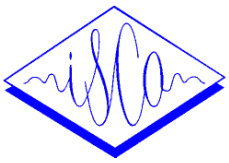

2nd International Conference on
Spoken Language Processing (ICSLP 92)
Banff, Alberta, Canada
October 12-16, 1992

A PC GRAPHIC TOOL FOR SPEECH RESEARCH BASED ON A DSP BOARD

Miguel A. Berrojo, Javier Corrales, Jesús Macías and Santiago Aguilera

Departamento de Ingeniería Electrónica. E.T.S.I. Telecomunicación
Universidad Politécnica de Madrid
Ciudad Universitaria s/n 28040 Madrid, Spain

ABSTRACT.

We introduce a speech analysis system that performs the complete recording-playback interface, and allows some edition facilities. The system can display parameters such as energy, zero-crossing rate, fundamental frequency (pitch contour), spectrogram and LPC envelope. Another capabilities are filtering of the pre-recorded speech, labelling of the different acoustic segments and the possibility of laser screen hardcopies.

The system needs a PC and some specific hardware. The PC must be AT or later, with a VGA graphic card and a serial mouse. The specific hardware is a DSP based board which was completely developed in the Department. The software is divided in two blocks: the software running on the iX86 processor, which controls the system and displays the speech information at screen, and the program running on the DSP, that performs the signal processing, the recording and the playback.

Since the system is oriented to not technical people (phoneticians, etc.) we focused on the user interface to make it as easy and friendly as possible. The user can configure the windows in the screen in a wide manner using a keyboard or the mouse driven menu system.

This speech analysis device takes advantage of all the standard PC features (video monitor, hard disk, printer, memory, etc.), what results in the lower cost compared with any other analysis tool in the market.

I. PREVIOUS WORKS

When we started to work in this field some years ago, as any speech research laboratory, we felt the need of an analysis tool. In those days, the best option was to develop our own system, which we could fit our needings with.

The first complete recording-analysis system was developed on a VAX computer [1], using a graphic TEKTRONIX terminal and a DSC analog interface. That system became obsolete as we needed something portable and versatile.

Based on this experience and the speech parametrization system for hearing despaired rehabilitation [1], we decided to implement a practical and powerful tool that could work on the widely extended IBM PC. Such a tool is what we present in this paper.

II. SYSTEM DESCRIPTION

This is a versatile window environment, in which user can choose windows position and size and what parameters and how many of them will be represented in every screen window. Each window has fixed width and variable height: the user can select any value in the vertical dimension except for sonogram windows in which this height is restricted to two different values, normal size and compressed. User can manage a maximum of four non overlapping windows in the screen. This number depends on their relative sizes.

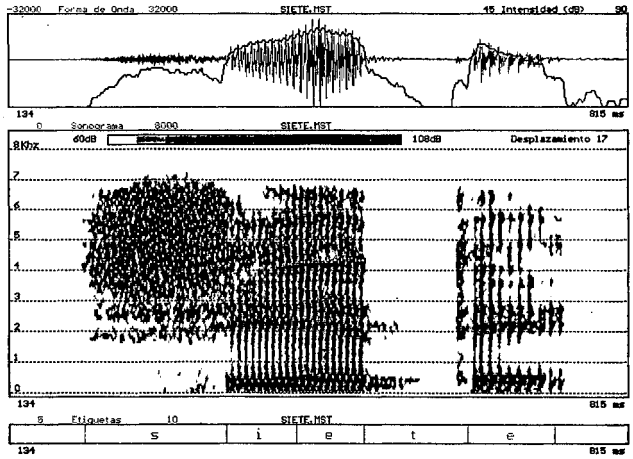

Fig. 1 Hardcopy that shows the waveform, the energy and the wideband spectrogram of the word "siete".

In order to easily relate different parameter sets, we allow to represent some of them in the same window when they belong to the same domain. The maximum number of parameters over each window is two, and this is restricted by the medium resolution of the VGA graphic card. User can change the scale of both x and y axes or use a default set. Time-domain windows can be synchronized, and shifted to the desired instant of time, simply, by pointing to the chosen position, or using the keyboard; if this point is not visible at screen, we can find it by moving the marker to the window extremes: every time we do it, the window will be automatically shifted a third part of its wide. All these operations are made in a graphic way, using the PC mouse. Of course, once that a screen configuration is completed, it can be saved on the hard disk, and this way the user has a set of predetermined window configurations for later use.

10.21437/ICSLP.1992-406

Everything can be managed from a, sometimes redundant, hierarchical menu environment where any needed option is always available. The keyboard and the mouse can both drive the selection and the adequate option search. Related help is always available too.

