Available online at www.sciencedirect.com

ScienceDirect

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

Original Research Article

Automatic voice pathology detection and classification using vocal tract area irregularity

Ghulam Muhammad [a,*], Ghadir Altuwaijri [a], Mansour Alsulaiman [a], Zulfiqar Ali [a,b], Tamer A. Mesallam [c,d,e], Mohamed Farahat [c,d], Khalid H. Malki [c,d], Ahmed Al-nasheri [a]

[a] Speech Processing Group, Department of Computer Engineering, College of Computer and Information Sciences, King Saud University, Riyadh, Saudi Arabia
[b] Centre for Intelligent Signal and Imaging Research, Department of Electrical and Electronic Engineering, Universiti Tekhnologi PETRONAS, Tronoh, Perak, Malaysia
[c] ENT Department, College of Medicine, King Saud University, Riyadh, Saudi Arabia
[d] Research Chair of Voice, Swallowing, and Communication Disorders, King Saud University, Riyadh, Saudi Arabia
[e] ENT Department, College of Medicine, Al-Menoufiya University, Shebin Alkoum, Egypt

ARTICLE INFO

Article history:
Received 7 September 2015
Received in revised form
11 January 2016
Accepted 13 January 2016
Available online 28 January 2016

Keywords:
Voice pathology detection
Vocal tract area
Voice disorders
Support vector machine

ABSTRACT

In this paper, an automatic voice pathology detection (VPD) system based on voice production theory is developed. More specifically, features are extracted from vocal tract area, which is connected to the glottis. Voice pathology is related to a vocal fold problem, and hence the vocal tract area which is connected to vocal folds or glottis should exhibit irregular patterns over frames in case of a sustained vowel for a pathological voice. This irregular pattern is quantified in the form of different moments across the frames to distinguish between normal and pathological voices. The proposed VPD system is evaluated on the Massachusetts Eye and Ear Infirmary (MEEI) database and Saarbrucken Voice Database (SVD) with sustained vowel samples. Vocal tract irregularity features and support vector machine classifier are used in the proposed system. The proposed system achieves 99.22% ± 0.01 accuracy on the MEEI database and 94.7% ± 0.21 accuracy on the SVD. The results indicate that vocal tract irregularity measures can be used effectively in automatic voice pathology detection.

* *Corresponding author at*: Speech Processing Group, Department of Computer Engineering, College of Computer and Information Sciences, King Saud University, P.O. Box 51178, Riyadh 11543, Saudi Arabia.
E-mail address: ghulam@ksu.edu.sa (G. Muhammad).

http://dx.doi.org/10.1016/j.bbe.2016.01.004

1. Introduction

Voice pathology detection refers to a detection procedure of the pathology in the vocal folds from an input voice. The detection can be subjective or objective. In subjective detection, an experienced physician hears the voice and assesses whether the voice is normal or pathological based on his or her previous knowledge and experience. However, this type of assessment may vary from physician to physician depending on the experience [1]. Therefore, an objective evaluation of voice pathology is gaining more attention day by day from both medical and engineering personnel.

Lieberman proposed one of the first acoustic voice parameters in pathological voice analysis in 1961 [2]. Voice perturbation and quality measures, such as jitter (pitch perturbation) and shimmer (amplitude perturbation), depend on accurate extraction of fundamental frequency and the amplitude of various waveform types. However, measuring fundamental frequency (F_0) is a very difficult task; especially in the case of a pathological voice [3].

Most of the work in voice pathology detection relies on the measurement of a sustained vowel, particularly, /a/ sound, uttered by the subject. /a/ is easy to pronounce for a patient with voice pathology; its formants are clearly distinguishable, and peaks are prominent. These three attributes of the sustained /a/ make it a proper choice for voice pathology detection [4]. Popular acoustic features extracted from sustained vowel to detect voice pathology are shimmer and jitter, spectral centroid and spectral flatness [5], cepstral peak prominence [6], etc. Vasilakis and Stylianou developed a method for short-time jitter and found the area under curve (AUC) reached 94.82% [7]. Though these popular features are widely used in the literature, many researchers found different relationships (poor to high) between these features and severity of voice pathology [8,9]. Little et al. found that nonlinear measures, such as recurrence period density entropy and detrended fluctuation analysis, were more stable and reliable than the classical measures on voice pathology quantification [10].

