

# A PC-BASED TOOL FOR HELPING IN DIAGNOSIS OF PATHOLOGIC VOICE

*Juan Ignacio Godino Llorente\**, *Santiago Aguilera Navarro\**, *Sira Palazuelos Cagigas\**, *Alberto Nieto Altuzarra\*\**, *Pedro Gómez Vilda\*\*\**

\*E.T.S.I. Telecomunicación (U.P.M.). Ciudad Universitaria s/n. 28040 Madrid. Spain.

e-mail: {godino, aguilera, sira}@die.upm.es

\*\*Universidad de Alcalá de Henares \*\*\*Facultad de Informática (U.P.M.)

## ABSTRACT

We have developed a diagnosis tool for the voice clinic which runs on a personal computer. The application records different registers, providing the clinicians with helpful information to assist them in the evaluation and diagnosis of pathologic voice, making the process quicker and more accurate.

In order to do this, we need to capture and store four different signals (in synchronous real time).

The signals to be captured and stored are the following [1] [2]: videoendoscopic video recorded with fibroscope or telarlaringoscope, electroglottographic signal during fonation, voice signal and air flow signal.

All these different signals are recorded with specific transducers and standard digitisation signal boards, using microphone input and input lines simultaneously.

Several systems to help in the diagnostic have been previously developed, but none of them captures the four previously mentioned signals simultaneously. All of them are highly interesting from the clinical point of view, and assist the expert in making the decision. [3] [4]. The main advantage, related to other systems, is that looking at the videoendoscopic record, clinicians are able to label voice registers with their associated pathologies.

We can emphasise the fact that the four signals are captured in a synchronous way, so that they can be edited (cut and paste), taking the same time reference. Because of this, an unlimited duration exploration can be carried out and, later, the most interesting time window can be selected and stored.

Another feature is the possibility of calculating the different acoustical parameters of the recorded signals.

On the other hand, an application of this system is using it as a tool to create a pathological voice database.

The workstation consists of a Windows™ application, and a standard hardware including a personal computer (Pentium 200, 32Mbyte RAM), 2 audio boards (Sound-Blaster™), and a video board.

## 1 INTRODUCTION

There are several tools developed to digitalise, record and extract the parameters of the speech signal. Such utilities are helpful for screening the speech signal in both, temporal and spectral domains. Also, using digital signal processing,

different parameters may be calculated, e.g.: jitter, shimmer, HNR, Fo...

Such tools allow the clinicians to evaluate the voice quality, and are an objective support for the diagnosis. They help the clinicians to diagnose vocal and voice diseases. The main advantage they provide is allowing objective evaluation of the patient after a surgical operation, matching the different registers both before and after the operation.

However, the specialist cannot evaluate from the voice register alone. An acoustical analysis is not enough and a visual register is needed to support diagnosis. This visual register is recorded using videoendoscopic techniques. It is even usual in the clinic, to record the voice signal together with the electroglottographic signal, the air flow signal and the videoendoscopic signal. Screening, capturing and recording of these signals were carried out independently by using expensive devices. In this paper, we propose a cheap and useful PC-based tool capable of simultaneously screening and recording this set of signals. This is called HISPAVOZ.

The main advantage of HISPAVOZ is that recording tasks and diagnosis may be carried out by different people with different qualifications, so the specialists may focus their work on diagnosis. Likewise, consulting different specialists without the presence of the patient is easier.

The system is based on a personal computer and several standard signal digitalisation boards. It has a low cost, so it may easily be introduced into ENT clinics and consulting rooms, even though there is no specialist. Then, the task of the specialist becomes screening already captured registers and evaluating them.

We also have to keep in mind that one of the research fields in speech technology is the automatic diagnosis of pathologies related to fonation. Such a task only may be performed if we provide a wide data base of well labelled registers, and HISPAVOZ is a useful tool for recording such a wide well labelled database.

## 2 BACKGROUND

As we said previously, there are several systems [3][4][5] that allow the screening and recording of the speech signal, i.e.: Speech Viewer (by IBM), CSL (by KAY Elemetrics, Inc.), Dr. Speech (by Tiger Electronics, Inc.) and VISHA system. The last one, was developed by the Dept. of Electrical Engineering

of the Technical University of Madrid. The VISHA [5][7] system uses a data acquisition board based on DSP (developed by the same department) and software that allows the screening, editing, recording and extracting of parameters. Such a system, just like the others mentioned above, only allow the treatment of speech signal, so the clinicians have to use different devices to record and to screen other interesting signals. The background received by the users of such a system was used to develop HISPAVOZ.