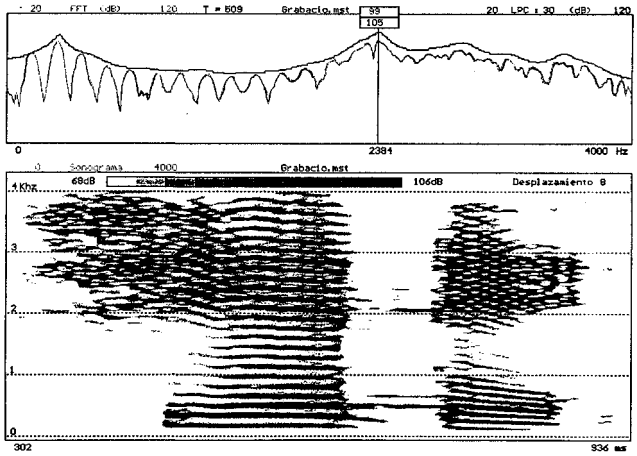

Fig. 2 Hardcopy that shows the FFT spectrum, the LPC envelope and the narrowband spectrogram of the word "siete"

The set of parameters that you can choose are: Intensity-Energy, Zero-crossing rate, FFT spectrum, Sonogram, Pitch contour and LPC envelope. When computing the FFT spectrum or sonogram, user can select the window (Hamming, Hanning or Squared), the pre-emphasis coefficient and the analysis filter width. The Sonogram is painted in both gray scale or thermic scale, and the number of LPC coefficients used to compute the LPC envelope can be adjusted between three and thirty.

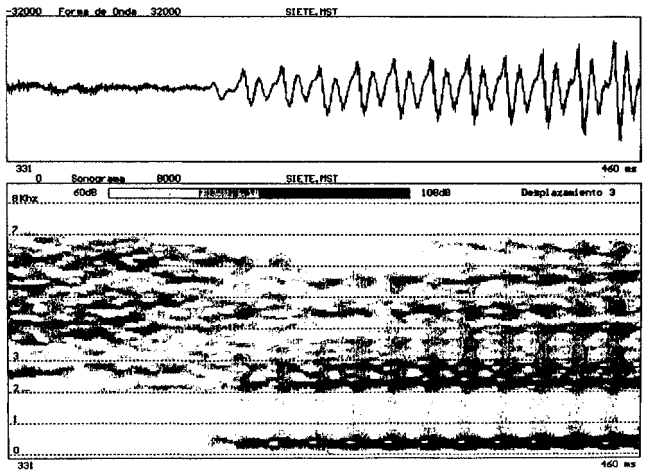

Fig. 3 Hardcopy with a zoomed zone of the same speech file.

Other possibility is the phonetic labelling of speech segments, by moving synchronized markers. The user can always hear the segments between these markers to help this segmentation. The segments and labels are saved in an ASCII file (one per speech file), so user has fast access to this information and can use it with other programs. The phonetic alphabet is also saved in an ASCII file, allowing creation of alphabets, modification of symbols, etc. The default is the standard CPA (Computer Phonetic Alphabet).

In the recording menu, input sound source (microphone or line), record duration and sampling frequency (between 8 and 16 Khz) can be selected. Due to memory restrictions, the maximum duration to be selected is 8 seconds when recording at 16 Khz. The signal envelope is displayed to guide the speaker and to avoid low level or saturation.

The system includes the File Segmentation function to make easier the acquisition of data bases. This function extract speech files from a recording file that includes some utterances, words, sentences, etc., creating these files with the names given by a list file. This file extraction utility is useful when the whole session was recorded in an audiotape system.

The edition facilities we mentioned above are basically cut and paste functions, where user can insert another speech file in the current one, etc., using the markers.

To analyze singular values of a parameter set, the program provides a special marker that displays the values of x and y axes: every time you place the graphic cursor inside a window, the corresponding x value appears, using lines, in all the windows with the same x-axis. This x value and the corresponding y value are displayed too. When moving the mouse along the window, values are actualized, automatically. User also can use this marker to choose the time instant where the window is centered for FFT and LPC analysis.

Filtering is performed, when selected, on the segment between markers. The kind of filtering is pre-emphasis, de-emphasis (both coefficients can be selected) and user defined FIR filtering (designed with a different program supplied with this software).

Most of these options can be selected in a configuration menu that includes input and output directories, default file names, the type of analysis to be performed and some filter coefficients and window sizes.

