Biologically Inspired Cognitive Architectures (2016) 15, 10–18

Available at www.sciencedirect.com

ScienceDirect

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

LETTER

Voice pathology detection based on the modified voice contour and SVM

Zulfiqar Ali [a,b,*], Mansour Alsulaiman [a], Irraivan Elamvazuthi [b], Ghulam Muhammad [a], Tamer A. Mesallam [c], Mohamed Farahat [c], Khalid H. Malki [c]

[a] *Digital Speech Processing Group, College of Computer and Information Sciences, King Saud University, Saudi Arabia*
[b] *Centre for Intelligent Signal and Imaging Research (CISIR), Department of Electrical and Electronic Engineering, Universiti Teknologi PETRONAS, Perak, Malaysia*
[c] *ENT Department, College of Medicine, King Saud University, Riyadh, Saudi Arabia*

Received 16 September 2015; received in revised form 1 October 2015; accepted 25 October 2015

KEYWORDS
Artificial intelligence;
Voice disorder detection;
Modified voice contour;
Continuous speech;
Simpson's rule

Abstract
In this study, a novel method based on the voice intensity of a speech signal is used for automatic pathology detection with continuous speech. The proposed method determines the peaks from the speech signal to form a voice contour. The area under the voice contour allows us to discriminate between normal and disordered subjects. In the case of disordered subjects, the calculated area under the voice contour is lower than that for a normal subject due to the malfunctioning of vocal folds, which makes the voice weaker and breathier. Some long-term features such as shimmer and jitter are based on the accurate estimation of fundamental frequency, which is itself a difficult task, especially for disordered speech signals. The proposed feature does not need to estimate the pitch period or fundamental frequency during the calculation of the voice contour and they provide a single value for the whole utterance similar to other long-term features. The voice disorder database used in this study includes 71 normal subjects and the same number of disordered subjects. Each disordered subject has one of the following voice disorders: vocal folds cysts, laryngopharyngeal reflux disease, vocal folds polyps, unilateral vocal folds paralysis and sulcus vocalis. The accuracy of the proposed method is 100%.

* Corresponding author at: Centre for Intelligent Signal and Imaging Research (CISIR), Department of Electrical and Electronic Engineering, Universiti Teknologi PETRONAS, Perak, Malaysia.
E-mail addresses: zuali@ksu.edu.sa, zulfiqar_g02579@utp.edu.my (Z. Ali), msuliman@ksu.edu.sa (M. Alsulaiman), irraivan_elamvazuthi@petronas.com.my (I. Elamvazuthi), Ghulam@ksu.edu.sa (G. Muhammad), tmesallam@ksu.edu.sa (T.A. Mesallam), mfarahat@ksu.edu.sa (M. Farahat), kalmalki@ksu.edu.sa (K.H. Malki).

http://dx.doi.org/10.1016/j.bica.2015.10.004
2212-683X/

Introduction

Vocal folds disorders (VFDs) occur due to the excessive use of the voice. Therefore, teachers, singers and lawyers have more risk of being affected by voice pathology. In the USA, the occurrence of voice problems in teachers is 57.7% during their lifetime and for other professions, it is 28.8% (Roy et al., 2004). The voice produced by a VFD-affected person differs from a normal person due to the malfunctioning of the vocal folds. Due to the incomplete closure and irregular vibration of the vocal folds, the speech signal of a pathological person becomes more transient. The speech signals produced by a pathological and a normal person are shown in Fig. 1. The difference between the amplitudes of the two signals can be observed, and they do not have the same voice intensity. In this paper, an automatic voice pathology detection (AVPD) system based on the voice intensity of a speech signal is developed. Moreover, continuous speech samples are used for the differentiation of normal and pathological samples.

Because of its noninvasive nature, the automatic detection of voice pathology can be considered as a primary screening tool for the clinician. It would be of great help to an ENT specialist if an automatic system could discriminate between normal and pathological samples. The detection of VFDs with a sustained vowel has been well investigated by the research community during recent years (Lee, Kang, Choi, & Son, 2013; Lee, Kim, & Kang, 2014; Markaki & Stylianou, 2011; Muhammad & Melhem, 2014), and it is a comparatively easy task than detection with continuous speech. Parsa and Jamieson (2001) concluded that the analysis of continuous speech is more challenging because of the inherent non-stationary of the signal. Moreover, other studies emphasize the value of continuous speech over the sustained vowel in obtaining acoustic characteristics. Askenfelt and Hammarberg (1986) reported the importance of pitch and loudness variation in continuous speech as indicators of abnormal voice quality. Some of the common and most spreading voice disorders are vocal fold nodules, keratosis, vocal folds paralysis, and adductor spasmodic dysphonia (Muhammad et al., 2012).

