Biomedical Signal Processing and Control 10 (2014) 209–223

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

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

An efficient voice pathology classification scheme based on applying multi-layer linear discriminant analysis to wavelet packet-based features

Ali Akbari [a,*], Meisam Khalil Arjmandi [b]

[a] Department of Computer Engineering, Firoozkooh Branch, Islamic Azad University, Firoozkooh, Iran
[b] Department of Electrical Engineering, Firoozkooh Branch, Islamic Azad University, Firoozkooh, Iran

ARTICLE INFO

Article history:
Received 18 June 2013
Received in revised form 2 September 2013
Accepted 6 November 2013
Available online 8 December 2013

Keywords:
Voice disorders classification
Discrete wavelet packet transform
Energy and entropy
Multi-class linear discriminant analysis

ABSTRACT

In this work, we are interested in developing an efficient voice disorders classification system by using discrete wavelet packet transform (DWPT), multi-class linear discriminant analysis (MC-LDA), and multilayer neural network (ML-NN). The characteristics of normal and pathologic voices are well described with energy and Shannon entropy extracted from the coefficients in the output nodes of the best wavelet packet tree with eight decomposition level. The separately extracted wavelet packet-based features, energy and Shannon entropy, are optimized with the usage of multi-class linear discriminant analysis to reduced 2-dimensional feature vector. The experimental implementation uses 258 data samples including normal voices and speech signals impaired by three sorts of disorders: A–P squeezing, gastric reflux, and hyperfunction. The voice disorders classification results achieved on Kay Elemetrics databases, developed by Massachusetts Ear and Eye Infirmary (MEEI), show average classification accuracy of 96.67% and 97.33% for the structure composed of wavelet packet-based energy and entropy features, respectively. In these structures, feature vectors are optimized by multi-class linear discriminant analysis and, finally classified by multilayer neural network. The obtained results from confusion matrix and cross-validation tests prove that this novel voice pathology classification system is capable of significant classification improvement with low complexity. This research claims that the proposed voice pathology classification tool can be employed for application of early detection of laryngeal pathology and for assessment of vocal improvement following voice therapy in clinical setting.

1. Introduction

Voice health has been playing a significant role in promoting the quality of daily activities in many aspects. Therefore, considering precise and preventive health care programs is essential for taking care of voice health. In many cases, the origin of the voice disorders arises from unknown causes. Sometimes lasting cold or flu, a continuing virus or bacteria, and vocal abuse or misuse may result in chronic laryngitis and other types of voice diseases [1]. Automatic voice pathology assessment can be employed in the field of communication science and disorders as a complementary tool to screen individuals for early cases of laryngeal pathology [2,19]. This paper mainly focuses on introducing an automatic algorithm in classification of voices developed by impairments in the vocal folds.

The human apparatus concerned with speech production is complex and involves many important organs. This mechanism involves three subsystems: air pressure system, vibratory system, and resonating system [1]. The vocal tract and the vocal cords are the most important components of speech system which their characteristics have the highest influences in the resulted voice [3]. Vibration of vocal folds is the result of providing and regulating air pressure by air pressure system – composed of diaphragm, chest muscles, ribs, abdominal muscles, and lungs – which is reflected in the concept of "pitch" [14]. Finally, a person's recognizable voice is produced by resonating system that involves throat, oral cavity, and nasal cavity [3]. The precise handling may be achieved by internal feedback in the brain. Actually, the central nervous system (CNS) coordinates voice production process through specific nerve connections and signals. The larynx muscles are stimulated by recurrent laryngeal nerve (RLN) and superior laryngeal nerve (SLN) [1]. The left and the right vocal folds, housed in the larynx, include tree distinct layer: mucosa, vocal ligament, and body. These

* Corresponding author at: Islamic Azad University, Firoozkooh Branch, Firoozkooh, Tehran 1787733741, Iran. Tel.: +98 2133491324; mobile: +98 9126900258.
E-mail address: irabka@gmail.com (A. Akbari).

1746-8094/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.bspc.2013.11.002

soft and fragile structure of the vocal fold makes it very sensitive and vulnerable [3].

The cycle of the vocal folds vibration creates a repeating undulation known as mucosal wave. The regularity of the mucosal wave is a sign of normal voice [1]. Also, the pliability of the mucosa depends on the integrity of the layer beneath it; superficial lamina propria. In normal voice, loudness which is the volume of sound resulted from the pressure of the air blown past the vocal folds, is in normal range. Pitch – the frequency of the mucosal wave – is an operative measure to evaluate laryngeal function which can be produced rapidly and precisely by who has a normal voice. Softness of voice is dependent on vocal fold closure and the integrity of the superficial lamina propria. Generally speaking, sign of “hoarseness”, “nasality”, “breathiness”, and “harshness” is sufficiently weak in normal voice.

Abnormalities and impairments in the vibratory system are termed “voice disorders” which may cause breathiness, harshness, and hoarseness [3]. Stiffness in the vocal folds caused by swelling develops hoarseness and, also air leakage caused by partial nerve input loss increases breathiness. Position of the vocal folds is controlled by laryngeal muscles, arytenoid cartilages, RLN, SLN, and mass of vocal folds. A laryngeal pathology, such as A–P squeezing or gastric reflux, generally produces asymmetrical changes in the mass, elasticity and tension of the vocal folds, leading to deviant vibration [1]. In the case of A–P squeezing and hyperfunction, laryngeal muscle tension is damaged by voice abuse and misuse. Also, swelling and inflammation caused by gastric reflux and viral laryngitis affect the elasticity of vocal fold and, therefore it changes the normal status of air pressure passing form glottis and the resistance of vocal folds. Abnormality in the vocal folds can be detected in terms of harshness, breathiness, hoarseness, effortful phonation, vocal fatigue and vocal fry [1].

