Journal of University of Science and Technology Beijing
Vol.8(2001), No.3, p.229

Information

High Performance Speech Compression System

Ke Liu, Zhichun Mu, Zhong Wang

Information Engineering School, University of Science & Technology Beijing, Beijing 100083, China
(Received 2000-12-10)

Abstract: Since Pulse Code Modulation emerged in 1937, digitized speech has experienced rapid development due to its outstanding voice quality, reliability, robustness and security in communication. But how to reduce channel width without loss of speech quality remains a crucial problem in speech coding theory. A new full-duplex digital speech communication system based on the Vocoder of AMBE-1000™ and microcontroller ATMEL 89C51 is introduced. It shows higher voice quality than current mobile phone system with only a quarter of channel width needed for the latter. The prospective areas in which the system can be applied include satellite communication, IP Phone, virtual meeting and the most important, defence industry.

Key words: digital signal processing; digital speech compression; digital communication; full-duplex; coding rate

1 Introduction

Previous speech compression algorithms such as LPC (Linear Predictive Coding), CELP (Code Excited Linear Predictive) and their variations can not work very well at low data rates, e.g. 9.6 kbps, either voice quality drops or channel error increases dramatically. However, in some cases such as military applications, data rate as low as 2.4 kbps or 3.2 kbps is necessary. To ensure voice quality at such low rate, proper speech compression algorithm has to be used.

The most commonly accepted speech model is [1]:

$$S_w(w)=H_w(w)E_w(w) \qquad (1)$$

where $S_w(w)$ is the Fourier transformation of the sample data after windowed; $H_w(w)$ the transfer function of the speech generation system; $E_w(w)$ the function of the stimulation. All the three are in frequency domain [2].

MBE (Multi-Band Excitation) algorithm differs with previous algorithms in the way of stimulation signal expression. It does not take a speech frame as either periodical signal or not. MBE separates the speech spectrum into several bands and make V/U (Voice/Unvoice) judgements separately. The overall stimulation is composed of the stimulations of all bands. Another difference is that MBE takes the advantage of the Masking Effect of human ear to pursue excellent perception quality rather than wave quality.

IMBE (Improved MBE) and AMBE (Advanced MBE) improved MBE by introducing analysis and synthesis of speech and new algorithm of quantized coding. AMBE is stronger in resisting both background noise and channel error than IMBE. It is currently the best solution available for low data rate. Comparisons among these algorithms are shown in **figure 1**.

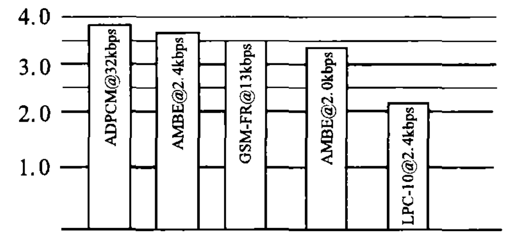

Figure 1 MOS of several algorithms.

The most widely accepted evaluation standard is MOS (Mean Opinion Score). Full score is 5. A score between 4.0–4.5 is called network quality, namely high quality. A score within the range near 3.5 is called communication quality. AMBE@2.4 kbps and GSM@13 kbps (current standard for mobile phone) are both within this range while the former is better [3]. Algorithm with MOS between 2.0 and 3.0 is acceptable for understanding but may be difficult to distinguish the speaker.

To realize speech processing there are three choices: mainly by software, mainly by hardware and the mixed. Hardware solution with AMBE-1000 as its core is adopted here due to reasonable price, high performance and reliability. This system is composed of two same PCBs (Printed Circuit Board), one at each side of the speakers, linked by wireless channel. Each side can be viewed as containing two parts: an encoder and a decoder. The encoder receives an 8 kHz sampled stream of data (16-bit linear) and outputs a stream of channel data at the desired rate. The decoder receives a stream of channel data and synthesizes a stream of speech data.

Their timing is asynchronous. The overview of the system structure is shown in **figure 2**.

