Identification of Voice Disorders Using Long-Time Features and Support Vector Machine With Different Feature Reduction Methods

Meisam Khalil Arjmandi, Mohammad Pooyan, Mohammad Mikaili, Mansour Vali, and Alireza Moqarehzadeh,
Tehran, Iran

Summary: Identification of voice disorders has a fundamental role in our life nowadays. Therefore, many of these diseases must be diagnosed at early stages of occurrence before they lead to a critical condition. Acoustic analysis can be used to identify voice disorders as a complementary technique with other traditional invasive methods, such as laryngoscopy. In this article, we followed an extensive study in the diagnosis of voice disorders using the statistical pattern recognition techniques. Finally, we proposed a combined scheme of feature reduction methods followed by pattern recognition methods to classify voice disorders. Six classifiers are used to evaluate feature vectors obtained by principal component analysis or linear discriminant analysis (LDA) as feature reduction methods. Furthermore, individual, forward, backward, and branch-and-bound methods are examined as feature selection methods. The performance of each combined scheme is evaluated in terms of the accuracy, sensitivity, specificity, and area under the receiver operating characteristic curve (AUC). The experimental results denote that LDA along with support vector machine (SVM) has the best performance, with a recognition rate of 94.26% and AUC of 97.94%. Additionally, this structure has the lowest complexity in comparison with other architectures. Among feature selection methods, individual feature selection followed by SVM classifier shows the best recognition rate of 91.55% and AUC of 95.80%.
Key Words: Voice disorder identification–Feature reduction–Support vector machine–Linear discriminant analysis–Individual feature selection.

INTRODUCTION

In modern communities, voice disorders because of severe daily activities become a major issue. Hence, invasive techniques, such as stroboscopy, laryngoscopy, and endoscopy, are used by physicians to diagnose voice disorders, especially those that impair vocal fold mechanism. The health condition and functionality of vocal folds have some effects on the quality of voice. If the vocal folds become inflamed, some growths may develop on them or they become paralyzed and, as a result, the speech production process may fail. In cases with disordered voice, speech samples carry symptoms of disorder from their origin. Therefore, any abnormality in larynx often affects the quality of voice signal. The common disorders are vocal fold paralysis, vocal fold edema, adductor spasmodic dysphonia, anterior and posterior (A-P) squeezing, and others. Disorders usually show up in speech signal in the form of acoustic perceptual measures, such as hoarseness, breathiness, harshness, and inability to project the voice loudly.[1,2] The voice utterance results from three components of voice production: voiced sound, resonance, and articulation. Voiced sound produced by vocal fold vibration is amplified and modified by the vocal tract resonators (the throat, mouth cavity, and nasal passages), and finally, vocal tract articulators (the tongue, soft palate, and lips) modify the voiced sound; therefore, recognizable words are produced.[3] Vocal folds are in the form of two elastic hands of muscle tissue located in the larynx (voice box) directly above the trachea, and most of the disorders result from their malformation. In this context, we encounter several types of organic voice disorders, such as abductor spasmodic dysphonia, A-P squeezing, arytenoid dislocation, bowing, cyst, and others.

Researches show that digital processing of voice signal can be used as a noninvasive technique and an objective diagnosis tool to assess voice impairments in a research setting.[4–6] There are a large number of studies mainly focused on the accurate measurement of the fundamental parameters of the speech signal to detect different types of voice disorders. In previous researches, many long-time parameters, such as fundamental frequency (F_0), jitter, shimmer (Shim), amplitude perturbation quotient (APQ), pitch perturbation quotient, harmonics-to-noise ratio (HNR), normalized noise energy, voice turbulence index, soft phonation index, frequency amplitude tremor, glottal-to-noise excitation ratio,[2,4,7–15] and many others have been recommended to evaluate the quality of voice. For example, Hadjitodorov and Mitev,[5] on the basis of some widely used long-time acoustic parameters, such as Shim, jitter, and HNR, developed a computerized system for automatic analysis of pathological voices. On the other hand, Chen et al[16] designed a study to investigate the capability of several long-time features in the discrimination of disordered voice samples.

Furthermore, previous studies indicate that voice disorder identification can be done by extracting Mel frequency cepstral coefficients (MFCC) as a short-time nonparametric method. For example, Godino-Llorente and Gomez-Vilda[6] proposed a voice recognition algorithm based on short-term cepstral parameters and neural network (NN). In this article, the effectiveness of artificial NN and linear vector quantization (LVQ) networks has been evaluated for feature vectors, which include

