Biomedical Signal Processing and Control 7 (2012) 3–19

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

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

An optimum algorithm in pathological voice quality assessment using wavelet-packet-based features, linear discriminant analysis and support vector machine

Meisam Khalil Arjmandi*, Mohammad Pooyan

Faculty of Engineering, Shahed University, Tehran, Iran

ARTICLE INFO

Article history:
Received 7 December 2010
Received in revised form 14 March 2011
Accepted 29 March 2011
Available online 5 May 2011

Keywords:
Voice disorders identification
Wavelet packet analysis
Energy and entropy
Linear discriminant analysis
Support vector machine

ABSTRACT

Voice impairments, attention to increased unhealthy social behavior and voice abuse, have been increasing dramatically. Therefore, diagnosis of voice diseases has an important role in the opportune treatment of pathologic voices. This paper presents an extensive study in identification of different voice disorders which their origin is in the vocal folds. Firstly, a qualitative study is applied based on short-time Fourier transform (STFT) and continuous wavelet transform (CWT) in order to investigate their aptitude in the presentation of discriminative features to identify disordered voices from normal ones. Therefore, wavelet packet transform (WPT) for their ability to analyze scrutinizingly a signal at several levels of resolution is chosen as strong speech signal parameterization method. The ability of energy and entropy features, obtained from the coefficients in the output nodes of the optimum wavelet packet tree, is investigated. Linear discriminant analysis (LDA) and principal component analysis (PCA) are evaluated as feature dimension reduction methods in order to optimize recognition algorithm. The performance of each structure is evaluated in terms of the accuracy, sensitivity, specificity, and area under the receiver operating curve (AUC). Eventually, entropy features in the sixth level of WPT decomposition along with feature dimension reduction by LDA and a support vector machine-based classification method is the most optimum algorithm that leads to the recognition rate of 100% and AUC of 100%. Proposed system clearly outperforms previous works in both respect of accuracy and reduction of residues; which may lead in full accuracy and high speed diagnosis procedure.

1. Introduction

Nowadays, due to the importance of social behavior in daily activities, the health in the field of voice becomes a major issue. Therefore, identification of voice disorders, in the shadow of severely abuse or misuse of speech production tissues, has a significant role. Measures of normal voice can be stated as: (1) loud enough to be heard; (2) hygienic voice production; (3) pleasing vocal quality; (4) flexible enough to express emotion; and (5) represent speaker age and gender [1]. Voice disorders fall into three main categories: organic; functional; and combination of the two categories. Organic voice disorders divided in two groups: structural and neurogenic [1]. Structural disorders involve something physically wrong with the voice mechanism that often infects tissues or fluids of the vocal folds. Neurogenic disorders are caused by a problem in the nervous system. Also, a functional disorders means the physical structure is normal, but the vocal mechanism is being used improperly. Voice diseases can be arisen by neurogenic, traumatic, organic, and psychogenic insufficiency that impair vocal mechanism and, therefore result in several types of voice disorders. Since vocal folds are one of the most sensitive tissues of the speaker's vocal apparatus, many types of the voice diseases are caused by a special failure in the vocal folds.

The health condition of larynx can be evaluated by three criteria: breathiness; harshness; and hoarseness [1]. Breathiness mentions audible air escapes approximating edges of glottis fail to make contact. Harshness represents amount of unpleasant, hard, rough or metallic quality in phonation. Hoarseness is the measure of both harshness and breathiness. Some measures such as pitch (frequency), loudness (intensity), quality (waveform complexity), air wastage (airflow rate), analysis of ENT (Ear, nose and throat) report, clinical facilitation techniques, analyzing of video endoscopic data, observing patient behavior and analyzing electroglotographic data have been being considered to evaluate voices. Also, glottal fry, register variation, pitch breaks, and phonation breaks are some perceptual judgments of voice quality.

From the medical diagnosis viewpoint, diagnosis of the pathological voices is based on the symptoms which appear in the

* Corresponding author.
E-mail address: msmarjmandi@gmail.com (M.K. Arjmandi).

1746-8094/$ – see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bspc.2011.03.010

outcome voice or physical impairments in the speech system. Diagnostic methods are often invasive like Laryngoscopy, Stroboscopy, and Endoscopy [2]. In these methods, the health condition of vocal production organs is monitored using the imaging facilities and therefore diagnostic decision is done. These approaches require a well-equipped speech laboratory and a well-trained speech pathologist. Such laboratories are expensive to setup and operate, and therefore, available in only a limited number of clinics. An alternative tool based on analysis of the speech waveform would enable preliminary and more accurate detection in any clinical setting. On the other hand, the engineers from the viewpoint of acoustical analysis have done several researches to improve the voice disorders recognition task. Acoustic analysis of speech has been used as a complementary tool in the incipient diagnosis of voice disorders, and also it is a non-invasive method. This study is designed for voice diseases caused by malformation of laryngeal vocal folds.

Voice disorders detection using acoustic signal processing can being pursued both qualitatively and quantitatively. Laryngeal abnormalities usually are reflected in the quality of speech signal [2,3]. Diseases impress vocal folds and vocal tract are often detectable through standard acoustic measures such as breathiness, harshness, hoarseness and weakness in the output sound [1,4]. Also, the noise and turbulence in the airflow passing through glottis, which caused by damage in the vocal folds, can be appeared as abnormal noise components in the spectrum of speech signal [1]. The main idea in some researches is based on the applying a method for modeling nonlinear effects in the excitation source (vocal folds) and the vocal tract [5,6]. Extraction of glottal excitation signal by using inverse filtering techniques is widely considered to study of diseases that affect the vocal folds [3,6]. Speech signal contains information about glottis excitation signal and vocal tract filter. These methods model speech system according to their original speech signal by applying auto-regression equation. However, due to the fact that inverse filtering is based on the assumption of a linear model, such methods do not efficient when pathology is present regarding to non-linearites introduced by pathology.

