Automatic Detection of Parkinson's Disease using Cepstral Features derived from Speech Signals

Thesis submitted in partial fulfillment
of the requirements for the degree of

Master of Science
in
Electronics and Communication Engineering
by Research

by

Monica Ponnam
201131227
monica.ponnam@research.iiit.ac.in

International Institute of Information Technology, Hyderabad
(Deemed to be University)
Hyderabad - 500 032, India

January 2024

International Institute of Information Technology
Hyderabad, India

CERTIFICATE

It is certified that the work contained in this thesis, titled ***Automatic Detection of Parkinson's Disease using Speech Processing*** by *Monica Ponnam* has been carried out under my supervision and is not submitted elsewhere for a degree.

Date	Advisor: Dr. Anil Kumar Vuppala

Acknowledgements

I want to sincerely thank my advisor, Dr.Anil Kumar Vuppala for his continuous support and encouragement. I want to express my sincere regards and gratitude for this opportunity to do research under his guidance and valuable advice during my entire journey starting from the basic course of Digital Signal Processing to the presentation of my paper at an IEEE conference.

This thesis would not be possible without Purva Ma'am who was with me at every step with advice, suggestions. She explained the basics of the various tools for the required analysis with great patience and enthusiasm. I want to extend my sincere appreciation for all her help and support. Also, I want to thank all the people of the Speech Processing Lab who supported me unseen by their insights during the same period.

I am very grateful to my parents and my husband, who are always encouraging and believed in me and without whom this jorney would not be envisioned. I cannot thank them enough for their love, care, and support.

I want to thank all my friends and many otherpeople whose names might not be here but whose support and encouragement would be cherished.

Abstract

A Parkinson's Detection System refers to a set of tools designed to assist in the early diagnosis of Parkinson's Disease, a progressive neuro-degenerative disease that affects movement control. In literature about Parkinson's Disease (PD) detection systems, various speech analysis tools were used which included features derived from the voice harmonics, jitter and shimmer, speech rate, etc. While these features describe the changes in the voice patterns of the patient, we found that statistical data of the cepstral coefficients are a more effective classification criteria. This resulted in our use of cepstral coefficients derived from two different methods in our system and the results are compared to a baseline model which uses MFCCs.

The first method is focused on the transformation of the speech signal through the time frequency treatment of a wavelet transform to make use of the multi resolution property of the wavelet transform. It is followed by a cepstral analysis in order to extract the cepstral coefficients. The resultant feature dimensions are classified using Support Vector Machine (SVM) with the help of different kernels. This model has been applied to PC-GITA database of diadokinetic word /pa-ta-ka/. The results showed us a performance of 65%.

The second method is the Zero-Time Windowing of the speech signal and it allows us a higher spectro-temporal resolution. The cepstral coefficients derived from this signal are used as as the basis for training the model for detection of Parkinson's Disease. The dataset and classifiers are same as the first method and resulted in 70% when the derivatives of the cepstral features are also used. Both the proposed systems compared favourably to the baseline used which is MFCC based detection system.

Contents

List of Figures

List of Tables

Abbreviations

ASR	Automatic Speech Recognition
AUC	Area under receivier operating Characteristic Curve
CNN	Convolutional Neural Network
CWT	Continuous Wavelet Transform
DaTSCAN	dopamine transporter chemical scan
DFT	Discete Fourier Tansform
EMG	electromyography
LPCCs	Linear Prediction Cepstral Coefficients
LPCs	Linear Prediction Coefficients
LSTM	Long-Short Term Memory
MFCCs	Mel Frequency Cepstral Coefficients
MRI	Magnetic Resonance Imaging
NNN	Narrow Neural Network
PD	Parkinson's Disease
ReLU	Rectified Linear Function
RNN	Recurrent Neural Network
SFCCs	single frequency filtering cepstral coefficients
STFT	Short-Time Fourier Transform
SVM	Support Vector Machine
UPDRS	Unified Parkinson's Disease Rating Scale
WT	Wavelet Transform
WTCCs	Wavelet Transformation Cepstral Coefficients
ZTW	Zero-Time Windowing
ZTWCCs	Zero-Time Windowing Cepstral Coefficients

Chapter 1

Introduction

1.1 Parkinson's Disease

PD is a progressive neurological condition and is the second most prevalent neurological disorder in the world. While the particular cause of PD was not established, it is found that in the patients suffering from PD, the nerve cells in the basal ganglia, an area of the brain that controls movement, become impaired or die. Normally, these neurons, produce dopamine.

