EARLY DETECTION OF VOICE DISEASES VIA A WEB-BASED SYSTEM

F. Amato[1], M. Cannataro[1], C. Cosentino[1], A. Garozzo[2], N. Lombardo[2], C. Manfredi[3], F. Montefusco[1], G. Tradigo[1], P. Veltri[1]

[1]School of Biomedical Engineering, Università degli Studi Magna Graecia di Catanzaro, Catanzaro, Italy
[2]School of Otorhinolaryngology, Università degli Studi Magna Graecia di Catanzaro, Catanzaro, Italy
[3]Department of Electronics and Telecommunications, Università degli Studi di Firenze, Firenze, Italy

Abstract: **Voice is the result of the coordination of the whole pneumophonoarticulatory apparatus. The analysis of the voice allows the identification of the diseases of the vocal apparatus and currently is carried out from an expert doctor through methods based on the auditory analysis. The paper presents a web-based system for the acquisition and automatic analysis of vocal signals. Vocal signals are submitted by the users through a simple web-interface and are analyzed in real-time by using state-of-the art signal processing techniques, providing first-level information on possible voice alterations. The system offers different analysis functions to the doctors that may analyze suspected cases in detail. The system is currently being tested in the otorhinolaryngologist setting to carry out mass prevention via screening at a regional scale**.

Keywords : **Voice Analysis, Otorhinolaryngology**

I. INTRODUCTION

Voice is the result of a complex mechanism involving different organs of the pneumophonoarticulatory apparatus. In particular, it is the result of the vibration of the upper part of the mucosa covering the vocal cords. Such vibration determines the production of a sound, the larynx-fundamental tone, that is enriched by a set of harmonicas, generated by the resonance cavities in the upper part of the larynx. Any modification of this system may cause a qualitative and/or quantitative alteration of the voice, defined as dysphonia. Dysphonia can be due to both organic factors (organic dysphonia) and other factors (dysfunctional dysphonia).

Dysphonia is one of the major symptoms of benign laryngeal diseases, such as polyps or nodules, but it is often the first symptom of neoplastic diseases such as laryngeal cancer as well. Spectral "noise" is strictly linked to air flow turbulences in the vocal tract, mainly due to irregular vocal folds vibration and/or closure, causing dysphonia. Such symptom requires a set of endoscopic analysis (by using videolaryngoscope, VLS) for accurate analysis.

However, clinical experience has pointed out that dysphonia is often underestimated by patients, and sometimes even by family doctors. As widely reported in literature [1, 2], an early detected glottis tumour (T1, T2 stadium) can be solved in 100 % of cases with surgical intervention. Thus, the screening of voice alteration is extremely important in larynx diseases.

Several experiences of using algorithmic approaches for the automatic analysis of signals exist. Software tools (commercial and freely available) allow manipulating voice components in an efficient way (e.g. WinPitch[1], VOICEBOX[2]) and permits specialists to manipulate and analyze voice signals. Many automatic systems are based on voice signal processing whereas others combine signal processing with machine learning and data mining algorithms. The problem is that most of them are usable only locally and none of them offers remote collection and analysis as well as storing in central data bases for further use. The system described in [3] is one of the few remote data analysis systems. The problem is that voice is loaded by using telephone standard, which is known having low signal quality that decreases quality of classification.

However, in our knowledge, no systems of remote screening is available, that allows setting up a data base of voice signals, at the same time giving disabled patients a simple test for voice screening, without the need of moving to the laboratory.

The paper presents the architecture and the first implementation of REVA (Remote Voice Analysis), a web based system for the acquisition and automatic analysis of vocal signals. The system consists of a client module where a user, after registration is driven into a test phase where voice signal is registered, after a verification of the minimum hardware requirements. The voice signal, cleaned from noises, is sent through the Internet to the remote server which is in charge of analyzing it; the server will return to the client the signal analysis results and the possible voice anomalies will be related to potential diseases. After testing in the University of Catanzaro Hospital, the system will be

[1] http://www.winpitch.com/
[2] http://www.ee.ic.ac.uk/hp/staff/dmb/voicebox/voicebox.html

Models and analysis of vocal emissions for biomedical applications: 5th international workshop: December 13-15, 2007: Firenze, Italy, ed. by C. Manfredi,
ISBN 978 88-8453-673-3 (print) ISBN 978-88-8453-674-7 (online)

finalized for diagnostics in the otorhinolaryngologist setting, in particular to carry out mass prevention via screening at a regional/national scale.

