

A HMM-Based Method for Vocal Fold Pathology Diagnosis

Vahid Majidnezhad¹ and Igor Kheidorov²

¹ Department of Computer Engineering, Shabestar Branch, Islamic Azad University
Shabestar, Iran

² Department of Computer Engineering, Shabestar Branch, Islamic Azad University
Shabestar, Iran

Abstract

Acoustic analysis is a proper method in vocal fold pathology diagnosis so that it can complement and in some cases replace the other invasive, based on direct vocal fold observation's methods. There are different approaches for vocal fold pathology diagnosis. This paper presents a method based on hidden markov model which classifies speeches into two classes: the normal and the pathological. Two hidden markov models are trained based on these two classes of speech and then the trained models are used to classify the dataset. The proposed method is able to classify the speeches with an accuracy of 93.75%. The results of this algorithm provide insights that can help biologists and computer scientists design high-performance system for detection of vocal fold pathology diagnosis.

Keywords: *Vocal fold pathology diagnosis, Hidden Markov Model (HMM), Mel-Frequency-Cepstral-Coefficients (MFCCs), Fundamental frequency, Vector Quantization (VQ), LBG Algorithm.*

1. Introduction

Vocal signal information often plays an important role for specialists to understand the process of vocal fold pathology formation. In some cases vocal signal analysis can be the only way to analyze the state of vocal folds. Nowadays diverse medical techniques exist for direct examination and diagnostics of pathologies.

Laryngoscopy, stroboscopy, electromyography and videokimography are most frequently used by medical specialists. But these methods possess a number of disadvantages. Human vocal tract is hardly-accessible for visual examination during phonation process and that makes it more problematic to identify pathology. Moreover, these diagnostic means may cause patients much discomfort and distort the actual signal that may lead to incorrect diagnosis as well [1-4].

Acoustic analysis as a diagnostic method has no drawbacks, peculiar to the above mentioned methods. It possesses a number of advantages. First of all, acoustic analysis is a non-invasive diagnostic technique that allows

pathologists to examine many people in short time period with minimal discomfort. It also allows pathologists to reveal the pathologies on early stages of their origin. This method can be one of great interest for medical institutions.

The presence of pathology in a vocal tract inevitably leads to voice signal distortion. Depending on pathology severity the distortion may be more or less significant. Among all sounds that are produced by vocal tract, sustained vowels and some sonorant consonants are most easily distorted if pathology is present.

One of the modern methods of speech modeling is hidden Markov models, successfully used for speech recognition [5-6]. This method presents speech as a stochastic random process and determines statistical structure of speech.

2. Material and Methods

In this section the dataset and the tools, which are used for the implementing of the proposed method, are explained.

2.1 Dataset

The database was created by specialists from the Belarusian Republican Center of Speech, Voice and Hearing Pathologies. We have selected 40 pathological speeches and 40 healthy speeches randomly which are related to sustained vowel "a". All the records are in PCM format, 16 bits, mono, with 16 kHz sampling frequency.

2.2 Mel-Frequency-Cepstral-Coefficients (MFCCs)

MFCCs are widely used features to characterize a voice signal and can be estimated by using a parametric approach derived from linear prediction coefficients (LPC), or by the non-parametric discrete fast Fourier transform (FFT), which typically encodes more information than the LPC method. The signal is

windowed with a hamming window in the time domain and converted into the frequency domain by FFT, which gives the magnitude of the FFT. Then the FFT data is converted into filter bank outputs and the cosine transform is found to reduce dimensionality. The filter bank is constructed using 13 linearly-spaced filters (133.33Hz between center frequencies,) followed by 27 log-spaced filters (separated by a factor of 1.0711703 in frequency.) Each filter is constructed by combining the amplitude of FFT bin.

The MFCCs are used as features in [7] to classify the speech into pathology and healthy class. We have also used MFCC coefficients as the observation sequences of models. The Matlab code to calculate the MFCC features was adapted from the Auditory Toolbox (Malcolm Slaney).

2.3 LBG Algorithm

An algorithm for a scalar quantizer was proposed by Lloyd (1957). Later, Linde et al. (1980) generalized it for vector quantization. This algorithm is known as LBG or generalized Lloyd algorithm (GLA). It applies the two necessary conditions to inputting data in order to determine optimal vector quantizers [8].

Given inputting vector data $x_i, i=1, \dots, n$, distance function d , and initial codewords $c_j(0), j=1, \dots, m$, the LBG iteratively applies two conditions to produce a codebook with the following algorithm:

- a) Partition the inputting vector data $x_i, i=1, \dots, n$, into the channel symbols using the minimum distance rule. This partitioning is stored in an $n*m$ indicator matrix S whose elements are defined as the following

$$S_{ij} = \begin{cases} 1, & \text{if } d(x_i, c_j(k)) = \min_p d(x_i, c_p(k)) \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

- b) Determine the centroids of the Voronoi regions by channel symbol. Replace the old codewords with these centroids:

$$c_j(k+1) = \frac{\sum_{i=1}^n S_{ij} x_i}{\sum_{i=1}^n S_{ij}}, j = 1, \dots, m \quad (2)$$

- c) Repeat steps a and b until no $c_j, j=1, \dots, m$ changes anymore.