In this work, the studied voice disorders for classification are impaired by three types of pathology: A–P squeezing, gastric reflux, and hyperfunction. The tense voice is often described as a voice produced with poor breath support, elevation of the hyoid bone and larynx, mandibular restriction, unnecessary tongue tension, and overall vocal tract constriction [3]. Part of such tract constriction was described by Colton and Casper [47] as anterior–posterior laryngeal squeezing (A–P squeezing) where the epiglottis and the arytenoids approach each other during phonation. A–P squeezing is demonstrated as a reduction in the supraglottic space as the arytenoid cartilages move anteriorly toward the petiole of the epiglottis, which results in varying degrees of visual obstruction of the adducted true vocal folds [3]. A–P squeezing is believed to be caused by excessive strain and tension in the laryngeal area. Sound of voice ranges from normal to extremely squeezed and tight sounding. The voice may sound rough if the squeezing causes irregular vibration of the vocal folds.

Irritation and swelling in parts of the voice box due to backflow of stomach fluids into voice box area or laryngopharyngeal reflux usually result in gastric reflux [4]. Patients who suffer from gastric reflux generally complain of hoarseness, frequent throat cleaning, sensation of lump in the throat, and cough or sore throat. To diagnose, physicians determine the cause of reflux laryngitis - backflow of stomach fluids to the voice box or laryngopharyngeal reflux-using rigid or flexible laryngoscope.

Hyperfunction involves increased laryngeal or supralaryngeal muscle tension, regardless of the presence of a vocal fold lesion. It is most often caused by abuse or misuse of the voice or constant throat clearing as a result of things such as gastro esophageal reflux. Resulted voice may leads to vocal fatigue, constant throat clearing, ulcer on vocal process, breathy voice, hoarseness, audible inhalation, and growth on vocal fold. An ear–nose–throat (ENT) specialist (otolaryngologist) or physician voice specialist (laryngologist) should be consulted in the case of voice disorders. Typically, laryngoscopy and stroboscopy are performed to provide elaborate information from larynx [2]. These technologies allow examiners to review the images of the larynx frame-by-frame, and to re-review images with members of the voice care team. However, these methods are expensive, invasive, and also bring discomfort.

On the other hand, an automatic acoustic analysis for diagnosis of pathologic voices would greatly aid in clinical evaluation of laryngeal function and early detection/classification of pathologic voice. The advantage of acoustic analysis is its nonintrusive nature and its potential for providing quantitative data with reasonable expenditure of analysis time [2]. During the last decades, various approaches to efficient voice pathology assessment have been adopted. The most recent techniques in the area of pathologic voice detection/classification have been explored for some techniques to model accurately the variations of a group of features in normal and disordered subjects. Among various pathological voice detection algorithms, wavelets [5], fractals [6], neural maps and networks [7] has been mostly employed. In addition, many parameters have been suggested such as pitch, jitter, shimmer, amplitude perturbation quotient (APQ), pitch perturbation quotient (PPQ), harmonic to noise ratio (HNR), Glottal to Noise Excitation Ratio [8–18] and normalized first harmonic energy (NFHE) [19] to improve voice pathology detection system. These techniques necessitate an accurate estimation of the fundamental frequency, a fairly complex task in the presence of voice pathologies. In recent studies, several alternative approaches such as Mel-frequency cepstral coefficients (MFCC) and linear prediction cepstral coefficients (LPCC) have been proposed based on conceptual modeling for appropriate parameterization of speech signal to detect voice pathology. Godino-Llorente and Gomez-Vilda introduced voice disorders identification techniques based on MFCC speech parameterization fed to neural network [20] and Gaussian mixture model (GMM) [21]. In these works, the ability of F-ratio and Fisher's discriminant ratio was examined in reducing the dimension of MFCC-based features. In the best case, efficiency of 94.04% was reported for selected dataset. Different modifications for Mel-frequency cepstral coefficients (MFCC) and linear prediction cepstral coefficients (LPCC) have been suggested to improve voice disorders detection. For example, Costa et al. [22] proposed weighted cepstral coefficients in order to account for the sensitivity of the low-order cepstral coefficients to overall spectral slope and the sensitivity of the high-order cepstral coefficients to noise. Little et al. [23] studied two non-linear features of return period density entropy (RPDE) and fractal self-similarity based upon the biophysics of speech production for speech pathology detection. It proved that this proposed measures can be both simple and robust by achieving detection rate of 91.4% to discriminate normal and pathological voices.

On the other hand, time-frequency transforms including wavelet and discrete wavelet packet transforms (DWPTs) are non-parametric estimation methods that have been recently proposed for speech parameterization to detect/classify voice disorders. DWPT is generally acknowledged to be useful for studying non-stationary phenomena and, in particular, have been shown or claimed to be of value in the detection and characterization of transient signals which is prevalent in pathologic voices. Local discriminant bases (LDB) algorithm and wavelet packet decomposition have been employed by Umapathy and Krishnan [24] to demonstrate the significance of identifying the signal subspace that contribute to the discriminatory characteristics of normal and pathological speech signals. In this research, a database of 212 speech signals (51 normal and 161 pathological) was used to evaluate the proposed system. Using the Daubechies wavelet with order 4 (*db4*) classification accuracy up to 96% for a two-class classification task (normal/pathological) and 74% for a four-class classification (male normal/female normal/male