Artificial intelligence (AI) in biomedical engineering is as vital as in other fields of science for providing solutions to the problems that are computationally expensive or hard to solve by following traditional methods. Computational AI deals with pattern recognition, and different studies have used fuzzy logic (Aghazadeh & Heris, 2009), artificial neural network (Paulraj, Yaacob, & Hariharan, 2009) and support vector machine (SVM) (Al Mojaly, Muhammad, & Alsulaiman, 2014) to successfully implement AVPD systems. In this paper, SVM is used as a binary classifier to make the decision about a speech sample.

Like other long-term features such as shimmer, jitter (Arjmandi, Pooyan, Mikaili, Vali, & Moqarehzadeh, 2011) and cepstral peak prominence (CPP) (Heman-Ackah et al., 2003), the proposed feature also provides a single value for a whole speech signal. The measurement of long-term acoustic features normally involves the accurate estimation of the pitch period, which is a very difficult task, especially in pathological samples. The proposed feature does not need to estimate the pitch period or fundamental frequency during the calculation of the modified voice contour (MVC). Arjmandi et al. (2011) used 22 long-term features including shimmer and jitter and the obtained accuracy was 91.5%. Watts and Awan (2011) reported that an accuracy of 91% with CPP was obtained for an AVPD system. The objective of this study is to propose a new long-term feature that may provide better results than existing long-term features.

Short-term features are also used in different studies to develop different AVPD systems. The most commonly used short-term features for pathology detection are LPC (Linear Prediction Coefficients), LPCC (Linear Prediction Cepstral Coefficients) and MFCC (Mel-frequency Cepstral Coefficients). LPC and LPCC have been used in many

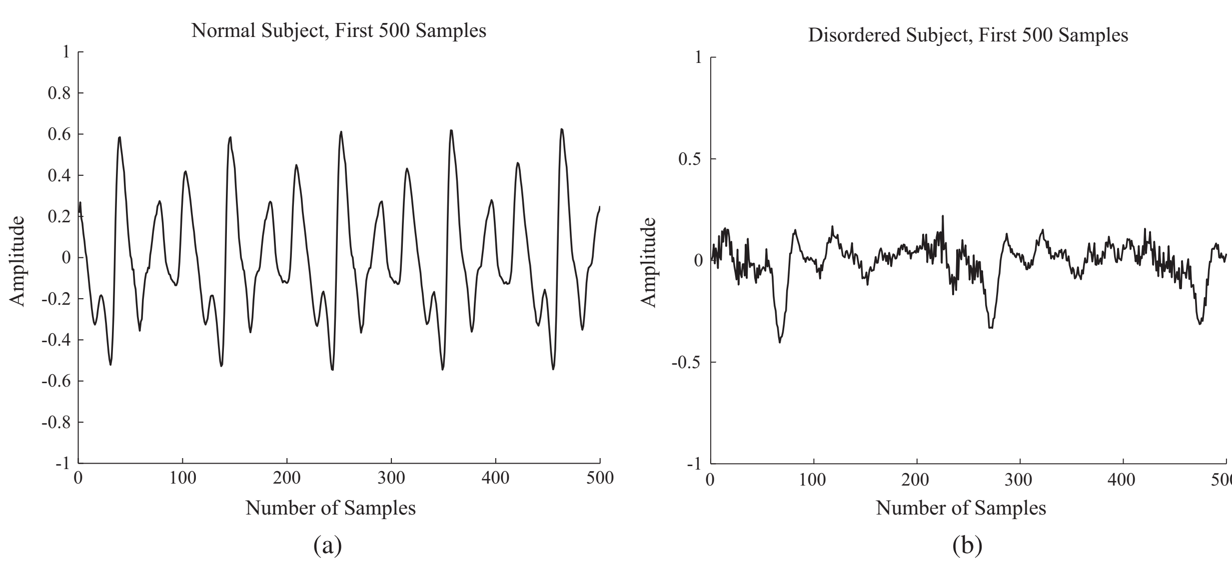

Fig. 1 Amplitudes in voice samples of (a) normal subjects and (b) disordered subjects.