

## ARABIC TEXT-TO-SPEECH : SINGLE BOARD

M. Ouadou, A. Rajouani, M. Najim, M. Zyoute and P. Baylou<sup>\*</sup>

### ABSTRACT

The purpose of this paper is to describe hardware realization of a real time Arabic text-to-speech system based on a diphone dictionary and LPC technique. The board consists of two processors : a general purpose processor for the control of operations and a speech synthesizer processor. The board is connected to the host computer through an RS-232 port. The whole synthesis system allows the conversion of any typed Arabic text into intelligible speech in real time.

### INTRODUCTION

Of the many systems that generate synthetic speech, all use either time-domain or frequency-domain algorithms. The former afford the fastest data rate - from 64 Kbits/s to several thousand a second - and thus the best production. The latter, operating at several hundred bits a second, generates good, intelligible sound and holds out the facility of unlimited vocabularies. However the technology evolution of semiconductor memories permits the use of a high capacity low cost memories. And the realization of a low cost time-domain unlimited vocabulary systems became a reality. The linear predictive coding (LPC) synthesis technique - used in our case - is a scheme that encodes parameters of a fixed mathematical mode, which defines the shape of the vocal tract at fixed analysis intervals. Typically the speech signal is analyzed to extract the parameters of the vocal tract model and the excitation source ones. The excitation source is modeled such as two generators, one for voiced sound, the other for unvoiced sounds. The synthesis filter is usually a 10-pole lattice filter, referred to as an LPC-10 filter. The filter bank simulates the contractions and expansions of the vocal tract. LPC work as slowly as 1200 bits/s, through most systems usually run at 2400 bits/s.

The study and realization of Arabic LPC diphone dictionary has been accomplished with 16 bits coded values and 14 parameters for each frame (energy, pitch and 12 reflection coefficients) /1//2/. The dictionary has been obtained by 10 KHz sampling frequency, 250 samples analysis window and 100 sample frames.

All parameters are recoded to create a new dictionary for our specific case

### HARDWARE DESCRIPTION

The system functional block diagram is shown in Fig.1. It consists of two processors : one for the control tasks, the other one for speech synthesis.

The 64 Kbytes memory address range of the 8 bits control processor is divided into :

LEESA, Faculté des Sciences, B.P. 1014 Rabat, MAROC

\* ENSERB, 351 Cours de la Libération, F33405 Talence, FRANCE.

1. 32 Kbytes of RAM as job memory;
2. 32 Kbytes of ROM containing the system firmware including :
  - interrupt routines;
  - grapheme-to-phoneme transform routine;
  - address calculation routine;
  - read routine from 192 Kbytes Arabic diphone's ROM;
  - prosodic routine;
  - speech synthesis control routine.

Speech synthesis processor : The TI TMS5220C digital voice synthesis processor (VSP) /3/ was selected for the realization of the single-board hardware. The TMS5220C is a linear predictive coding (LPC) speech synthesis device enabling verbal communication with a microcomputer based system. It has been designed to minimize the data rate required to produce intelligible speech and to simplify the interface with the host CPU.

Speech data that has been compressed using pitch-excited LPC, is supplied to the VSP by the CPU. The VSP decodes this data to simulate a time-varying digital filter model of the vocal tract. This model is excited with a digital representation of either glottal air impulses (voiced sounds) or the rush of air (unvoiced sounds). The output of this model is converted by an eight bit digital-to-analog converter to produce a synthetic speech waveform.

The CPU serves the device by responding to interrupt service requests generated by the VSP. A simplified block diagram of the VSP is shown in fig.2.

Coded speech parameter data is fed serially from the FIFO to the parameter input register, then the controller unpacks the data and performs various tests (i.e. is the repeat bit set, is pitch zero, is energy zero). The unpacked coded parameter data is stored in RAM to be used as the index value to select the appropriate value from the parameter Look-Up ROM. The outputs of this ROM are the target values for the interpolation logic to reach in this frame period. During each of the eight interpolation periods the interpolation logic sends new pitch and energy parameters to the signal generator which produces the filter excitation sequence, and new K-parameter values to the LPC lattice network. So, at the end of each sample period there is a new value of digitized synthetic speech available to the D/A converter.

Each of 12 synthesis parameters (pitch, energy and reflection coefficients K1-K10) occupies between 3-6 bits. These coded values select a 10-bit actual parameter from the parameter Look-Up table. Table 1 summarizes parameter coding for the TMS5220C.

