Biomedical Signal Processing and Control 42 (2018) 134–144

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

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

A methodology for voice classification based on the personalized fundamental frequency estimation

Laura Verde [a], Giuseppe De Pietro [b], Giovanna Sannino [b,*]

[a] Department of Engineering, University of Naples "Parthenope", Centro Direzionale, Isola C4, Naples, Italy
[b] National Research Council of Italy, Institute of High Performance Computing and Networking (CNR-ICAR), Via Pietro Castellino, 111 Naples, Italy

ARTICLE INFO

Article history:
Received 30 May 2017
Received in revised form
21 December 2017
Accepted 17 January 2018

Keywords:
Voice disorders
Signal processing
Fundamental frequency analysis
m-Health system

ABSTRACT

Nowadays, the incidence of voice disorders is increasing rapidly, with about a third of the population suffering from dysphonia at some point in their lives. Dysphonia is a disorder that alters vocal quality and can impair and reduce the quality of life. The structural or functional alteration of the phonatory apparatus, unhealthy lifestyles or an excessive use of the vocal cords for work activities (e.g. teaching) can cause voice disorders. Unfortunately, people who suffer from dysphonia often underestimate its symptoms and therefore delay consulting a speech therapist for accurate voice assessment and treatment. Voice disorder evaluation involves a series of tests, including an acoustic analysis. This quantifies the measurements of voice quality through the evaluation of certain characteristic parameters, for example the fundamental frequency (F_0). In this paper, a personalized methodology for the estimation of the F_0 is presented. The personalization is accomplished by taking into account two of the main factors that influence the F_0, the gender and age of the subject. The estimation of the F_0 is crucial for the classification of the voice signal, because the discrimination of a healthy voice from a pathological one is achieved by evaluating the inclusion of the F_0 value within the healthy range. To evaluate the presented methodology, we have carried out a set of tests by using some voice signals selected from an available database in order to compare the classification ability of the proposed methodology with other algorithms existing in the literature. The numerical results obtained show that the proposed methodology provides a good accuracy, sensitivity, and specificity, respectively of over 77%, 72% and 81%, values better than those achieved by the most frequently other used and cited fundamental frequency estimation algorithms. Additionally, a statistical analysis to evaluate whether or not a statistically significant difference exists between the accuracy, sensitivity and specificity has been carried out. The outcome of the ANOVA tests and of the *t*-tests confirms that there is a significant difference between the proposed methodology and the other algorithms. Finally, the presented methodology could be embedded in a portable and simple m-health application that could be useful for the monitoring of the state of vocal health and the prevention of voice disorders.

1. Introduction

Dysphonia is a voice disorder that affects about 29% (considering a lifetime prevalence) of the population, impacting on their social, psychological and professional life [1], with in general a prevalence in the population aged 60 years or more [2]. It is an alteration of the sound structure of the voice that can be understood as a reduction in the acoustic energy, or a variation of the melodic component or tonal harmonic structure.

There are several potential causes of voice disorders. Although they may be hereditary, some voice disorders are caused by disease, for example by a common infectious illness (e.g. sinusitis or bronchitis), a chronic medical disease (e.g. sarcoidosis, hypothyroidism or rheumatoid arthritis), an inflammatory condition resulting, for example, from smoking or a neurological condition (e.g. vocal cord paralysis or Parkinson's disease). Usually the etiology of the voice disorder can be related to an incorrect use of the vocal cords, or an inappropriate speaking or breathing technique, possibly leading to voice alterations [3,4].

People who routinely use their voice in their jobs are particularly prone to vocal disorders. As reported in many studies [5],

* Corresponding author.
E-mail addresses: laura.verde@uniparthenope.it (L. Verde), giuseppe.depietro@icar.cnr.it (G. De Pietro), giovanna.sannino@icar.cnr.it (G. Sannino).

https://doi.org/10.1016/j.bspc.2018.01.007
1746-8094/

professional voice users have a greater possibility of contracting dysphonia, aphonia, edema, polyps or nodules, compared to non-vocal professionals (51.2% vs 27.4%). Teachers represent the main category of voice users seeking medical help for voice problems. Between 20% and 80% of teachers have reported various vocal symptoms, yet only 13.5% are aware of the risks of dysphonia during education or training activities. In fact, many teachers underestimate their vocal symptoms and do not necessarily take appropriate countermeasures.

In addition to its impact on health and quality of life, severely limiting communication at work and affecting all social aspects of daily activities, dysphonia results in the sufferer making frequent medical appointments with a consequent loss of productivity due to absenteeism which may undermine work functions. Teachers, for example, miss a significantly high number of working days throughout their career due to vocal problems. A European study [6] reports that one out of five teachers (19.2%) reported missing at least one day of work because of voice-related dysfunction. As reported in [7], approximately 10% of the U.S population has experienced dysphonia and it is estimated that the social cost incurred by problems such as unemployment caused by dysphonia amounts to $2.5 billion each year for the teaching profession alone.