### 3 HARDWARE

The workstation is made up of a personal computer and a set of peripherals (transducers) that allow the capturing of different signals.

The PC has to have the following next minimum features:

- Processor Pentium 200 MMX MHz
- 32 Mbytes RAM
- 1 Gbyte HD
- 17" colour monitor 17". SVGA video card.
- 2 Sound-Blaster™ audio boards
- 1 video board compatible with PAL video system and with MJPEG hardware encoding.
- O.S. Windows 95

Needed peripherals:

- Microphone.
- Electroglottographic system (electroglottographic signal measure)
- Videoendoscopic system.
- Pneumotachograph system (air flow measure).
- Printer.

### 4 RECORDED SIGNALS

#### 4.1 Electroglottographic Signal

Electroglottography allows the screening of the movement of the vocal folds. This technique measures impedance fluctuations produced while an accent current passes through the neck. The measurement is carried out using two electrodes placed on both sides of the cartilage thyroids. The signal obtained is denoted as electroglottographic signal (EGG). The waveform is very similar to a sinusoid with frequency being the fundamental frequency of the speech. The bandwidth is around 200 Hz.

#### 4.2 Air flow signal.

The air flow signal is measured using a pneumotachograph. It is connected to a mask that is held in place with the teeth and

it measures differences of pressure produced by the air exhaled when breathing. The bandwidth is around 60-70 Hz.

#### 4.3 Videoendoscopic signal

This technique allows visual screening of the larynx. A video camera, a magnetoscope, and an endoscope is needed to generate the videoendoscopic signal. We obtain a real time video of the vocal folds during fonation.

There are two different kinds of videoendoscopes:

- Fibroscope (flexible endoscope)
- Telelaringoscope (rigid endoscope)

Both methods can be used with continuous or with stroboscopic light [6].

#### 4.4 Voice signal

Voice signal during fonation.

Captured using a microphone that acts as transducer. It is sampled at 44.1 KHz.

### 5 DESCRIPTION OF THE SYSTEM

HISPAVOZ runs in a personal computer with O.S. Windows™ 95. The visual interface is designed according to the Windows™ standard, offering an option set that can be accessed by mouse or by keyboard. Items can also be accessed by pop-down menus, and the main features can be reached by the button task bar.

#### 5.1 Recording Protocol

Registers have to be recorded in a simultaneous way in order to reach a perfect synchrony.

<b>STAGE 1st</b>	Electroglottographic Signal Voice Signal
<b>STAGE 2nd</b>	Electroglottographic signal Voice signal Videoendoscopic signal recorded with a telelaringoscope
<b>STAGE 3rd</b>	Electroglottographic signal Voice signal Air flow signal Videoendoscopic signal recorded with a fibroscope

**Table 1:** The three different recording stages

There are three different recording stages. Each one simultaneously records a different combination of signals. In **Table 1** the different stages and the different signals to be recorded are shown.

Each stage must be treated independently. The application provides utilities for screening the different signals in the different recording stages. It also provides several controls for adapting the signal level and to avoid saturation.

It is very important to screen the signals in the three stages before recording. Screening helps the operator to adjust the different levels of the signals and helps to adjust all the parameters of the system. E.g.: wrong contact between electrodes, the lenses becoming misty, lenses are wrongly focused.

## 5.2 Configuration Options

We use this option set to configure the hardware installed and the parameters that can be adjusted in the interface. We can configure the audio boards (sample rate, number of quantisation bits...), the video board (compression rate, number of frames...), printers, fonts, colours, parameters for screening and adjusting the signal level, etc. Default: minimum compression rate in video board, and maximum sampling rate in audio boards.

## 5.3 Dealing with Sessions

We define a session as the recording of all the signals in the three defined different stages. A session also stores textual information: ID number, S.S number, christian, surname, date of birth, exploration date, doctor that performs the exploration, diagnosis, treatment, medical information, etc. These data are requested by the operator just at the moment that you create a new session, and must be completed by the specialist with both the diagnosis and treatment. Such data are stored in a local database. The christian name and the surname name are used as a key to identify a session.

