A Review on Dysarthria speech disorder

Miss. Yogita S. Mahadik , Prof. S. U. Deoghare ,

1 Student, Dept. of ENTC, Pimpri Chinchwad College of Engg Pune, Maharashtra, India

2 Professor, Dept. of ENTC, Pimpri Chinchwad College of Engg Pune, Maharashtra, India

ABSTRACT

Communication is a bridge between people which enables them to share facts, ideas, feelings etc. with each other. Speech communication is easier and simpler compared to other types of communication. But speech disorders affect one's ability to communicate. Dysarthria is a neuro-motor disorder, where one losses his/her ability to articulate words normally due to tongue/muscle weakness or stroke etc. It results in distorted speech which is hard to understand. Dysarthria affect more on the articulation of consonants (stops in particular) than on the articulation of vowels. Intelligibility varies greatly depending on the extent of neurological damage.

Keyword : - *MFCC, LPC, ANN, SVM*

1. INTRODUCTION

The problem of speech disorder has become an important topic in modern society. We express thoughts and feelings through speech. The process by which thoughts are translated into speech involves the use of articulators for producing various speech sounds. Speech plays a major role in an individual participation in the society. Spoken communication is one of the most used and most effective ways of communication. Dysarthria is a verbal communication disorder. A person suffering from dysarthria has poor synchronization between lip and tongue. Hence, patient is not able to communicate his thoughts in an effective manner. Unfortunately, 170 people out of 100,000 suffer from dysarthria. A patient with severe dysarthria has low intelligible speech and vice-versa. Speech with severe dysarthria is completely unintelligible. For medium degree dysarthria, only half of the speech is understandable and speakers show breathlessness and some sounds are over aspirated. Patient with mild dysarthria has comparatively better speech quality except for few words. Dysarthria is a speech disorder that is caused due to lack of control of various organs that are used in the production of speech. The mechanism involved in speech production includes respiration, phonation, resonance, articulation and prosody. It affects the speed and strength of the speech signal that is produced. Dysarthria is caused mainly due to neurological disorders like stroke, brain injury, multiple sclerosis and brain tumor which results in weakness or inability to control the articulatory organs. In this paper we will see different methods to detect dysarthria speech disorder.

2. Speech Disorder

Speech disorders or speech impediments are a type of communication disorder where 'normal' speech is disrupted.

Types of Speech Disorder

Apraxia

Apraxia of speech may result from stroke or progressive illness, and involves inconsistent production of speech sounds and rearranging of sounds in a word ("potato" may become "topato" and next "totapo"). Production of words becomes more difficult with effort, but common phrases may sometimes be spoken spontaneously without effort.

Dysarthria

Dysarthria is a motor speech disorder resulting from neurological injury of the motor component of the motor-speech system and is characterized by poor articulation of phonemes. Neurological injury due to damage in the central or peripheral nervous system may result in weakness, paralysis, or a lack of coordination of the motor-speech system, producing dysarthria. These effects in turn hinder control over the tongue, throat, lips or lungs; for example, swallowing problems (dysphagia) are also often present in those with dysarthria. Dysarthria does not include speech disorders from structural abnormalities, such as cleft palate, and must not be confused with apraxia of speech, which refers to problems in the planning and programming aspect of the motor speech system.

Stuttering

Stuttering affects approximately 1 percent of the adult population. Stuttering is a speech disorder in which sounds, syllables, or words are repeated or prolonged, disrupting the normal flow of speech. These speech disruptions may be accompanied by struggling behaviors, such as rapid eye blinks or tremors of the lips. Stuttering can make it difficult to communicate with other people which often affects a person's quality of life.

Muteness

Muteness is complete inability to speak.

Cluttering

Cluttering is a speech and fluency disorder characterized primarily by a rapid rate of speech, which makes speech difficult to understand.

Causes of speech disorder:

1. Vocal cord damage

2. Brain damage

3. Muscle weakness

4. Respiratory weakness

5. Strokes

6. Vocal cord paralysis

Dysarthria Speech Disorder Types

1. Flaccid Dysarthria

This result from damage to the lower motor neurons (cranial nerves) involved in speech. Mono-pitch and mono loudness may both result from vocal fold paralysis.

