Feature Analysis for Automatic Detection of Pathological Speech

Alireza A. Dibazar[1], S. Narayanan[2], T. W. Berger[3],

[1,3] Biomedical Engineering Department, University of Southern California, CA, USA
[2] Electrical Eng. Department, University of Southern California, CA, USA
Email: dibazar@usc.edu

Abstract- **This study focuses on a robust, rapid and accurate system for automatic detection of normal and pathological speech. This system employs non-invasive, non-expensive and fully automated measures of vocal tract characteristics and excitation information. Mel-frequency filterbank cepstral coefficients and measures of pitch dynamics were modeled by Gaussian mixtures in a Hidden Markov Model (HMM) classifier. The method was evaluated using the sustained phoneme /a/ data obtained from over 700 subjects of normal and different pathological cases from the Massachusetts Eye and Ear Infirmary (MEEI) database. This method attained 99.44% correct classification rates for discrimination of normal and pathological speech for sustained /a/. This represents 8% detection error rate improvement over the best performing classifier using carefully measured features prevalent in the state-of-the-art in pathological speech analysis.**

***Keywords* – pathological speech, HMM, Mel frequency filters**

I. INTRODUCTION

There are numerous medical conditions that adversely affect the voice. Many of these conditions have their origins primarily in the vocal system and many tools available for detection of speech pathologies are invasive or require expert analysis of numerous human speech signal parameters. So, a reliable, accurate and non-invasive automatic system for recognizing and monitoring speech abnormalities is one of the necessary tools in pathological speech assessment.

In previous studies, several methods for assessing speech pathologies have been introduced. In general, these methods, based on features they use, fall into two groups: Spectral envelope measures and temporal dynamic measures. In the spectral analysis methods, researchers have tried to keep track of the spectral variations of signal such as amplitude, bandwidth and frequency of formants including subband processing methods. In time domain, authors have employed two major methods: 1) Methods based on temporal measurements of signal and their statistics, such as average pitch variation, jitter, shimmer, etc 2) analysis of residual of inverse filtering [1] of the speech signal, which corresponds to an estimate of the source excitation, to distinguish between normal and pathological speech. However, there are some difficulties associated with these methods. As it has been reported in published articles, these methods have accuracies between 80 and 90% [2] , typically with a small limited number of subjects. In addition, there are robustness, consistency and comp lexity difficulties for measuring those features including the degree of human intervention needed in the measurements.

In this study, to successfully achieve the assessment of pathological speech, spectral envelope and pitch information have been employed. The aim of this work is to classify speech signals in terms of being normal or pathological.
This paper is organized as follows: in the next section, the employed method and database are discussed. In section three, the experimental results are described and finally, in section four, the conclusions of this study are reported.

II. METHOD AND DATABASE

In our implementation, two requirements were imposed. First, the features had to be efficient in terms of measurement cost and time. Second, both the vocal tract and excitation source information, had to be included. The block diagram of the proposed algorithm is shown in Fig. 1. The cepstral features of a mel frequency filter bank outputs were obtained by a standard short-term speech analysis along with fra me-level pitch estimates. The normalized cross-correlation [3] based method was used for pitch estimation. The algorithm assumes a monophonic signal. The method follows the assumption that the signal has a periodicity corresponding to the fundamental frequency or pitch. Then, a HMM based classifier was applied wherein these features were modeled by mixture Gaussians.

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 (MFCCs) and pitch. Detection of pathological speech can be regarded as computing:

$$\arg \underset{i}{MAX}\{P(w_i \mid O)\} \qquad (1)$$

where, $W_i = \{normal\ pathology\}$. In practice, if a parametric production model such as the Markov model is assumed, then computing the joint probabilities, which are necessary for solving (3), can be replaced by estimating the Markov model parameters.

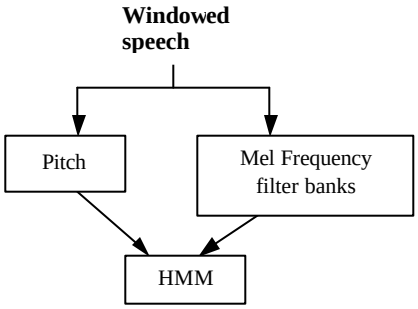

Figure 1: Block diagram of the method.

For training HMM, the hidden [4] Markov model toolkit (HTK) was modified to accommodate the fundamental frequency. Twelve Mel frequency Cepstral coefficients using 10 msec Hamming windowed frames were extracted. The pitch frequency was also computed in the same window. The zero order energy and F0 were included with the above-mentioned features. In order to take the advantage of pitch and spectral dynamics, the velocity (delta) and acceleration parameters were also added to the feature space.