Once the session is completed, it must be stored definitively in a high storage device (optical disk or CD-ROM).

The system allows the user to create, open, modify and delete sessions. So, a session is something atomic and the operator does not need to worry about the complexity of the created file system: names and localisation of the files and directories needed to create a new session are automatically assigned by the software.

## 5.4 Hardware Interconnection

	STAGE 1 <sup>st</sup>	STAGE 2 <sup>nd</sup>	STAGE 3 <sup>rd</sup>
1 <sup>st</sup> Audio board	Microphone input: voice signal. Line input: electroglottogra phic signal	Microphone input: voice signal. Line input: electroglottogra phic signal	Microphone input: voice signal. Line input: electroglottogra phic signal
2 <sup>nd</sup> Audio board	-----	-----	Microphone input: voice air flow signal.
Video board	-----	Video input: videoendosco pic signal.	Video input: videoendosco pic signal.

**Table 2:** Board configurations.

Most computers do not have four free slots to keep four digitalisation boards. We minimise the number of audio

boards using the sound card mixer features: we mix microphone input and line input in a stereo file. So, we are recording two audio monophonic signals (one each channel) and we can record up to four monophonic signals with only two boards. Signals are stored in standard Windows™ files (\*.WAV for audio, and \*.AVI for video). The video signal is recorded with a video card. The configuration is shown in Table 2.

## 6 FUTURE WORK

A large amount of data is stored in a session (200-300 Mbytes). This is because the main that a session takes 10 to 15 minutes. But only a part of this large amount of data is interesting for the ENT specialist. So, we have to develop an integrated editor that allows the operator to cut and paste the most significant 5-10 seconds of the recordings in a simultaneous way and with the same temporal reference.

Once time signals are edited, they must be stored in a centralised remote database. We will develop a database for this purpose according to the client/server paradigm.

The system has to be complemented with automatic diagnosis techniques. We will apply digital signal processing to calculate several parameters of the voice signal, and we will develop algorithms for the automatic detection of pathologies.

Access via INTERNET: the system will use INTERNET allowing co-operative voice analysis, data and software exchange, and remote support of the diagnostics.

## 7 CONCLUSIONS

The main advantages of the system we are proposing are:

- Joint research: registers can be shared between specialists.
- ENT specialists can focus their work in diagnosis: recording may be performed by less qualified.
- Low cost: just a personal computer and three digitalisation boards are needed.
- Remote diagnosis: registers can be sent through the network, so different specialists could evaluate them.
- It allows the capturing of a wide database of pathological voices. This is the first step for researching in automatic diagnosis tools.
- Off line consultation.

## 8 ACKNOWLEDGEMENTS

This project has been financed and supported by the Health Ministerial of Spain (IMSERSO) : TER 96-1938-C02-01.

## 9 REFERENCES

1. R. García-Tapia, I. Cobeta. "Ponencia Oficial del XVI congreso nacional de Diagnóstico de los Trastornos de la Voz". Ed. Garsi.
2. Baer, T.; Titze, I.R.; Yosioka, H.: "Multiple simultaneous measures of vocal fold activity". San Diego, California: College Hill Press, 1983 p.p. 229-237
3. Teston, B.; Galindo, B. "A diagnostic and Rehabilitation aid workstation for speech and voice pathologies". ESCA Eurospeech 95 p.p. 1883-1886.
4. Weizhong Zhu, Yoshinobu Kikuchi. "An integrated acoustic evaluation system of pathologic voice." ICSLP94 p.p. 1983-1986.
5. Aguilera ,S; Pescador, F; Godino; J.I.; Rodriguez A." Improvement of a Spanish Speech Processing System" AAATE Conference 1997, p.p.. 115-119
6. M. Hirano; D.M. Bless, "Videostoboscopic examination of the larynx" Singular Publishing Group (1993)
7. Mateos, J.F.; Macarrón A.; Aguilera, S. "A P.C. card for rehabilitation of deficient auditive people" Proc. V European Signal Processing Conference (EUSIPCO-90), p.p. 1175-1178