2. Spastic Dysarthria

This is due to exaggerated stretch reflexes, resulting in increased muscle tone and in coordination. Vocal quality is harsh. Sometimes the voice of a patient with spastic dysarthria is described as strained or strangled. Pitch is low, with pitch breaks occurring in some cases. Bursts of loudness are sometimes noted.

3. Ataxic Dysarthria

This disorder is due to damage to the cerebellar control circuit. It can affect respiration, phonation, resonance and articulation, but its characteristics are most pronounced in articulation and prosody.

4. Hyperkinetic Dysarthria

Hyperkinetic dysarthria is usually thought to be due to lesions of the basal ganglia. Its predominant symptoms are associated with involuntary movement. Vocal quality may be described as harsh, strained, or strangled. Voice stoppages may occur in dysarthria associated with dystonia.

5. Hypokinetic Dysarthria

This is associated mainly with Parkinson's disease. Hoarseness is common in Parkinson's patients. Also, low volume frequently reduces intelligibility.

6. Mixed Dysarthria

Characteristics will vary depending on whether the upper or lower motor neurons remain most intact.

Types of Dysarthria Disorders

Col Name	Col Name	Col Name	Col Name
1	Flaccid	Lower motor neuron	Weakness
2	Spastic	Upper motor neuron	Spasticity
3	Ataxic	Cerebellar control circuit	Incoordination of muscles
4	Hypokinetic	Basal ganglia control circuit	Rigidity
5	Hyperkinetic	Basal ganglia control circuit	Involuntary movement
6	Mixed	More than one	More than one

What are the risk factors?

Stroke is a major risk factor of dysarthria. Some risk factors of stroke include:

1. High blood pressure

2. Diabetes mellitus

3. Cigarette smoking

4. Ageing

5. Heart diseases

3) Specific methods used in study

1) Method to Detect Dysarthria Speech Disorder using prosodic features

For dysarthric speech recognition there are different methods are used to diagnose the dysarthria.

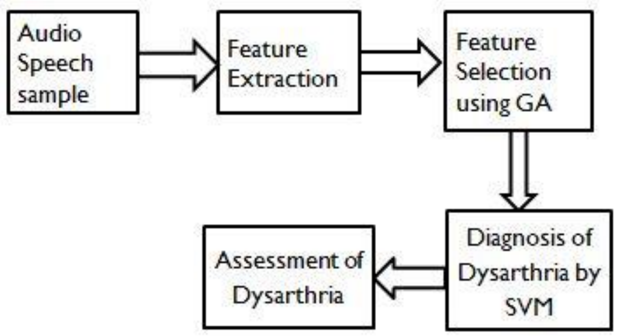

Fig -1: The basic block Diagram of detection of dysarthria.

1.1 Feature Extraction

Features play a major role in identifying the voice and making a decision that highly depends on how much we are successful in extracting useful information from the voice in a way that enables our system to differentiate between voices and identify them according to their feature. This is accomplished by extracting Mel frequency Cepstral Coefficients (MFCC) features. Figure shows the block diagram of extraction of Mel Frequency Cepstral Coefficients (MFCC).

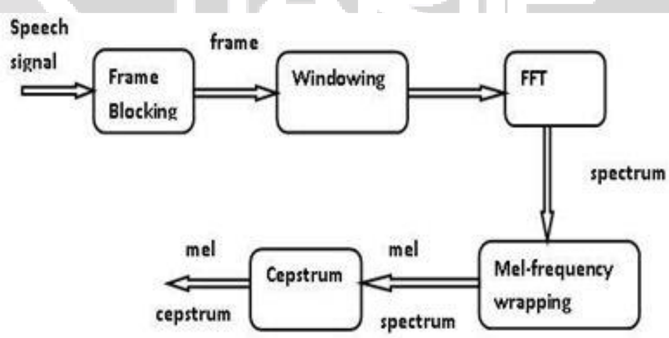

Figure. MFCC feature extraction

1.2 Feature Selection Using GA

The main aim of using the feature selection algorithm is to use fewer features to attain the same or better classification accuracy. Genetic algorithm is employed on the complex feature set. Features which are more discriminant for dysarthric speech are selected. Hence, the speech prosodic features have been selected. It is observed from the experiments that MFCCs, or skewness and formants are most prominent features for diagnosis of dysarthric speech. The feature selection deals with the task of selecting the best feature set which reduces the classifier training time and as well as increase the classification accuracy. From a given set of the features, the feature selection algorithm selects a subset of size m which increases the classification accuracy.

