WORKSTATION AND SIGNAL PROCESSING SOFTWARE FOR EXPERIMENTAL PHONETICS

M. Scheffers and W. Thon

Institut für Phonetik und
digitale Sprachverarbeitung (*ipds*)
Kiel, Germany

ABSTRACT

We present a description of the PAS (Phonetische Arbeitsstation) workstation, developed at *ipds*. We will focus on what, from our experience, are the hardware and software requirements for a workstation for phonetics research and education.

1. INTRODUCTION

Based on the experience with the speech signal processor software SSP [1] running on a minicomputer, we decided some years ago to implement a network of PAS (Phonetische Arbeitsstationen) workstations with an advanced speech signal processor (ASSP) package optimized for phonetics research and education. While this configuration provides standard functions (signal acquisition, display and analysis) nowadays implemented in commercially available products, it offers additional facilities that cannot easily be obtained:

- Manipulation of signal and parameter files.
- Computer-controlled listening experiments and measuring of reaction times.
- Variability in processing parameters for different signals and for educational purposes.
- Support for novice users.

Though relying upon hardware and software standards (VMEbus, 'C' programming language) for easy upgrading, the software system as a whole is not easily portable to different workstations.

2. HARDWARE

2.1. PAS structure

A relatively compact device contains nearly all hardware components needed to perform these tasks. Fig. 1 shows the PAS hardware structure. The VMEbus based processor core is a single board computer with a MOTOROLA 68000 CPU, a NATIONAL 32081 floating point processor, 1MB DRAM, an SCSI host adapter and two serial ports. It will be replaced by a CPU board with 68030 CPU, 68882 numeric coprocessor, 4MB DRAM and the same peripheral controllers. A separate VMEbus memory board provides addi-

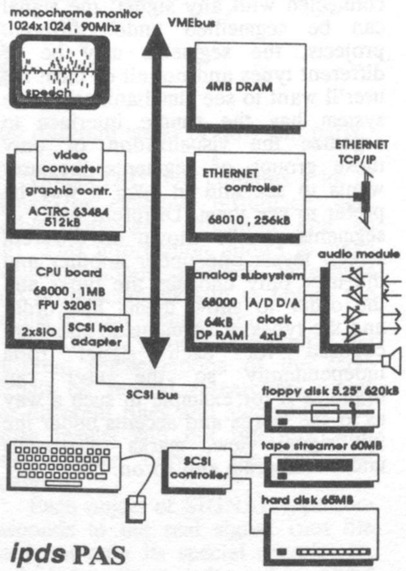

Fig. 1 PAS hardware structure

tional 4MB DRAM. The SCSI bus controls a 5.25" floppy disk drive, a 60MB QIC02 cartridge tape streamer and a 65MB hard disk. One serial port is used for the keyboard, the second one is available for a printer. A mouse or a trackball may be connected to the keyboard.

A graphic controller with an ACRTC HITACHI 63484 and 512kB video frame buffer provides graphic and alphanumeric display. An *ipds* designed high resolution converter generates a monochrome video signal with 1024 x 1024 pixels and 90MHz video dot clock, which is displayed on a 15" monitor.

The analog subsystem is a VMEbus board made up of two parts: the industrially available part contains a separate 68000 CPU with 64kB of 'dual ported' RAM, which is part of the memory map of the main CPU. The *ipds* designed part contains two 12 bit A/D-converters, two 12 bit D/A-converters, four low pass filters with software controlled cut-off frequencies and a software controlled sample clock generator common to all converters. The subsystem processor firmware developed at *ipds* allows for setting of sample rate and low pass cut-off frequency and for continuous conversion and data transfer from or to the hard disk for two channels with a sample rate of 20kHz max., each. These may be two input channels, two output channels or one input and one output channel.

The *ipds* designed audio module provides a user frontend to the analog subsystem. It contains all further analog components like amplifiers, a monitor loudspeaker and an operation front panel with channel switch, amplitude controls, input/output jacks. A digital operation state display is controlled by the analog subsystem. Hard- and software of the PAS analog interface provide features similar to a stereo tape recorder.

2.2. Networking

The six PASs implemented now are connected to the ETHERNET LAN of the *ipds* covering the complete premises of the department. The PAS's ETHERNET controller features a separate 68010 processor and 256kB of buffer memory. The networking protocol TCP/IP implemented as onboard firmware allows for easy transfer of files between the PASs and other nodes of the LAN even with different operating systems. At present, besides the six PASs running OS9/68k as operating system, an APOLLO DN3500 (UNIX SYSTEM V and BSD 4.3) and two AT compatibles (MS-DOS) are connected.

2.3. Supplements and expansions

For listening tests, an *ipds* designed device to acquire reactions and reaction times from up to 16 listeners simultaneously may be connected to each PAS. During a listening test a number of systematically varied speech signals are offered to the listeners through the PAS's analog interface. The subjects are asked to make a decision by pressing one of several buttons. The reaction time device, controlled by the PAS, starts a time measurement with reference to the speech signal at times specified by the experimenter. It monitors the keys of any listener up to a specified maximum time. At the first reaction of a listener, time and decision are stored. The data collected are transferred to the PAS for further processing at the end of the test.

The PAS is designed to provide the basic computing capabilities for phonetic speech signal processing. The widespread networking protocol TCP/TP allows for incorporation of virtually any computer designed for special signal processing tasks into the network, thus making its computing capability available to the users of the PAS.