Dopamine allows messages to be sent to the parts of the brain that co-ordinate movement. As the dopamine levels decrease, symptoms start appearing and are divided into two categories: motor and non-motor symptoms. Motor symptoms are related to the related to movement issues while the non-motor symptoms cannot be easily seen and include sleeplessness and depression [1].

The major symptoms that affect the patient include: impaired balance and movement (leading to falls); tremor in hands, arms, legs, jaw, or head; muscle stiffness, where muscle remains contracted for a long time; slowness of movement. Other symptoms are: depression and other emotional changes [2], [3], [4]; difficulty in swallowing, chewing, and speaking; urinary problems or constipation.

1.2 Diagnosis of PD

The present diagnostic methods are based on identifying the physical and chemical changes in the patient using Magnetic Resonance Imaging (MRI) brain scanning, dopamine transporter chemical scan (DaTSCAN), lumbar puncture (a procedure to test the the spinal fluid that surrounds the brain), electromyography (EMG) ,which tests the health of muscles and the motor neurons of the urethral or anal sphincter. To obtain comprehensive data, the patient is subjected to detailed medical history check (to recognise medication side-effects),a thorough physical examination to evaluate the effects of PD in the patient. All these tests are invasive, expensive, yet necessary, to diagnose Parkinson's Disease accurately due to the similarities of many neurological disorders.

The motor symptoms of PD include tremors, rigidity, bradykinesia (slowness of movement), difficulty in breathing and swallowing which affects the speech of the person. The results include reduced articulation and inability to induce expressions into speech, etc. These symptoms can be used to detect

PD [5], [6]. As Parkinson's is generally found in the elderly and is accompanied with a host of other difficulties, detection using speech processing can be very useful as it is economical, convenient, and helps in earlier detection so that the patient can lead an independent life with the help of medication.

1.3 Challenges Involved

As previously discussed, the use of speech processing in PD detection system, while convenient, has its own challenges. One of the major challenge is that not every PD patient show the same symptoms. Also, progression of the disease and the effect on the patient varies from individual to individual. Earliest signs that a person is affected by Parkinson's Disease occur gradually and cannot be easily recognised by the medically untrained. For example,people might experience mild tremors or have issues with their balance. Another change might be that the patient may speak too softly or in bursts, or that their handwriting is uneven and looks cramped or small, or even difficulty in following conversational cues. The people in close contact with the patient may be the first to notice early changes in someone with Parkinson's. They may see the change in a person's expression, or that the person does not have their normal gait. These are symptoms that cannot be evaluated and used for diagnosis by a medical professional who is a near-stranger, without additional tests.

So, while we can devise a detection system using speech processing, it can be useful for only those with changes in speech pattern. Secondly, due to PD, the quality of speech of the test subjects suffers. This makes recording and utilizing their speech patterns for analysis difficult. It may be that the different laryngeal diseases exhibit similar symptoms [7]. Another difficulty is the need for a broad and varied database, representing various accents, similarity with a different category (healthy or PD of sounds from another language, age, stage of PD etc. Due to this, devising a generic system for the detection of PD is difficult. So, for the sake of convenience, PD detection systems, for the present, are researched on using only one single language. This thesis, for example, focuses on a database of native Spanish speakers.

1.4 Objective and Scope of the Thesis

The main objective of this thesis is to develop a system, which helps in detection of PD, such that it is convenient to use it remotely and multiple times. This is a very important requirement because as the Parkinson's disease is a complex condition that can also manifest with various non-motor symptoms which can significantly impact the quality of life for individuals with Parkinson's disease in addition to the mobility issues. These include cognitive changes which involve difficulties with memory, attention, and other cognitive functions; mood disorders including depression (feelings of sadness, hopelessness, and loss of interest), anxiety (excessive worry or fear), apathy: Lack of interest, enthusiasm, or motivation. As PD is a progressive disease, early diagnosis can mark the difference between an independent life with the help of medication and a depressed dependent who has difficulties with even breathing. An evaluation/diagnosis using an automatic diagnostic system can be used as a precursor to initiate comprehensive tests, resulting in increased screening during the initial stages of PD, which in turn improves the prognosis of the patient [8].