1.3 Diagnosis of Dysarthria speech by SVM

SVM is used for classification. SVM can be classified into binary SVM and multi SVM. In binary SVM, we can determine whether the person is recognized or not. Binary SVM compares the features of two speakers. But multi SVM compares the features of more than two speakers. It comes under supervised classifier. Basic of SVM is to create a hyper plane. This hyper plane differentiates the features. In binary SVM features are classified into two classes, each class for recognized and non -recognized speaker.

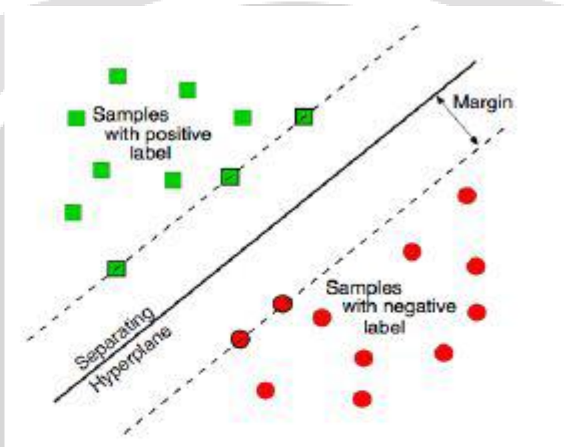

Figure. Principle of SVM

1.4 Assessment of dysarthric speech

Once the Dysarthric speech is detected, it is important to label the severity of dysarthria also so that a SLP can treat the patient according to his severity level. The levels of severity are decided by the intelligibility of the speech. A patient with high intelligible speech has low severity of dysarthria. Similarly, low intelligible speech is characterized as high severity dysarthric speech. The multi class support vector machines are used to label the speech intelligibility and severity.

2. Method to Detect Dysarthria Speech Disorder using Extended Feature Extraction and Neural Networks Classification

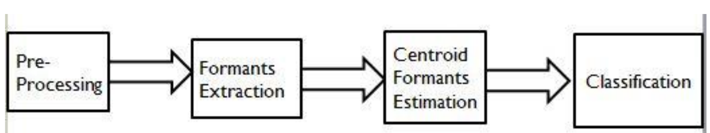

Figure: The basic block Diagram of detection of dysarthria.

AS shown in Figure, The first stage is pre-processing followed by feature extraction. After extraction of the speech features, the next stage is classification based on neural network techniques.

2.1 Formant in speech signal

A formant is a concentration of acoustic energy around a particular frequency in the speech wave. Formants are bands of resonance in the frequency spectrum of a speech signal. These bands of resonance are the significant representation of the signal. The formant extraction algorithm, in this proposed technique, is based on the Linear Prediction Coding (LPC) analysis. The LPC analysis gives a smoothed approximation of the power spectrum of the original signal.

2.2 Centroid Formants Estimation by LPC

The centroid formants are the weighted averages of the formants in each frame in the short time frequency spectrum. The spectral centroid is a measure of where the power in the frequency spectrum of an audio signal is centralized. The centroid formant is very sensitive the rapid changes in pitch and intonation. This means that the high pitch variability in dysarthria can effectively and efficiently be tracked using centroid formants. Any sudden change in pitch or intonation is reflected.

2.2.1 Linear Predictive Coding

LPC technique for determining the basic parameters of speech and provides precise estimation of speech parameters and computational model of speech. LPC methods provide extremely accurate estimates efficiently of the speech parameters. It is desirable to compress signal for efficient transmission and storage. For medium or low bit rate coder, LPC is most widely used. The LPC calculates a power spectrum of the signal. It is used for formant analysis. LPC is one of the most powerful speech analysis techniques and it has gained popularity as a formant estimation technique. While we pass the speech signal from speech analysis filter to remove the redundancy in signal, residual error is generated as an output. It can be quantized by smaller number of bits compare to original signal. So now, instead of transferring entire signal we can transfer this residual error and speech parameters to generate the original signal. A parametric model is computed based on least mean squared error theory, this technique being known as linear prediction (LP).