One of the most prevalent voice disorders is Reinke's edema, a chronic, diffuse, inflammatory disease of the vocal folds [8]. As demonstrated in the study detailed in [4], Reinke's edema is the third most common cause of dysphonia with an incidence of 14%, and appears almost exclusively in chronic smokers older than 40. Its main impact is on women [9], causing the voice to become hoarse and low in pitch. Several studies [10,11] have confirmed the widespread prevalence of this disorder, a prevalence that has boosted the interest of the medical partner involved in this project, encouraging us to focus this study on voices suffering from this disease.

The lack of awareness of dysphonia, and of the causes of these disorders, is a potential barrier to appropriate treatment meaning that in some circumstances it is not possible to achieve a complete resolution [12].

Based on these considerations, this paper presents a personalized methodology, an improvement of the currently existing one, described in [13], for the performance of a smart, easy and quick screening of the voice to discriminate between a possibly pathological and a healthy voice that could be embedded in a smart mobile phone.

Currently, the proposed methodology evaluates only the Fundamental Frequency (F_0) of the speech signal, the main parameter estimated to evaluate voice disorders, variations of which are indicative of the patient's state of health, characterizing quali-quantitatively a specific vocal dysfunction.

The innovation of the methodology presented is the personalization of the algorithm to calculate the F_0, that takes into account two of the main factors that influence this parameter, the gender and age of the subject [14–16]. In more detail, we have finalized a method, based on the exhaustive search algorithm [17], a more accurate evaluation of the F_0. More accurate means that, considering this evaluation of the F_0, we obtain have had the opportunity to classify in a more accurate way (so obtaining fewer false positives and false negatives compared with other classifiers) the speech signal as healthy or pathological.

2. Background

Voice alteration is rarely associated with disease symptoms but often with a temporary alteration due to voice usage (e.g., by teachers). Teachers in general, for example, tend to underestimate the seriousness of the condition. Clinical voice alteration analysis, performed by an otorhinolaryngological expert, is an important instrument for early detection. It is based on a visual inspection of the vocal tract by means of a laryngoscopy associated with vocal signal analysis using an appropriate software, like PRAAT [18,19], a software system that takes its name from the imperative form of "praaten" ("to speak" in Dutch), or the Multi-Dimensional Voice Program (MDVP) [20].

These analyses useful for the clinical evaluation of the voice are required by the SIFEL Protocol (Società Italiana di Foniatria e Logopedia) [21], a protocol containing the guidelines for the clinical and instrumental investigations essential for the evaluation of voice disorders proposed by the Italian Society of Phoniatrics and Logopedics.

The SIFEL protocol provides a series of different examinations to evaluate the presence of voice pathologies, such as the anamnestic evaluation to analyze the behavioral characteristics and vocal attitudes of the subject and the familial history of the disorder, the laryngovideostroboscopic examination to detect any physiological or morphological laryngeal alteration, the acoustic analysis, and the subjective self-assessment of the voice.

While the laryngoscopy requires a subjective identification of problems in the larynx and vocal folds, resulting in a qualitative assessment of these structures, the acoustic analysis is used to give an objective quantification of the health of the speech signal through an evaluation of characteristic parameters calculated from a recording of the vowel /a/ of five seconds in length. The acoustic analysis is a useful instrument to evaluate the presence of a possible laryngeal alteration that can cause a deviation in vocal quality. The association between laryngeal functionality and acoustic measures has been shown in several studies in literature [22–24].

The first important parameter calculated for the acoustic analysis is the F_0 [25,26], defined as the frequency of the opening and closing of the glottis. The rate of vibration of the vocal folds is an important index of laryngeal function. Any abnormality of the larynx can alter the speech production system, resulting in a deterioration of voice quality. Unfortunately, this parameter is influenced by several conditions due to physiological and non-physiological factors, including most significantly, the age and the gender of the subject [14–16]. In fact, if, on the one hand, the voice production and vocal health are influenced by people's lifestyle including factors such as smoking, an incorrect diet or an excessive alcohol intake, on the other hand, the main differences are linked to the anatomical differences in laryngeal systems between men and women during their life. Although the pre-puberty laryngeal systems of males and females are quite similar, in adulthood gender differences in laryngeal geometry affect the voice production. In fact, the female vocal folds are shorter and thinner than male ones and these differences contribute to the different F_0 between women and men. Additionally, the higher F_0 in women increases their risk of developing voice disorders because a higher F_0 results in a greater number of vocal fold oscillations and collisions with a resulting higher stress on the thinner vocal folds.

3. Fundamental frequency estimation methodologies: state of the art

As specified in the SIFEL protocol, the F_0 is the first and the most important parameter that has to be evaluated during the acoustic analysis. Moreover, F_0 retrieval is at the basis of the other parameters calculated in the acoustic analysis and most noise estimation methods [27,28]. Such methods evaluate the noise that characterizes the pathological voices caused for example, by an incomplete closure of the vocal folds with corresponding random variations of the speech signal. A critical issue is that there is no standard algorithm to calculate the F_0 of a voice signal. It is important to