

TESCOS - AN INTEGRATED WORKSTATION TO COLLECT LARGE SPEECH DATABASES ON THE TELEPHONE NETWORK

Canavesio F., G. Castagneri, G. Di Fabrizio, A. Massone

CSELT - Centro Studi e Laboratori Telecomunicazioni S.p.A.,
Via G. Reiss Romoli 274, 10148 Torino, Italy

ABSTRACT

The goal of this paper is to describe a workstation designed to collect speech databases from the Public Switched Telephone Network: the TESCOS workstation, that has been used to collect large speech databases from more than 2000 telephone customers. It implements different features like: easy design of user interaction, call progress analysis, error detection, on-line recognition testing. Different TESCOS versions have developed to fulfil growing needs of telephone data.

1. INTRODUCTION

The growing availability of speech recogniser designed to be used on the telephone network and the very new fields of application linked to the audio-tex services, have increased the need of telephone speech database both for developing new algorithms and to test new devices.

The collection of these large telephone speech database requires different kinds of knowledge to solve peculiar problems due to the characteristics of the audio channel and to the lack of control over the speaker behaviour. Better performances of the workstation in inferring what is happening from the man-machine interaction analysis, will decrease the final data discard rate. The recording workstation should be carefully designed according to the network characteristics and the specific task: to this purpose the TESCOS (TELEphone Speech COLLECTION System) project has been planned and accomplished during the last year.

2. WORKSTATION DESIGN

The TESCOS workstation has been designed to collect speech databases over the telephone network ensuring a low rate of data discharge. The project has followed different steps according to the present needs, however the basic characteristics of the system were already implemented in its first version; the main features were:

- full automation of recording procedures;
- voice driven user interface;

- possible identification of individual speakers;
- use of automatic speech recognition;
- speech material collected according to the specifications defined in the ESPRIT Project SAM [1];
- workstation activities documented by a Relational Database Management System (RDBMS);
- strict control over the user behaviour in order to avoid most frequent errors.

Hardware

The system hardware is based on an IBM compatible Personal Computer equipped with mass-memory devices, proportional to the daily data load, and the following boards:

- a Telephone Interface Board Dialogic D/41D
- an isolated words recogniser
- a high quality 16 bit acquisition board OROS AU21,
- an Ethernet board
- a four channel programmable attenuator for level adaptation
- 2 serial ports.

The block diagram of the workstation is reported in Fig. 1. The system was able to answer the call, to play messages and prompts (PCM 16 bits), to recognise DTMF digit sequences, for caller identification, and to store speech tokens at 8000 Hz sampling rate.

TESCOS has been designed to work continuously and to perform a series of automatic checks on input speech. Furthermore the workstation can be connected with a speech recogniser through a serial line. The system be equipped with an automatically activated backup unit.

Software

TESCOS is managed by an event-driven program written in C language. The program allows to manage parallel processes of data acquisition and telephone call progress analysis. All information available during acquisition is stored in database tables directly produced by the program; fig. 2 reports the entity-relationship diagram of the data structure.

