Biomedical Signal Processing and Control 31 (2017) 156–164

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

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

Voice pathology detection using interlaced derivative pattern on glottal source excitation

Ghulam Muhammad [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], Mohamed A. Bencherif [a]

[a] Digital Speech Processing Group, Department of Computer Engineering, College of Computer and Information Sciences, King Saud University, Riyadh 11543, Saudi Arabia
[b] Centre for Intelligent Signal and Imaging Research (CISIR), Department of Electrical and Electronic Engineering, Universiti Teknologi PETRONAS, Tronoh 31750, 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, Menoufiya University, Shebin Alkoum, Egypt

ARTICLE INFO

Article history:
Received 19 April 2015
Received in revised form 29 July 2016
Accepted 2 August 2016

Keywords:
Interlaced derivative pattern (IDP)
AVPD
SVD
MEEI
Voice pathology detection
Glottal source excitation

ABSTRACT

In this paper, we propose a voice pathology detection and classification method using an interlaced derivative pattern (IDP), which involves an *n*-th order directional derivative, on a spectro-temporal description of a glottal source excitation signal. It is shown previously that directional information is useful to detect pathologies due to its encoding ability along time, frequency, and time-frequency axes. The IDP, being an *n*-th order derivative, is capable of describing more information than a first order derivative pattern by combining all the directional information into one. In the IDP, first-order derivatives are calculated in four directions, and these derivatives are thresholded with the center value of each directional channel to produce the final IDP. A support vector machine is used as a classification technique. Experiments are conducted using three different databases, which are the Massachusetts Eye and Ear Infirmary database, Saarbrucken Voice Database, and Arabic Voice Pathology Database. Experimental results show that the IDP based features give higher accuracy than that using other related features in all the three databases. The accuracies using cross-databases are also high using the IDP features.

1. Introduction

Automatic detection of vocal fold pathologies is an interest to the researchers of speech or voice community, as well as the respective medical community. This is due to its non-invasive nature, free from subjective biasness, and relatively low cost. There have been a lot of researches to detect voice pathology by analyzing voice. The main emphasis was to develop a feature or a feature vector that can effectively distinguish between normal and pathological voices. The features can broadly be divided into two groups, one imported from speech or speaker recognition applications, and the other from voice quality measurements. The features that came from speech or speaker recognition applications include Mel-frequency cepstral coefficients (MFCC), linear prediction cepstral coefficients (LPCC), and relative spectra perceptual linear prediction (RASTA-PLP) [1–3]. On the other hand, the features coming from voice quality measurements are, among others, shimmer, jitter, harmonic-to-noise ratio, and cepstral peak prominence [3]. Recently, features from audio and image processing applications are also integrated in voice pathology detection. These features include MPEG-7 audio features [4], fractal analysis [5], modulation spectrum [6], formants [35]. The use of nonlinear features have studied too in the field of voice pathology detection [29–32].

Though there exists a lot of related works, most of them use the Massachusetts Eye and Ear Infirmary (MEEI) database [7], where normal and pathological samples are recorded in two different environments. Therefore, for a classification technique, it is not clear whether the classifier is classifying pathologies or environments. There are some works that use other databases; however, many of them are using only one database at a time. The question arises: are the existing voice pathology detection techniques database dependent? For example, Markaki and Stylianou

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

http://dx.doi.org/10.1016/j.bspc.2016.08.002
1746-8094/

showed that normalized modulation spectral features achieve more than 92% detection rate in the MEEI database, but it drops below 80% in a cross-database experiment [8]. The same features set without normalization have a detection rate of 94.1% in the MEEI database, and 62.3% in cross-database. The normalization increases the accuracy in the cross-database scenario at the expense of a decrease in a single database experiment. Therefore, there is a need to develop such a features set that achieve a high accuracy in both single database and cross-database experiments. These features should also visually justify the detection.

In this paper, a set of features are proposed to detect voice pathology from a glottal source excitation signal. A voice signal is a convolution between a source signal (originated from the lung) and the vocal tract filter. In a voiced signal, the vocal folds open and close in a periodic manner that produces pitch of the signal. The vocal tract is simply shaping the signal to produce a desired phone. In this study, we refer voice pathology to an abnormality in voice caused by pathology in the vocal fold(s). Vocal folds contribute to the glottal source excitation, and therefore, any abnormality in the glottal source excitation is an indication of voice pathology. There are some researches of voice pathology detection using the glottal source excitation [9–12]. For example, glottal formant frequency and bandwidth, spectral balances, and center of gravity are extracted from the glottal source signal to detect voice pathology in [10]. Features from dynamics of average glottal source and mucosal wave spectrum are used to detect voice pathology in [11]. In [9], a power spectral density envelop of the glottal source is utilized in a form of a specific harmonic-to-harmonic relationship. Back in 1975, Koike and Markel first showed the use of inverse filtering (residue output) in voice pathology detection in an experiment involving only 10 normal persons and 10 patients having voice pathology [12]. Residue signal is also used in [13] for voice pathology detection. There are several techniques to estimate the glottal source excitation from the voice signal. A good review of this topic can be found in [14].