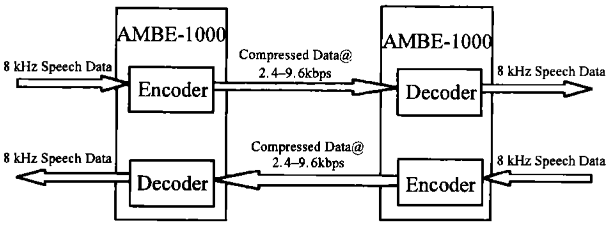

Figure 2 System structure.

2 Basic Operation of AMBE-1000 [4, 5]

2.1 Channel interface

AMBE-1000 has many combinations of working modes: parallel vs. serial; framed vs. unframed; passive vs. active. The most flexible and powerful mode is parallel framed passive mode, especially when a microcontroller is available. The overview of the channel interface is shown in **figure 3**.

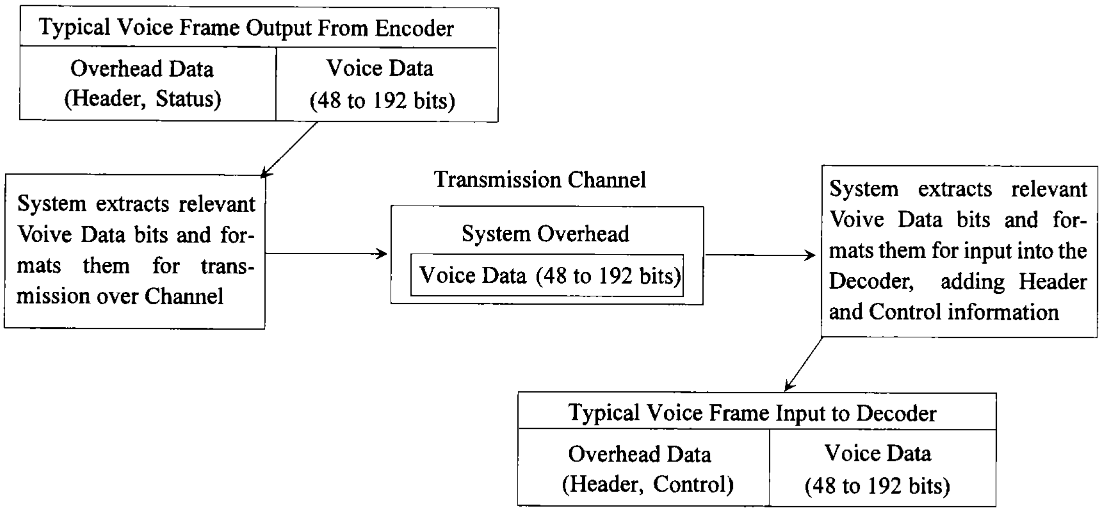

Figure 3 Channel interface overview.

2.2 Data frame

Both input frame and output frame consist of a 5-word overhead and 12-word data, totally 272 bits. READ or WRITE of a frame needs 20 ms. This implies that in parallel mode, in every 20 ms the hardware, connected to the AMBE-1000, has to perform READ or WRITE operations for 34 times, regardless of the voice coding rate. The parallel interface runs asynchronously to any clock and is controlled by ATMEL 89C51 under passive mode [6, 7].

2.3 Timing for READ and WRITE

In passive mode, the READ and WRITE control strobes are generated externally, in this system, by 89C51.

Timings for READ and WRITE in passive parallel mode are shown in **figures 4** and **5** respectively.

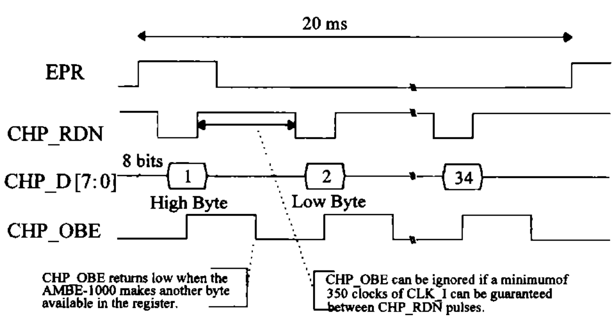

Figure 4 Timing for passive parallel mode READ.

