IJCSE International Journal of Computer Science and Engineering Open Access
Research Paper Volume-2, Issue-2 E-ISSN: 2347-2693

Analysis of Speech Algorithms in Disease affected Voice Patterns

Karunanithi. D[1*], Dr. Paul Rodrigues[2]

[1]Information Technology, Hindustan University,Chennai, India, karunanithid@gmail.com
[2] Dean Research, Velammal College of Engineering,Chennai, India, drpaulprof@gmail.com

www.ijcaonline.org

Received: 15 Jan 2014 Revised: 15 Feb 2014 Accepted: 26Feb 2014 Published: 28 Feb 2014

***Abstract* —** The speech is the most important and effective method in the Human Interaction. Speech signal is the basis of Human Computer Interaction and Human Communications Technology. Many Human Diseases like Parkinson Disease, Cerebellar Demyelization, Stroke and many Neurological Diseases are analyzed using the their speech patterns for diagnosing the disease. To analyze these pathological voice patterns algorithms are used. In this paper, various algorithms used to analyze the voice patterns are analyzed and also comparative study of Parkinson and Larynx Diseases are analyzed using the voice patterns.

***Index Term*—** Speech Patterns; Speech Algorithms; Speech Analysis; Speech Diagnosis

I. INTRODUCTION

There are numerous medical conditions that adversely affect the human voice. Many of these conditions have their origins primarily in the vocal system and the available tools for detection of speech pathologies are either 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 facilities in pathological speech assessment. Automatic voice analysis for pathological speech have several advantages: 1) It has quantitative and non-invasive nature. 2) Allows identification and monitoring of the onset of vocal system diseases. 3) Reduces analysis cost and time.

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 sub-band processing methods. In time domain, authors have employed two major methods: 1) Methods based on temporal measurements of signal and their statistics, such as of the source excitation, to distinguish between normal and average pitch variation, jitter, shimmer, etc 2) analysis of residual of inverse filtering [1] of the speech signal, which corresponds to an estimate 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 to 90% [2], typically with a small limited number of subjects. In addition, there are robustness, consistency and complexity difficulties for measuring those features including the degree of human intervention needed in the measurements.

Corresponding Author: *Karunanithi. D*

It has been shown that spectral measurements can identify diseases like some kinds of Cysts, Polyps and other diseases, which indicate vocal fold malfunctioning. It is has also reported that one of the characteristics of speech signals under a non-healthy [3] condition is an irregular periodicity. Since this is essentially a function of excitation of the speech production system, it is necessary to incorporate measures that focus on excitation characteristics.

In the production of short vowels, the poor control of respiratory system is not significant [4]; hence, vowels phonated in a sustained fashion with comfortable levels of pitch and loudness are interesting and useful from a clinical point of view. The results of study, which conducted by Vieira et al [5], showed that there is consistency between electro glottal graph (EGG) parameters and acoustic signal features of sustained vowel /a/. According to their report, this is because of the larger and sharper peaks of time domain acoustic signal of /a/ with respect to the other vowels.

In this study, to successfully achieve the assessment of pathological speech, spectral envelope and pitch information have been considered. The first aim of this work is to classify speech signals in terms of being normal or pathological then, based on the results of this step, recognizing different abnormalities. We focus on two methods for pathological speech assessment. The first method is based on classification using an assortment of utterance level parameters, reflecting those used in the state of art in practice (as provided by the multi dimensional voice analysis – MVDP -- program), which have been derived from time domain analysis of speech signal. The second method is based on short-term analysis of spectral envelope and temporal dynamic measures. The results of these two methods and influence of different classifiers in the classification rate of sustained vowel /a/ are also investigated.

II. SPEECH ANALYSIS ALGORITHMS

A. Dynamic Time Warping Algorithm

Dynamic Time warping (DTW) is an algorithm for measuring similarity between two temporal sequences which may vary in time or speed. For instance, similarities in walking patterns could be detected using DTW, even if one person was walking faster than the other, or if there were accelerations and decelerations during the course of an observation. DTW has been applied to temporal sequences of video, audio, and graphics data — indeed, any data which can be turned into a linear sequence can be analyzed with DTW. A well known application has been automatic speech recognition, to cope with different speaking speeds. Other applications include speaker recognition and online signature recognition also it is seen that it can be used in partial shape matching application. DTW algorithm is used in analyzing speech patterns of Parkinson's disease.

B. Hidden Morkov Model Algorithm

The Hidden Markov Model have a wide range of application in temporal pattern recognition such as speech, handwriting, gesture recognition, part-of-speech tagging, musical score following, partial discharges and bioinformatics.