1.5 Organisation of the Thesis

- Chapter 2 of this thesis, presents an overview of methods involved in detection of PD using speech processing techniques along with their focus, baselines considered in this work, classifiers, and the evaluation metrics.
- Chapter 3 of this thesis, presents the use of wavelet transformation of the speech signal to obtain the Mel Frequency Cepstral Coefficients (MFCCs). This in turn is used to train a classification system to discriminate between healthy controls and Parkinson's Disease that can be used for PD detection.
- Chapter 4 of this thesis, presents a classification system for detection of PD using Zero-Time Windowing (ZTW) based cepstral feature.
- Chapter 5 of this thesis, presents conclusions and future scope of the work.

Chapter 2

An overview of PD Detection System using Speech Processing

2.1 Introduction

Parkinson Detection using speech processing is very useful and economical due to its non-invasive nature and allowance for remote and repeated testing. This resulted in research into it using various approaches. The detection of PD from speech has been investigated using many approaches: phonatory, articulatory, prosodic, and linguistic. All these methods focus on different characteristics of the speech signal. The features extracted from the speech signal are then used to train a classification model with binary results: healthy (HC) and parkinson's (PD).

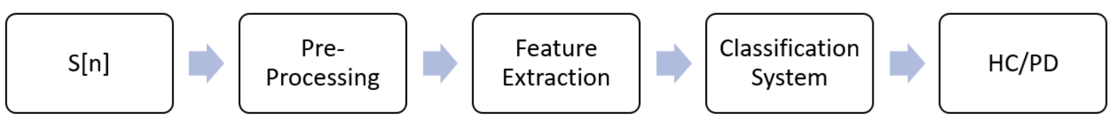

Figure 2.1 Sequential Steps of PD detection System.

?? gives us the block diagram of a PD detection system. The speech signal of interest is pre-emphasised before relevant features are extracted. The extracted features are then fed to a classifier that gives us the PD detection results.

We shall now discuss in detail the various methods of extracting the features.

- Articulation and phonation analysis which focuses on the glottal source and larynx changes. The studies based on the phonatory approach are few in number and also not easily scalable due to the difficulty in estimating the glottal source from the speech signals of a patient suffering from Parkinson Disease. Dysarthria (difficulty in articulating) Detection, even though cannot be used to diagnose PD definitively , is very useful to analyse articulation patterns and phonation characteristics and provides valuable insight to the patients speech patterns.

- Acoustic Analysis which includes Pitch variation and Speech rate analysis as it observed that PD patients have reduced pitch variability and changes in speaking speed rhythm. In other words, the PD affected generally speak in monotonously and in bursts and stops.

- Voice Quality Analysis which include Jitter (fundamental frequency variation) and Shimmer (amplitude variation), Harmonic-to-noise ratio (HNR).

- Machine Learning and Pattern Recognition is by far the most prevalent method. It includes extraction of relevant features from the speech signals, such as pitch, intensity, formants and other parameters and using classification algorithms to classify speech samples as either healthy or affected by PD [9]. There are many speech processing tools which help us obtain the characteristic features and their derivatives to use in the above method: MFCCs, Linear Prediction Cepstral Coefficients (LPCCs),etc.

- The last major method which uses speech processing in devising diagnostic systems of PD use Deep Learning. It includes Recurrent Neural Network (RNN)s and Long-Short Term Memory (LSTM) networks(for sequence modelling capturing temporal dependencies), Convolutional Neural Network (CNN) [10], [11] to apply over feature extraction from spectrograms, etc.