The database [5] developed by Massachusetts Eye and Ear Infirmary (MEEI) Voice and Speech Laboratory was used. It contains voice samples of 710 subjects. Included are sustained phonation speech samples from patients with a wide variety of organic, neuralgic, traumatic, and psychogenic voice disorders, as well as 53 normal subjects.

Along with sustained samples of the vowel /a/, the database also contains acoustic speech sample files: readings of "Rainbow passage". The Rainbow dataset consists of 657 pathological subjects from the same patients who provided samples of the sustain vowel /a/. There are also 53 recordings of up to 12 second of the Rainbow passage from the same normal subjects included in normal sustained /a/.

Utterance level results of analyzing each of the vowels by the Multi-Dimensional Voice Program (MDVP) were also included in the database [5]. These acoustic parameters, which include 34 time domain features of speech signal, represent a superset of the most popular measures employed in pathological speech analysis.

III. RESULTS

In order to evaluate effectiveness of the method and features, data from 657 abnormal and 53 healthy subjects from the MEEI database were used. The 12 cepstral coefficients, first order energy, and fundamental frequency were obtained using frame by frame analysis, dividing the input signal into a sequence of frames of 10 msec length and with 2.5 msec overlap.

The features derived from MDVP analysis of sustained vowel /a/, were classified using different classification methods. The linear discriminant classifier (LDC), nearest mean classifier (NMC) and Gaussian mixture modeling (GMM) classifier were applied to these features to distinguish between normal and pathological subjects. The correct classification results in both training and test phase are shown in table 1.

In addition, two 3-state, 3-mixture, left to right HMMs were formed based on 42 features obtained from the spectral and fundamental frequency information. The EM algorithm was used to train the HMM and a series of experiments were carried out with this HMM topology. In all of the experiments of this study, seven training iterations were enough for good convergence of model likelihoods. The database was divided into two equal groups, over which training and testing took place.

In the first experiment, HMMs were trained and tested using sustained vowel /a/ data to discriminate healthy and pathological speech. In the second experiment, the HMMs were trained and tested with the spoken utterances ("Rainbow passage"). The training and testing for HMMs was performed with and without F0 information. The results of these experiments are shown in table 2.

Method	Training %	Test %
LDC	95.64	95.93
NMC	67.15	65.26
GMM	97.97	97.67

Table1: Correct classification rates for MDVP parameters for sustained /a/.

	Vowel /a/		Rainbow Passage	
	Training	Testing	Training	Testing
MFCC	98.59	97.75	98.03	97.46
MFCC + Pitch	99.44	98.30	98.59	97.75

Table 2: Correct classification rates for training and testing of HMMs using vowel /a/ and Rainbow passage database.

IV. DISCUSSION AND CONCLUSION

In this paper two methods for pathological speech assessment were discussed. The first method was based on classification of MDVP parameters, which were derived from time domain analysis of speech signal. As the results of table 1 show, using the GMM classifier, the correct classification rate was 97.97%. The best correct classification rate for sustained vowel /a/ was 99.40% using spectral and pitch features with an HMM. In addition, the best correct classification rate for the Rainbow passage was 98.59%. As illustrated by the results of this study, the spectral and pitch features, which are low cost and fully automatic, showed better classification rate with respect MDVP features so it is possible to make a low cost, accurate, and automatic tool for pathological speech assessment using spectral and pitch information. In addition, the results showed that using just a sustained vowel /a/ provides fairly reliable detection; the use of lexical utterances such as the rainbow passage are still reliable albeit with some performance degradation. More research is needed to develop methods for automatic recognition of specific speech pathological types.

REFERENCES

[1] M. D. O. Rosa, J. C. Pereira, M. Grellet, "Adaptive Estimation of Residue Signal for Voice Pathology Diagnosis", IEEE Trans. Biomedical Eng. Vol. 47, No. 1, Jan. 2000.

[2] L. G. Ceballos, and H. L. Hansen, "Direct Speech Feature Estimation Using an Iterative EM Algoritm for Vocal Fold Pathology Detection", IEEE Trans. Biomedical Eng. Vol. 43, No. 4, April. 1996.

[3] D. Talkin, W. B. Klejin, and K. K. Paliwal, "A Robust Algorithm for Pitch Tracking", Speech coding and synthesis, Elsevier, New York, 1995.

[4] S. Young, D. Kershaw, J. Odell, D. Ollason V. Valtchev, P. Woodland, "The HTK book", Microsoft Corporation, July 2000.

[5] "Disorder Dtabase Model 4337" Massachusetts Eye and Ear Infirmary Voice and Speech Lab, Boston, MA, Jan. 2002.