Intelligent Pathological Voice Detection

Akbar Ali[1], Prof. Sanjay Ganar[2]
[1]*Research Scholar, Anjuman College of Engineering & Technology*
[2]*Assistant Professor, Anjuman College of Engineering & Technology*

Abstract- **Pathology is the study and diagnosis of disease. Due to the nature of job, unhealthy habits and voice abuse, the people are subjected to the risk of voice problems. The diagnosis of vocal and voice disorders shouldbe in the early stage otherwise it causes changes in the normal signal. It is well known that most of vocal fold pathologies cause changes in the acoustic voice signal. Therefore, the voice signal can be a useful tool to diagnose them. Acoustic voice analysis can be used to characterize the pathological voices. This paper presents the detection of vocal fold pathology with the aid of the speech signal recorded from the patients. We are going to recognize the disordered voice for vocal fold disease by focusing on the classification of pathological voice from healthy voice based on acoustic features. The method includes two steps. The first step is the extraction of feature vectors based on MFCC, LFCC, ZCR, etc . The second is the classification of feature vectors using ANN. The extracted acoustic parameters from the voice signals are used as an input for the MFCC. The main advantage of this method is less computation time and possibility of real-time system development. This report introduces the design and implementation of the proposed system for recognizing pathological and normal voice.**

Index Terms- ANN, Pathological voice, Disorder, LFCC, MFCC.

I. INTRODUCTION

Speech is one of the ancient ways to express. Pathology is the study and diagnosis of disease. In this paper we are going to recognize the disordered voice for vocal fold disease. Vocal fold disease can affect the quality of the sound which is produced from vocal cord. The presence of pathologies in vocal folds causes significant changes in the normal vibratory patterns, which will results in the quality of voice production. The problems in the production of voice are due to the 1) functional disorder (due to the abuse or wrong use of the anatomical and physiologically intact voice system) or 2) Laryngeal pathologies (nodules of vocal folds. polyps, ulcers, carcinomas and paralysis of the laryngeal nerve. Some of the more common vocal cord disorders include laryngitis, vocal nodules, vocal polyps, and vocal cord paralysis. Diagnosis of pathological voice is one of the most important issues in biomedical applications of speech technology. In the past 20 years, a significant attention has been paid to the science of voice pathology diagnostic and monitoring. Normally physicians often use invasive technique like Endoscopy to diagnose the symptoms of vocal fold disorder[2]. Furthermore, the irregular vocal fold oscillations can be observed by means of a digital high-speed camera using image processing techniques in order to extract the vocal fold edges, estimate the minimum glottal area defined by the vocal fold positions, and compute the distance between the glottal midline and the vocal fold edges extracted at medial position in real-time. Voice pathologies may be assessed by either perceptual judgments or an objective assessment. The perceptual judgment resorts to qualifying and quantifying the vocal pathology by listening to patient's speech. Although this is the most commonly used method by clinicians, it suffers from several drawbacks. First of all, the perceptual judgment has to be performed by an expert jury in order to increase its reliability. Second, due to the lack of universal assessment scales and the dependence on expert's professional back-ground and experience or the knowledge of patients history, the perceptual judgment may involve large intra and inter-variability. Third, the perceptual analysis is very costly in time and human resources and cannot be planned regularly. Nowadays an increasing use of objective measurement-based analysis as a non-invasive technique for supporting diagnosis in laryngeal pathology has been observed [10]-[12]. Objective measurement-based analysis qualifies and quantifies the voice pathology by analyzing acoustical, aerodynamic, and physiological measurements. These measurements may be directly

extracted from patient's speech utterance using a simple computer-based system or may require special instruments. The purpose of this work is to help patients with pathological problems for monitoring their progress over the course of voice therapy. Currently, patients are required to routinely visit a specialist to follow up their progress. Moreover, the traditional ways to diagnose voice pathology are subjective, and depending on the experience of the specialist, different evaluations can be resulted. Developing an automated technique saves time for both the patients and the specialist can improve the accuracy of the assessments. Through acoustic analysis, finding out which factors that affect the human voice production mechanism can lead to the noninvasive diagnosis of disease. Developing an automatic pathological voice classification is training a classification system which enables to automatically categorize any input voice as either normal or pathological. Once the signal features are extracted, if the extracted features are well defined, even simple classification methods will be good enough for classification of the data.

The objective of this work is the search for a technique that will allow the quantification of a speaker's voice quality by means of an audio samplebased on the short term amplitude spectrum from speech (phonemes).