2.2 Literature Review

The detection of PD from speech has been investigated in many studies, [12], [13], [14], [15]. These previous studies can be divided into four categories depending on the approach which has been used to analyse the speech signal [12], [14]. These four categories of the studies are: (1) phonatory, (2) articulatory, (3) prosodic, and (4) linguistic. The studies based on the phonatory approach focus on the changes in the glottal source and in the larynx. There are only a few studies in this group due to the difficulties in estimating the glottal source from disordered speech signals. Studies based on the articulatory and prosodic approaches are much more prevalent as there exist several speech processing tools for deriving features such as MFCCs [16], [17], [18], pitch, duration, etc. [19], [20]. In addition, the prevalence of the articulatory approach is explained by results shown in several studies indicating that articulation is greatly affected in PD [19], [20], [21]. Finally, the studies belonging to the linguistic approach examine the use of vocabulary, phrase construction and repetition of words by Parkinsonian speakers [22], [23]. For investigating the linguistic approach, previous studies have used Automatic Speech Recognition (ASR) in representing linguistic units from speech. These studies have used classical features such as bag of words and term frequency-inverse document frequency [24]. For the analysis and detection of PD, many feature extraction methods of speech have been used to express Parkinsonian speech signals in parametric forms. For capturing phonatory aspects, features quantifying variations in speech periodicity have been investigated [25], [26], [27], [28]. This parameterization approach is justified because the extent of variations in the vocal folds vibration in Parkinsonian speakers is more likely to deviate from healthy speakers [29], [30]. Features such as jitter (which is defined as perturbation in fundamental frequency) and shimmer (which is defined as perturbation in amplitude) are the most commonly used measures for capturing variations in vocal fold vibrations [13], [31]. In [31], [13], [32], various nonlinear voice production -based features (such as the recurrence period density entropy) were investigated. Recently in [19], glottal source features (such as the quasi-open quotient, the normalized amplitude quotient, the harmonic richness factor) were investigated in the analysis of newly diagnosed Parkinsonian patients. To measure the amount of noise in voice due to incomplete glottal closure (with symptoms like breathiness and harshness) [33], the harmonics-to-noise ratio and the noise-to-harmonics ratio have been used [13], [31]. For capturing articulatory variations, different feature extraction methods have been widely investigated in the areas of speech, speaker, and language recognition [34], [35], [36]. Among the popular features used in these areas, MFCCs, along with their first and second derivatives, have been found to be effective in the detection of PD [37] [35], [38], [39], [40]. Examples of other popular feature extraction methods, which have been first used particularly in ASR but later also in the detection of PD, are the Linear Prediction Coefficients (LPCs),LPCCs and perceptual linear prediction cepstral coefficients (PLPCCs) [41], [42], [43]. To capture the prosodic aspects, features such as duration of voiced sounds, intonation, loudness, speaking (syllable) rate are typically investigated in the detection of PD. More details about the various types of features used in the literature can be found in [35], [44], [45], [46], [47], [48], [49]. In [35], [41], [50], [51], it was observed that cepstral coefficients such as MFCCs outperformed conventional features such as phonatory and prosodic features. So, in this thesis, we are using cepstral coefficients derived from two different methods and comparing to the baseline. The first includes the statistics derived from the cepstral coefficients of a wavelet transformed speech signal. The second compares the same of cepstral coefficients of Zero-Time Windowed signal.

2.3 Baseline

2.3.1 Baseline Feature

In this thesis, from the above literature survey, a PD detection system based on MFCCs is considered as a baseline model. MFCCs model the vocal tract system based on human perception. It is very difficult for humans to differentiate between higher frequencies compared to the lower frequencies. Making use of this fact, Mel scale reduce the redundancy of the signal information by using a log scale, where the rate of information preserved is higher in the lower frequencies compared to higher frequencies.

The block diagram Figure 2.2 explains the steps of MFCC feature extraction. First, the power spectral density (PSD) of the windowed speech signal (20ms speech segment multiplied by window) which represent energies at different frequencies. Then, it is passed through the bank of filters which are spaced by Mel scale distance. Further, these filter bank energies are converted to log scale and discrete cosine cosine transform (DCT) is applied to decorrelate energies. The first 13 coefficients obtained after the application of DCT are considered as MFCC features. In addition to MFCCs, the first and second derivatives of MFCCs, i.e., MFCC+Δ and MFCC+Δ+$\Delta\Delta$ have also been combined and a total of 39 features.