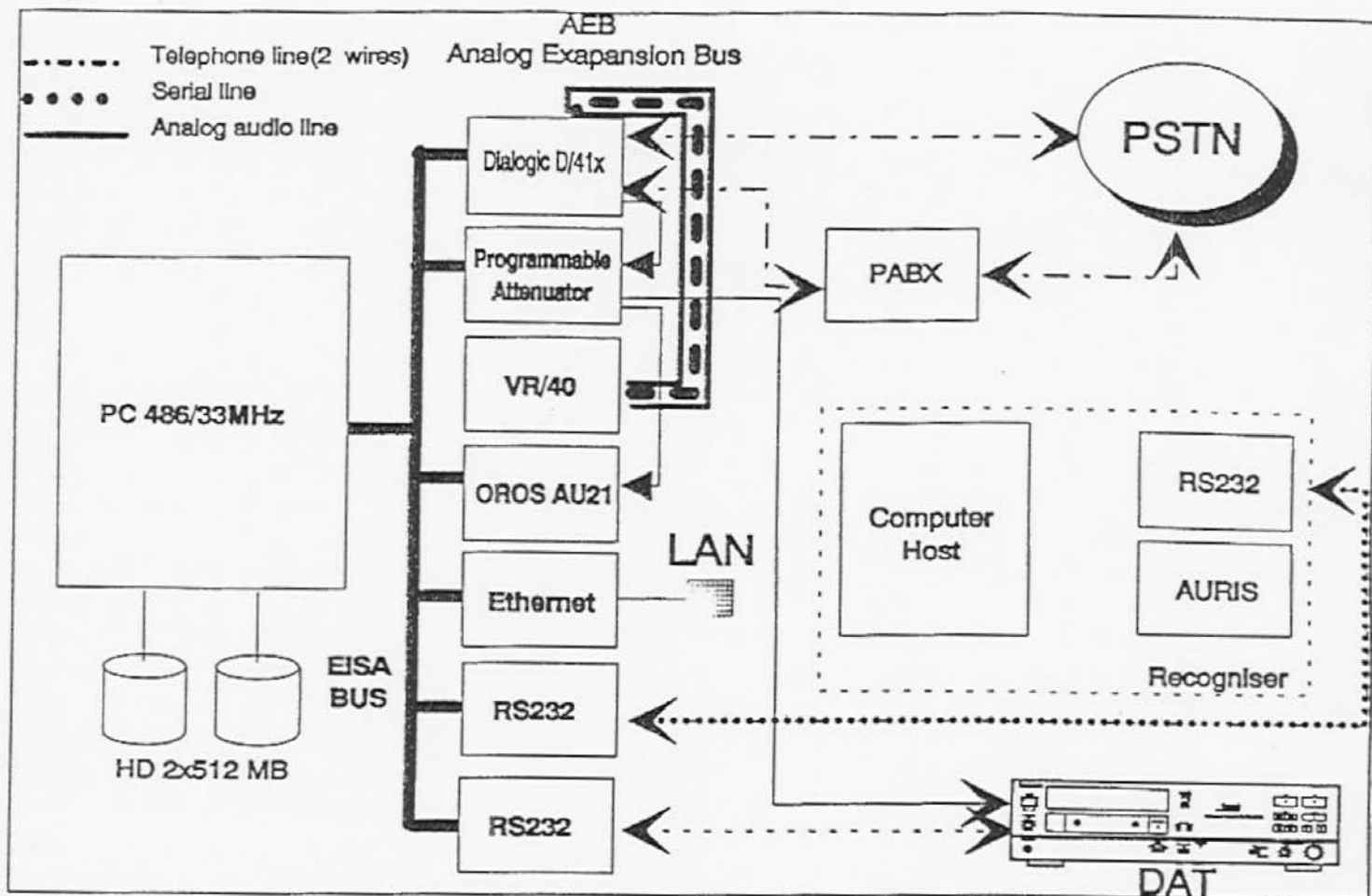


Fig. 1. TESCOS block diagram

Acquisition

The audio signal was low-pass filtered with a telephone bandwidth filter at 3.4 kHz, and digitally recorded (PCM 16 bits) by a sampling rate of 8000 Hz. The gain of the acquisition channel was held constant at the value which maximised the input dynamic.

Three different acquisition modes were implemented:

- isolated word with speech detection and automatic end-point detection;
- continuous speech with speech detection and programmable window duration;
- fixed acquisition window.

The workstation computes the maximum value out of the samples of each recorded word in real time, in order to check whether any saturation occurred.

A procedure for the periodical calibration of the workstation has been developed. The audio channel characterisation highlighted the following values:

- Signal to Noise Ratio: 57 dB
- Maximum input range: 1.3 dBm
- Total Harmonic Distortion: 0.55 %

3. SYSTEM EVOLUTION

The TESCOS project was started to collect speech databases on the PABX and has evolved to face the growing

needs of collecting speech material over the telephone network. In the following the main phases of this project are described.

