a_0 = \sum_{i=1}^n c_i a_{n-i} \pmod{2} \quad (1)$$

Move a_n to the right and substitute $a_n = c_0 a_n (c_0 = 1)$ into the above equation. The above equation can be rewritten as

$$0 = \sum_{i=0}^n c_i a_{n-i} \quad (2)$$

Define a polynomial corresponding to the above equation

$$F(x) = \sum_{i=0}^n c_i x^i \quad (3)$$

The polynomial is called as characteristic polynomial of the linear feedback shift register. The i represents the corresponding location of the elements. When $F(x)$ satisfies the following conditions, m sequence can be generated.

- (1) $F(x)$ is the most simple polynomial.
- (2) $F(x)$ can be divisible by $x^p + 1$ (where $p = 2^n - 1$).
- (3) $F(x)$ can not be divisible by $x^q + 1$ (where $q < p$).

The m sequences can be generated with shift register and the feedback structure which has simple structure, stable performance and fast synchronization. Furthermore, it can achieve longer period and has good randomness and hamming correlation. In the system, the entire spectrum of the acoustic transducer is divided into 16 frequencies. The m sequence is generated by 4-linear shift register. Every four frequency is divided into a group. As a result, every four Symbols can be regarded as a cycle corresponding to the FH sequence. Frequency hopping pattern is shown in Figure 5.

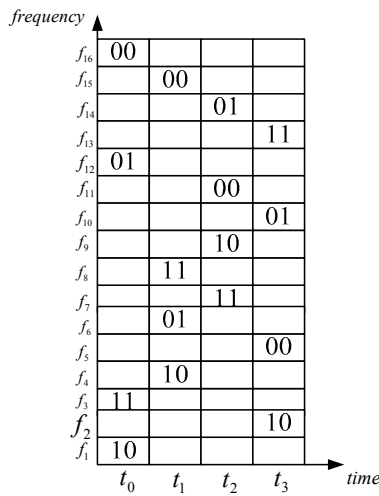


Figure 5. Frequency hopping pattern

VI. SYNCHRONIZATION

The system uses non-coherent demodulation. In order to improve the accuracy, coarse synchronization and accurate synchronization are both adopted. In the modulation side, each frame signal is consisted of the coarse synchronization and the accurate synchronization signal. The coarse synchronization adopts a fixed frequency to adjudge the start of frame. A linear frequency modulation (LFM) signal is used to accurate synchronization. In the receiver, the received signal is processed with FFT in the beginning. Then the demodulation system catches the coarse synchronization signal. If the coarse sync signal has been caught, the demodulation system continues to capture the accurate synchronization signal otherwise capture the coarse synchronization signal instead. A LFM signal which is the same as the one of transmitter is generated by the local. The demodulation system calculates the correlation of these two LFM signals and finds the accurate frame start. Finally, the demodulation system starts to de-hop the hopping sequence and demodulate the signals.

VII. FRAME STRUCTURE

The system adopts 300bps vocoder and 4FSK modem algorithm. In order to accurately locate the header of frame, coarse synchronization and accurate synchronization are both adopted. The accurate synchronization uses LFM signal so that the receiver can improve the accuracy of synchronization. Some protection symbols are inserted between sync symbols and data symbols in order to prevent the inter-symbol interference and the protection signal is Zero-frequency DC signal. The frame structure is shown in Figure 6.

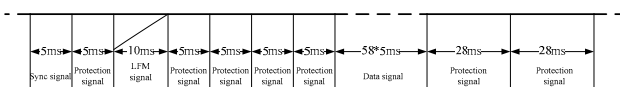


Figure 6. Frame Structure

VIII. HARDWARE IMPLEMENTATION

The system hardware consists of DSP module, memory module, audio codec module, power module, crystal oscillating circuit and reset circuit. The DSP module is the core module of the whole system, and it is responsible for processing and control algorithms. The memory module is composed of SDRAM and flash memory. The SDRAM provides space for program running and the flash memory can permanently store program in the chip. The audio codec module is responsible for the acquisition, playback, and ADC, DAC of the original speech signal. Power module provides stable and reliable power supply for the hardware platform. Crystal oscillator circuit provides stable and reliable clock signals to the entire hardware system. The role of the reset circuit is reset the whole system or restart the system. The system hardware architecture is shown in Figure 7.

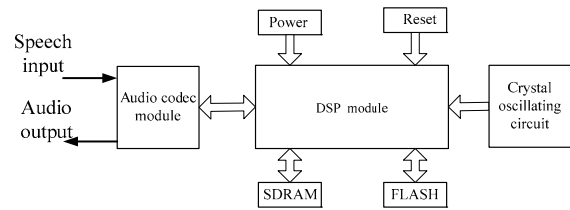
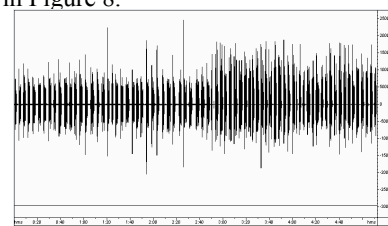


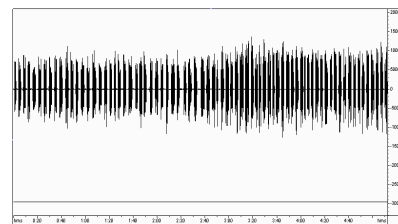
Figure 7. System hardware architecture

IX. SIMULATION RESULTS

The experiment is done in the pool. Standard speech material (8 kHz sampling, 16 bit quantization) is adopted in the experiment. The tested speech length lasts 5 minutes. Experiments show that the bit error rate of the system 10-2 and the received signal has great intelligibility. The results are shown in Figure 8.



(a)The input speech waveform of transmitter



(b) The output speech waveform of receiver

Figure 8. Analysis of experimental results

X. SUMMARIES

This paper presents a real-time underwater speech communication system. The source coding is based on a 300bps vocoder whose intelligibility is up to 80%. Modem part of the system employs FH-4FSK which has strong anti-interference ability. The signal processing chip adopts TMS320VC5509A. The simulation results show that the system is suitable for long-range underwater acoustic communication and has broad application prospects.

REFERENCES

- [1] Xu tianzeng, Xu lufen. Underwater Acoustic Digital Communication. Ocean Press, Beijing China, 2010.
- [2] NATO STANAG 4591. MIL-STD-3005 Analog-to-Digital Conversion of Voice by 2,400 bit/second Mixed Excitation Linear Prediction (MELP). The Hague: NATO, 1999.
- [3] Wang T, Koishida K, Cuperman V, et al. A 1200 bps coder based on MELP. IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP-2000. Istanbul, Turkey: IEEE Press, 2000. 1375~1378.
- [4] NATO STANAG 4591 ANNEX M Ratification Draft 1. MELPe variation for 600 bit/s NATO narrow band voice coder. The Hague: NATO, 2003.
- [5] Wang T, Koishida K, Cuperman V et al. A 1200/2400 bps coding suite based on MELP. IEEE Speech Coding Workshop. Proceedings 2002. Ibaraki, Japan: IEEE Press, 2002. 90~92.
- [6] Li ye et al. High quality speech coder at 300 bit/s. Audio Engineering, 2010,34 (12):51~53.