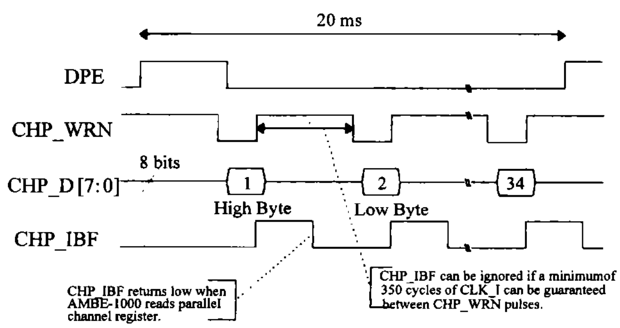

Figure 5 Timing for passive parallel mode WRITE.

2.4 Clock and reset

AMBE-1000 requires the clock frequency between 26 and 30 MHz. A 28 MHz oscillator is selected.

A valid reset signal must be active-low and last at least 6T (width of a clock pulse). It is generated by 89C51 with proper software.

3 Peripheral Chips

Some chips are used in the system to perform AD-

DA conversion, amplification, power rectification and monitoring, etc.

3.1 AD-DA conversion

TI TLC32046 is used to construct the analog interface circuit. It has a 14-bit resolution AD converter and a 14-bit resolution DA converter.

This chip also offers a powerful combination of options under DSP control. Three operating modes: dual-word (telephone mode), word and byte; two words format: 8-bit and 16-bit; timing: synchronous and asynchronous.

If the master clock frequency f_m is equal to 5.184 MHz, according to equation (2), the AD-DA conversion rate R will be 8 kHz, which fits AMBE-1000 very well such that auxiliary command word is not necessary.

$$R=\frac{f_m}{T(A)\times 2\times T(B)} \tag{2}$$

where channel parameters $T(A)=T(B)=18$ at reset.

However, only a 5.000 MHz oscillator is actually available. This gap results in a shift of 3-dB roll-off points and switched-capacitor frequency.The final AD-DA conversion rate is 8 kHz×(5.000/5.184) = 7.716 kHz.The working mode is detailed in **figure 6**.

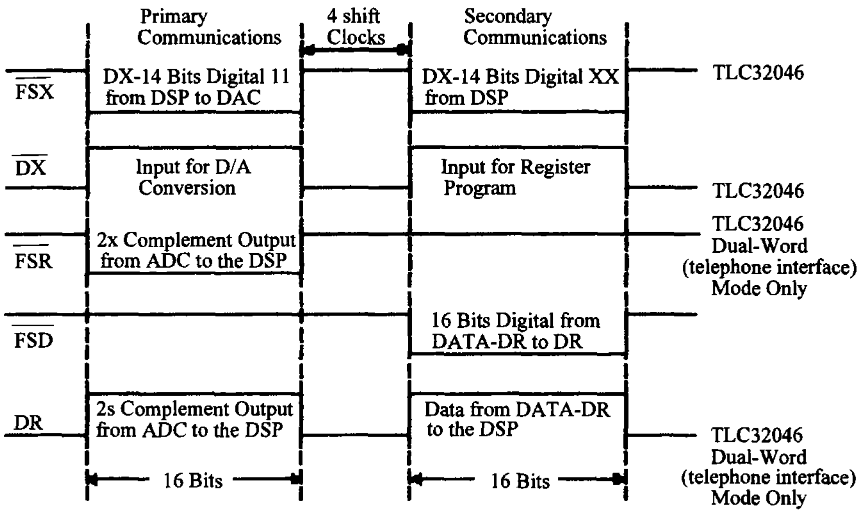

Figure 6 Working mode of AD-DA.

A combination of capacitors and diodes is used to eliminate the coupling effect between the analog and digital circuits, especially the two grounds. The circuit of AD-DA is shown in **figure 7**.