The rest of the paper is organized as follows. Section 2 describes the system architecture. Section 3 presents the first prototype implementation. Section 4 points out the benefits and the Section 5 concludes the paper and sketches future work.

II. SYSTEM ARCHITECTURE

The REVA system employs a client/server architecture deployed as a web based application.

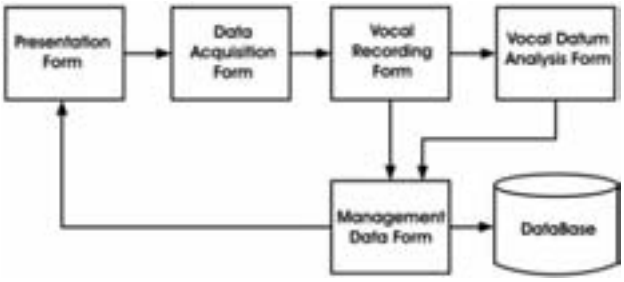

Fig. 1 REVA Architeture

The main modules of the system are shown in Fig. 1:

1. The Presentation Module represents the web interface between the system and the user. It is used to allow interaction with both the final users and the doctors. It contains the system description and the disease description. Its main tasks consist of giving the instructions for the system use and returning the analysis result to the user. Moreover, a specialized interface for the doctors is also provided.
2. The Data Acquisition Module is in charge of managing user personal data. After data is collected, the user is guided to the voice recording phase.
3. The Vocal Data Registration Module acquires the vocal samples and, after checking whether they are suitable for the analysis, sends the audio files to the server.
4. The Vocal Data Analysis Module, after the audio file has been received, extracts key signal parameters and performs the analysis for classification. It returns the result to the Administrator module.
5. The Data Administrator Module saves the data in the database and generates the response for the Presentation Module. The response is also sent by e-mail.
6. The Database Module contains data acquired through client interface. Voice signals are stored both in the raw data format as well as in a preprocessed format, where the main parameters related to the signal are stored.

III. PROTOTYPE

The REVA system has been implemented by using the Java technology. In particular the client is implemented through a Java Applet, while the server functions have been implemented by using the Java Server Pages. The database is implemented by using the open source relational MySQL DBMS.

In the following a brief description of the system functionalities is provided, both from the client and server sides.

A. Client module

The minimal system requirements for the remote user consist of a PC with Internet connection, web browser, audio card and microphone. The user visits the web site where the remote diagnostic service is available. The main page of the site provides a detailed description about the service offered, the scopes and effectiveness of the service itself.

To make a test, the user accesses the registration page entering his/her data and other information useful for the diagnosis. When the registration phase is completed, the user enters the testing phase, accesses a new page (see Fig. 2), which drives him/her through the acquisition of vocal samples.

Fig. 2 Patient view: voice recording.

The file containing the audio registration is analysed on the client, for a preprocessing phase (for instance, to exclude empty, inconsistent or too long files or to reduce noise). If the registration is validated, the audio file is sent to the server through the Internet. The result of the signal analysis is transmitted to the user both via a new webpage and via e-mail.

Note that patient's personal (name, surname, etc.) and clinical data are collected into an XML document that is sent to the server together with the audio file. Metadata are stored with audio files into the database such that for those patients periodically accessing to the service, the medical specialist and the system are able to monitor and analyze the voice signals.

B. Server module

The server hosts a listener process waiting for the connections of the remote users. It receives from the client both an XML file, containing the metadata of users asking for service, and the related audio files (in WAVE format), obtained from registration. Vocal files and metadata are archived into the database.

The server executes a preliminary elaboration of the signal (preprocessing) to extract from the audio sample various information useful for classification. At this point the classification procedure of the vocal signal is run by using the parameters defined in a preliminary phase with the doctors and based on their experiences and using a statistical study of available samples (see the next subsection). For a returning user a comparison with the previously registered samples is foreseen, to evaluate the temporal evolution of the user voice.

