Discrimination of Pathological Speech Using Hidden Markov Models

Jianglin Wang* · Cheolwoo Jo*

ABSTRACT

Diagnosis of pathological voice is one of the important issues in biomedical applications of speech technology. This study focuses on the discrimination of voice disorder using HMM (Hidden Markov Model) for automatic detection between normal voice and vocal fold disorder voice. This is a non-intrusive, non-expensive and fully automated method using only a speech sample of the subject. Speech data from normal people and patients were collected. Mel-frequency filter cepstral coefficients (MFCCs) were modeled by HMM classifier. Different states (3 states, 5 states and 7 states), 3 mixtures and left to right HMMs were formed. This method gives an accuracy of 93.8% for train data and 91.7% for test data in the discrimination of normal and vocal fold disorder voice for sustained /a/.

Keywords: pathological speech, HMM, MFCCs

1. Introduction

The diagnosis of pathological voice is a hot topic that has been recently received considerable attention. There are several medical diseases that adversely affect our human voice [1]. There are numerous sorts of diseases which can be happened in our vocal cord. The doctor can use the available apparatus for detection of pathological voice. Generally medical doctors can use various apparatuses like EMG(electromyography), electroglottograph and videoendoscopy to assess and diagnose the malfunctioning vocal fold. However, it is invasive and requires an expert analysis of numerous human speech signal parameters. Automatic voice analysis for

* SASPL, School of Mechatronics, Changwon National University

pathological speech has its advantages, such as having its quantitative and non-invasive nature, allowing the identification and monitoring of vocal system diseases and reducing analysis cost and time [2]. In the pathological voice classification, based on the voice of a patient, the goal is to make a decision whether it is normal or pathological. Successful pathological voice classification will enable an automatic device to diagnose and analyze the voice of the patient [3].

There are many diseases which can be occurred on vocal cord. These diseases affect the quality of our speech signal. Speech quality is very subjective and can be used to denote distortion in a signal, clarity of the spoken words, intermittent loss of data etc. Human listeners, doctors or even plain peoples in some degree, can distinguish the differences between normal and pathological speech. Our experiment is focusing on that point. Applying HMMs to speech signal, it can extract certain feature details to construct a model that is characteristic of the speech signal. Research has been conducted to investigate the use of such models to estimate the quality of speech signal [4]. In this application, the HMM of the test speech sample is compared with that of a known, pure speech sample, to compute a distance. This distance denotes the deviation of the test sample from the source speech sample and is used as an indication of speech quality. The measured distance is inversely proportional to speech quality.

In previous studies, several methods for classifying pathological voice have been introduced. The paper described in [5] provides information about how acoustic analysis can be used to measure parameters in motor speech disorders. Research has also been conducted to automatically detect speech pathology. Dibazar et al. applied HMM to classify the pathological speech. A good accuracy of that study has been achieved with this method. Simple measures have been described that correlate well with perceptual features like vowel durations, imprecise consonants and others. Acoustic correlates have also been developed to detect certain speech characteristics like breathiness in a subject [6]. In other researches using artificial neural network, such as classification of pathological voice including severely noisy cases [7], pathological voice quality assessment using ANN [8] and using short-term cepstral parameters and neural network based detectors for automatic detection of voice impairments [9], have recently been applied to various kinds of pathological classification tasks. Generally, ANN has been widely used because we need not think

about the details of the mathematical models of the data and relatively easy to train and has produced a good pathological recognition performance. The major drawback to ANN method is that it highly depends on the data set and the ANN method can not show us the good performance if the new data is added.

In this paper, to successfully achieve the classification of pathological speech, Mel-frequency filter cepstral coefficients have been employed. The HMM-based method was used to classify the pathological voice into normal and pathological voices. This paper is organized as follows: Section 2 introduces the experimental database. Section 3 presents the theory of MFCC, Hidden Markov Toolkit, Hidden Markov Models. Section 4 shows the experimental results and discussion. Finally, the conclusion is included in section 5.

