Biomedical Signal Processing and Control 40 (2018) 318–323

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

journal homepage: www.elsevier.com/locate/bspc

Acoustic analysis of voice signal: Comparison of four applications software

Susana Vaz-Freitas[a,b,c,*], Pedro Melo Pestana[d], Vítor Almeida[e], Aníbal Ferreira[e]

[a] Speech Pathology Unit of Otolaryngology Service – Centro Hospitalar do Porto, Portugal
[b] Speech Therapy Department – Faculty of Health Sciences, Universidade Fernando Pessoa, Porto, Portugal
[c] Biomedical Engineering Department – Faculty of Engineering, Universidade do Porto, Portugal
[d] PMP terapia – Largo do Pelourinho 11, 4740-214 Esposende, Portugal
[e] Electrical and Computer Engineering Department – Faculty of Engineering, Universidade do Porto, Rua Doutor Roberto Frias, 4200 Porto, Portugal

ARTICLE INFO

Article history:
Received 7 December 2016
Received in revised form 7 September 2017
Accepted 30 September 2017

Keywords:
Voice
Dysphonia
Acoustic
Analysis
Software
Correlation

ABSTRACT

Objectives: To describe the results of the acoustic analysis of a database of 90 voice samples with distinct dysphonia levels, using four different – commercial and open source – software programs.
Study design: Exploratory, transversal.
Methods: The samples were analyzed by four different types of software programs that perform acoustical evaluation – one open source software (Praat) and three commercial ones (*Multi Dimensional Voice Program* – MDVP by Kay Elemetrics; *VoiceStudio* by Seegnal; and *Dr. Speech* by Tiger Electronics) – for comparison among the most commonly used acoustic measures (frequency, perturbation and noise measures).
Results: There is a moderate to strong correlation, positive and statistically significant among the software programs. The mean F0 is not statistically different among the used applications. The other acoustic measures revealed statistically significant differences.
Conclusion: Even though it is easier to access software programs and there are numerous proposals for acoustic measures, not all of them are statistically representative nor have numeric semblance among the different applications.

1. Introduction

The voice, as well as its disorders, is multifaceted. The complaints of the patients who usually search for clinical support are the bottom line of the therapy process.

Usually, the patient describes his voice as hoarse and it is up to the clinicians to understand the etiology and the best and fastest way to normalize the situation. Along with therapy, a complete evaluation is mandatory, so that the diagnosis can be established.

Voice is a phenomenon that implies several variations [1] and depends on the complex activity of the muscles and the integrity of structures of the phonatory apparatus [2]. It is produced in the vocal tract; its onset is in the larynx, caused by air passage through the vocal folds and its modification is by movements of the phono-articulatory system [3].

The concept of voice quality is directly related to physiological, perceptive and acoustic characteristics [2–5]. Dysphonia or voice disorder is a pathological condition of oral communication, in which the voice does not fulfill its aim of transmitting verbal and emotional messages [2].

Taking these aspects altogether, it is clear that it is difficult to find a single method that can assess voice quality and its variations. A multifactorial analysis is necessary so that a broad knowledge can be built regarding the laryngeal function and quality of the voice signal [4,6–8].

In addition, the voice quality analysis relies on the signal recording characteristics. Titze [9] published a paper based on the *Workshop on Acoustic Analysis*, in which he suggests three types of sound signals. Recently, a fourth signal was added which corresponds to a stochastic noise signal [10].

The acoustic analysis – when used in the assessment and therapy of voice disorders – allows the quantitative characterization of rel-

* Corresponding author at: Serviço de Otorrinolaringologia, Centro Hospitalar do Porto, Largo Prof. Abel Salazar, 4099-001 Porto, Portugal.
E-mail addresses: svazfreitas@gmail.com (S. Vaz-Freitas), melopestana@gmail.com (P.M. Pestana), vitoralmeida.et@gmail.com (V. Almeida), ajf@fe.up.pt (A. Ferreira).

https://doi.org/10.1016/j.bspc.2017.09.031
1746-8094/

evant aspects of human voice perturbations in a non-invasive way [2], [11–16]. The acoustic method provides the so-called objective measures, which are extracted from the data, or not, in an automatically or manually assisted way using computer processing. The method facilitates the association between audio perceptive assessment and voice physiology. It also promotes a comprehensive knowledge about the voice production process: It gives an indirect estimation of the vibratory patterns of the vocal folds, the format of the supra glottal tract and its modification [17–19]. That is achieved by analyzing the most relevant acoustic parameters of the signal – periodicity, amplitude, duration and spectral composition [4,20] – which characterize the voice's physical attributes in time, frequency and sound pressure domains. It may also involve other complex measures that intersect all of these domains [18,21–24].