On the server side a different web interface allows doctors and specialists to analyze the stored voice samples. Data coming from user submissions are automatically stored into the database where a simple Electronic Patient Record (EPR) stores voice samples, signal parameters, metadata and information about patients. Using such interface the doctor can:

- visualize the last entered voice samples requiring attention;
- load, listen and compare voice signals (e.g. for patients that had a surgery intervention);
- analyze them with the implemented voice analysis module (see Fig. 3).

C. Vocal signals analysis techniques

The classification of the vocal samples requires a suitable elaboration, to extrapolate from the audio registration a set of significant parameters. For such a purpose, computations are usually performed mapping signal data into the frequency domain [4].

The main parameters of clinical interest, considered for the evaluation, are:

- Fundamental frequency tracking (linked to laryngeal and vocal folds pathologies) as well as irregularities in vocal folds oscillation (jitter and shimmer).
- Measures of dysphony (voice quality indexes, based on "noise" estimation, as caused by irregularities and pathologies producing turbulences in the air flow from the glottis).

The pitch estimation is performed via two approaches which use respectively the Average Magnitude Difference Function (AMDF) and Simple Inverse Filter Tracking (SIFT) [5, 6].

In the first approach the estimate of the fundamental frequency value (f_0) is obtained by filtering the signal with a proper Continuous Wavelet Transform (CWT) and extracting its time periodicity by means of the AMDF method [7].

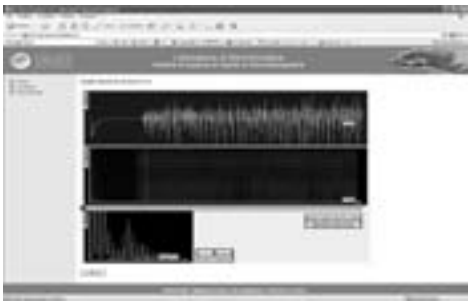

Fig. 3 Doctor view: spectral analysis of voice samples.

Given a signal frame of length M: $\{x(k)\}, k = 0,\ldots,M$, the AMDF is defined as:

$$AMDF(\eta) = \frac{1}{M-\eta}\sum_{i=1}^{M-\eta}[x(i) - x(i+\eta)], \quad \eta = 0,\ldots,M-1$$

The scale factor $(M-\eta)^{-1}$ eliminates the decreasing trend of the AMDF method, due to the truncated sum. For noisy signals, the AMDF minimum is usually greater than zero. Hence, in order to recover f_0, one has to select the η value that gives the minimum of the AMDF function ($\eta = F_s / f_0$, where F_s is the sampling frequency [7]).

The method is appealing, due to the low computational burden, but it is more sensitive to the noise than other approaches.

In the second approach, which relies on Linear Prediction (LP) analysis of data, the vocal tract is described through an Auto Regressive (AR) model. The following procedure is implemented on each data frame of length $M = 1/F_{\inf}$, where $F_{\inf}$ is the lowest value in the frequency range of interest for f_0:

- estimation of the correct order p of the model AR by means of Singular Value Decomposition (SVD) approach;
- computation of the AR parameters, which enable to determine the varying vocal tract inverse filter IF, through the forward – backward algorithm [8];
- estimation of the residual sequence by applying the signal to the filter IF;
- band – pass filtering of the residual sequence in the range 50 – 1.5 KHz and evaluation of the maximum of the autocorrelation sequence (AS) of the residuals in the frequency range of 60 – 250 Hz ($f_0 = F_s / \tau$, where τ is the index corresponding to the maximum of the AS).

The computational complexity is rather high, but this procedure is one of the most robust and accurate.

A measure of the dysphonic component of the voice spectrum related to the total signal energy is evaluated by using the NNE index [9]. Given the speech signal $x(n) = s(n) + w(n)$, where $s(n)$ is the periodic component

and $w(n)$ is the additive noise component, let $X(k)$, $S(k)$ and $W(k)$ be the discrete fourier transform (DFT) of $x(n)$, $s(n)$ and $w(n)$, respectively. The adaptive NNE (ANNE) is defined as

$$ANNE(k) = 10\log\left[\frac{\sum_{m=N_L}^{N_H}\left|\widetilde{W}_m(k)\right|^2}{\sum_{m=N_L}^{N_H}\left|X_m(k)\right|^2}\right], \quad k = N_L,\ldots,N_H$$

where $N_L = [Nf_L T]$, $N_H = [Nf_H T]$, $N =$ number of DFT points, $L =$ number of frames in the analysis interval, and f_L and f_H respectively the lowest and the highest frequencies of the frequency band of interest. $\left|\widetilde{W}_m(k)\right|^2$ is an estimate of the unknown noise energy $\left|W_m(k)\right|^2$, $\left|X_m(k)\right|^2$ is the signal energy and T is the sampling period. At lower ANNE values, the noise energy is larger on that signal frame. The signal is more noisy for ANNE values close to zero.