A Hidden Markov Model can be considered a generalization of a mixture model where the hidden variables (or latent variables), which control the mixture component to be selected for each observation, are related through a Markov process rather than independent of each other.

Hidden Markov Models (HMMs) is used to search for interesting sequences in a database of DNA sequences. The models are used to create a cost map for each sequence in the database. These cost maps can be searched rapidly for subsequences that have significantly lower costs than a null model.

A. A well-tuned HMM generally provides better compression than a simple Markov model, allowing more sequences to be significantly found.
B. The models are fairly readable (at least when drawn rather than just listed). A high-quality model for REPs(Repetitive Extragenic Palindromic sequences) (compressing previously unseen REPs to about 1.25 bits/base) may have around 200 states and 300 edges, rather than the counts of the order-8 simple Markov model. The low ratio of edges to states means that large parts of the model are simple straight-line sequences, which are easy to draw and to understand.

The HMMs can be used for generating alignments, with each state of the machine corresponding to one column in the alignment. The best path found by the Viterbi algorithm identifies a state for each position, and that in turn can specify the column. HMMs are a bit more powerful than alignments, since the same state can be used repeatedly in a path, but each column can only be used once in an alignment. This results in ambiguous alignments if a column alignment model is used, but can be quite convenient for describing phenomena like random numbers of repeats of a short subsequence.

C. Mel-Frequency Cepstral Coefficients

Mel-frequency cepstral coefficients (MFCC) [1] have been dominantly used in speaker recognition as well as in speech recognition. This is counter intuitive to many researchers since speech recognition and speaker recognition seek different types of information from speech, namely, phonetic information for speech recognition and speaker information for speaker recognition. MFCC was first proposed for speech recognition and its mel-warped frequency scale is to mimic how human ears process sound. Its spectral resolution becomes lower as the frequency increases. Therefore, the information in the higher frequency region is down-sampled by the mel scale. However, based on theory in speech production [2][3], speaker characteristics associated with the structure of the vocal tract, particularly the vocal tract length, are reflected more in the high frequency region of speech.

III. COMPARISON OF SPEECH AND NON-SPEECH SOUNDS

Research into speech recognition began by reviewing the literature and finding techniques that had previously been used for speech/speaker recognition. It was found that six techniques are commonly used for speech/speaker recognition or have been used for this domain in the past. These were:

- Dynamic Time Warping (DTW)
- Hidden Markov Models (HMM)
- Vector Quantization (VQ)
- Mel - Ceptral Co-efficient
- Artificial Neural Networks (ANN)

That compared the different feature extraction and classification methods used by each of these six techniques. Looking at these comparison tables, we can begin to examine whether any of these speech recognition or speaker identification techniques can be used for no speech sound recognition [5].

From looking at the comparison tables, it appears that some of these techniques, by their very nature, cannot be used for non-speech sound recognition. Any of the techniques that use sub word features will not be able to be used for non-speech sound identification. This is because environmental sounds lack the phonetic structure that speech does.

There is no set “alphabet” that certain slices of non-speech sound can be split into, and therefore sub word features (and the related techniques) cannot be used. Due to the lack of an environmental sound alphabet, all of the Hidden Markov Model (HMM) based techniques that are shown in the table cannot be used. Since HMM techniques are the main techniques now used in speech and speaker recognition, this leaves only a few other techniques[6][8]. After discounting HMM, the remaining five techniques were tested for their ability to classify no speech sounds. This was done in two ways.

First, benchmarking is performed, using these techniques, on non-speech sounds and data on the parameters, the resulting time taken and the final correct classification rate is recorded. Then, these results are compared with statistics and benchmark results reported in the literature for the performance of these techniques on speech.
Based on this information, in this paper we make a comparison between Vector Quantization and Artificial Neural Networks as possible techniques for non-speech sound recognition. As an initial test, eight sounds were used, each with six different samples.
Data set size was kept as small as possible due to the time it takes to train larger data sets. The sounds used for this test are detailed and are some typical sounds that would be classified in a sound surveillance system.