Most of the recent researches, developed in the field of voice disorders identification, are designed to pursue the variation of a group of specific features in normal and pathological voice samples. Therefore, many long-time parameters such as pitch, jitter, shimmer, amplitude perturbation quotient (APQ), pitch perturbation quotient (PPQ), harmonic to noise ratio (HNR) and other parameters have been introduced as measures to determine the degree of abnormalities in the voice signal [7–16]. Hadjitodorov and Mitev [17] introduced the normalized first harmonic energy (NFHE) as a new parameter to evaluate pathological voice disorders. Godino-Llorente et al. [18] evaluate capability of the Glottal to Noise Excitation Ratio for the screening of voice disorders. In this study the normative values are presented for screening purposes. To evaluate this parameter, the effect of the bandwidth of the Hilbert envelope and the frequency shift were analyzed. A good discrimination was obtained with a bandwidth of 1000 Hz and a frequency shift of 300 Hz. The results confirmed that the Glottal to Noise Excitation Ratio provides reliable measurements in terms of discrimination among normal and pathological voices, comparable to other classical long-term noise measurements found in the literature, such as Normalized Noise Energy and HNR. However, some of these parameters are based on an accurate estimation of the fundamental frequency, a fairly complex task in the presence of voice pathologies [19].

Short-time analysis has been used as an efficient tool to achieve more accurate characteristics for pathology in speech signal. In short-time analysis, voice disorders detection is based on the frames obtained by windowing of speech signal and therefore the final decision about being pathologic or normal is made with placing a threshold on the number of frames which are classified as normal and abnormal. A well-known method to extract proper features from speech signal is extracting the Mel-frequency cepstral coefficients (MFCC) [20]. This method is a parametric method that seeks to extract information from speech signal based on a physiological hypothesis. Godino-Llorente et al. have done several researches in the field of voice disorders recognition by employing MFCC analysis. The MFCC features, their energy, first and second derivatives evaluated by applying to multilayer perceptron neural network (MLP) and linear vector quantization (LVQ) [21]. Also, performance of F-ratio and Fisher's discriminant ratio as feature dimension reduction methods and Gaussian mixture model (GMM) were examined to improve the voice disorders identification rate in [19]. The results indicated that LVQ network had the best recognition rate, although its size is larger than the MLP. The interesting point in these results is that adding the first and second derivatives, especially second derivative, has not significant effect in the rate of recognition. Theoretically this issue demonstrated, in the identification of voice disorders, there are not considerable information in the dynamic of speech signal. Totally, the best recognition rate was achieved by 24 MFCC coefficients fed to Gaussian mixture model with 6 mixtures with the efficiency of 94.04%. A new scheme based on the modified cepstrum has been suggested by S.C. Costa et al. [22]. The calculation method for each modified cepstrum-based feature is discussed in [22] in detail. In this study, a LVQ network is applied to measure the distortion, and then detection of voice disorders was performed by using hidden markov model (HMM). In all cases HMM led to greatly improved results and, therefore the efficiency higher than 95% was obtained.

In some studies, automatic screening of pathological voice was investigated based on wavelets [23], fractals [24], neural maps and networks [25]. Titze et al. [26] suggested understanding of voice disorders with nonlinear dynamic concepts and analysis methods. Matassini et al. [27] have analyzed voice signals in a feature space including quantities from chaos theory (like correlation dimension and first lyapunov exponent) besides conventional linear parameters among which nonlinear parameters have been reported to have a clear separation between normal and sick voices.

On the other hand, discrete wavelet analysis has been employed in the field of voice disorders recognition because of its ability to extract time-frequency and localized information. J. Nayak et al. [28] proposed a voice disorders identification algorithm by applying energy of coefficients, which obtained by discrete wavelet transform (DWT) analysis, to a neural network. The system used to identify disorders such as vocal fold paralysis. Efficiency of 90% is achieved. Also, E.F. Fonseca et al. [29] employed Daubechies discrete wavelet transform, linear prediction coding and support vector machine as a novel method to identify disordered voices. Wavelets with different support-sizes and three LS-SVM kernels were compared. This algorithm presents classification accuracy over than 90%, and also it showed a low order of computational complexity respect to the speech signal's length.

The first step in the recognition of voice disorders is the awareness about the effects of abnormalities in the characteristics of speech signal. Therefore, in this study, we investigate qualitatively normal and pathological signals with time-frequency analysis by employing short-time Fourier transform and continuous wavelet transform. After that, on the basis of mentioned analysis, a quantitative method is introduced that provides preliminary decision about abnormalities in the larynx. Wavelet packet transform is employed as strong feature extraction strategy. Through the present paper six classification approaches applied to the automatic detection of voice disorders will be studied. Classifiers examined are quadratic discriminant classifier, nearest mean classifier, parzen classifier, k-nearest neighbor classifier, multilayer neural network, and support vector machines fed using energy and entropy feature vectors, calculated accordingly to the wavelet-packet transform parame-