Table 1. Parameter coding for the TMS5220C.

PARAMETER	LEVELS	CODE BITS
ENERGY	15	4
PITCH	64	6
K1-K2	32	5
K3-K7	16	4
K8-K10	8	3

A full set of coded parameters with 100 samples/frame would require a data rate  $100 \text{ Hz} \times 50 \text{ bit} = 5 \text{ Kbits/s}$ . However by using the repeat bit, only 4

parameters for unvoiced frame and no parameters when energy equal to zero the data rate is reduced to 3Kbits/s.

Due to the slow device characteristics of the VSP relative to the MP bus cycle time, the peripheral device PIA is used to control operations between MP and VSP. The interrupt signal of VSP activates the interrupt signal of the PIA and the MP respond to the interrupt of the PIA as the interrupt of VSP device.

The output of the VSP is fed through an analog low-pass filter with cut-off frequency at 5KHz to cancel high frequency due to the A/D converter. The output of the filter is then fed to analog amplifier.

Host serial interface : The interface that provides for communication with the host is connected to system bus as shown in fig.1. The 3.6 MHz system clock of the control processor is divided by means of a frequency divider circuit to provide the UART with its receiver and transmitter clocks. Eight data rates select switches are provided to adjust the UART data rate from 300 to 9600 Bd. For most applications, a 9600 Bd rate is used to minimize downloading time. Due to the slow device characteristics of the UART relative to the MP bus cycle time, it is necessary to insert TTL latches for both input and output parallel data as well as for the UART status bits between the MP and the UART.

The CPU serves the device interruption according to the priority level of the device in the case of multiple interrupt at the same time.

#### SYSTEM FIRMWARE DESCRIPTION

Sequence of operations: While the control processor begin to receive Arabic text from a host system, the grapheme-to-phoneme transform routine is activated at the same time to ensure real time synthesis. When the first sentence is transformed to the diphone elements with there addresses on the dictionary ROM, the read routine - from this ROM - loads parameters of each diphone in the RAM memory. At the same time the processor continue to receive the remainder part of the text through UART interrupt request. The prosodic routine makes changes on pitch values for each frame according to the type of synthesized sentence and then activate TMS5220 interrupt request which controls synthesis process and transmission of parameters between the CPU and VSP. On the time out of interrupt routine the CPU continues to activate for each sentence the same routine until the end of the text.

#### CONCLUSION

The intelligibility of the synthetic speech is highly accepted. To avoid the limitations due to the RS-232 connection including a separate box and a power supply, a new board compatible with PC is being realized basing on a different architecture and using the TI50C42 speech synthesis processor . The new system will be a multilanguage speech synthesis system.

#### REFERENCES

- /1/ A. Mouradi, M. Najim and A. Rajouani, Proc. PWSPA, Porto, B1/4/1(1982)
- /2/ A. Mouradi, A. Rajouani and M. Najim, Proc. 4th Int. Conf. Digit. Sign. Proc. Signals in Communications, Loughborough, 329,(1985).
- /3/ T.I. Inc, TMS5220C Voice Synthesis Processor Data Manual, Houston (1984).