IV. ANALYSIS OF PARKINSON’S DISEASE

Parkinson’s disease is a neurodegenerative disorder of central nervous system that causes partial or full loss in motor reflexes, speech, behavior, mental processing, and other vital functions [1]. In 1817, PD was described as “shaking palsy” by Doctor James Parkinson [2]. It is generally observed in elderly people and causes disorders in speech and motor abilities (writing, balance, etc.) of 90% of the patients [3].
IN this study, using speech data from subjects is expected to help the development of a noninvasive diagnostic. The studies based on the PD focus on symptoms like slowness in movement, poor balance, trembling, or stiffness of some body parts [8]–[11] but especially voice problems. The main reason behind the popularity of PD diagnosis from speech impairments is that telediagnosis and telemonitoring systems based on speech signals are low in cost and easy to self-use [6], [12]. Such systems lower the inconvenience and cost of physical visits of PD patients to the medical clinic, enable the early diagnosis of the disease, and also lessen the workload of medical personnel [6], [12]–[14]. People with Parkinsonism (PWP) suffer from speech impairments like dysphonia (defective use of the voice), hypophonia (reduced volume), monotone (reduced pitch range), and dysarthia (difficulty with articulation of sounds or syllables). Even though there are many studies aiming at diagnosing and monitoring PD using these impairments, the origin of these studies leans to diagnose basic voice disorders.
Voice disorders can be measured by acoustic tools simply using a periodic vibrations in the voice.In their study, sustained vowel “a” phonations were recorded from 31 subjects of which 23 were diagnosed with PD. Then, dysphonia measures were extracted from the phonations to identify the grade of the disease by telemonitoring.
The dataset collected in this study contains multiple voice samples per subject such as sustained vowels, numbers, words, and short sentences. In this paper, we also compare the success of alternative cross-validation techniques that can be used with such datasets in a classification algorithm built for PD diagnosis.
As classification algorithms, they use k-nearest neighbor (k-NN) and support vector machines (SVM) and evaluate the success of the models in discriminating the healthy subjects from subjects with PD according to their accuracy, specificity, sensitivity, and Matthews correlation coefficient (MCC) scores.

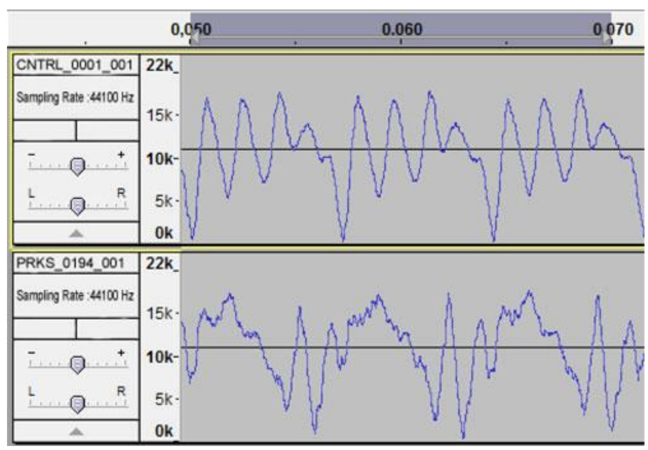

Fig. 1. Waveform of a voice sample belonging to a PWP (bottom) and a healthy individual (top). Amplitude of the signal (y-axis) is plotted against time duration (x-axis)

V. ANALYSIS OF LARYNX DISEASE

An unambiguous recognition and precise evaluation of the proposed set of parameters considered essential for representing pathological speech samples was difficult as phonetic data taken from the subjects differed from one another also with respect to aspects not undergoing an analysis (e.g., different speaking rates). Similarly reference speech material of healthy people with correct (standard) articulation varied. Our attention was focused on preliminary transformation of convert speech waveforms into a set of parameters whose values constituted a basis for a diagnostic description of the patient's disease. To construct models of speech deformation, it is necessary for a developed set of parameters describing the recorded phonetic samples to be arranged in an appropriate structure, a vector of features. In the model, various parameters of the acoustic signal and parametric space constituting sound patterns are assumed to be features or symptoms for detection of a given glottal disease.

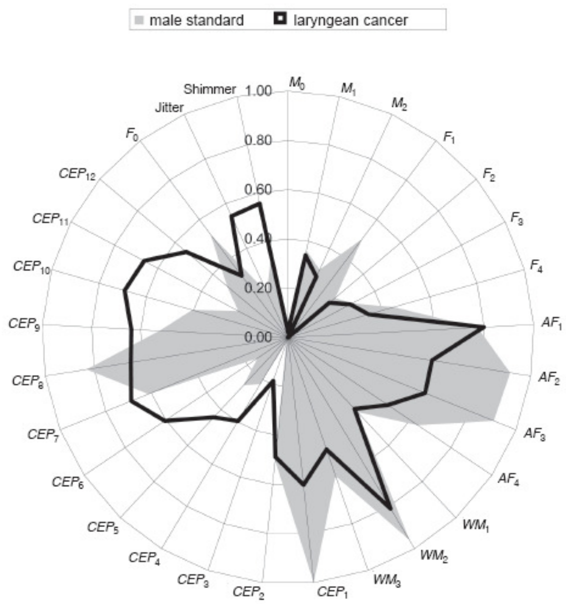

Fig .2 Visualization of the feature vector of the deformed speech recorded from a male patient diagnosed with laryngeal cancer

Reference values of feature vectors for normal speech are marked as shadow areas in Figures .Individual feature vectors of pathological speech signals are shown with solid lines. Such a display is convenient for diagnostic purposes: it is possible to make an instant comparative visual analysis of features specific for a disease.

