AN IBM-PC BASED SPEECH RESEARCH WORKSTATION

Michael Wagner(*) and John Fulcher(**)

(*) Department of Computer Science
University College, ADFA
University of New South Wales

(**) Department of Computing Science
University of Wollongong

ABSTRACT - A microcomputer-based speech research workstation with analog-to-digital and digital-to-analog conversion hardware and software, a specially designed antialias filter and software packages for speech editing, homomorphic analysis and linear predictive analysis and synthesis is described in this paper.

INTRODUCTION

Most speech research laboratories require a workstation with certain basic facilities, irrespective of whether the particular area of research is speech analysis, speech synthesis, automatic speech recognition, speech perception, speech coding or any of the other current topics in speech research.

These requirements generally comprise digitisation of speech signals at rates of least 10,000 samples/s and at a precision of at least 12 bits/sample; high-quality anti-alias filtering; large-capacity secondary storage for storing large amounts of speech data; fast numerical processing for computation-intensive analysis methods, e.g. Fourier transforms; and medium- to high-resolution graphics capability to display time and frequency representations of speech signals.

In recent years, the developments in the area of microcomputers have brought the necessary basic tools for speech research well within the reach of those who are constrained by a relatively small budget. This paper describes a speech research workstation which is based on an IBM PC-AT personal computer. The analog-to-digital and digital-to-analog conversion capability is provided by a commercially available signal interface adapter for the IBM PC while a high-quality anti-alias filter has been designed and manufactured in-house at a fraction of the cost of commercially available filters.

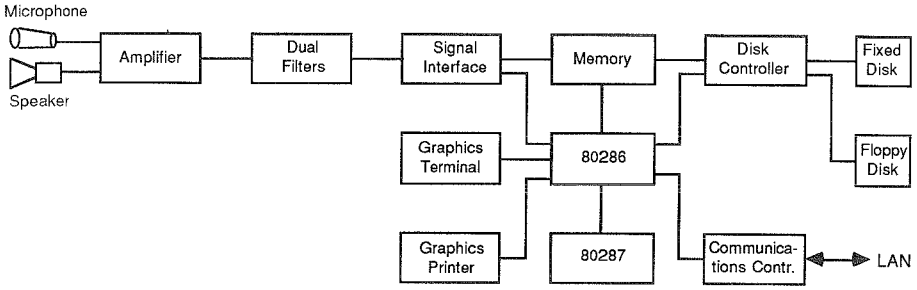
While the availability of a reasonable program development environment has been a problem for microcomputer users for many years it seems that this problem is gradually being remedied. Fortran, Pascal and C compilers are now available for PC-DOS and software for the basic functions of speech digitisation, homomorphic speech analysis and linear predictive speech analysis and synthesis has been developed using Fortran, C, and assembly language.

HARDWARE CONFIGURATION

The speech research workstation is centred around an IBM PC-AT personal computer and is currently configured as shown in Figure 1.

The Intel 80286 processor is currently used only in its non-protected mode. An 80287 mathematical coprocessor provides fast parallel floating-point

operations at 80-bit precision. The system is configured with 640Kbytes of memory, one 1.2Mbyte floppy-disk unit and a 20Mbyte hard disk. The workstation provides for monochrome graphics at a resolution of 640x350 pixels using the IBM Extended Graphics Adapter (EGA). An IBM Proprinter provides hardcopy at a rate of 100 characters/s. Through the Communications Adapter the workstation is connected to a Local Area Network and is able to access a variety of larger machines for speech data processing.



Memory and secondary storage capacities are of particular interest to speech research as they are directly related to the amount of speech that can be digitised or played back by the system. The comparative storage times for digitised speech signals are shown in comparison in Table 1.

Memory	Floppy disk	Hard disk
640Kbytes	1.2Mbytes	20Mbytes
0:32 min	1:01 min	17:45 min

Table 1. Storage capacity in minutes of speech, digitised at 10,000 samples/s and 2 bytes/sample.

Using the 80286 in protected mode, the system could be extended up to a maximum capacity of 1Gbyte of virtual memory mapped onto a physical memory space of 16Mbytes.

ANALOG-TO-DIGITAL AND DIGITAL-TO-ANALOG CONVERSION

The main advantage of building a speech research laboratory around a popular personal computer is the fact that not only the computer itself is much less costly but that all the supporting hardware and software costs very much less than that of a minicomputer system, often by an order of magnitude or more. Accordingly, the signal interface for the system is provided by the Data Translation DT2801-A Analog and Digital I/O System, a board which is available on the personal computer market and fits into 1 full slot in the IBM PC.

Analog-to-digital conversion can be done at speeds of up to 27,500 samples/s at 12 bits/sample. Up to 8 differential or 16 single-ended channels may be used for A/D conversion at a throughput rate which is correspondingly smaller. The A/D converter gain is programmable for values of 1, 2, 4 and 8. Digital-to-analog conversion is provided at a maximum speed of 33,000 samples/s. 2 D/A channels may be used simultaneously at

16,500 samples/s on each channel. The sample resolution for D/A conversion is also 12 bits/sample. Both A/D and D/A conversion are triggered by an on-board clock with a frequency of 800kHz which must be divided down to the required conversion frequency.

