A CLINICAL WORKSTATION SOFTWARE FOR VOICE QUALITY ASSESSMENT

J. Cai¹, A. Alpan¹, T. Dubuisson², I. Verduyckt³, F. Grenez¹, J. Schoentgen¹,⁴

¹ Laboratory of Images, Signals & Telecommunication Devices, Université Libre de Bruxelles, Brussels, Belgium
² Service de Théorie des Circuits et Traitement du Signal, Faculté Polytechnique de Mons, Mons, Belgium
³ Service ORL-ORLO, Cliniques Universitaires UCL de Mont-Godinne, Yvoir, Belgium
⁴ National Fund for Scientific Research, Belgium

Abstract: This paper presents the design and implementation of a clinical workstation software for analyzing voice disorders. The software is developed by using Java technology and MySQL database system. A variety of vocal cues, e.g. jitter and shimmer, that describe irregularities of speech cycles in sustained vowels can be automatically derived by the system. For assessing voice disorders in connected speech, a vocal cue called signal-to-dysperiodicity ratio is evaluated by carrying out a generalized variogram analysis. In the development, special attention has been paid to software engineering conventions and the principles of architectural design of software structures to achieve good quality attributes such as developmental simplicity and modifiability. Preliminary tests have shown that the system provides satisfactory usability and performance for clinical applications.

Keywords: Pathological voice assessment, disordered voice analysis, software engineering, Java application

I. INTRODUCTION

Disordered voice timbres are usually caused by improper vibrations of the vocal folds, as a consequence of pathological changes of the larynx. Recently, voice disorders have been observed more frequently and more extensively than before because of an increasing number of professional voice users. Therefore, reliable and efficient means of evaluating pathological voice quality are required for the assessment and prevention of laryngeal problems.

In clinical voice evaluation, acoustic assessment methods have been used to facilitate the clinical documentation of vocal problems because previous experiments have established that vocal cues exist that are clinically relevant [1, 2]. Furthermore, these acoustic-based methods have the advantages of non-invasiveness and quantitativeness. In this paper, the design and implementation of a clinical workstation software for analyzing pathological voice signals are presented. A variety of vocal cues such as jitter, shimmer, and harmonics-to-noise ratios in temporal and spectral domains, which describe irregularities of speech cycles in voiced speech can be automatically obtained for sustained vowels by means of the system. In addition, connected speech quality assessment is also included. The reason is that in clinical practice, people consider connected speech to be more informative than sustained vowels. Moreover, the perceptual evaluation of voice quality is likely to be based on both connected speech and sustained vowels uttered by the same patient. Variogram-based analysis is carried out to track dysperiodicities in connected speech, and thus a signal-to-dysperiodicity ratio value is obtained as a vocal cue of voice disorders [3].

Graphical means, e.g. spectrogram, phonetogram, and spider charts, are available to visualize the analysis results. Java technologies have been utilized to build the application system, mainly for the purpose of facilitating portability on different operating-system platforms. Software engineering conventions and the principles of architectural structures design have been used to guide the design and development of the system, to achieve developmental simplicity, modifiability, and other quality attributes.

II. METHODS

A. Disordered Voice Analysis

Because of technical feasibility, voice analysis is usually performed by providing different vocal cues of voice disorders for sustained vowels only. In this clinical software, the voice disorder analysis is versatile in terms of that not only sustained vowels but also connected speech segments can be assessed. For sustained vowels, features such as mean and standard deviation of fundamental frequency, jitter, shimmer, and harmonics-to-noise ratios in both time and spectral domains [4, 5] can be obtained to describe voice disorders.

A technique called speech sample salience analysis [6] has been used to perform voice cycle detection. Conventional voice cycle detection relies on the selection of signal peaks from several candidates. The selection technique usually assumes that the signal peaks are regularly spaced in time so that they can be determined one by one based on an a priori estimation of the typical
fundamental period. This assumption does not apply to disordered voices, though it is valid for modal ones.