TESCOS base

The first version has been designed to collect a speech database from few hundred speakers over the internal PABX. The system, besides the standard features, was able to manage a Digital Audio Tape in order to record the audio signal on tape. The DAT was completely controlled by a serial interface using the communication protocol SMPTE (Society of Motion Picture and Television Engineers). The DAT was fully integrated in the workstation that was able to identify the number of the tape available on the device by a DTMF code. The tape number, the exact address and the time code of each word was recorded by the system in order to allow the direct access and play of individual tokens.

The presence of the DAT recorder in the system allowed both compatibility with the recogniser training environment and the availability of the speech signal analog-digital backup. Unfortunately the high maintenance cost, due to the substitution of the DAT tape every two hours, discouraged the use of the DAT option in the collection of larger speech databases.

- Real time DC offset subtraction

TESCOS LAN

A substantial modification of the LITE version became necessary in order to allow the collection of a speech database in an intercontinental exchange in Rome.

The distance between the recording site and the laboratory made necessary a remote monitoring of the workstation activity to ensure current system maintenance (back-up verification, software up-date).

TESCOS LAN configuration was composed by two TESCOS workstations connected to a network server by a Local Area Network. The server was equipped with a modem and a remote-login software that allowed a complete remote check of the workstations by:

- daily transfer of database files for monitoring of the received call;
- transfer of samples of speech files for quality check of recorded material (background noise, offset);
- remote control of system status (disk space available, error log file) and file managing.

Using the LAN, the system was able to perform the daily automatic back-up for both the stations.

This configuration allowed to collect more than 1200 telephone calls with only one on-site operation caused by an hardware impairment of an hard-disk.

TESCOS ML

The experiences gathered in developing TESCOS, the availability of telephone interfaces for the Italian Digital Telephone Network and new requirements on the format of telephone speech databases allowed the design of the ML (Multi Line) version. This new version can manage more than one line (typically 8-12 lines) both analog and digital. The audio signal is digitally recorded in PCM 8bit A-law format, directly from the network (using the E1 interface) or after its conversion performed by the telephone interface itself.

TESCOS ML uses the RDBMS for complete monitoring of the man-machine interaction and is able to provide measurements on the workstation performances (number of calls, number of errors, mean call duration for each line). In the ML version, the embedded speech recogniser allows to identify the caller identification code. Furthermore, the simplified structure of the system, facilitates the workstation maintenance.

4. MAN-MACHINE INTERFACE

The variety of naive speaker behaviours, possible during collection of large databases, require a workstation with flexible tools to design and modify the man-machine dialogue. TESCOS has been designed to fulfil this requirement and the possibility of defining all the interaction writing a script file, resulted in a substantial improvement of the workstation performances. Simple changes of instructions, delays and prompts caused a significant error reduction.

Subjects were often asked to read list of words synchronised with a beep to decrease the mean duration of the calls. This feature was easily implemented because