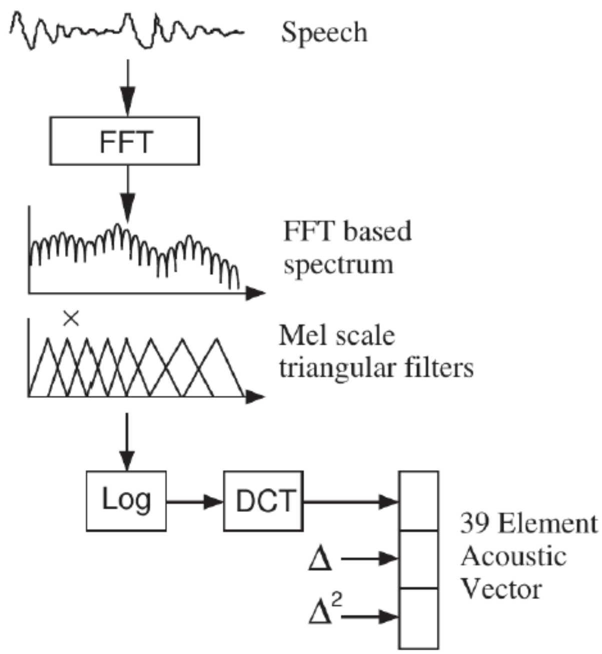

Figure 2.2 MFCC Feature Extraction

2.3.2 Database

The PC-GITA database [52] includes speech recordings of 70 people with PD and 70 healthy controls, 35 men and 35 women on each group. All the speakers are Columbian Spanish native speakers. The mean age of each gender and health is around 60-61 years. Therefore, the database is well balanced in terms of age and gender. The recordings were captured using a professional audio card with up to 24 bits and such that supports up to 96 Kbps of sampling rates. All the patients were diagnosed by neurologist experts and were labelled according to the Unified Parkinson's Disease Rating Scale (UPDRS) and H&Y scales. Also, none of the healthy controls had symptoms associated to PD or any other neurological disease.

2.4 Evaluation Metrics

In this thesis, three different evaluation metrics are used. The primary one is classification or detection accuracy, and the secondary include AUC and F1 score.

2.4.1 Classification Accuracy

The classification accuracy is the ratio of the number of correct predictions to the total number of input samples.

Table 2.1 Possible combinations of a Prediction vs Ground truth

Prediction/real class	Class 1	Class 0
Class 1	True positive (TP)	False positive (FP)
Class 0	False negative (FN)	True negative (TN)

The possible combinations of the actual data and the results are shown in the table 2.1. The accuracy is calculated by taking the ratio of the true predictions versus all entries as (TP+TN)/(TP+TN+FP+FN).

2.4.2 Area under the Curve

Area under receivier operating Characteristic Curve (AUC) is the area under the receiver operating characteristic curve, which is a plot of FPR versus TPR (i.e., False Positive Rate defined as FP/(FP+TN) and False Negative Rate defined as FN/(FN+TP)). The higher the AUC, the better is the performance of the classification system at predicting the true class.

2.4.3 F1 score

F1 score is an evaluation metric that measures the model accuracy. It combines the precision and recall scores of a model by using their harmonic mean. When the database is class balanced, i.e., the number of samples in each class is the same, F1 score maximises both precision and recall metrics. As the database used in this thesis is class balanced, F1 score is used.

Chapter 3

WAVELET TRANSFORMATION BASED PARKINSON'S DISEASE DETECTION SYSTEM

3.1 Introduction

As previously discussed, most of the works present focus on spectral, prosody features with spectral studies providing better results. So, in this thesis, we use cepstral coefficients derived from two different approaches. In this chapter, cepstral coefficients derived from wavelet transform which will be used and explained subsequently.

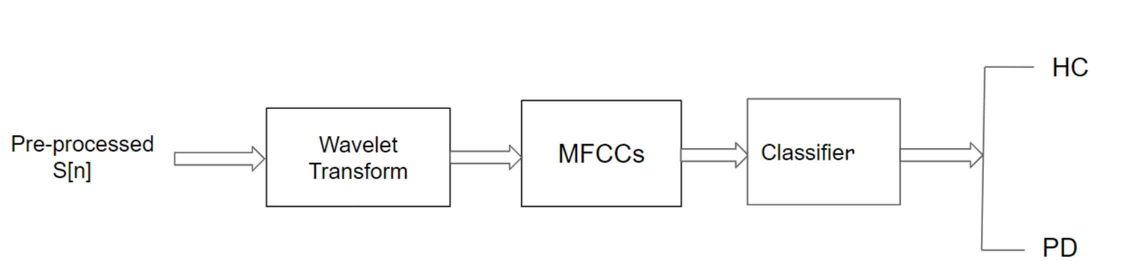

Figure 3.1 Block diagram of PD Detection System

3.2 Wavelet Transform

Wavelet Transform (WT) represents signals in terms of localized variations in both time and frequency. This gives us the ability to analyse different parts of a signal at different scales. The wavelet transform uses wavelets—short, well-localized waveforms or functions compared to the sinusoidal functions of the Fourier Transform. It decomposes a signal into different frequency components at different resolutions. This decomposition allows the representation of both high and low-frequency information in a signal. To achieve this, the wavelet transform decomposes the signals from a mother wavelet on a family of wavelets dilated by a coefficient of scale "a" inversely proportional to the frequency that enables to get different versions: dilated ones or compressed ones of the window, and translated by a translation coefficient "b" which characterizes the displacement of the window along the time axis. The Continuous Wavelet Transform (CWT) of a signal s(t) is defined by