The DT2801-A board also provides for 16 channels of digital input/output which can be usefully employed for the prompting of speakers or listeners, for push-button responses etc.

For the connection of the analog inputs and outputs and the digital outputs to the signal interface board, a connector box has been constructed which provides 2 RCA sockets for A/D channels 0 and 1 and another 2 RCA sockets for D/A channels 0 and 1. These channels can therefore be connected immediately to most audio equipment by way of standard audio cables. The digital outputs D0 - D3 of the signal interface board drive an array of 4 small relays in the connector box each of which is connected to a pair of banana sockets. Prompting lights or other low-power devices can thus be conveniently connected to the 4 digital outputs.

ANTI-ALIAS FILTER

A high-quality anti-alias filter is one of the essential requirements of a speech research laboratory. Commercially available filters which typically feature a 24dB to 96dB per octave roll-off, user-selectable gain and tuning over a large frequency range are comparable in price with the personal computer itself and therefore present a major problem to the researcher with a somehow constrained budget. We therefore decided to design and build the anti-alias filter in-house according to the following specifications:

a) As very few different sampling frequencies are used for speech digitisation, the low-pass filter has a fixed cut-off frequency;

b) Since the gain can usually be set satisfactorily by the audio amplifier, the filter has a fixed gain of 1;

c) The filter is designed as an 8th-order Butterworth filter with a maximally flat passband and 48dB per octave roll-off;

d) The filter is realised in form of a printed circuit board with pins for SIGNAL IN, SIGNAL OUT, +12V IN, -12V IN and GROUND;

e) Filter boards are housed in a box which provides 2 card-edge sockets for easy replacement of filter boards, 4 RCA sockets for the input and output signals of 2 filters and power for 2 filters.

The 8th-order filter is realised by cascading 4 second-order stages each of which is implemented as a biquad circuit as shown in Figure 2. The biquad circuit has excellent stability and tuning features at the cost of using 3 operational amplifiers for each 2nd-order stage (Johnson, 1976).

The transfer function of the biquad circuit in Figure 2 is

$$H(s) = b / (s^2 + a*s + b) \quad (1)$$

with

$$a = 1 / (R2 * C1) \quad (2)$$

and

$$b = 1 / (R1^2 * C1 * C2) \quad (3)$$

For a second-order stage of the Butterworth filter

$$a = -2\,s\,|s_i|\,\cos\vartheta_i \qquad (4)$$

and

$$b = |s|^2 \qquad (5)$$

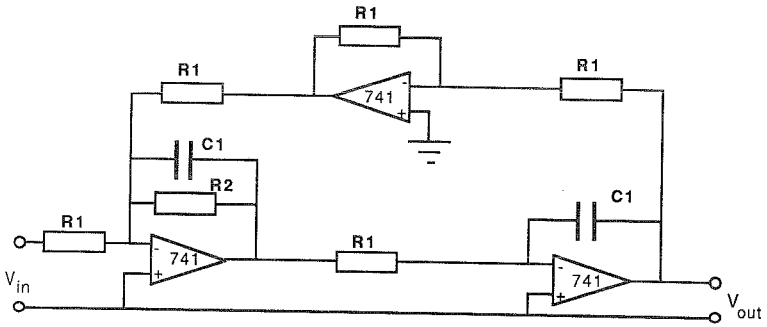
where

$$|s_i| = 2\pi f_{cut} \qquad (6)$$

and

$$\vartheta_i = 191.25^\circ;\ 213.75^\circ;\ 236.25^\circ;\ 258.75^\circ \qquad (7)$$

are the magnitudes and phases of the Butterworth poles for the 4 stages (Oppenheim and Schafer 1975).



The component values for each biquad were determined by

a) selecting 2 available capacitors C1 and C2, each of a nominal capacitance of 10nF;

b) measuring the actual capacitances of C1 and C2;

c) computing the resistance values R1 and R2 from (2-5).

Since the filter is used for audio frequencies only, the standard uA741 operational amplifier is used in conjunction with standard capacitors (5% tolerance) and resistors (1% tolerance). The resulting 8th-order lowpass filter has been built at a very much lower cost than that of similar commercial filters.

DIGITISATION SUPPORT SOFTWARE

One of the most tedious tasks in a speech research laboratory is the digitisation of speech data, either directly from a microphone or, more often, from endless reels of prerecorded audio tape. It is therefore essential that support software for the digitisation process is optimised for convenient, efficient and effective operation on long stretches of speech data.

All the programs described here use the direct-memory-access facility of the Data Translation board to convert speech data to or from one half of a large memory buffer while simultaneously transferring the other half of that buffer to or from the hard disk. This means that the duration of the longest digitisable signal is limited by the size of the hard disk only.