VI. ANALYSIS OF ALGORITHMS WITH THE HUMAN DISEASE

Using the above Speech Analysis Algorithms and methods the Parkinson's Disease and Larynx Disease are analyzed.It is shown clear that the rest of the disease also can be analyzed using the speech patterns. Hence with the same procedure, research is to be made for the general diseases in Human. Hence the voice samples of normal human being are collected and stored in a Database. Thus finding the average frequency of the normal voice is found and stored.

To do this research 50 control points are choosed for voice analysis. 50 control point data consist of 25 Human male voices and 25 female voices are collected. The age group of 20 to 40 years is collected. The raw voice waveforms are analyzed This is done using PRATT and MATLAB to do the analysis[1].

In similar way the pathological voice of patients are stored and analyzed using the appropriate algorithm. Which will show the difference in the voice frequency of normal and pathological voice of disease affected Humans which will be more useful for the diagnosis of diseases in human.

VII. CONCLUSION

The voice patterns of normal human and pathological voice patterns are analyzed using the algorithm using MATLAB. It is analyzed the difference in voice pattern of disease affected human under gone the medication and non disease affected human. Thus we can find the difference between voice pattern using our self developed software. Dynamic Time Warping Algorithm is used for finding the similarity between the voices and identify the median frequencies of pathological voices. Hence the research is going on to identify the frequency range of each disease of human through their pathological voice for diagnosing the disease.

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 Database Model 4337" Massachusetts Eye and Ear Infirmary Voice and Speech Lab, Boston, MA, Jan. **2002**.

[6]. M. A. Little, P. E. McSharry, E. J. Hunter, J. Spielman, and L. O. Ramig,"Suitability of dysphonia measurements for telemonitoring of Parkinson's disease," IEEE Trans. Biomed. Eng., vol. 56, no. 4, pp. **1010–1022**, Apr.**2009**.

[7]. National Collaborating Centre for Chronic Conditions, Parkinson's Disease, London, U.K.: Royal College of Physicians, **2006**.

[8]. L. Cunningham, S. Mason, C. Nugent, G. Moore, D. Finlay, and D. Craig, "Home-based monitoring and assessment of Parkinson's disease," IEEE Trans. Inf. Technol. Biomed., vol. **15**, no. **1**, pp. **47–53**, Jan. 2011.

[9]. G. Rigas, A. Tzallas, M. Tsipouras, P. Bougia, E. Tripoliti, D. F. D. Baga, S. Tsouli, and S. Konitsiotis, "Assessment of Tremor activity in the Parkinson's disease using a set of wearable sensors," IEEE Trans.

Inf. Technol.Biomed., vol. **16**, no. **3**, pp. **478–487**, May **2012**.

[10]. S. Marino, R. Ciurleo, G. Lorenzo, M. Barresi, S. De Salvo, S. Giacoppo, A. Bramanti, P. Lanzafame, and P. Bramanti, "Magnetic resonance imaging markers for early diagnosis of Parkinson's disease," Neural Regeneration Res., vol. **7**, no. **8**, pp. **611–619**, **2012**.

[11]. Z. Dastgheib, B. Lithgow, and Z. Moussavi, "Diagnosis of Parkinson's disease using electrovestibulography," Med. Biol. Eng. Comput., vol. **50**, no. **3**, pp. **483–491**, **2012**.

[12]. M. A. Little, P. E. McSharry, S. J. Roberts, D. A. E. Costello, and I. M. Moroz, "Exploiting nonlinear recurrence and fractal scaling properties for voice disorder detection," Biomed. Eng. Online, vol. **6**, no. **23**, **2007**, doi: **10.1186/1475-925X-6-23**.

[13]. A. Tsanas, M. A. Little, P. E. McSharry, J. Spielman, and L. O. Ramig,"Novel speech signal processing algorithms for high-accuracy classification of Parkinson's disease," IEEE Trans. Biomed. Eng., vol. **59**, no. **5**, pp. **1264–1271**, May **2012**.

[14]. A. Tsanas, M. A. Little, P. E. McSharry, and L. O. Ramige,"Nonlinear speech analysis algorithms mapped to a standard metric achieve clinically useful quantification of average Parkinson's disease symptom severity," J. Royal Society Interface, vol. **8**, pp. **842–855**, **2011**.

[15]. A Vibroacoustic Model of Selected Human Larynx Diseases, International Journal of Occupational Safety and Ergonomics (JOSE) , Vol. **13**, No. **4, 367–379.**