2. Database

To collect voice data, collection system was installed in a room of the otolaryngology department of hospital. The recording process was executed semi-automatically with the intervention of operator to control the quality and procedure. Also the voice materials from the different male speakers were collected using DAT (Digital Audio Tape) [10]. The sampling rate was 50 KHz and the resolution 16 bits.

The collection was conducted in the soundproof room of a hospital. All the subjects were asked to pronounce a sustained vowel /a/. Total voice data included 41 normal cases and 111 pathological cases (108 relatively less noisy voices and 3 severely noisy voices) after removing invalid data from the raw data sets. The vocal diseases considered in the relatively less noisy pathological cases consisted of Vocal Polyp, Vocal Cord Palsy, Vocal Nodule, Vocal Cyst, Vocal Edema, Laryngitis and Glottic Cancers. The database of vocal fold diseases is shown in Table 1.

Table 1. Vocal Fold Diseases

Disease	No. of Case

Cyst	5
Edema	10
Laryngitis	5
Nodule	10
Palsy	10
Polyp	10
Glottic Cancer	58
Total	108

3. Methodology

In this section, the parameters which used in our study will be introduced and the analysis procedures and methods are described.

3.1 Parameters

The cepstral coefficients have been widely applied in speech recognition applications. A distinctive advantage of the cepstral analysis is that correlation between coefficients is extremely small so that simplified modeling assumptions can be applied. In particular, the effect of inserting a transmission channel on the input speech is to multiply the speech spectrum by the channel transfer function [11].

The MFCCs are calculated from the log filterbank amplitudes $\{m_j\}$ using the discrete cosine transform

$$c_i = \sqrt{\frac{2}{N}} \sum_{j=1}^{N} m_j \cos\left(\frac{\pi}{N}(j-0.5)\right) \tag{1}$$

where N is the number of filterbank channels.

The delta coefficients are computed using the following regression formula

$$d_t = \frac{\sum_{\theta=1}^{\Theta} \theta(c_{t+\theta} - c_{t-\theta})}{2\sum_{\theta=1}^{\Theta} \theta^2} \tag{2}$$

where d_t is a delta coefficient at time t computed in terms of the corresponding static coefficients $c_{t-\Theta}$ to $c_{t+\Theta}$. The value of Θ is set using the configuration parameter DELTAWINDOW. Since the equation 2 relies on past and future speech parameter values, some modification is needed at the beginning and end of the speech. The default behavior is to replicate the first or last vector as needed to fill regression window.

This end-effect problem was solved by using simple first order differences at the start and end of the speech, that is

$$d_t = c_{t+1} - c_t, \quad t < \Theta \tag{3}$$

and

$$d_t = c_t - c_{t-1}, \quad t \geq T - \Theta \tag{4}$$

where T is the length of the data file.

3.2 HTK

HTK is a toolkit for building Hidden Markov Models (HMMs). HMMs can be used to model any time series and the core of HTK is similarly general-purpose. However, HTK is primarily designed for building HMM-based speech processing tools, in particular speech recognizers.

HTK consists of a set of library modules and tools available in C source form. The tools provide sophisticated facilities for speech analysis, HMM training, testing and results analysis. The software supports HMMs using both continuous density mixture Gaussians and discrete distributions and can be used to build complex HMM systems [12].

3.3 HMM

The HMM uses a Markov chain to model the changing statistical characteristics that exist in the actual observations of speech signals. The Markov process is therefore a double stochastic process in which there is an unobservable Markov chain defined by a state transition matrix, and where each state of the Markov chain is associated with either a discrete output probability distribution (discrete HMM) or a continuous output probability density function (continuous HMM). The double stochastic processes enable modeling of not only acoustic phenomena, but also time scale distances [13].

