Investigation of the Effects of Speech Signal Length on Vocal Disorder Sorting Done Via Dynamic Pattern Modeling

***Vida Mehdizadehfar, *Farshad Almasganj, and †Farhad Torabinezhad,** *†*Tehran, Iran*

Summary: Objectives. Development of a noninvasive method for separating different vocal fold diseases is an important issue concerning vocal analysis. Due to the time variations along a pathologic vocal signal, application of dynamic pattern modeling tools is expected to help in the detection of defects that occur in the speech production mechanism.
Materials and Methods. In the present study, the hidden Markov model, which is a state space model, is employed to sort some of the vocal diseases. Moreover, this research mainly investigates the effects of the processed vocal signal lengths on the mentioned sorting task. To this end, the signal lengths of 1, 3, and 5 seconds of different disorders are used.
Results. The experimental results show that some pathologic conditions in vocal folds such as cyst, false vocal cord, and mass are more evident in continued voice production, and the recognition accuracies gained via dynamic modeling of pathologic voice signals with more lengths are considerably improved.
Key Words: Vocal disorder–Dynamic modeling–Hidden Markov model–Wavelet packet–Classification.

INTRODUCTION

Recognition of speech production systems' anomalies has recently attracted the attention of researchers in the fields of medical engineering and signal processing. Normal speech is composed of regular and alternating fluctuations, which can express the dynamic behavior of the vocal tract in the speech production period. The presence of pathogenic factors in the larynx leads to variations in the fluctuating and dynamic behavior of the vocal cords and anomalies in their regular performance. Resultantly, this finding changes the dynamic characteristics of the produced speech signal.[1] As a result, to achieve better curative methods for assessing vocal pathologies, it is required that the trend for the changes in natural fluctuations of the vocal cords be investigated.

The main focus of researchers in recent years has been on vocal parameters and on using standard classification techniques to achieve high accuracy in the pathologic assessment of the generated voices.[2] However, separation of different diseases from each other is not properly considered yet.[3,4] In most studies in this field, nondynamic features are extracted from vocal signals and indeed only reflect general information about the signal structure.[5–7] For instance, in Fraile et al,[8] to separate normal voices from pathologic ones, a multilayer perceptron neural network is employed. In this work, the mel-frequency cepstral coefficients (MFCCs) are the features fed to the neural network input layer, yielding a classification accuracy percentage of 88.3%. In Shama et al,[9] a linear discriminant analysis-based classifier is used for separating 53 normal speakers from 163 diseased speakers. The used features are selected from the time-frequency ones via which a classification accuracy percentage of 93.4% is obtained.

In some rare studies, application of dynamic processing methods in the recognition of vocal folds pathologies is considered. For instance, in Dibazar et al,[10] the hidden Markov model (HMM) is employed to recognize normal from pathologic voices. The used features involve 12 MFCC coefficients and their first and second derivatives, along with the voice basic frequency and energy. Two distinctive left-to-right HMM models are learned to model the normal and pathologic cases. For this purpose, the *HTK* software is employed. By applying this approach, the highest percentage of the recognition accuracy is reported to be 99.44%. In Lachhab et al,[11] a classifier based on the monophone HMM and Gaussian mixture model is applied to assess the pathologic voices. The proposed continuous speech recognition method is implemented over a French database. The obtained results showed a recognition rate of 63.59%. In Jothilakshmi,[12] the obtained accuracy for separating normal from pathologic voices is reported as 94.44%. The pathologic voices are then sorted, using the Gaussian mixture model applied to the MFCCs evaluated from the pathologic voices. In Dibazar et al,[13] sorting of five types of vocal fold diseases is considered. In Dibazar et al's work, the MFCC features are fed to a HMM-based classifier, and the sorting accuracy is reported as 71%. In Arias-Londoño et al,[14] pathologic assessment of voices is considered. The involved feature vector consists of short-time noise parameters and the MFCC features. In Arias-Londoño et al's work, a novel feature space transform technique is proposed. The steps in feature space transformation and classification are synchronously performed; the model parameters are calculated by considering a criterion that in turn minimizes the involved classification error. The experimental results show a recognition accuracy percentage of 96.67%.

Generally speaking, vocal signals have really a dynamic entity at all. Therefore, for a better investigation of such signals, their dynamic nature should be carefully extracted and studied. Besides, to classify vocal diagnosis or vocal abnormalities more realistically, the modeling or representation of voice signal variations in time could be efficiently exploited, for example, by introducing

Accepted for publication December 16, 2016.
From the *Faculty of Biomedical Engineering, Amirkabir University of Technology, Tehran, Iran; and the †Department of Speech therapy, School of Rehabilitation, Iran University of Medical Science (IUMS), Tehran, Iran.
Address correspondence and reprint requests to Vida Mehdizadehfar, Faculty of Biomedical Engineering, Amirkabir University of Technology, 424 Hafez Ave, Tehran, Iran, 15875-4413. E-mail: mehdizadeh@aut.ac.ir
Journal of Voice, Vol. 31, No. 4, pp. 515.e1–515.e8
0892-1997
© 2017 The Voice Foundation. Published by Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.jvoice.2016.12.009

an assigned time series. In some situations, the static features cannot afford to reflect exact information regarding such signals structures.