Markaki and Stylianou suggested a modulation spectra method to identify pathological voices and achieved an AUC of 96.26% [11]. Arias-Londono et al. introduced a method, which is based on extracting eleven features using nonlinear analysis of time series, and gained a classification accuracy of 98.23% using classifier combination [12]. Godino-Llorente et al. proposed a detection system of pathological voice by means of Gaussian mixture models and short-term mel cepstral parameters complemented by frame energy together with first and second derivatives [13]. All of the three methods in [11–13] were applied on a common subset of the Massachusetts Eye and Ear Infirmary (MEEI) database [14]. Muhammad and Melhem proposed MPEG-7 audio low-level descriptor based voice pathology detection and classification [15]. A framework of voice pathology detection in the cloud was proposed in [16]. Some research was conducted with continuous speech to detect voice pathology [17–19], where the authors claimed that continuous speech was more practical than sustained vowel in everyday life.

Voice pathology detection is still an open problem. The difficulty of detecting voice pathology is closely associated to the severity of the pathology. A patient with less severe pathology can pronounce a vowel without much difficulty; however, it is not the same with the patient having severe voice pathology; for example, a big cyst in the vocal fold prevents proper closing and opening of the fold while pronouncing a vowel sound.

In this paper, we proposed a voice pathology detection (VPD) system based on the vocal tract area features. Voice pathology is related to a vocal fold problem, and hence the vocal tract area which is connected to the vocal fold should exhibit irregular patterns over frames in case of a sustained vowel for a pathological voice. Typically, it is believed that there is no infralaryngeal (tracheobronchial tree) effect on the vocal tract while producing a vowel on the assumption that the voicing source has infinite impedance. However, a finely detailed analysis must recognize that the infralaryngeal structures exert an influence on the vocal tract; the articulatory configuration in the vocal tract interacts with the articulation in the vocal folds [20]. Therefore, additional vocal tract-related information is expected to assist in detecting the characteristics of the vocal folds, especially during phonation [21]. We quantified this irregularity in the form of variances along time. Several methods based on glottal noise measures, such as signal to noise ratio (SNR), harmonic to noise ratio (HNR), normalized noise energy (NNE), pitch amplitude (PA), and spectral flatness ratio (SFR), were proposed in [22]. However, calculation of SNR, HNR, and NNE required correct estimation of F_0. PA and SFR were based on linear prediction (LP) modeling, where PA needed computationally expensive autocorrelation, and SFR needed windowing and Fourier transform after LP. On the other hand, our proposed system is much simpler than the PA and SFR in the sense that it uses LP coefficients directly to calculate the area of different tubes in the vocal tract. Four features, which are average, variance, skewness, and kurtosis, are extracted from the vocal tract area to feed into a classifier.

2. Materials and methods

2.1. Data

To prove the proposed concept, we used two different databases. First, we utilized a common database, which is the MEEI database [14], in the experiments. Campbell and Reynolds stated that ‘‘the main benefit of using standard corpora is that it allows researchers to compare performance of different techniques on common data, thus making it easier to determine which approaches are most promising to pursue’’ [23]. To compare the proposed system with others, we used the same subset of the MEEI database which was used in [11–13], denoted as $\text{MEEI}_{\text{subset}}$; this subset has sustained vowel /a/ recordings for 53 normal and 173 pathological speakers. The criteria of selecting this subset took into consideration a wide variety of voice disorders and a uniformly distributed gender and age between both normal and pathological classes [22]. A distribution of the samples of the $\text{MEEI}_{\text{subset}}$ in terms of gender and age is shown in Table 1.