3. SOFTWARE

3.1. Program structure

The software package ASSP has been developed to give both experienced and inexperienced users easy access to the processing facilities described in the introduction. The package consists of a main program and several subprograms. Since, in our experience, graphical display and interactive manipulation of signals play a central role in nearly every task, these functions are included in the main program. The subprograms perform functions such as analysis, synthesis, analog I/O, etc.

The main program is characterized by three levels: A graphical, menu-driven user-interface, an interpreter for the problem-oriented command language ELK (Easy Language Kit), and a toolbox with general file-handling and signal processing routines. Menu items can be selected using the mouse or the cursor keys and by single keystrokes. Some frequently used functions such as setting and deleting time markers and acoustic output of (parts of) displayed speech signals are directly accessable via function keys. The dialogue with the user leads to the generation of ELK commands such as 'DISPLAY 3,(0.1,2.3)' (display the selected parameters of all files allocated to window 3 in the time range from 0.1 to 2.3 seconds). The command line interpreter either executes such a command itself, using routines from the toolbox, calls a subprogram to do so, or issues a system command. Macros may be defined e.g., to design a series of manipulations with identical structure. For this the ELK language provides predefined variables, jump labels, and IF and GOTO statements to build loops.

The sub-programs have a standardized user-interface, very similar to the one in the main program. They can, however, also be called with an argument list. If all arguments are specified, the user-interface is not invoked so that these programs can also run as a background process.It may be clear that the sub-programs can also be run without starting the main program. This implies that new processing may be designed and tested outside the package. The data structure, the toolbox, and the standard user-interface greatly reduce the overhead and permit accessing existing data files and creating new ones that can be handled by the package. The modular structure of the package facilitates inclusion of new functions.

3.2. Data structure

Data files are grouped in so-called AREAs (subdirectories of the user's directory ASSP). Each AREA contains a configuration file with processing parameters such as sampling frequency, LPC order, analysis frame size and shift, etc. This configuration concerns all data files in the AREA, thus relieving the user of repeatedly having to specify or acknowledge them. A user may define and store several configurations optimized for the signals he is working with (e.g. physiological signals rather than speech) and select one of these when creating a new AREA instead of using the default settings. Furthermore, each data file contains a header, specifying the data type and all basic information for handling and display. Thus file handling and display routines in the toolbox could be made very general and small in number. It also means that introduction of a new data type generally requires modification of only one routine, viz. the one for creating a data file.

3.3. Summary of features

The package provides standard LPC analysis (autocorrelation method, reflection coefficients), formant analysis based on root-solving of the LPC polynomial, and F0-analysis. Intermediate analysis results, such as autocorrelation coefficients, may be stored for educational purposes. As an aid for segmentation/labeling and for loudness manipulation, an energy analysis (short-term rms) can be performed. For manipulation of speech rate during synthesis, a file may be created and filled with the standard frame duration values. Routines for converting LPC parameters (e.g. cepstral coefficients to area functions) are available.

The synthesis program requires definition of an F0 and a filter file. Using the information stored in the file header, the programm will automatically convert the filter parameters if necessary and select the appropriate synthesis routine. Optionally, an energy and a frame duration file (see above) may be specified. Synthesis results may be written to a new file or appended to an existing one.

All data files can in principle be displayed. Both time and y-scales may freely be adjusted by the user. Up to 10 graphical windows can be defined to which up to 64 files can be allocated. If more than one signal is displayed in a window, a vertical shift may be defined to ease comparison. Contours may be plotted on a linear or logarithmic y-axis. For multiparameter files such as formants, a selection can be made on a subset to be displayed. FFT and LPC spectra may be blended in. Per default, the windows have a common time axis, meaning that if the time range in one window is changed, the other windows will automatically be redisplayed with this new range. It is, however, possible to decouple a window from this common axis to provide a zoom or an overwiew window or to compare signals in different time ranges.

Interactive manipulation on displayed signals include CUT, COPY, APPEND, and INSERT of signals, MULTIPLY, DIVIDE, ADD, and SUBTRACT of constant values, SETting single values, DRAWing contours, SMOOTHing, and INTERPOLATION. File-to-file manipulations include COPY, INSERT, APPEND, ADD, and SUBTRACT of signals, MAPping (selective copying to adjust time range), and MASKing (transferring voiced/unvoiced information). Processing such as filtering, up/down sampling, and automatic stylization of contours is also provided. All manipulations are available for all data files even if they make little sense. Manipulations may be performed on all parameters or on a subset.

Whereas for analog input the duration of the recording is limited by the available disk space (typically 15 min. at 16 kHz), this was deemed undesirable for acoustic output. Since a sequence of output calls yields pauses between the stimuli that cannot be controlled very well and disrupt reaction time measurement, a batch list option has been developed. A batch list contains the sequence of segments, pauses, and marking signals that should be output. In the output program, a pre-processor will check the list, open the data files that contain the segments, generate the marking signals, and create a local command list for the analog sub-system. The firmware of this system can process this command list without creating gaps while simultaneously keeping the reaction time system synchronized. This means that there is virtually no limit on the duration of acoustic output.

4. REFERENCES

[1] BARRY, W. and KOHLER, K. (eds) (1982), *"Phonetic data processing at Kiel University"*, Arbeitsberichte des Instituts für Phonetik, Universität Kiel (AIPUK), *18*.