In Bayestehtashk et al,[15] the speech impairment in a number of patients with Parkinson disease is automatically assessed. The needed voices are elicited from the patients by implementing three different tasks: The sustained phonation task (10 seconds long), the diadochokinetic task (10 seconds long), and the reading task (4 minutes long). It is shown that, among all three kinds of recorded files, the reading task ones act significantly better at capturing cues needed to assess Parkinson disease. These works suggest that increasing the duration of voices recorded from patients could be potentially beneficial for automatic detection of vocal fold infections.

A normal person during voice production is expected to produce voice in its fixed form with an accepted tolerance; in this case, the changes in the basic frequency and intensity of the produced signals should be small enough. These features are typically checked by measuring parameters such as jitter, shimmer, and standard deviation of the pitch frequency in speech therapy and medicine. The vibration of vocal cords causes voiced signals; so, by increasing its vibration rate, the frequency, as an acoustic parameter, increases. Moreover, the perturbation of the vibrations leads to an increase in the jitter parameter but does not change the basic frequency of the generated signal.

In Arias-Londoño et al,[14] it was found that the estimated short-term jitter values evaluated for the recorded continuous speech show a more frequent presence of high jitter values in the case of pathologic voices as the voice production time increases. In Eskenazi et al,[16] the first 0.5 second of the signal samples is removed, and the remaining parts are analyzed. The number of voice breaks is then counted, whereas the locations of the voice breaks are ignored. It must be seriously investigated that in addition to the number of voice breaks, the times of their occurrences and some other probable dynamic properties could be also impressive. So, proper approaches must be developed to detect and utilize these internal changes of signals to clear the benefits of this hypothesis. This is the case in the present study, in which we are going to check out the potential of this idea using a dynamic model and signal samples with different lengths, short to long ones.

One salient feature of the dynamic modeling methods that distinguishes them from the static ones is that, in these methods, the variations existing in the temporal structure of the signal containing information are considered. In this condition, the assumptions are closer to reality and the speech production mechanism is more clearly stated.

The present study has sought to use the dynamic pattern modeling methods for recognition of speech disorders and to show the benefits of increasing the recorded speech signal length for this task. The main purpose of the present study was to investigate the effects of speech signal length on the classification of vocal disorders, taking into account the effect of voice breaks in longer time series. Clinically, the start time of the voice breaks is important; in severe voice disorders, the breaks start from the beginning of the generated voice, but in the mild ones, the breaks appear later. So, it seems that this feature can be efficiently used to discriminate these cases from each other. The rationale for this sort of acoustic analysis can be expressed briefly as follows:

(1) Due to some clinical observations, this hypothesis is strengthened by the finding that different voice disorders have different types of voice breaks, which could be an efficient measure for the separation of some diseases from each other. So, it is reasonable that this subject be seriously investigated in different properly designed studies.
(2) In speech therapy clinics, using a typical voice analysis software tool, this issue is somehow followed up and appears in other forms, for example, via an evaluation of several types of the jitter phenomenon; in the current study, this issue is indirectly investigated by applying the HMM to the prolonged voices.

The remaining part of the paper is organized as follows. In the "Materials and Methods," the database, the feature extraction method, and the used classifier are introduced. The third section of the paper shows the experimental results, and finally, the results and discussion are provided in the last section.

MATERIALS AND METHODS

The problem of detecting vocal signal dynamics and using dynamic features for sorting different kinds of vocal fold diseases is nearly a newly proposed approach in this field. In addition to considering the dynamic nature of a voice signal and using the information contained in the time pattern of variations across the voice signal for abnormality assessment, the present study investigates the possibility of separating different speech diseases from each other, using this approach. In this way, the wavelet entropy features are first extracted from vocal signals, and a dynamic pattern modeling method is then used to sort the involved diseases. In this section, we explain the theory basics of the implemented methods. After that, the tools exploited in the present study will be introduced.

Database

One of the most commonly used databases in the field of speech processing is the KAY database (Kay Elemetrics Corporation, Lincoln Park, NJ), developed by the Massachusetts Eye and Ear Infirmary Voice and Speech Lab. The samples of this database are mostly 1 second long, which is a short period for the present study viewpoint. Investigation of the effect of the vocal signal length on sorting involved diseases requires a dataset containing signals with higher time lengths.[17]

In the current work, the used vocal signals were chosen from a set of Native Persian Speakers' Voices (briefly called NPSV) including 61 samples of pathologic vocal signals. These samples were collected and labeled in the Language and Speech Pathology Laboratory of the Rehabilitation Faculty of Iran University of Medical Sciences. All signals were recorded in an acoustic noise-free room with a speech studio record system with a sampling frequency of 16 kHz. The individuals were asked to express the vowel /a/ for a short while. Each vocal sample had a disease label, designated based on the larynx stroboscopy and clinical treatment done by two expert laryngologists.

This database consists of the voice samples from patients with different abnormalities such as the false vocal cord, functional,

Download English Version:

https://daneshyari.com/en/article/5124208

Download Persian Version:

https://daneshyari.com/article/5124208

Daneshyari.com