2.4 Hidden Markov Model

A Hidden Markov Model (HMM) [9] is a finite state machine which has some fixed number of states. It provides a probabilistic framework for modeling a time series of multivariate observations. Hidden Markov models were introduced in the beginning of the 1970's as a tool in speech recognition. This model based on statistical methods has become increasingly popular in the last several years due to its strong mathematical structure and theoretical basis for use in a wide range of applications. In recent years researchers proposed HMM as a classifier or predictor for speech signal recognition, DNA sequence analysis, handwritten characters recognition, natural language domains etc. It is clear that HMM is a very powerful tool for various applications.

HMM is specified by a five-tuple (S, O, π , A, B) [10].

- a) $S = \{1, 2, \dots, N\}$ Set of hidden states.
 N : the number of states, S_t : the state at time t .
- b) $O = \{O_1, O_2, \dots, O_M\}$ Set of observation symbols.
 M : the number of observation symbols.
- c) $\pi = \{\pi_i\}, \pi_i = P(S_0=i), 1 \leq i \leq N$
 The initial state distribution.
- d) $A = \{a_{ij}\}, a_{ij} = P(S_t=j | S_{t-1}=i), 1 \leq i, j \leq N$
 State transition probability distribution.
- e) $B = \{b_i(k)\}, b_i(k) = P(X_t=O_k | S_t=i), 1 \leq i \leq N, 1 \leq k \leq M$
 Observation symbol probability distribution in states.

Given a HMM, $\lambda = (A, B, \pi)$, and a sequence of observations, $O = \{O_1, O_2, \dots, O_L\}$, the objective in the final classification phase is calculate the probability $P(O|\lambda)$ (the probability of the model that generates these observations).

3. Proposed Method and Results

In this section a HMM-based method for vocal pathology diagnosis is proposed. The block diagram of the proposed method is shown in the Fig. 1. First two subsets of dataset, the test and train set, are made so that the train set includes 50% of records of dataset randomly and the test set includes whole records of dataset. Then for each record of train set, the MFCC coefficients along with the fundamental frequencies are extracted and assigned as the observations and states of model respectively. Since these values are continuous, Vector Quantization (VQ) must be applied on them. For this purpose, by the means of LBG algorithm these continuous values are converted to the discrete values and are feed to the HMM model. Finally, two models for healthy and pathological speech signals are trained.

For evaluating the performance of the proposed method, for each record of test set, the MFCC coefficients are extracted and assigned as the observations of model. Then by the means of LBG algorithm these continuous values are converted to the discrete values and are feed to the healthy and pathology trained HMMs. Then the likelihood of both trained models for the given observation sequence are calculated. The maximum likelihood identifies the type of signal which is healthy or pathological.

The proposed method is simulated in Matlab 7.11.0. In the proposed method, the Kevin Murphy toolbox is used for training the HMMs [11]. In both models the number of observations is $N=4$ and the number of states is $M=2$.

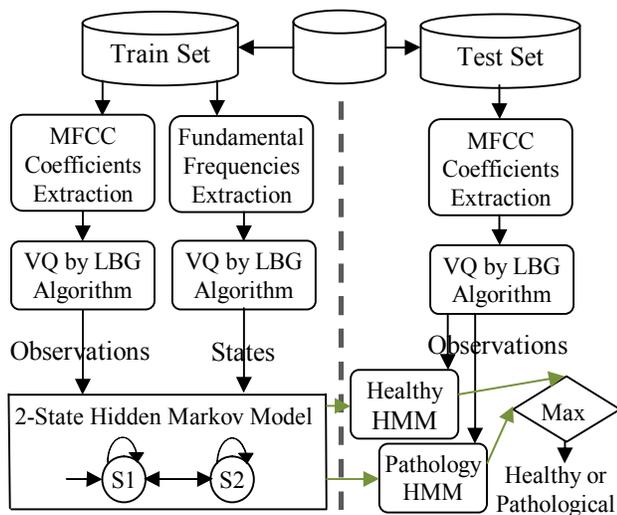


Fig. 1 The block diagram of the proposed method.

For displaying the results we have used these indicators: true positive rate (TP), also called sensitivity, is the ratio between pathological files correctly classified and the total number of pathological voices. False negative rate (FN) is the ratio between pathological files wrongly classified and the total number of pathological files. True negative rate (TN), sometimes called specificity, is the ratio between normal files correctly classified and the total number of normal files. False positive rate (FP) is the ratio between normal files wrongly classified and the total number of normal files. The final accuracy of the system is the ratio between all the hits obtained by the system and the total number of samples. Also for displaying the results, the ROC curve (AUC) [12] is used which is a very common way in medical decision systems.