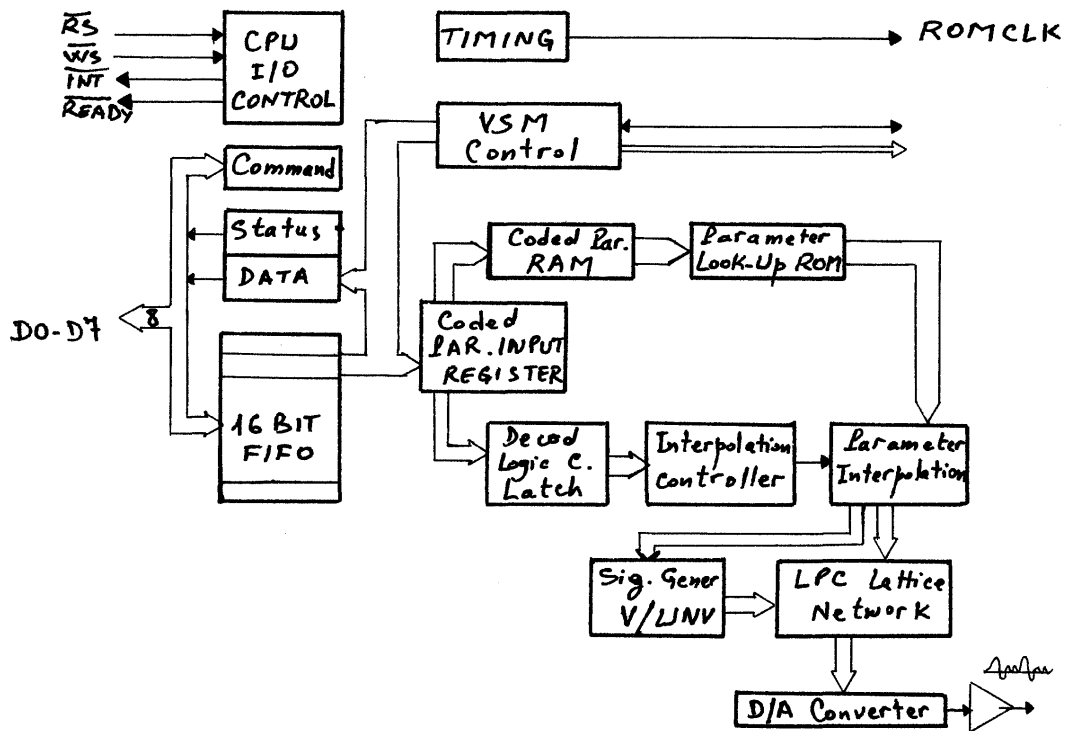


Fig.2. Voice synthesis processor block diagram.

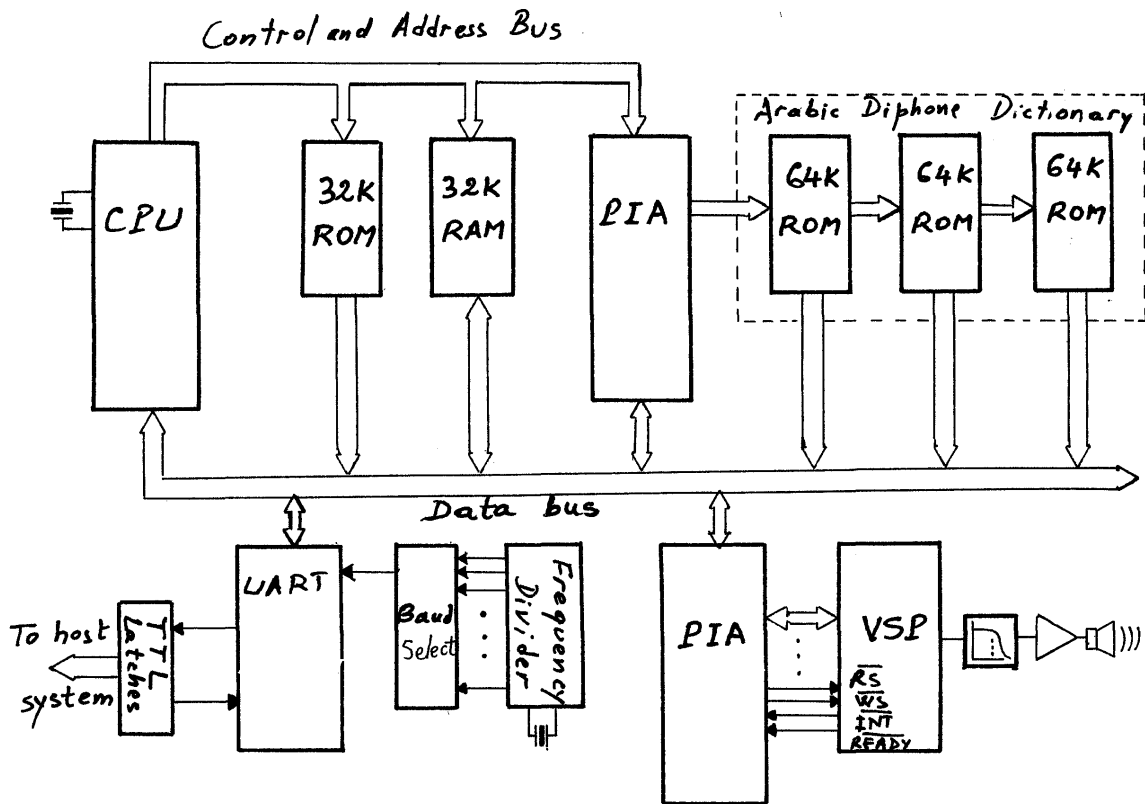


Fig.1. System functional block diagram.