You can obtain screen hardcopies on a HP LaserJet printer. This is a black and white printer, so colors are coded in gray levels. This is a slightly slow process. So, a growing bar shows the percentage of conversion made, which can be interrupted by a key stroke if necessary.

The system stores all the information in two different files: a binary one, ILS format speech file and an ASCII one, containing information about the ILS one such as sampling frequency, number of samples in the file, labels, graphemic representation, speaker characteristics, etc.

III. HARDWARE.

We commented above that the system runs on an IBM AT or compatible and needs some specific hardware. This specific hardware is a low cost DSP based board which uses the AT&T DSP32C processor and includes 128 Kbytes SRAM, a programmable, 14 bits resolution, linear codec and analog amplifiers and filters.

Main modules on this board are (Fig. 4):

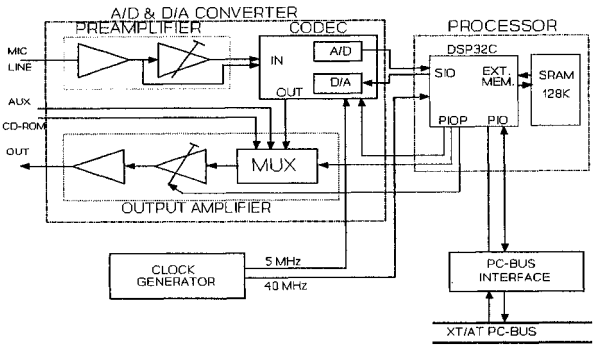

Fig. 4. DSP board hardware diagram.

a) A/D and D/A conversion.

Formed by input preamplifier, codec and output amplifier. Input signal will be supplied to the MIC/LINE inlet, which admits two signal levels: Low level signals (some mv) from microphones, and high level signals (hundreds of mv) from amplifiers, audiotape recorders, etc.

The low noise input preamplifier adapts the signal level to the A/D conversor margins. The codec is the sampling frequency programmable Texas Instruments TLC32044. It includes the antialiasing and reconstruction filters, and offers a global SNR of 59 dB.

A/D circuit has a programmable gain amplifier, a commuted capacitors anti-aliasing filter, a sample and hold step and a commuted capacitors technology 14 bit linear converter. The programmable amplifier can vary MIC/LINE input sensitivity from 0 to 12 dB. D/A circuit includes a 14 bit D/A converter, a reconstruction filter, a sin(x)/x correction filter and an output buffer.

This codec provides a serial interface with the main processor and voltage references for the A/D and D/A converters.

All the serial interface signals and the codec internal frequencies are generated from a 5 MHz master clock. Six different values for filters cutoff frequency can be programmed between 3.6 and 8.1 Khz, and 30 values of sampling frequency can be selected between 7.7 and 19.5 KHz.

The analog output circuit includes an analog multiplexer and a power module. We can select the signal to be amplified by using this multiplexer. Possible choices are: pre-amplified MIC/LINE input signal, AUX and CD-ROM signals or codec output signal. We can vary the gain of the circuit when codec output signal is selected (0, -20, -40, -60 and $-\infty$ dB). The maximum output power is 1 Watt when the loudspeaker impedance is 4 Ohm. This is enough to drive a loudspeaker, headphones.

b) Processing module.

The main processing module is the AT&T signal processor WE-DSP32C [3], working at 40 MHz and with 128 Kbytes external memory. DSP32C is a 32 bit floating point processor, with two arithmetic units, a floating point data arithmetic unit (DAU) and a fixed point control arithmetic unit (CAU). It also has 6 Kbytes 0 wait states SRAM internal memory, a parallel port (PIO), a serial port (SIO) and the external memory interface. Both arithmetic units work in parallel, resulting in 10 Mips and 20 Mflops.

The serial port (SIO) performs the data transfer between the processor and the codec, and the parallel port (PIO) maintains the communication with the PC bus. The general purpose PIO register PIOP is used to manage the output amplifier analog multiplexer.

External memory is SRAM, that needs 1 wait state and is organized in four 32 Kbytes banks. Internal and external memory are mapped in the same addressing space and contains data and executable code indistinctly.

c) PC bus interface.

This board has been designed for the 8 bits PC-XT bus, so it also works on any AT, ISA or EISA bus computer. Communication between the signal processor DSP32C and the PC bus is stablished through the parallel port (PIO) in the DSP. The board is mapped in 16 addresses into the PC bus I/O map, and the base address of this 16 can be hardware selected. Through this interface, the PC program can control the DSP transfer mode, download the DSP code and transfer the application data.