II. LITERATURE SURVEY

Various analogous research work are reviewed and analyzed to understand the nature and circumstances of the work. The purpose has been well studied and the need for proposed work is identified based on the literature survey on pathology voice detection is described below. Gaganpreet Kaur et al.,

[1] Reviewed the research done in the area of speaker recognition. The different methods used for feature extraction and feature classification had been discussed. Some techniques preferred over others such as MFCC for feature extraction had better performance rather than LPC or LPCC, because MFCC were most consistent with human hearing due to Mel scale representation. Thus, it was concluded that feature extraction of GMM performs better as they require fewer amounts of data to train the classifier. It also decreases the memory usage of the system. Aman Ankit et al.,

[2] Proposed ASR techniques and had put forth some of the essential information. The Speech Recognition System and various approaches used in ASR developed for various languages. Hidden Markov Model and Hidden Markov Model Toolkit (HTK) had been used in this paper. It describes the methods used and comparative study of the performance system developed. In this paper Hidden Markov Model (HMM) was used as a classifier and Mel Frequency Cepstral Coefficients (MFCC) as speech features were the most common technique. ASR implemented by using Hidden Markov Tool kit (HTK) are more efficient than the other systems implemented by using other tools. Shweta Vijay Dhabarde et al.,

[3] Presents LDB algorithm for audio classification. This helps to achieve high classification accuracy. LDB also uses simple dissimilarity measures for selecting the nodes and features. A database of 213 audio signals were used.LDB performs well for all signals while MFCC works well for music. Combination of MFCC and LDB also gives promising results. Zvi Kons et al.,

[4] Examines the performance of state-of-threat methods and investigates their weaknesses. Methods examined include features in time, frequency, perturbations, noise and spectral structure. Those features were evaluated by different machine learning techniques. The database contains samples from 719 subjects (320 male and 339 female) that were recorded at the Department of Otolaryngology, Kaunas University of Medicine, and Kaunas, Lithuania Since the glottal source extraction algorithm depends on the existence of the GCIs. Initial results show that by using - scalar quality measures against fixed thresholds and achieved 75% and 70% correct classification rates for healthy and severity-1 cases, respectively, using samples from over 100 healthy and over 50 severity-1 human subjects. Vahid et al.,

[5] Suggested an initial study of feature extraction and feature reduction in the task of vocal fold pathology diagnosis. A new type of feature vector, based on wavelet packet decomposition and MelFrequency-Cepstral-Coefficients

(MFCCs), was proposed. Also, Principal Component Analysis (PCA) was used for feature reduction. An Artificial Neural Network was used as a classifier for evaluating the performance of proposed method. The database was created by specialists from the Belarusian Republican Center of Speech, Voice and Hearing Pathologies. The selected 75 pathological speeches and 55 healthy speeches randomly which are related to sustained vowel "a". All the records were wave files in PCM format. The algorithm gives the best result of accuracy. V.Srinivasan et al.,

[6] explored a method of finding the ability of acoustic parameters in discrimination of normal voices from pathological voices that were analyzed and classified. The classification of pathological voice from normal voice was implemented using support vector machine (SVM) and the classifiers were trained and tested. The dataset was recorded by speech utterances of a set of Tamil phrases containing speech samples of 10 distinct subjects (5 normal, 5 pathological children). The speech signals were analyzed and were extracted. A Genetic Algorithm (GA) based feature selection has improved the classification accuracy of this work. Support vector machine shows better performance in terms of classification accuracy.

III. SCOPE OF THE WORK

My work is to classify normal and pathological voice using acoustic feature MFCC,LFCC, ZCR etc and to classify the signal using ANN. In this project we are going to classify different disease. This is done by extracting the features from the signal. Feature extraction is the first step in any speaker recognition system. MFCC is one of the most popular feature vectors. The cepstral representation of the signal allows us to characterize the vocal tract as a source-filter model and the Mel frequency characterizes the human auditory system which perceives the sound in a nonlinear frequency binning. ANN was used as classifier for speaker identification application.

IV. MATERIALS AND METHODS

The patterns for training the MFFCs were obtained from recordings of people's voices with normal voice and patients with pathologies of-the vocal system. Each signal is a recording of the sentence. 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 MFCC features were obtained by a standard short-term speech analysis, along with frame-level pitch, to form the feature vectors. Then, set of ANN classifiers were applied for the assessment of feature vectors. The architecture of the proposed system is given in figure 1.