The main focus of this study is the binary classification of the speech signal. Let each of subjects be represented by a sequence of feature vectors O, which are the Mel frequency cepstral coefficients. Detection of pathological speech can be regarded as computing:

$$\arg MAX\{P(w_i|O)\} \tag{5}$$

where w_i={normal, pathology}.

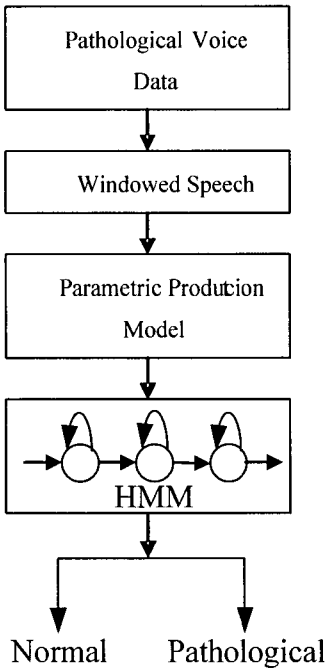

Figure 1. Flow of HMM-based classification

The discrimination of pathological voice used in this study was based on HMM-based approach, which consists of 3 steps. First, the pathological voice data using 10 msec Hamming windowed frames were extracted. Second, the MFCCs were obtained using the parametric production model. Third, the HMM models using different states (3 states, 5 states and 7 states), 3 mixtures and left to right HMM models were trained in order to obtain an optimal classification rate. The flow of HMM-based method is shown in Fig 1.

4. Results and Discussion

In this section, we discuss some experimental results obtained from the proposed analysis methods. Once the test data has been processed by the designed recognizer, the next step is to analyze the results. The percent correct is defined as

$$\text{Percent Correct} = \frac{N-D-S}{N} * 100\% \quad (6)$$

where S: the number of substation errors,

D: deletion errors,

N: the total number of labels in the reference transcriptions.

The correct classification rates using HMM-based method are shown in Table 2. In Table 2,

Table 2. The correct classification rates (%) in HMM.

States	Train	Test
3	90.4	88.5
5	92.5	87.5
7	93.8	91.7

the respective results for train data and test data are shown. And it was observed that the classification rates were increased with the rising number of HMM states. And the rate is the biggest when the number of states is 7. Increasing number of states over 7 didn't show better results. Also current results are compared with those of our previous work using ANN (artificial neural network) method [7] for reference. In Table 3, the classification results of ANN-based method are shown. Comparing Table 2 to Table 3, the HMM method did not show better results for train data and test data. But direct comparison between two cases is not proper due to the differences of parameters used. With ANN 6 parameters (Jitter, Shimmer, NHR, SPI, APQ and RAP) were used. The results are not better than those of reference [3] (Train 98.59%, Test 97.75%). The possible reason can be different set of data used.

Table 3. The correct classification rates (%) in ANN.

No. of Hidden Layers	Train	Test
6	98.0	90.2
9	97.4	90.2
12	98.0	94.3

The confusion matrix can show us how many correct classification rates have been identified. The best result (7 states) is presented through confusion matrices. In the Table 4, True positive (TP) is the ratio between normal voice correctly classified and the total number of normal voices. False negative (FN) is the ratio between wrongly classified normal voices and the total number of normal voices. True negative (TN) is the ratio between pathological voices correctly classified and the total number of the pathological voices. False positive (FP) is the ratio between pathological voices wrongly classified and the total number of pathological voices. From table 4, HMM classification shows a highly correct classification rate for normal and pathological voices. The FP=6.7% means that some normal speech data were misclassified as pathological case; however, the TN=100% demonstrates that all of the pathological speech data were correctly classified. From the result, we can reason that the pathological speech voice owns so distinctive characteristics that HMM can recognize the data according to its characteristic parameters.

Table 4. The confusion matrix for HMM classification.