Speech salience analysis can be performed without any a priori knowledge of the typical cycle length. Thus, it is well suited for tracking vocal cycles in pathological voices. For a speech signal $x(n)$ of $M$ samples, the salience $S(k)$ of the $k$th sample ($0 \leq k < M$) is defined as the length of the longest interval within which that sample is a maximum. Based on a sliding analysis window technique, which eliminates the bias related to the arbitrary position of the signal origin, salience analysis is performed sample by sample to determine a speech cycle sequence which minimizes the standard deviation of the durations of all cycle candidates. Based on this sequence, jitter, the vocal cue for describing the small random perturbation of voice cycle lengths, can be calculated.

For assessing the voice quality in connected speech, the segmental signal-to-dysperiodicity ratio (SDR) [7, 8], which is based on generalized variogram analyses, is used to summarize vocal perturbations. For a stationary signal $x(n)$, the variogram, which is defined in Eq. (1), is a measure of the departure from periodicity over an interval of length $T$:

$$v(T) = \sum_{n=0}^{N-1} \left( x(n) - x(n-T) \right)^2,$$

where the variable lag $T$ satisfies $-T_{\text{max}} \leq T \leq -T_{\text{min}}$ and $T_{\text{min}} \leq T \leq T_{\text{max}}$. $T_{\text{min}}$ and $T_{\text{max}}$ are, in number of samples, the shortest and longest acceptable glottal cycle lengths. They are fixed to 2.5 ms and 20 ms, respectively (i.e., 50 Hz $\leq f_0 \leq$ 400 Hz). For voiced speech sounds, the lag $T_{\text{opt}}$, which minimizes (1), is interpreted as a multiple of the speech cycle length.

Speech signals are expected to be locally stationary at best. A weighting coefficient can be inserted to account for slow changes in signal amplitude. Therefore, the variogram computation defined in Eq. (1) can be modified as follows:

$$v(T) = \sum_{n=0}^{N-1} \left( x(n) - \alpha x(n-T_{\text{opt}}) \right)^2,$$

where the coefficient $\alpha$ is defined so as to equalize the signal energies in the current and shifted analysis windows:

$$\alpha = \frac{E}{\sqrt{E_T}},$$

where $E$ and $E_T$ are the signal energies of the current and the lagged frames. The frame length $N$ and frame shift length are set to the value of 2.5 ms, which guarantees that each signal frame is included exactly once in the analysis. The instantaneous value of the dysperiodicity is estimated as follows:

$$\sigma(n) = x(n) - \alpha x(n-T_{\text{opt}}), 0 \leq n \leq N-1$$

$$\sigma_{\text{max}} = \max_{0 \leq n \leq N-1} \sigma(n)$$

where $T_{\text{opt}}$ is equal to the lag which minimizes the variogram for the current frame position. For a given fragment of connected speech, the analysis interval is divided into $K$ blocks of length $M$ and the SDR of each block of length 20 ms can be computed as follows:

$$\text{SDR}(k) = 10 \log_{10} \frac{\sum_{n=M(k-1)}^{Mk} x^2(n)}{\sum_{n=M(k-1)}^{Mk} \sigma^2(n)}, 0 \leq k \leq K-1$$

### B. Software Architecture Design

The objective is to develop a workstation application system which runs on a normal PC platform and supports pathological voice quality assessment. Modifiability, usability, and portability are the software quality attributes which are emphasized throughout the development of this system. To achieve portability, Java programming language and MySQL database system (MySQL Community Server) are selected as the development infrastructure because of their cross-platform availability.

**Fig. 1 Module decomposition View of the System**

Architecture Tradeoff Analysis Method [9] is employed to devise the architectural structures of the system. To guarantee the clinical usability of the system, different stakeholders in the system’s development, including developers of disordered voice analysis methods, otolaryngologists, speech therapists, and software engineers, have been involved in designing the functionalities. Based on the use case views [9] which specify the behavior of both the system and the users, a module decomposition view (see Fig. 1) is determined to assign the functionalities to different modules of the system. All the functionalities of a particular data-processing type are organized in a specific module. For
instance, the computation and graphical display of different vocal cues of voice disorder are grouped into the Technical Assessment Module.