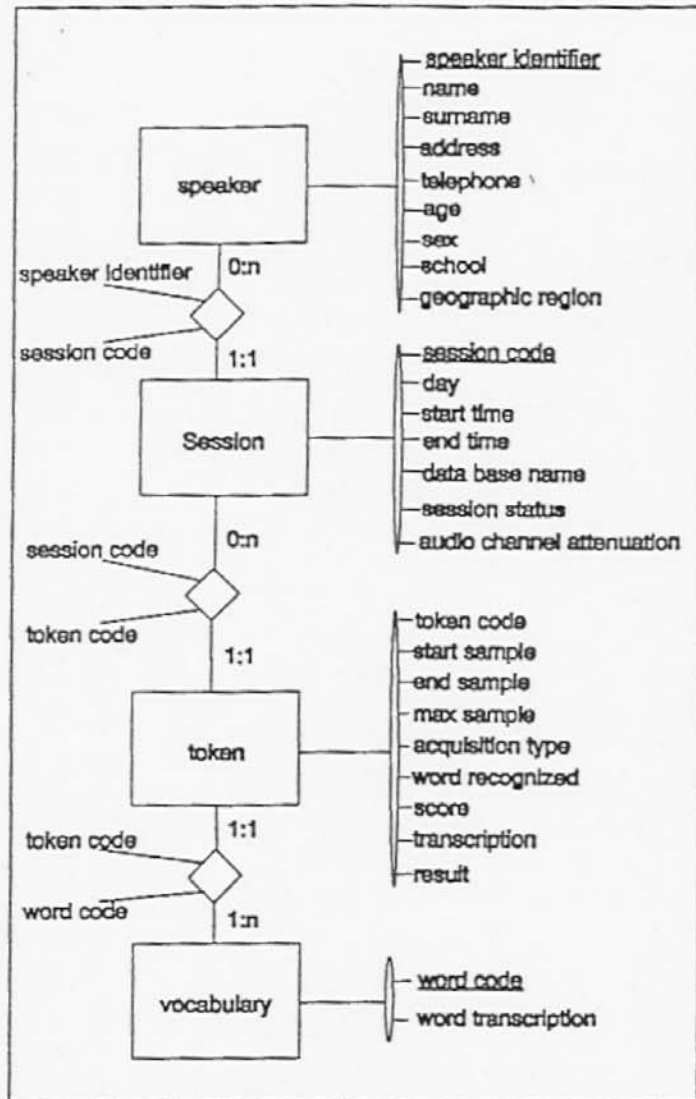


Fig. 2. Entity -Relationship diagram.

TESCOS LITE

The LITE version of TESCOS, designed to collect large speech databases over the Telephone Network [2], implemented significant improvements with respect to the previous version, like:

- Fast design and modification of the user interface by a *script file*. The whole structure of the application is described in a script file containing the definition of each message used in the application and specifying each action to be performed. The script file enables to change completely the application dialogue without any code manipulation. The script file structure is described in Appendix 1.
- Possibility of changing the acquisition parameters (attenuation, timeout, etc.) by defining their values in a configuration file (see Appendix 2).
- Check of the signal level in real time for verification of the saturation.
- Possibility to connect a speech recogniser to TESCOS either by RS232 port or by the internal BUS. This feature allows testing recognisers during the speech collection process.

TESCOS incorporates an effective "speech detection" implemented using the embedded recogniser. The workstation was able to prompt the speaker with an appropriate error message asking for repetition whether a response occurred over the beep or whether no signal was detected. This procedure allows a substantial reduction of missed or truncated words.

As the possible range of average speech level over the Italian Telephone Network can exceed 20 dB and the gain of the audio channel has been calibrated in order to avoid recording of extremely low signals, TESCOS controlled the saturation for each recorded word asking for repetition if any saturation occurred. This solution has been chosen because the saturation of the DAC is an experimental artefact depending on the available board and is not an intrinsic characteristic of the signal. On the other hand, the controlled solution is coherent with the normal behaviour of a real recogniser that prompts speakers with appropriate error messages when receiving out-of range speech signals.

TESCOS was also able to detect the background noise level; interrupting the recording whether it was too high and asking the speaker for a later call in a more convenient situation.

All these checks allowed to minimise the number of discarded words in the collected databases.

5. ON LINE RECOGNISER TESTING

A further feature of TESCOS enables to test a speech recogniser during collection of speech databases.

The workstation is equipped with a recogniser directly connected to the DIALOGIC telephone interface. During recording of a speech database the recogniser can be activated by an appropriate command contained on the script file.

It is also possible to control a remote recogniser through a serial port by the same procedure. The device receives the same signal as recorded on the workstation and sends its answers through the RS232 to TESCOS. The workstation displays the percentage of correct recognition and monitors the whole activity of the recogniser. The results are stored

on files according to the specifications of the ESPRIT Project SAM, for the performance evaluation on a given vocabulary.

This feature is particularly important because it allows to characterise a device in a situation similar to the final application environment. It also offers the unique opportunity of characterising the laboratory test environment [3]. It becomes possible to calibrate the laboratory testing environment in order to get the same results obtained during the database collection.

6. CONCLUSIONS

TESCOS project evolved according to the growing needs of telephone speech databases. The availability of new technologies and the evolution of the state-of-the-art of the telephone database collection techniques allowed to refine the most important features of this workstation.