		Decision	
		Normal	Pathological
In	Normal	TP=93.3%	FP=6.7%
	Pathological	FN=0%	TN=100%

5. Conclusions

A pathological voice classifier has been designed in this project using discrete HMMs. Speech is classified into normal and pathological case. With the 7 states Hidden Markov

Model using MFCC parameters, the accuracy of 93.8% for train data and 91.7% for test data are achieved. Comparing HMM-based method to ANN-method, the latter one can show us good results for the same pathological data. However, characteristic parameters are different in each experiment. So the direct comparison is not proper. The confusion matrix for HMM classification shows that the pathological voice including distinctive characteristic can be accurately classified by statistical method and the result is comparable to those of human experts and considered to be useful for various cases such as online screening test. Because the total amount of voice data is still not enough to generalize the performance and characteristic, more data collection is required. In the future work, more characteristic parameters and classification approach should be studied in the classification method.

References

[1] Reynolds D. A. & Rose R. C. 1995. "Robust Text-independent Speaker Identification Using Gaussian Mixture Speaker Models", *IEEE Transactions on Speech and Audio Processing*, vol. 3, 72-83.

[2] Cheolwoo Jo & Daehyun Kim. 1998. "Diagnosis of Pathological Speech Signals Using Wavelet Transform", *Proceedings of ITC-CSCC'98*, 657-660, Sokcho, Korea.

[3] Alireza Afshordi Dibazar & Shikanth Narayanan. 2002. "A System for Automatic Detection of Pathological Speech", *36th Asilomar Conf. Signals, Systems & Computers.*

[4] Talwar. G. & Kubichek, R. 2003. "Output-based speech quality measurement using hidden Markov models", Int.l Signal Processing Conference, Dallas TX.

[5] Rabiner L. R. & Juang B. H. 1986. "An introduction to hidden Markov models", *IEEE ASSP Magazine*, Vol. 3, 4-16.

[6] Forrest, K. & Weismer, G. 1997. "Acoustic analysis of dysarthric speech", *Clinical Management of Sensorimotor Speech Disorders*, Thieme Medical Pub.

[7] Tao Li, Cheolwoo Jo, Soo-Geon Wang, et al. 2004. "Classification of pathological voice including severely noisy cases", in *Proc. 8th International Conference on Spoken Language Processing (INTERSPEECH 2004-ICSLP)*, vol. 1, 77-80, Jeju, R. O. Korea.

[8] Frohlich, M., Michaelis, D. & Strube, H. 1998. "Acoustic breathiness measures in the description of pathological voices", *ICAASP*, vol. 2, 937-940.

[9] Ritchings, R. T., McGillion, M. A. & Moore, C. J. 2002. "Pathological voice quality assessment using artificial neural network", *Medical Engineering & Physics*, vol. 24, no. 8, 561-564.

[10] Cheolwoo Jo, Kwangin Kim, Daehyun Kim, Soogeon Wang & Gyerok Jeon. 2001. "Comparisons of Acoustical Characteristics between ARS and DAT Voice", *2001 International Conference on Speech Processing (ICSP'2001)*, Taejon, Korea.

[11] Young, S., Kershaw, D., Odell, J., Ollason, D., Valtchev, V. & Woodland, P. 2000. "The HTK book", Microsoft Corporation. http://htk.eng.cam.ac.uk/

[12] Huang, X. D., Ariki Y. & Jack M. A. 1990. "Hidden Markov Models for speech recognition", *Edinburgh Information Technology Series*, 79-80.

received: July 31, 2006
accepted: August 28, 2006

Jianglin Wang
SASPL, School of Mechatronics, Changwon National University
9 Sarim-dong, Changwon, Kyungnam, Korea (641-773)
Tel: +82-55-279-7559 Fax: +82-55-262-5064
E-mail: xiaowangyc@hotmail.com

Cheolwoo Jo
SASPL, School of Mechatronics, Changwon National University
9 Sarim-dong, Changwon, Kyungnam, Korea (641-773)
Tel: +82-55-279-7552 Fax: +82-55-262-5064
E-mail: cwjo@sarim.changwon.ac.kr