Accepted for publication August 25, 2010.
From the Department of Electrical and Electronic Engineering, Faculty of Engineering, Shahed University, Tehran, Iran.
Address correspondence and reprint requests to Meisam Khalil Arjmandi, Unit 2, Number 9, Sefid Dast Alley, Emam'e Ali Boulevard, Mahallati Square, Shahid Mahallati Highway, Tehran, Iran. E-mail: msmarjmandi@gmail.com
Journal of Voice, Vol. 25, No. 6, pp. e275-e289
0892-1997/$36.00

doi:10.1016/j.jvoice.2010.08.003

MFCC parameters with different lengths. Feature vectors included MFCC features and their first and second derivatives. Results of this research demonstrated that speech signal parameterization with MFCC leads to accurate discrimination of disordered voice from normal voice, and that first and second derivatives of MFCC parameters did not improve significantly the classification rate. Also, LVQ network showed better performance compared with Multilayer Percepteron network. In their continued investigation, they had designed an automatic voice disorder detection system using the MFCC features, *F* ratio, and Fisher's discriminant ratio, as feature reduction methods, along with Gaussian mixture model to improve recognition rate. In the best case, the classification accuracy of 94.07% is achieved.[17] In some studies, linear prediction coding (LPC)[18] and discrete wavelet transform (DWT) analysis are applied to voice utterance to extract appropriate features from speech signal to identify pathological voices from normal ones. For example, Nayak and Bhat[19] reached an accuracy of 90% by applying DWT to extract efficient features and with NN as classifier. Furthermore, a novel scheme based on Daubechies' discrete wavelet transform, linear prediction coefficient, and least-square support vector machine (SVM) is suggested by Fonseca et al[20] that results in more than 90% accuracy in the detection of voice disorders in Brazilian Portuguese-language database.

In this article, some long-time acoustic features are applied to classify normal and pathological voices. Long-time parameters are generally calculated by *averaging* local time perturbations measured from a moving window over a long segment of speech signal, thus providing estimations of the degree of normality.[17] The parameters used in this study (Table 1) are proper measures of hoarseness, harshness, tremor, the intensity of amplitude and frequency noise (turbulence), abnormal noise, irregularity, variation of fundamental frequency, relative energy level of high frequency, and spectral density (frequency content) in the analyzed voice signal. These measures indicate that the detection of voice alternations can be carried out by means of the aforementioned long-time parameters. In addition to voice disorder diagnosis, the evaluation of the performance of laryngitis treatment and surveying pharmacological treatment and rehabilitation effects are other goals of this research. Acoustic parameters enable each individual voice utterance to be quantified by a single vector (feature vector). In this article, the ability of the aforementioned acoustic features in the identification of voice disorders is investigated.

This research attempts to evaluate different algorithms for classification and also attempts to introduce an efficient scheme to improve recognition rate and reduce complexity to design an automatic voice improvement system. Twenty-two long-time features, developed by the Massachusetts Eye and Ear Infirmary (*MEEI*) Voice and Speech Labs,[21] are used to discriminate pathological voice from normal ones. To improve voice disorder identification algorithm, effects of different feature reduction methods are examined. Through this study, the automatic detection of voice impairments is carried out by means of two feature extraction methods—as pattern recognition techniques—and four feature selection techniques. This research scrutinizes the strength of principal component analysis (PCA) and linear discriminant analysis (LDA) as feature reduction methods to achieve optimum feature set and to improve the classification rate. Furthermore, individual feature selection (IFS), forward feature selection (FFS), backward feature selection (BFS), and branch-and-bound feature selection (BBFS)—as feature selection methods—are evaluated in comparison with PCA and LDA methods. The efficiency of these methods is examined by six classifiers: quadratic discriminant classifier (QDC), nearest mean classifier (NMC), Parzen classifier (PC), K-nearest neighbor classifier (K-NNC), SVM, and multilayer NN (ML-NN). In this process, each structure, including certain long-time features, a feature reduction method, and a classifier, is examined to find the best algorithm to identify disordered voices. Finally, a novel and efficient approach is proposed to discriminate pathological voices from normal ones. This article demonstrates that only by using suitable long-time features with efficient feature reduction method and an appropriate classifier, voice disorder identification can be improved considerably. In the following parts of this article, the implementation of the methods will be discussed.

MATERIAL AND METHODS

Database

In this study, the voice samples are selected from the disordered voice samples,[21] model 4337, Version 1.03 (Kay Elemetrics Corporation, Lincoln Park, NJ), developed by the MEEI Voice and Speech Lab. The database comprises more than 1400 voice sample of approximately 700 subjects and includes speech samples from patients with a wide variety of organic, neurological, traumatic, and psychogenic voice disorders. In this research, the selected database contains 50 normal voice samples (32 male and 18 female) and 50 pathological voice samples (35 male and 15 female) with ages from 16 to 85 years (mean ± standard deviation = 44 ± 19). These data have been divided into two subsets, in which 70% were used for training and the remaining for validation. Every voiced speech signal was sampled from a sustained vowel /a/ (1-second long). All voice recordings were made in a soundproof booth on a digital audio tape (DAT) recorder at a sampling frequency of 16 kHz with 16-bit resolution for each subject. In addition, the database includes 33 parameters that were calculated from each of these voice signals. For calculation of long-time parameters, the records were framed using 32-millisecond windows.[21] These acoustic features are mentioned in Table 1. Totally, we used 22 long-time acoustic parameters of each sample, because other parameters either do not reflect voice quality or they are not reported for some voices. These acoustic parameters were derived from traditional *Multi-Dimensional Voice program* (*MDVP*; Kaypantax).[21]

Feature reduction

From the viewpoint of pattern recognition, there are two general approaches to perform feature reduction: feature extraction and feature selection. Feature extraction strategy is transforming a subset of the existing features into a lower dimensional space, whereas the feature selection method selects a subset of the