A layered architecture is adopted to allocate the functional modules into two tiers, shown in Fig. 2. The data access logic implemented by the Database Accessing Module is contained in a separate Data Tier which is dedicated to database communication. Above the tier, there is the Application Component Tier that packs all the other modules carrying out the presentation logic and application-relevant logic. This layered architecture forms the basis for the development project’s organization; i.e., the source code files are organized into packages according to the module decomposition and the layered architecture. The module decomposition also ensures that possible changes of the system are localized to only one or a few small modules, enabling a large part of the system’s modifiability.

![Fig. 2 Two-Tier Software Architecture](image)

**C. Functionality Design**

The Identity Module is used to maintain the identity information of ORLs, speech therapists, and patients. Also, the information related to each clinical visit, e.g. the diagnosis, the type and state of the therapy, can be recorded by using this module. Fig. 3 shows the user interface for visit registration. Since a clinical assessment is typically performed at each clinical visit, each recorded voice and the relevant assessment results are uniquely associated with a specific visit. Therefore, the registration of visit information is designed as the first step which triggers the other events in the assessment, as well as creates a data slot in the database to store the visit information and the assessment results. To achieve flexible data management, a Patient Search sub-module is designed to enable the user to find the clinical data of patients by using different search conditions such as name, pathology, or a certain ORL.

A specific module, the Non-Technical Assessment Module, is designed to facilitate the patient interview and perceptual voice quality evaluation, which constitutes the beginning of the assessment process. To inquire about the case history, a questionnaire format [10] has been chosen to perform the interview. A VHI (Voice Handicap Index) questionnaire sub-module is designed to facilitate the assessment of the patient’s perception of discomfort, handicap, and distress resulting from voice difficulties. For singing voice and speaking voice, different sets of questionnaires have been designed to derive the VHI score.

![Fig. 3 User Interface for Visit Registration](image)

The Voice Processing Module enables the user to record patients’ voices, including sustained vowels and connected speech fragments, with a high quality microphone and a built-in sound card which performs 16 bit, 44.1 kHz sampling. All the voice signals can be stored as files in WAV format. With the toolbox in this module, the user also can playback the recorded voice and view its waveform or spectrogram. Voice samples can be selected for subsequent computation of the vocal cues of voice disorder.

All the vocal cues are computed in the Technical Assessment Module. For a particular voice sample, seven vocal cues can be computed the values of which can be visualized by means of spider charts. The software can display the spider charts for two different voice samples in the same axes system in an overlapped manner. Fig. 4 depicts such an example. The spider chart of the voice sample for the pre-therapy assessment of a patient is plotted in dark color, while the spider chart of another voice sample related to the post-therapy assessment of the same patient is plotted in a bright color. By using overlapped spider charts, it is easy to compare the post-therapy vocal cue values with the pre-therapy ones for a certain patient, and to compare the vocal cue values of two voices from different patients. Therefore, this kind of display provides a useful means of evaluating the effectiveness of therapy. Besides the spider chart, another graphical tool – the phonotgram – is implemented in the system to depict the dynamic ranges of both the pitch and the intensity of the voice.
III. RESULTS

Java programming language has been used to build the software for the sake of its portability to embedded systems. MySQL Community Server has been chosen as the database management system because of its speed, flexibility and reliability. The system is organized in a layered structure which therefore supports the modifiability of the system and the work assignment in the development team. The ergonomic design of the system makes the user interface easy to use owing to the conventional sequencing of the tasks which clinicians are expected to perform when assessing laryngeal function. All the acoustic cues of disordered voices and the relevant diagrams can be generated in real-time or quasi-real-time by the software running on a normal PC platform. Preliminary tests have shown that the system has a satisfactory usability and performance, though further clinical tests and development remain to be carried out to establish its suitability for pathological voice assessment.

IV. CONCLUSION

The Java-based voice assessment workstation software can be deployed on almost all PC platforms. A variety of vocal cues of voice disorders can be provided by the system for both sustained vowels and connected speech to support the clinical voice quality assessment. By using the principles and methods of architectural structures design in the development of the system, the quality attributes of the software, such as the developmental simplicity, modifiability, and usability can be well achieved.

ACKNOWLEDGEMENTS

This work was supported by the “Région Wallonne”, Belgium, in the framework of the “WALEO II” program.

REFERENCES