The classification results, including ROC curve and classification results, are illustrated in table 1 and Fig. 2.

Table 1: The classification results of the Proposed Method

	TP	FP	TN	FN	Accuracy
Results (%)	97.5	2.5	90	10	93.75

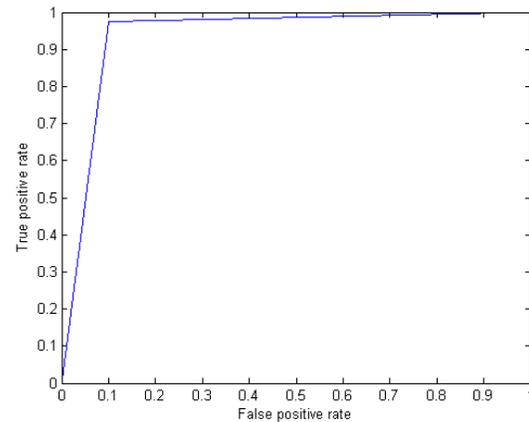


Fig. 2 The ROC curve of the proposed method.

4. Conclusion and Future Work

In this article, two hidden markov models for healthy and pathological speech signals are created. For this purpose, the Mel-Frequency-Cepstral-Coefficients (MFCC) and the fundamental frequencies are used as the observations and states of models. Since these values are continuous, they are converted to the discrete values by the means of LBG algorithm. The experiments results show that the proposed method has rather good classification accuracy.

It may be possible to try to build a complete multiclass classification system with multiple HMMs so that detection of different type of pathological speech will be possible. In other words, one HMM for each class of speeches will be trained. So, authors suppose it as the future work.

Acknowledgment

This work was supported by the speech laboratory of NASB in Belarus. The authors wish to thank the Belarusian Republican Center of Speech, Voice and Hearing Pathologies by its support in the speech database.

References

- [1] J. B. Alonso, J. Leon, I. Alonso and M. A. Ferrer, "Automatic Detection of Pathologies in the Voice by HOS Based PArAmeters", *EURASIP Journal on Applied Signal Processing*, Vol. 2001, No. 4, 2001, pp. 275-284.
- [2] L. Gavidia-Ceballos, J. Hansen and J. Kaiser, "A Non-Linear Based Speech Feature Analysis Method with Application to Vocal Fold Pathology Assessment", *IEEE Trans. Biomedical Engineering*, Vol. 45, No. 3, 2005, pp. 300-313.
- [3] L. Gavidia-Ceballos, J. Hansen and J. Kaiser, "Vocal Fold Pathology Assessment Using AM Autocorrelation Analysis of the Teager Energy Operator", *ICSLP-1996 Proc.*, 1996, pp. 757-760.
- [4] C. Adnene and B. Lamia, "Analysis of Pathological Voices by Speech Processing", *Signal Processing and Its Applications 2003 Proc.*, 2003, 1(1), pp. 365-367.
- [5] G. Hemakumar, "Vowel-Plosive of English Word Recognition using HMM", *IJCSI International Journal of Computer Science Issues*, Vol. 8, No. 6, 2011, pp. 362-365.
- [6] M. Dua, R.K. Aggarwal, V. Kadyan and S. Dua, "Punjabi Automatic Speech Recognition Using HTK" *IJCSI International Journal of Computer Science Issues*, Vol. 9, No. 4, 2012, pp. 359-364.
- [7] L. Ji-Yeoun, J. SangBae and H. Minsoo, "Classification of pathological and normal voice based on linear Discriminant analysis", *Proceedings of the Eighth International Conference on Adaptive and Natural Computing Algorithms*, 2007, Vol. 4432. pp. 382-390.
- [8] F. Shen and O. Hasegawa, "An adaptive incremental LBG for vector quantization", *Neural Networks*, 2005, pp. 1-11.
- [9] S. C. Costa, B. G. A. Neto and J. M. Fechine, "Pathological Voice Discrimination using Cepstral Analysis, Vector Quantization and Hidden Markov Models", *8th IEEE International Conference on Informatics and BioEngineering*, 2008, pp. 1-5.
- [10] L. R. Rabiner, "A Tutorial on Hidden Markov Models, and Selected Applications in Speech Recognition". *Proceedings of the IEEE*, Vol. 77, No. 2, 1989, pp. 257-286.
- [11] <http://www.cs.ubc.ca/~murphyk/Software/HMM/hmm.html>
- [12] J.A. Hanley and B.J. McNeil, "The meaning and use of the area under a receiver operating characteristic (ROC) curve", *Radiology*, Vol. 143 , No. 1, 1982, pp. 29-36.