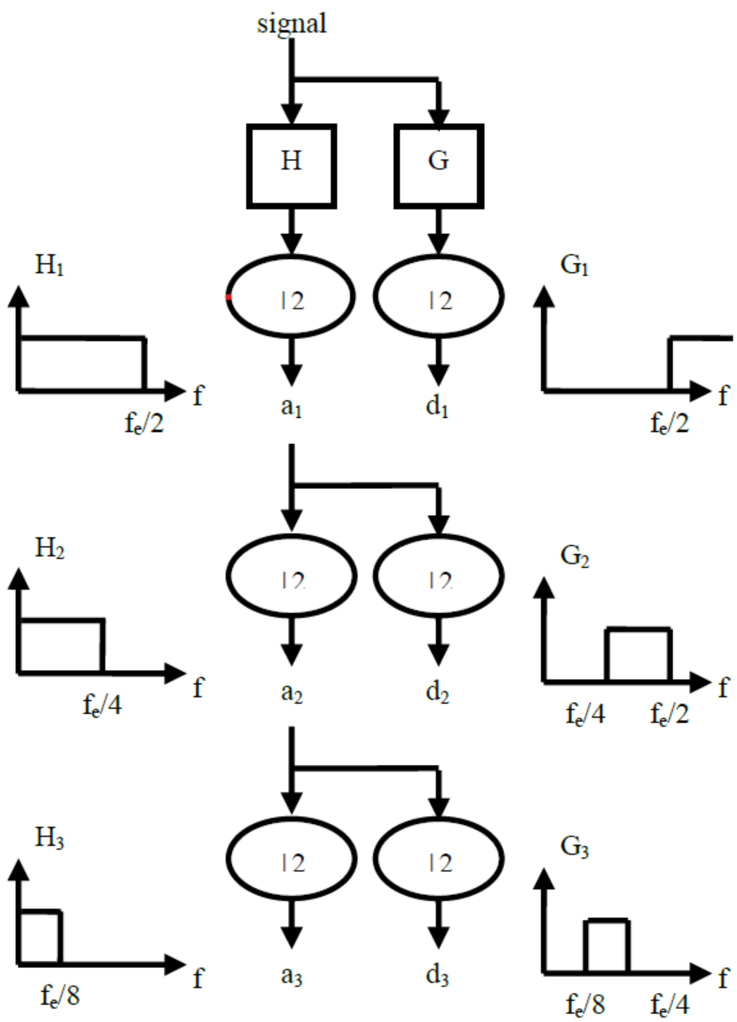

Figure 3.2 Multi-Resolution Analysis

$$W(a,b) = \int_{-\infty}^{\infty} x(t) \cdot \psi^* \left(\frac{t-b}{a} \right) dt \tag{3.1}$$

where a is the scale factor, and b is the translation parameter. As we need discrete values due to the fact that we are using sampled speech signals, we use discrete wavelets(DWT). This is achieved by Mallat algorithm for the wavelets coefficient calculation which is based on multi-resolution analysis by a sequence of filter application [53]. Given a signal, we can separate high frequency components(details) and low frequency components(approximations) by using a pair of filters H and G which are a complementary low pass filter and a high pass filter. The low pass filter is a scaling function while the high pass filter is a wavelet function.

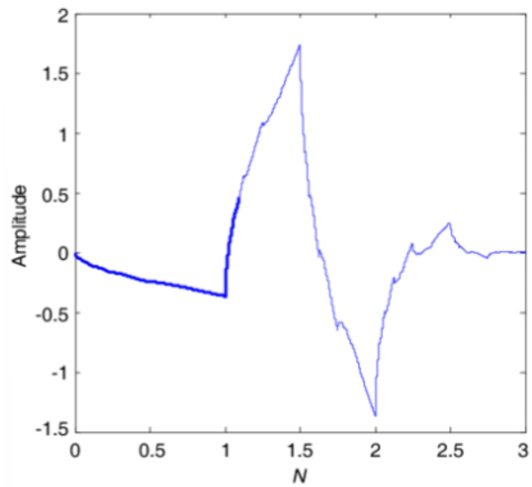

Figure 3.3 Daubechies wavelets with N=10

3.3 Experiment setup

Multi-scale decomposition of the starting signal is done by separating into the low frequencies (approximations) and the high frequencies (details) of the signal. The process is repeated recursively on the approximation coefficients, creating a binary tree structure. In other words, at each level, the approximation coefficients are split into two parts using the same scheme. The result is a decomposition of the original signal into different levels of approximation and detail coefficients. This allows for a multi-resolution analysis. In this thesis, we are using three levels of decomposition with db3 [54] wavelet as shown in the figures below.