The reliability of the collected and analyzed information may be influenced by the confusion variables that are difficult to control in the clinical setting, such as recording conditions and procedures, storage conditions, and editing and analysis of the voice signal [4,18,21,23,25]. These factors must be controlled because the acoustic measures should reflect the true complex interactions between the glottal source and the resonance cavities of the vocal tract. This means that they depend on the biomechanical and aerodynamic forces of the larynx and supra glottal structures [3,26,27], as well as of the cortical neuro-motor control. If these components have anatomical and/or physiological disorders, the obtained results will be different from the expected results for a normal voice, and should be representative of vocal pathology and its severity [28–35].

Briefly, the objective measures of vocal quality [18,21,25] are also affected by the confusion variables, including: the conditions when recording the data, the specific characteristics of the hardware and software system, the protocols of voice recording and analysis, the acoustic and aerodynamic individual variability and the severity and type of voice disorder [11,19,36–39].

In this paper, we will consider speech as a product of the voice source, noises and resonance caused in the vocal tract, which is processed in the auditory system and integrated by the central nervous control [5].

In Portugal, the first software and hardware solutions for the voice signal assessment and analysis emerged in the 90's. Nowadays, the available brands and types of equipment are numerous. The inherent costs and future usability of the equipment should also be taken into consideration. On the other hand, the acoustic measures depend strongly on the software used. There are several voice analysis applications (open source and commercial), which differ in a number of aspects, mainly the processing speed, compatibility, available acoustic measures and associated costs. However, the use of this kind of software or equipment by Portuguese clinicians is increasing, mostly because of its advantages, some of which are given below [2,5,15,20,21,40]:

a) It allows a wider understanding of the acoustic voice output and promotes the integration of multiple assessment methods (audio perceptive and acoustic; laryngostroboscopy and acoustic);
b) It enables – quickly and easily – normative data for distinctive voice patterns (professional, linguistic and clinical);
c) It provides important data about the impact of the voice signal in the listener;
d) It delivers graphical and numerical data that are important not only to describe the vocal quality in the case of a professional voice user or a dysphonic patient under therapy, but also to support or confirm judicial or official forensic reports;
e) It provides images or graphs about the acoustic signal that are easily understood by the patient/speaker under assessment or therapy, allowing a better prognosis because of the commitment and motivation along the therapy process;
f) It monitors the treatment effectiveness and allows comparing of voice results using different methodologies, in distinctive therapy phases or medical approaches (surgery or drugs);
g) It allows tracking the development of a professional voice and supports its adjustment through lifespan and professional needs, namely with the possibility of using real time acoustic feedback systems;
h) It aims to be an early detection tool of voice and laryngeal problems, such as in screenings, by comparing the obtained measures with the reference values.

Few papers have objectively compared more than two software programs. As far as the authors know this is the first paper comparing the results of four different software programs and the only one using *Voice Studio* [41–54].

The major goal of this paper is to compare the results provided by the different software platforms for voice analysis which calculate the F0 using algorithms based on the autocorrelation method – commercial and open source – for the same group of acoustic measures.

2. Materials and methods

A database of 90 voices was collected from real clinical setting recordings. The authorization was obtained from the Ethics Committee and the Hospital Board of Directors from the *Centro Hospitalar do Porto*, Portugal. From the initial database, 21 voice recordings were considered normal (23.30%) and 69 showed some degree of voice disorder: 20 (22.22% mild), 47 (52.22 moderate) and 2 (2.22% severe). The recordings and assessments were conducted by the first author. The gender distribution was 28% male (n = 25) and 72% female (n = 65). All the subjects were adults, more than 18 years old.

The authors used the same protocol to record the voice signals: 44100 Hz sample rate; 16 bits per sample; a unidirectional (cardioid) table microphone SBC ME 400; the room noise level was always below 40 dB regarding sound pressure level (SPL), even though it was not acoustically treated. The mouth-to-microphone distance was 10 cm and the sustained vowel [α] was asked to be reproduced after illustration [12,55,56], at least during five seconds, in two attempts [12,54,55]. The last vowel that was produced while the subject was standing was considered in this study. These records were usually conducted during the phoniatric appointments in the hospital of the first author using the Dr. Speech software, version 4.0 (Tiger Electronics). Each vowel was analyzed in a region after the 2nd second [12,54,56], because this was considered the most stable region of each signal.

The following acoustic parameters were extracted for each voice sample: MeanF0 (Hz), S.D. F0 (HZ), Jitter (local) (%), Jitter (PPQ5) (%), Shimmer (local) (%), Shimmer (APQ5) (%), Harmonic-to-Noise Ratio [HNR] (dB). Four acoustic software programs were under analysis: the open source Praat and the commercial ones, Dr. Speech (Tiger Electronics), Voice Studio (Seegnal) and Multi Dimensional Voice Program [MDVP] (Kay Elemetrics).

All the programs provide for automatic F0 extraction.

The sampling rate of 44100 Hz was the same for Dr. Speech and MDVP; it was then adjusted to 22050 Hz for Praat and Voice Studio, using a MatLab script, in order to ease the analysis. The most stable regions of the recordings were selected and pre-processed, so that the desired sampling rates were obtained.

The statistical analysis was conducted on *Statistical Package for the Social Sciences* – IBM® SPSS® for Windows, version 19.0.