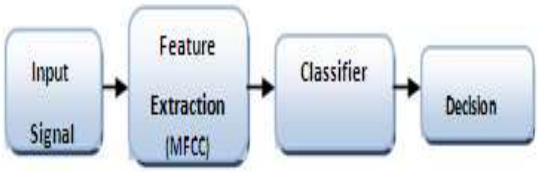

Fig 1: System Architecture

REFERENCES

[1] Gaganpreet Kaur Dr. Dheerendra Singh, "A Survey on Speech Recognition Algorithms", International Journal of Emerging Research in Management &Technology ISSN: 2278-9359 (Volume-4, Issue-5), May 2015.

[2] Aman Ankit, Sonu Kumar Mishra, Rinaz Shaikh, "A Survey Paper on Acoustic Speech Recognition Techniques", International Journal of Recent Advances in Engineering & Technology (IJRAET), ISSN (Online): 2347 - 2812, Volume-4, Issue -7, 2016.

[3] Shweta Vijay Dhabarde, P.S.Deshpande, "Feature Extraction and Classification of Audio Signal using Local Discriminant Bases", International Journal of Industrial Electronics and Electrical Engineering, ISSN: 2347-6982 Volume-3, Issue-5, May-2015.

[4] Zvi Kons, Aharon Satt, Ron Hoory,, "On Feature Extraction for Voice Pathology Detection from Speech Signals", Advances in Nonlinear Speech Processing", LNCS 5933, 35-42, edited by R. Goebel, J. Siekmann, and W.Wahlster, DOI: 10.1007/978-3-642- 11509-7_5, ISBN: 3-642-11508-X, Springer, New York.

[5] Vahid Majidnezhad, Igor Kheidorov, "An ANN-based Method for Detecting Vocal Fold Pathology", International Journal of Computer Applications (0975 – 8887) Volume 62– No.7, January 2013.

[6] V.Srinivasan, V.Ramalingam, V.Sellam, "Classification of Normal and Pathological Voice using GA and SVM", International Journal of Computer Applications (0975 – 8887) Volume 60– No.3, December 2012 Ce Peng., Wenxi Chen., Xin Zhu.,Daming Wei., and Baikun Wan., (2007). Pathological Voice Classification Based on a Single Vowel Acoustic Features. IEEE computer society. Seventh International Conference on Computer and Information Technology.

[7] Jagadish Nayak ., Subhanna Bhat , P.,(2003). Identification of Voice Disorders using Speech Samples ," IEEE Trans.Speech processing.,

[8] Schwarz, R., Hoppe, U.,Schuster, M., Wurzbacher,U., and Eysholdt., Lohscheller, J. (2006). Classification of unilateral vocal fold paralysis by endoscopic digital high-speed recordings and inversion of a biomechanical model. IEEE Trans. Biomed.Eng., vol. 53, no. 6, pp. 10991108.

[9] Constantine Kotropoulos ., and Gonzalo R. Arce. (2009). Linear Classifier with Reject Option for the Detection of Vocal Fold Paralysis and Vocal Fold Edema. EURASIP Journal on Advances in Signal Processing,

[10] Parsa, V., and Jamieson, D.G. (2003). Interactions between speech coders and disordered speech. Speech Communication.,vol.40, no. 7, pp. 365385,

[11] Hawley,M. S., Green ,P.,Enderby, P., Cunningham, S., and Moore, R.K., (2005). Speech technology for e-inclusion of people with physical disabilities and disordered speech. INTERSPEECH 05.,pp. 445448, Lisbon,Portugal

[12] Oscar Saz., Javier Simon.,Ricardo Rodriguez, W., Eduardo Lleida., and Carlos Vaquero., (2009). Analysis of Acoustic Features in Speakers with Cognitive Disorders and Speech Impairments EURASIP Journal on Advances in Signal Processing,

[13] Ruben Fernandez Pozo., Jose Luis BlancoMurillo.,Luis Hernandez Gomez., Eduardo Lopez Gonzalo.,JoseAlcazar Ramirez and Doroteo T. Toledano., (2009). Assessment of Severe Apnoea through Voice Analysis, Automatic Speech, and Speaker Recognition Techniques EURASIP Journal on Advances in Signal Processing,

[14] Alireza Afshordi Dibazar., Shikanth Narayanan.,A System for Automatic Detection of Pathological Speech (2002)

[15] Julian D. Arias-Londono., Juan I. Godino-Llorente.,Nicolas Saenz-Lechon.,Victor Osma-Ruiz., and German Castellanos-Dominguez., Automatic Detection of Pathological Voices Using Complexity Measures, Noise Parameters, and Mel-Cepstral Coefficients (2011) IEEE Trans. on biomedical engineering vol. 58, no. 2.,