In this paper, the first-order derivative of the glottal source excitation is utilized to extract features. It is well-known that a first-order derivative (without smoothing) generally decreases the signal-to-noise ratio. Any weak disturbances caused by irregular vibrations of the vocal folds (due to the location and the shape of the pathology, elasticity of the tissues, etc.) are emphasized in the first-order derivative signal. This derivative signal is further processed to have a spectrum. The spectrum is divided into several bands, whose center frequencies are 'Mel' spaced. An interlaced derivative pattern (IDP) [15,16] is extracted from the spectro-temporal representation of the first-order derivative of the glottal source. The IDP is a texture descriptor, and successfully applied to applications including gender recognition from faces [15] and automatic speech recognition [16]. A support vector machine (SVM) is applied for classification. Three different databases are used to evaluate the proposed method for voice pathology detection and classification. The major contributions of this paper are (i) to develop a method using the IDP to detect and classify voice pathologies, (2) to use first-order derivative in the glottal source excitation to lower the signal-to-noise ratio to enhance the weak noises arisen from the closing and opening of the pathological vocal folds, (3) to validate the performance of the proposed method using multiple databases, and using cross-database, and (4) to investigate the use of the voice signal and the glottal source excitation by an inverse filtering in the proposed method.

The rest of the paper is organized as follows. Section 2 describes the proposed method; Section 3 presents the experiments; Section 4 gives the discussion, and Section 5 draws some conclusions.

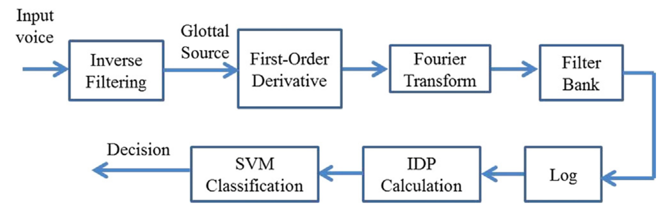

Fig. 1. Block diagram of the proposed IDP based voice pathology detection and classification method.

2. Proposed method

Fig. 1 shows a block diagram of the proposed IDP based voice pathology detection and classification method. First, a glottal source excitation signal is estimated using an interactive adaptive inverse filtering (IAIF) technique from an input voice signal. First-order derivative of the glottal signal is calculated for further processing. The Fourier transform is applied to the frames of the first-order derivative signal to get a spectrum. 24 band-pass filters (BPFs), whose center frequencies are spaced on a Mel scale, are utilized to get the contributions of 24 different frequency bands. The filter outputs are then log-compressed. The IDP encodes an n-th order directional derivative pattern of the spectro-temporal description (whose horizontal axis represents the frame number, and vertical axis represents the filter number). The IDP features are fed to an SVM classifier.

2.1. Glottal source excitation

There are several techniques to estimate the glottal source excitation. In this study, we used the IAIF algorithm [17], which is publicly available at http://users.aalto.fi/~traitio/research.html. In the IAIF algorithm, high-pass filtering is applied to remove low frequency fluctuations. Using the linear predictive coding, the effect of the vocal tract filter is estimated, and is canceled out through inverse filtering. The effect of the lip radiation is canceled through integration. This process is done iteratively to get a good estimation of the glottal source [17]. By inverse filtering, the effect of vocal tract filtering is eliminated. We do not need spectral information of vocal tract filtering, because we are concerned about glottal signal (which is affected by vocal fold pathologies), not speech characteristics. The spectral characteristics that we are interested in are produced by irregular vibrations of the vocal folds, not the spectral characteristics produced by the vocal tract shaping.

Once the glottal source waveform is estimated, a first-order derivative operation is applied to the waveform to get the derivative wave. This derivative wave has a lower signal-to-noise ratio, which is useful to find the weak noise contribution that arises from vocal fold pathologies. For a good phonation, energy is mostly concentrated in lower frequencies due to low glottal formant frequencies; however, because of irregular vibration of the vocal folds in case of pathology, energy is significant also in higher frequencies. Fig. 2 shows examples of the spectrums of the glottal source signal and its first-order derivative signal of a pathological sample. From the figure, we can see that in the higher frequencies there is more energy in the first-order derivative signal spectrum than in the original glottal source signal (see the zoomed-in frequency range [2000–4000] Hz in (e) and (f) of Fig. 2). This justifies the use of first-order derivative signal for further processing. The signal is framed, where the frame size is 30 milliseconds, and Hamming windowed. The frames are overlapped by 50%. The Fourier transform is applied to each frame to get a spectrogram from the first-order derivative signal. 24-BPFs are applied to the spectrogram. From experiments, we found that outputs of filters 17–24 are not dis-