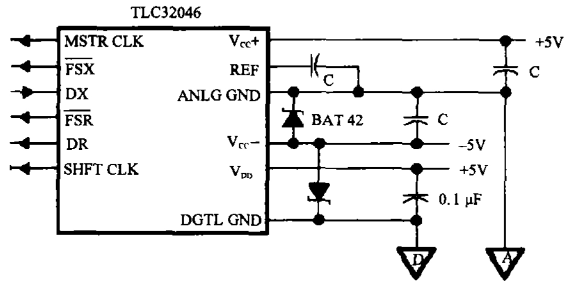

Figure 7 Typical circuit of AD-DA.

3.2 RS-232 interface

MAX232E cooperates with four 1μF capacitors to construct a standard RS-232 interface, acting as a bridge between the system and the wireless communication channel. MAX232E has two channels which can meet the demand of a board for sending and receiving data. Its data rate is up to 120 kbps.

The channel rate is 3.2 kbps,which is composed of 2.4 kbps coding rate and 0.8 kbps overhead information rate. The RS-232 interface channel rate varies from 2.4 kbps up to 120 kbps with every span of 2.4 kbps. It is finally set to 4.8 kbps.

3.3 μP supervisory circuit

The main purpose of using MAX813L is to perform the watchdog function. If not feed within 1.6 s, it will generate a reset signal to restart the system. Typical circuit is demonstrated in **figure 8**.

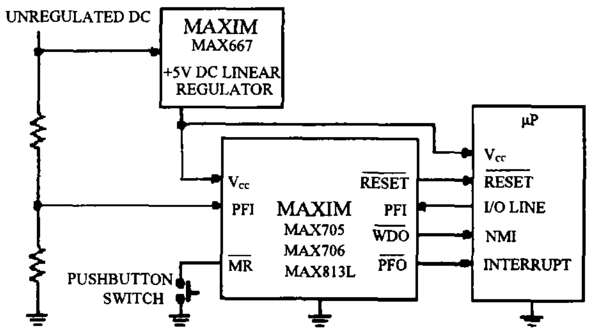

Figure 8 Typical operating circuit of MAX813L.

4 Software

4.1 Timing requirement

The microcontroller 89C51 must perform 34 READs and 34 WRITEs with AMBE-1000, feed dog, response to interrupt and send data to serial interface over each 20 ms. Each READ or WRITE requires 138 ns and the time between two READs is 350 ns. Feeding dog requires 2 ms. The time for interrupt processing and data sending should also be counted. 12 MHz frequency is just below the maximum. Considering the cooperation with serial interface whose clock frequency must be a multiple of 9, an 11.0592MHz clock source is used.

4.2 System allocation (Table 1)

The RAM is arranged as following:

0000H~0023H: Jumps;

from 0100H: Main program;

from 0200H: Subprogram of filling;

from 0300H: Subprogram of READ and WRITE with AMBE;

from 0400H: Subprogram of interrupt processing;

from 0500H: Data table.

Special data and registers:

02H: bit for judging if the serial input buffer is full;

R0: counter for writing to AMBE-1000;

R1: counter for reading from AMBE-1000.

Table 1 Allocation of buffer

Words / Bytes	Header, 5 words = 80bits.			Data, 12 words = 192bits.		
	1	···	5	6	···	17
Read	30H / 31H		38H / 39H	3AH / 3BH		50H / 51H
Write	53H / 54H		5BH / 5CH	5DH / 5EH		73H / 74H

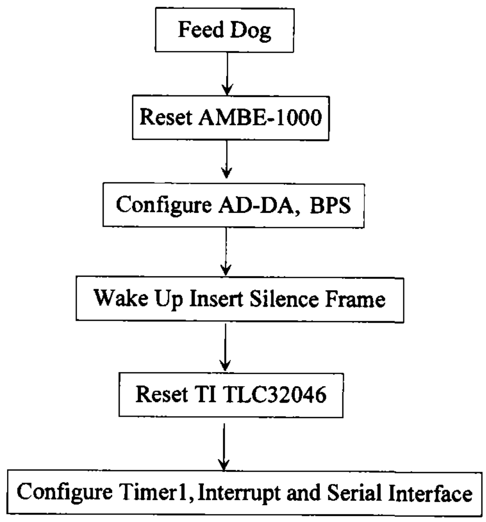

Figure 9 Program of RESET.

