

SESAM: A LOW COST WORKSTATION FOR SPEECH ASSESSMENT

Jean Claude CAEROU, Jean Marc DOLMAZON & Jean Michel LUNATI

Institut de la Communication Parlée
INPG/ENSERG - Université STENDHAL Unité de recherche associée au CNRS n°368
46, Avenue Félix Viallet 38031 GRENOBLE CEDEX
FRANCE

ABSTRACT

Feeling the urgent need of standardization for evaluation of devices in speech Technologies, european countries gathered their efforts to produce multilingual assessment procedures within the ESPRIT project **SAM** (Multilingual Speech Input/Output Assessment, Methodology and standardization). One of the numerous actions of the **SAM** project is the definition of a **common working tool** for all the tasks (whether for input or output Assessment) involved in the project. The project was initiated two years ago and this paper presents the main achievements for the task provision of workstation facilities and states the results of this experience in standardization for speech assessment procedures.

1. INTRODUCTION

Speech technology is fast emerging as an important future industrial activity and widely accepted standards will assist its industrial development and broad applications. For this purpose a project, funded as part of the European Community's ESPRIT programme, was raised for the establishment and use of multi-lingual European standards in the assessment of speech I/O devices. The basic material for this assessment is a collection of speech items spoken by selected speakers, and gathered into databases. To record and to use these databases within the various centers of the community, standards are needed. The implementation of these standards require, for dissemination and use, a common tools to share data and to run normalized tests. To meet these goals we defined a common workstation, widely spread, which allows :

- an easy access for non-specialists, through the use of a friendly user interface,
- the use in association with mainframes, through built-in networking facilities,
- the processing of input and output assessment in situation,
- the use of general speech processing algorithms,

leading to a very large distribution within the working groups of the **SAM** consortium.

The paper presents in the first paragraph the hardware aspect of the defined workstation. The next paragraph is concerned by the normalized software used on the workstation.

2. THE WORKSTATION

The speech workstation must provide all the computing facilities that technical professionals need in their daily work routines (including specialized applications such as speech processing or real time I/O as well as general purpose tools for programming, writing and sharing information through networks). To be extensively used inside the SAM community, it was necessary to consider the speech workstation as being a standard, e.g. widely available, in order to collect and to stimulate the dissemination of large amounts of work carried out by different centers. From the numerous discussions between SAM partners we had to consider the workstation at two levels: a low cost one for mainly production environments or OEM distribution, and a higher level workstation for research organizations engaged in speech technology. Naturally this two types of workstations have to be fully compatible and the procedures developed on the high level workstation have to be able of direct implementation on the low cost one. For speech processing requiring such very large amount of power that workstations will not share, the tasks are supposed to be carried out on a mainframe connected to the workstation by a local network.

The IBM-PC compatible machines have been found to correspond to the above mentioned requirements. This family has been chosen because of the very large number of types ranging from low costs XT models to high performances AT working up to 25 Mhz clock speed. This family of machines presents also the great

advantage to be widely open and to support a significant number of different types of boards. Due to the large number of IBM-PC compatible machines, SAM consortium has not recommended a given model, the only constraint on SESAM being to be compatible with the IBM AT3 model (although this model has medium performances and slightly oldfashioned). The hardware requirements to perform the functions of the SESAM workstation are listed below. Naturally some of these devices are optional depending of the station objectives:

- a hard disk with reasonable access time and capacity ranging from 70 MB up to 300 MB;
- a graphic screen for some friendly interface software: EGA or VGA;
- a mouse as pointing device (Microsoft serial compatible or parallel);
- a Local Area Network board (Ethernet and TCP/IP protocol);
- memory extension, which can stand as virtual disk for fast access, or as extended memory for large software (EMS standard);
- a CDROM reader and his associated interface board to access data bases recorded on compact disks. The available volume of one disk is about 550 Mbytes. Portions or whole databases are available on CDROM at the present time: EUROM.0 from the SAM project (Grice & Barry, 1989), BDSONS.0 from the French GRECO and TIMIT from Texas. Some other are under processing (7 disks from BDSON in fall 89, for example);
- a Digital Signal Processor board for fast speech processing (in parallel with the PC). The OROS AU21 and AU22 boards are already used in several SAM laboratories, and recommended. These boards support analog Input/output (with 16 bits ADC and DAC);
- two RS232 serial lines for the control of recognizers (configuration and results) and an "intelligent" keyboard connected by a parallel line adapter for listener or speaker responses during the various tests or during recording speech material;
- a Mass Data Storage device. A Digital Audio Tape recorder and its digital interface is under evaluation in our laboratory. This device could be used during database recording as a huge temporary storage before pressing a CDROM. The same procedure is planned for WORMs, magneto-optic cartridge disks and others high capacity supports.