IV. SOFTWARE

The DSP software has been developed with the AT&T tools. Most of it is written in C language except for the critical routines that are coded in DSP assembler language. The PC can access DSP program global variables through DMA, so they are used in the data transfer. The access to these variables is restricted by a flag (semaphore) so PC and DSP do not interfere each other.

We will comment some functions: Sonogram, LPC envelope, Pitch contour and Recording.

The sonogram (spectrogram) and FFT spectrum are computed by a typical radix-2 FFT algorithm. The signal is pre-emphasized and windowed (Hamming, Hanning or Squared) as selected in the configuration menu. FFT length can be 256, 512 or 1024. That upper bound is due to limitations on DSP memory, anyway we do not need more resolution. Independently on the FFT length, the window length can be set to any value below FFT number of samples. When drawing the sonogram in the screen, we only permit 256 or 128 rows window height, so interpolations are the easiest (we just duplicate or delete FFT points). This is one of the assembler coded functions since this is the most used feature.

LPC spectrum is computed by autocorrelation method over a 40 ms pre-emphasized, Hamming windowed segment. We use the autorregresive production model (AR) [4,5], i. e., an all pole transfer function system. The coefficients are computed using Levinson-Durbin's recursion [5]. The transfer function is sampled in 640 points, corresponding to the VGA resolution, computing the function module directly.

Pitch is derived from a method similar to the one used in the DIGITAL ISOTON system [1], from autocorrelation sequence. In this implementation, we improved the postprocessing since the system does not have the real time restrictions. Voiced-unvoiced decision as a function of energy and zero-crossing rate is made first and then the voiced segments are pre-processed.

Pitch is derived from a method similar to the one used in the DIGITAL ISOTON system [1], from autocorrelation sequence. In this implementation, we improved the postprocessing since the system does not have the real time restrictions. Voiced-unvoiced decision is made first as a function of energy and zero-crossing rate is made first. Then voiced segments are pre-processed (center clipping and low pass filtering) and we get a pitch estimate from autocorrelation. Post-processing this raw data eliminates the possible estimation errors. The DSP program is the responsible of all the processing while the PC program performs the post-processing.

The recording utility is the most critical task, since the PC program has to represent the recorded speech envelope to adjust the input level. This is made using 250 samples input buffers and computing the maximum and minimum over each buffer (DSP program). These two values are used to follow the signal envelope. The maximum length is limited by the amount of available memory.

The DSP program has to be downloaded using the parallel DMA facility of the signal processor. For this task DSP must be halted. After loading, the program is verified before unhalting the processor.

The program in the PC controls the DSP function and the data used to compute a set of parameters and implements the screen functions. All the windows are fixed (configuration menu) except for the sonogram and the shift is varied depending on the final VGA video resolution.

An important topic when working on a MS-DOS PC system is memory usage. In order to accelerate the processing, we should have the samples vectors on RAM, what limits the memory available and the functionality of the tool (overlapping windows, etc.). As a result of such a compromise, the maximum duration is 8 seconds when sampling at 16 Khz.

Other task that the PC has to implement is the screen hardcopy. In this mode, each screen pixel is coded in a 5x5 dot matrix, which makes the process slightly slow and the printer needs at least 1 Mbyte RAM.

V. ACKNOWLEDGMENT

We would like to thank Francisco Javier Jiménez de los Galanes and Javier Ferreiros López for their help and suggestions in the achievement of this paper.

VI. REFERENCES

[1] J. Barquero Goñi, "Herramientas de Visualización y Análisis de Voz". Proyecto Fin de Carrera, ETSI. Telecomunicación, 1990.

[2] J. F. Mateos et al. "A PC Card for the rehabilitation of deficient auditive people". EUSIPCO 90, pp. 1175-1178, 1990.

[3] WE DSP32 Digital Signal Processor. Information Manual. AT&T Documentation Management Organization, 1986.

[4] D. O'Shaughnessy, Speech Communication, Human and Machine. Ed. Addison-Wesley, pp. 336-379, 1987.

[5] T. W. Parsons, Voice and Speech Processing. Ed. McGraw-Hill, pp. 136-169.

[6] L. Robert Morris, "A PC-based Digital Speech Spectrograph". IEEE Micro, Vol. 8 number 6, Diciembre 1988.

[7] S. Furui, Digital Speech Processing, synthesis and Recognition. Ed. Marcel-Dekker Inc, 1989.

[8] A. V. Oppenheim, and A. S. Willsky, Signals and Systems. Ed. Prentince Hall, 1984.