By this method, the speech signal is approximated as a linear combination of its p previous samples. In this technique, the obtained LPC coefficients describe the formants. The frequencies at which the resonant peaks occur are called the formant frequencies. Thus, with this method, the locations of the formants in a speech signal are estimated by computing the linear predictive coefficients over a sliding window and finding the peaks in the spectrum of the resulting LP filter. We have excluded 0th coefficient and used next ten LPC Coefficients.

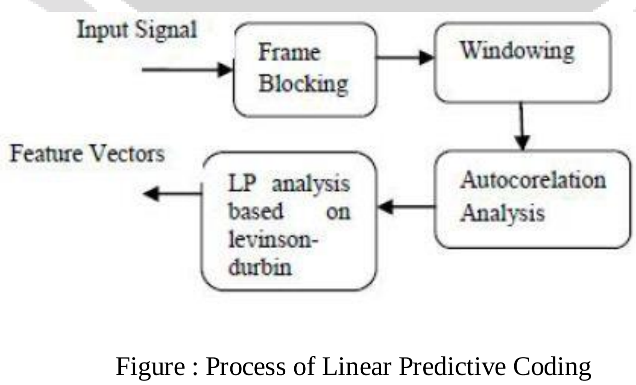

Figure : Process of Linear Predictive Coding

2.3 Classification by ANN

The Artificial Neural Networks is used for classifying the disordered speech. It provides fantastic simulation of information processing analogues to human nervous system. One of the commonly used machine learning methods is the neural network. This classification technique is robust and it combines pattern recognition with acoustic phonetic methods. In this artificial learning technique, knowledge of the acoustic phonetic characteristics of the speech is used to generate rules for classifiers. A multilayer neural network with one hidden layer was used for this classification. The excitations (inputs) are the centroid formants and the observations (outputs) indicates whether or not the corresponding audio sample is from the dysarthric speaker (0) or healthy speaker (1). Here single layer neural network with 10 neurons in a hidden layer was used.

3) Method to Detect Dysarthria Speech Disorder by using Automatic Speech Recognition

For an ASR system, a speech signal refers to the analogue electrical representation of the acoustic wave, which is a result of the constrictions in the vocal tract. Different vocal tract constrictions generate different sounds. The basic sound in a speech signal is called a phoneme. These phonemes are then combined, to form words and sentences. Each phoneme is dependent on its context. Each language has its own set of distinctive phonemes, which typically amounts to between 30 and 50 phonemes. An ASR system mainly consists of four components: pre-processing stage, feature extraction stage, classification stage and a language model.

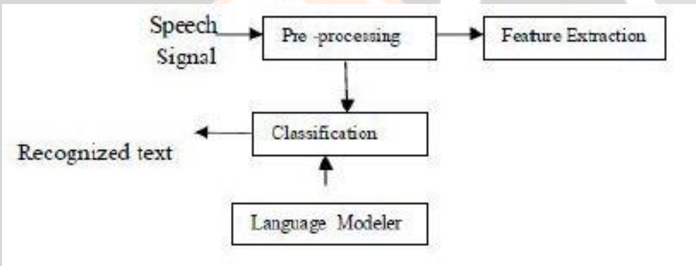

Figure : Automatic Speech Recognition System

3.1 Pre-processing

The pre-processing stage transforms the speech signal before any information is extracted by the feature extraction stage. The functions to be implemented by the pre-processing stage are also dependent on the approach that will be employed at the feature extraction stage. A number of common functions are the noise removal, endpoint detection, pre-emphasis, framing and normalisation.

3.2 Feature Extraction

After pre-processing, the feature extraction stage extracts a number of predefined features from the processed speech signal. These extracted features must be able to discriminate between classes while being robust to any external conditions, such as noise. Therefore, the performance of the ASR system is highly dependent on the feature extraction method chosen, since the classification stage will have to classify efficiently the input speech signal according to these extracted features. Over the past few years various feature extraction methods have been proposed, namely the MFCCs, and the linear predictive coding (LPC).