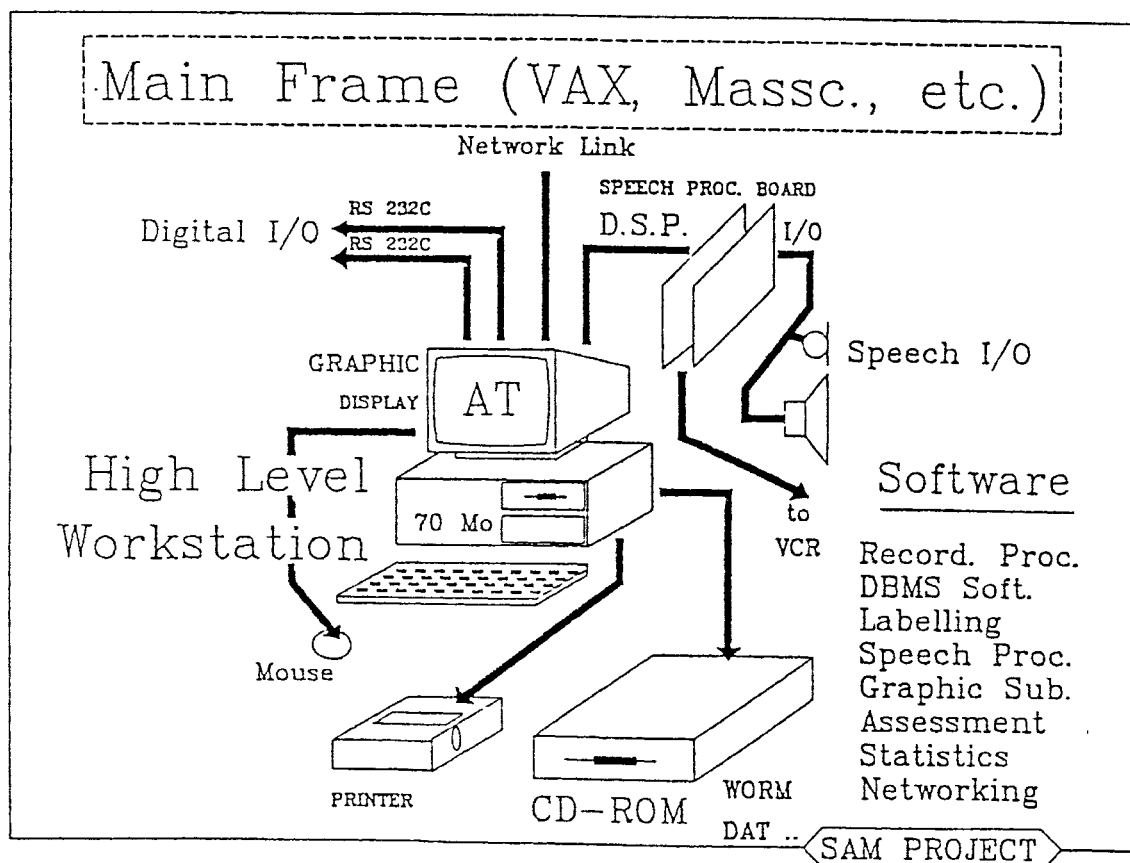


Figure 1: schematic of the high level workstation.

3. THE SOFTWARE

The software of the workstations has to deal with very different topics such as speech material recording, prompting the speakers, database management, speech input assessment, speech output assessment, scoring, statistic, word processing, etc. An important part of the workstation specification for the extension phase were the control files for input assessment, which were necessary for the input group tasks of assessing the prototype standard workstation and evaluating the EUROM.0 database as a tool for recognizer assessment. The specification of the control files also contribute to the main phase tasks recognizer testing and provision of basic workstation configuration. The assessment control files are a hierarchically organized suite of five files comprising the top-level **control file**, a recognizer-specific **configuration file**, two speech data oriented files the **training** and **testing** files, and finally the **response file** containing the recognizer test run (Tomlinson, 1989). These files are tightly connected to the assessment control software (Gauvain, 1989). The scoring process is a recognition performances measurements.

3.1 Operating system and languages

The choice of the operating system has a great effect on software development and execution efficiency. The basic workstation uses at present MS-DOS version higher than 3.1 as operating system. This system is simple enough to use for non specialists, but it has some severe weakness, the main being the 640 Kbytes bottleneck. It has also some flaw in memory management. To go beyond MS-DOS imperfection we have to worry about more powerful and more flexible operating systems. After fall 1989 we plan to port the basic software package under UNIX. This operating system joined to the X-Window package would increase the workstation power on networking, device independence, multitasking windowing and graphics system. For efficient exchanges between SAM partners, it is important to produce reliable, maintainable, portable and practical programs. To be uniform, software production and documentation rules has been produced. The programs should be written with structured languages and documented. Specialized applications are recommended to be written in "C" language, the standard used being MSC release 5.1.