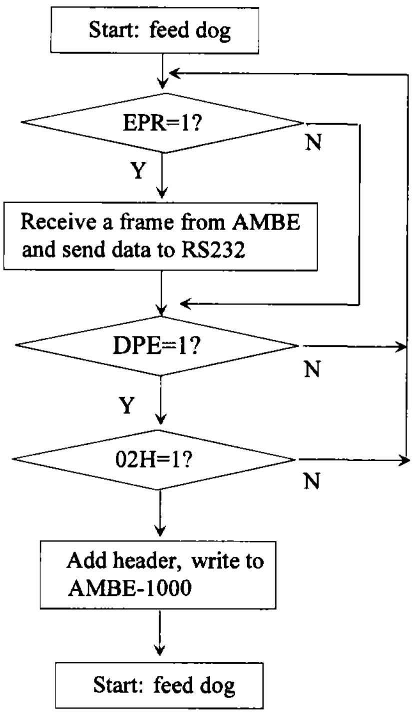

Figure 10 Main program.

4.3 Flow charts

The flow charts of RESET and main program are shown in **figures 9** and **10**, repsectively.

5 Block Diagram and Testing Result

5.1 Block diagram

Block diagram of this system is shown in **figure 11**. Apparently the system is symmetric.

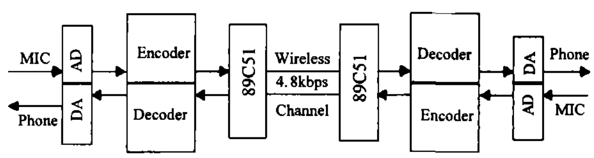

Figure 11 Block diagram of the system.

5.2 Testing result

The system is tested under different situations, shown in **table 2**. The error is generated by a special program running in microcontroller.

The system works very well at 2.4 kbps and 4.8 kbps, but the quality does not improve much above 4.8 kbps. This system is more suitable for low rate communication.

Table 2 Testing results

Data rate / kbps	Speech coding rate	FEC	Error rate	Reconstruction voice quality
2.4	2400	0	1/48	Speaker distinguishable
			2/48	Can not understand
	2350	50	1/48	Consistent
			2/48	Speaker undistinguishable
4.8	4800	0	1/96	Almost the same with original
			3/96	Somewhat worse
			5/96	Understandable
			10/96	Can not understand
	3600	1200	<4/96	Almost the same with original
			7/96	Understandable
			8/96	Can not understand
			9/96	Inconsistent
	2550	2250	<5/96	Almost the same with original
			8/96	Understandable
			9/96	Inconsistent
			10/96	No voice at all
9.6	9600	0	2/192	Almost the same with original
			10/192	Can not understand
			19/192	Can not understand
	4850	4750	2/192	Almost the same with original
			10/192	Somewhat worse
			19/192	Understandable

References

[1] X. Yang, H. Chi: *Digital Processing of Speech Signal.* Press of Electronics Industry, Beijing, 1995.

[2] P. Cheng: *Digital Signal Processing.* Press of Tshinghua University, Beijing, 1995.

[3] F. Li: *Journal of Post and Telecommunication University of Nanjing,* (1998), p. 186.

[4] *AMBE-1000 Vocoder Chip User's Manual.* Version 3.0, Digital Voice System Inc., 1997.

[5] *New DVSI AMBE+ Vocoder.* Digital Voice System Inc., 1999. 6.

[6] H. Li: *Manual of Interface of Serry 51 Single-Chip Microcomputer.* Press of Aeronautics and Aerospace University of Beijing, 1993.

[7] J.Fang: *Speech Synthesis Techniques and Single-Chip Microcomputer Integrated Systems.* Press of Aeronautics and Aerospace University of Beijing, 1993.