TESCOS has been widely tested during the process of collection of different speech databases. The use of speech recognisers both for signal detection and for on-line testing ensued a low rate of discarded utterances due to clipped signals, mutilated words or inadequate speech behaviour. On the other hand the recogniser on-line testing provided preliminary results as a basis for further laboratory tests.

REFERENCES

- [1] AA. VV. "User Guide to Input Assessment" Esprit SAM document SAM-UCL-G005
- [2] G. Castagneri, G. Di Fabbri, A. Massone, M. Oreglia "Sirva - A Large Speech Database Collected On The Italian Telephone Network" EUROSPEECH 93, Berlin, (1993).
- [3] F. Canavesio, G. Castagneri, G. Di Fabbri, F. Senia "An Overview on Recogniser Testing Activities Performed in CSELT", ESCA Workshop, Lautrac (1993).

SCRIPT FILE DESCRIPTION

ACQUISITION & PLAY operations

PLF: <file name>

Play file

ACQ: <m>, <n>, <p>, <mode>, <tm>

Inizializing acquisition. Where:

m = label group,

n = label file associated (Y or N),

p = number of acquisition,

mode= IIC I = Isolated Mode C = Continuous Mode

tm = msec. of silence following the acquisition

LBL: <word transcription>, <prompt file name>

definition of an item in a list of word to be recorded

ELB:

End label list.

PAU: <file1 >, <file2>, <file3> <i>

Pause with prompt file and play refrain file <i> times;

<file1> presentation,

<file2> refrain

<file3> conclusion.

If <i>=0 then the procedure wait for dtmf input.

WAI: <msec>

Wait msec.

DTMF operation

DTM: <m>, <n>

Get <n> digit DTMF into system variable number <m>

PLD: <m>

Play speech files contating digits memorized in variable <m>

DAT Commands

DTS:

DAT START RECORDING

DTP:

DAT PAUSE

Local Recogniser Commands

VOC: <file name>, <file name>, <subvoc>

Vocabulary bynary file name (default suffix is .VB0),

vocabulary text file name (default suffix is .VT0),

active subvocabulary number.

SUB: <subvoc>

Select New Active Subvocabulary.

ERA: <m>

Enable recognition and specify the modality:I=isolated, C=continue

DRA:

Disable recognition .

System Control Parameters

PTH: <environment> <path>

Specify environment variable and value associated to it.

CFG: <file name>

System configuration file name.

CFF: <file name>

Recognizer configuration file name.

Remote Recogniser Control via RS232

COM: <Port>, <Baud Rate>, <Parity>, <Data Bits>, <Stop Bits>

Set RS232. Initialise the specified serial port.

REC: <Recogniser Name>

Selects the initialisation procedure and communication protocol for specified recogniser

TST: <mode>

Enables the recogniser to work in the specified <mode>

<mode> -> IIC I = Isolated Mode C = Continuous Mode

ETS:

End Test. Archive the best hypothesis and scores into database and disables the recogniser.

CONFIGURATION FILE

CFG: <file name>

Specifies a file containing system parameter in SAM format.

Possible label are :

PAR: <parameter name>, <value>

The possible parameter name are:

OFFSET, <offset value>

offset value defined for the specific OROS board

SAMPLING, <sampling frequency [Hz]>

OROS Frequency Sampling Rate

FXDIR, <number>

Max file number per directory

DISK, <logical name>

hard disk logical name, one for every disk used in the system.

ATTEMPT, <n>

MAX number of attempt before call interruption.

PROMT_ACQ, <file name>

Acquisition prompt file (typically a shor beep)

PROMPT_ERROR: <file name>

Error prompt file (typically a short buz)

GAIN_OROS, <OROS gain value>

Gain of the A/D channel

MBYTE_FREE <memory space>

minimum amount of memory, on current hard disk, for correct operation. The value is in Mbyte

START_TIME, <hour>

backup start time

WORKSTATION, <workstation name>

DB, <database identifier>

SYSTEM_NAME, <data base name>

CITY, <acquisition location name>

PTH_BACKUP, <backup path name>

PTH_MSG, <message path name>

ECF:

end the system parameter file.