3.2 Software modules and files formats

The assessment workstation must be able to test all isolated and connected word recognizers. The recognizers are controlled through one or two RS232 interface. The assessment software is divided into up to ten modules: the RS232 control routines, the training program, the test program, the scoring program, the initialization program, the signal processing routines, the assessment session manager, the assessment file generator and the configuration generator.

The software running on the workstation uses or generates plenty of files: signal, parameters, assessment, configurations, scoring, etc.). To clarify exchanges the files format has to be well defined, concerning the naming and the contents. Different kinds of file formats has been defined: Label files (Tomlinson & al., 1989)(Barry, 1989), assessment control files (Tomlinson, 1989), etc. other are being defined.

3.3 Prompting software

For future databases recording special cares has to be taken. Due to the large amount of speakers involved in such an investment, automatic procedures prompting software has to be used. The purpose is to automate the recording instructions, the list presentations, the level detection and adjustment. It also has to automatically update some data-base files such as speaker characteristics, addresses of items, etc. Error recovery must be available for the speaker itself and also to the recording manager.

In the light of the French BDSONS recording conditions, we have defined a prompting software with the above characteristics, and some other such as optional silence cancelation. Packaged with this software, a query module for selected material is also available. This package has been distributed to partners for evaluation and review.

3.4 Database management software

Since all SAM partners agreed about the need of a Database Management System to build an integrated speech assessment workstation, some experiments were made to verify the suitability of a number of DBMS to our purposes. An accurate comparison between the DBMS showed that the best choice is ORACLE. The structure of data proposed is split into two parts. The first describes the speech database loaded on the workstation, in connection with the database used to test recognizer, or to label files. The second part is designed to document and archive all crucial information about tests performed on the workstation itself. This structuring of data will permit the addition of several descriptive levels with little modifications to the table set.

3.5 Basic software package

The first aim of this software was to develop the basic functions required to display and handle speech signal and related contexts. After completion of this goal, our major labor was devoted to manual labelling. This software, called "PTS", employ graphical representation and mouse control. PTS uses "working contexts", each of them is related to a window on the screen. Windows could be moved, resized, activated and wiped out by mean of the mouse. PTS has the ability to view speech signal in different ways (zoom, translation, etc.), edit, listen, compute sonagrams, put and edit labels, etc.

3.6 Scoring and statistics

The scoring procedure is considered to be a two-step procedure composed by a preprocessing and the actual scoring. The preprocessing turns the recognizer dependant response file(s) into a recognizer independent **result file(s)**. In the scoring, the application of result files is controlled by a **scoring control file** which is assumed in the long term to be DBMS generated. The **score file** is the output of the scoring and thus presents the output of the assessment task. The score file should be readable by the DBMS system for further registration of the results of the assessment task.

When dealing with scoring and other evaluation results, statistical packages are helpful. A research in literature shows that statistical applications software are numerous (Raskin & al., 1989).

3.7 Word processing

For an effortless exchange of informations, SAM partners have decided to use a standard word processing to create their texts inside the consortium. A set of encoding and decoding software is also used to exchange word processing files by Electronic-mail.

4. CONCLUSIONS

Within the SAM project a low cost workstation has been defined, including a family of hardware systems and a set of software to record, process and assess speech I/O devices.

The future will be involved in developing the non achieved software. The rough MS-DOS operating system has to be polished by software porting on other systems such as OS2 and XENIX.

Others tools has also to be included inside the workstation: semi-automatic labelling, specific software for speech output assessment, etc.

REFERENCES

- W.J. BARRY (UCL) (1989): "Proposals for Body of Labelling File", Esprit Project 1541 (SAM) Multilingual speech input/output: assessment, methodology and standardization. Extension Phase Final Report, pp. 194-197.
- J.L. GAUVAIN (LIMSI) (1989), "Proposal for Assessment Control Software", Esprit Project 1541 (SAM) Multilingual speech input/output: assessment, methodology and standardization. Extension Phase Final Report, pp. 241-243.
- M. GRICE, W.BARRY (UCL) (1989), "EUROM.0 Technical Description", Esprit Project 1541 (SAM) Multilingual speech input/output: assessment, methodology and standardization. Extension Phase Final Report, pp. 170-188.
- R. RASKIN & AL (1989), "Statistical Analysis", PC Magazine special issue, VOL 8, Number 5, March 14, 1989, pp. 103-199.
- M.J. TOMLINSON (RSRE) (1989), "Proposal for Assessment Control File", Esprit Project 1541 (SAM) Multilingual speech input/output: assessment, methodology and standardization. Extension Phase Final Report, pp. 231-240.
- M.J. TOMLINSON (RSRE), R. WINSKI (Logica), W.J. BARRY (UCL) (1989), "Label File Format Proposal", Esprit Project 1541 (SAM) Multilingual speech input/output: assessment, methodology and standardization. Extension Phase Final Report, pp. 189-193.