The voice signals analysis was implemented using the software Matlab.

IV. Discussion

The main goal of the proposed system is the realization of a web based system for the acquisition and automatic analysis of vocal signals. It is important to remark that the goal of the proposed instrument is neither to replace the doctor specialist, nor to provide a diagnosis; rather it is aimed to give a response about the potential presence of pathologies of the larynx or the vocal tract, and to advise potentially affected patients to go to a specialist for an accurate voice control.

The possibility to produce in a simple and rapid way the detection of voice alterations for a possible huge amount of users, is one of the main requirements of the system. This is an important goal required and suggested by clinical experiences, where patients with voice anomalies often delay specialist's controls, in most cases limiting treatments effectiveness. Thus, the idea behind the system raises from the need of educating patients to the auto diagnosis by using a simple, remotely accessible and user friendly system.

The system will be made completely and freely (prior to free registration) accessible from a web portal. This solution offers several advantages:

- elimination of the discomfort due to time and/or distance constraints, that often induce the patient to indefinitely postpone the specialist's visit;
- removal of a possible psychological block in presence of the doctor (due, for example, to the fear deriving from a possible investigation via endoscope);
- since the system can be freely accessed on the Internet, even the less wealthy patients may use it, also when the suspect of a pathology is very light.

The use of the client-server system would allow a diagnostic analysis from the client (patient) side and, at the same time, will allow populating a national scale database containing several types of vocal anomalies.

V. Conclusion

The paper presented a web-based system for the remote acquisition and automatic analysis of vocal signals. Vocal signals are submitted by the users through a simple web-interface and are analyzed in real-time by using state-of-the art signal processing techniques, providing first-level information on possible voice alterations.

Future work will regard the experimentation of the system in the Department of Otorhinolaryngology of our University for full clinical validation and for post-surgery control, i.e. for checking the status of patients after surgical intervention and during follow-out.

References

[1] J. C. Stemple, L. E. Glaze, and B. K. Gerdemann, *Clinical Voice Pathology: Theory and Management*, Thomson Delmar Learning, 2000.

[2] S. W. Smith, *The Scientist and Engineer's Guide to Digital Signal Processing*, California Technical Publishing, 1997.

[3] R. J. Moran, R. B. Reilly, P. de Chazal, and P. D. Lacy, "Telephony-based voice pathology assessment using automated speech analysis," *IEEE Trans. Biomed. Eng.*, vol. 53, no. 3, pp. 468-477, 2006.

[4] K. Umapathy, S. Krishnan, V. Parsa, and D.G. Jamieson, "Discrimination of pathological voices using a time-frequency approach," *IEEE Trans. Biomed. Eng.*, vol. 52, no. 3, pp 421-430, 2005.

[5] C. Manfredi, M. D'Aniello, P. Bruscaglioni, and A. Ismaelli, "A comparative analysis of fundamental frequency estimation methods with application to pathological voices," *Med. Eng. Phys.*, vol. 22, no. 2, pp. 135 – 147, 2000.

[6] C. Manfredi, and G. Peretti, "A new insight into postsurgical objective voice quality evaluation: application to thyroplastic medialization," *IEEE Trans. Biomed. Eng.*, vol. 53, no. 3, pp 442-451, 2006.

[7] J. R. Deller, J. G. Proakis, and J. H. L. Hansen, *Discrete-time Processing of Speech Signals*, New York: Maxwell McMillan, 1993.

[8] Marple SL., *Digital spectral analysis with applications*, Englewood Cliffs, NJ: Prentice-Hall, 1987.

[9] H. Kasuya, S. Ogawa, K. Mashima, and S. Ebihara, "Normalized noise energy as an acoustic measure to evaluate pathologic voice," *J. Acoust. Soc. Amer.*, vol. 80, no. 5, pp. 1329–1334, 1986.