Programs ADC and ADCSH

These programs are written in assembly language and create speech files up to 20Mbytes from speech signals which are digitised at up to 27,500 samples/s. The user interface allows the selection of file name and size, sampling frequency and amplifier gain. Program ADCSH allows the user to effectively edit the input signal as disk blocks are only written while the user depresses the SHIFT key on the PC keyboard.

Program DAC

DAC plays disk files back to the user through the digital-to-analog converter. The user interface allows the selection of file name, sampling frequency and the start and end blocks of the signal to be converted.

Programs EZSTAT and CUT0

This pair of programs provides automatic segmentation of digitised speech into periods of speech and periods of silence (Rabiner and Sambur, 1975). Program EZSTAT looks at the first few seconds of the signal to compute the signal mean (DC level), the energy extrema and the distribution of zero crossings. Program CUT0 then uses the EZSTAT information to automatically segment the entire speech file into speech and silence segments. Both programs work in approximately real time. Thus EZSTAT takes about 5 seconds to compute the necessary statistics and CUT0 segments a 5- minute speech file in approximately 5 minutes. Programs EZSTAT and CUT0 are essentially error-free for speech recorded in a quiet environment.

Programs CUT1 and CUT2

Program CUT1 allows the user to define a segmentation of a speech file by repeatedly listening to precisely specified 512-byte blocks of the speech file. Program CUT2 uses the segment information from either CUT0 or CUT1 to edit required segments out of a given speech file.

Homomorphic Analysis

Program HOM performs a complete homomorphic analysis on a digitised speech file (Oppenheim and Schafer, 1968). The user interface provides for the selection of the file names for the speech input, spectrum output, cepstrum output and smooth spectrum output files. The user is further requested to input the range of input file blocks to be processed, the width and shift of the Hamming window, the preemphasis (0.0-1.0) and the cepstrum cutoff quefrency. The program is mainly written in Fortran with 2 subroutines for the Hamming window computation and the fast Fourier transfom coded in assembly language for maximum execution speed.

Linear Predictive Analysis

Program LPA performs linear prediction analysis by means of the autocorrelation method on a digitised speech file (Markel and Gray, 1976). The user selects the file names for the speech input, LPC output, spectrum output and error signal output files. Similarly to HOM, the user enters the range of input blocks, width and shift of the Hamming window, preemphasis and the predictor order (0-20).

For each input frame the linear prediction polynomial and the corresponding

signal spectrum are computed. The non-preemphasised input signal is then inverse-filtered to obtain a filtered error signal. Autocorrelation analysis on this error signal finally yields estimates for the voicing parameter and the fundamental frequency of the signal. While the major part of LPA is written in Fortran, the routines WINDOW (multiplication by a Hamming window), CORR (computation of autocorrelation coefficients), LPC (computation of the linear predictor) and FFT (fast Fourier transform) are coded in assembly language for maximum speed.

Linear Predictive Synthesis and Parameter Manipulation

The linear prediction synthesis program LPS takes an LPC file as created by program LPA as its input. The user is requested to provide the voicing threshold for the voiced/unvoiced decision, and the gain and dc level for the output signal. For voiced frames, a sequence of single spikes with amplitude and frequency according to the excitation energy and fundamental frequency as determined by LPA is used as the signal source. For unvoiced frames, a noise source is simulated by a pseudo-random number generator. The source signal is filtered using the LPC coefficients as determined by LPA.

The program CHGPRM permits interactive manipulation of both source and filter parameters of an LPC file. The modified parameters can then be synthesised using program LPC (Barlow and Wagner, 1986).

CONCLUSION

A low-cost speech research workstation with full facilities for speech digitisation, speech analysis and resynthesis has been described. A high-quality low-cost anti-alias filter has been designed and constructed to fit the particular needs of the speech research workstation. A software package has been developed for convenient, effective and efficient speech data collection and for speech analysis, parameter manipulation and resynthesis.

REFERENCES

BARLOW, M. & WAGNER, M. (1986) "Acoustic Correlates of Speaker Characteristics", Proc. 1st Aust. Speech Sc. Techn. Conf.

JOHNSON, D.E. (1976) "Introduction to Filter Theory" (Prentice Hall, Englewood Cliffs).

OPPENHEIM, A.V. & SCHAFER, R.W. (1968) "Homomorphic Analysis of Speech", IEEE Trans. Audio Electroacoust., AU-16, 221-226.

OPPENHEIM, A.V. & SCHAFER, R.W. (1975) "Digital Signal Processing" (Prentice Hall, Englewood Cliffs).

MARKEL, J.D. & GRAY, A.H. JR. (1976) "Linear Prediction of Speech" (Springer Verlag, Berlin).

RABINER, L.R. & SAMBUR, M.R. (1975) "An Algorithm for Determining the Endpoints of Isolated Utterances", Bell Syst. Tech. J., 54, 297-315.