3.3 Language model

The next stage is the language model, which consists of various kinds of knowledge related to a language, such as the syntax and the semantics. A language model is required, when it is necessary to recognise not only the phonemes that make up the input speech signal, but also in moving to either trigram, words or even sentences. Thus, knowledge of a language is necessary in order to produce meaningful representations of the speech signal.

3.4 Classification

The final component is the classification stage, where the extracted features and the language model are used to recognise the speech signal. The classification stage can be used by discriminative approach. A model based on a discriminative approach finds the conditional distribution using a parametric model, where the parameters are determined from a training set consisting of pairs of the input vectors and their corresponding target output vectors. Two popular methods that are based on the discriminative approach are the ANNs and support vector machines (SVMs).

4. Future Scope

As future work, the classification of dysarthric voices from the normal voices using MFCC features and SVM. Mel-frequency cepstral coefficients are the widely used features to characterize voice signals. 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. The MFCC feature vectors are used by SVM or Artificial Neural Network to identify and classify the dysarthric voices.

5. CONCLUSIONS

Different speech disorder with Dysarthria is studied. Dysarthria speech disorders caused by weakness, paralysis, rigidity, spasticity, sensory loss, or incoordination of muscle groups responsible for speech. The different methods to detect and diagnosis of dysarthria are studied.

As human voice is nonlinear in nature, Linear Predictive Codes are not a good choice for speech estimation. MFCC are derived on the concept of logarithmically spaced filter bank, with the concept of human auditory system and hence had the better response compare to LPC parameters.

Early detection of dysarthria can help to diagnosis the patient.

6. ACKNOWLEDGEMENT

I wish to acknowledge the support of many people who have contributed to my seminar. I would like to thank my guide, Prof. S. U. Deoghare, Professor in Electronics & Telecommunication Department, for her encouragement and valuable guidance. She motivated me greatly which contributed tremendously to my work.
I would like to thank Dr. M. T. Kolte, PG coordinator for his insightful suggestions. I am also thankful to Dr. N. B. Chopade, HOD Electronics & Telecommunication Department and Dr. A. M. Fulambarkar, Principal, Pimpri Chinchwad College of Engineering for their support. I would also like to express my sense of gratitude to all the people who directly or indirectly supported me in completing this seminar work. At the last but not least, I am thankful to my parents, who encouraged and inspired me with their blessings.

7. REFERENCES

[1] Jun Ren , Mingzhe Liu , An Automatic Dysarthric Speech Recognition Approach using Deep Neural Networks", International Journal of Advanced Computer Science and Applications, vol. 8, No. 12.2017
[2] T B Ijitona , J J Soraghan , "Automatic Detection of Speech Disorder in Dysarthria using Extended Speech Feature Extraction and Neural Networks Classification," International Conference on Intelligent Signal Processing, 2017
[3] Garima Vyas, Malay Kishore Dutta , An Automatic Diagnosis and Assessment of Dysarthric Speech using Speech Disorder Specific Prosodic Features, International Conference on Telecommunications and Signal Processing , 2016.
[4] Taylor Spangler1 and N. V. Vinodchandran1 and Ashok Samal1 and Jordan R. Green2, Fractal Features for Automatic Detection of Dysarthria, 2017 IEEE.

[5] Seyed Reza Shahamiri, Siti Salwah Binti Salim , Multi-Views Multi-Learners Approach Towards Dysarthric Speech Recognition Using Multi-Nets Artificial Neural Networks," IEEE Transaction on neural system and rehabilitation Engineering, VOL. 22, NO.5, September 2014.
[6] Michal Novotn, Jan Rusz ,Automatic Evaluation of Articulatory Disorders in Parkinson's Disease, IEEE/ACM Transaction on audio, speech and language processing VOL. 22, NO, 2014
[7] Jun Wang , Myungjong Kim, Regularized Speaker Adaptation of KL-HMM for Dysarthric Speech Recognition, IEEE Transactions on Neural Systems and Rehabilitation Engineering, 2017.