This transformed signal is used to obtain cepstral coefficients which are used as input data to train the classifier system. In addition to Wavelet Transformation Cepstral Coefficients (WTCCs), the first and second derivatives of WTCCs, i.e., WTCC+Δ and WTCC+Δ+$\Delta\Delta$ have also been combined and a total of 39 feature to monitor the dynamic changes.

3.3.1 Baseline Feature

In this thesis, from the previous literature survey, a PD detection system based on MFCCs is considered as a baseline model.

3.3.2 Database

In this thesis, the PC-GITA database [52] is used as to evaluate the efficiency of the system prosed. It consists of the speech recordings of 70 people with PD and 70 healthy controls, 35 men and 35 women on each group. All the speakers are Columbian Spanish native speakers. The age of the men with PD ranges from 33 to 77 years old (of mean 62.2$\pm$ 11.2), the age of the women with PD ranges from 44 to

75 years old (mean 60.1$\pm$ 7.8). For, the case of healthy controls, the age of the men ranges from 31 to 86 (of mean 61.2 $\pm$ 11.3) and the age of the women ranges from 43 to 76 years old (mean 60.7$\pm$ 7.7). Therefore, the database is equal in terms of the number of each gender and the mean of each category is also similar to reduce the effect of the loss of dexterity caused by old age. To evaluate the articulation, we used the rapid repetition of the phonemes /pa-ta-ka/ (diadochokinetic evaluation).

3.3.3 Classifier

In the classification problems with small samples, the SVM is considered as one of the most powerful tool. It is a class of machine learning method (kernel learning method) and can transform a nonlinear separable problem into a linear separable problem with different kernel functions. Here, data is segregated into two classes, each represented by points in space separated by as large a distance as possible. A dividing boundary separates the classes. The choice of the boundary is taken as the one that maximizes the distance (the "margin") between the boundary and the closest point in each group. In other words, the best boundary is the one that cleanly divides the data but doesn't approach either group too closely. The data points lying closest the boundary are called the support vectors. They represent the points most difficult to classify and have a direct bearing on the optimal position of the boundary. Generally, the boundary is just a straight line or plane. This is performed by transformation of the data into a higher dimension using kernel function. The optimal boundary in SVM is calculated using a training set whose classifications and inputs are known. New cases are then classified as to which side of the boundary they fall.

Radial Basis Function (RBF) kernel, also known as the Gaussian kernel and obtained by:

$$K(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right)$$

To increase the accuracy of the classification, neural networks are used. They increase the number of layers between the input and the output and reduces the boundary values. In this thesis, we are using Narrow Neural Network and Medium Neural Network

3.4 Results and discussions

For the automatic detection of PD, in this study, we explored wavelet transformation of the speech signals and the cepstral coefficients WTCCs derived from them. The study of Parkinson's disease detection was conducted on PC-GITA and the performance of the system is compared with baseline MFCCs feature. The analysis is carried out using four features, namely static WTCCs (52-dimension feature vector), WTCCs along with its delta and delta-delta coefficients (156-dimension feature vector). For building the detection system, we used SVM classifier with an RBF kernel and narrow neural network were used. The results of this study is evaluated in terms of classification accuracy (in %), AUC, and F1-score for diadochokinetic word. From the tabulated results, it can be observed that proposed features outperformed the baseline MFCCs feature by approximately 10%. This improvement in the performance might be due to the multi-scale decomposition provided by WT transformation compared to Short-Time Fourier Transform (STFT) which is used for calculation of baseline MFCCs. From the results compiled of the detection system SVM classifier, static WT show better performance in terms of classification accuracy, AUC and F1-score of 64.3%, 0.7, and 0.73, respectively as compared to static MFCCs feature. It can be seen that performance is improved up to 9% improvement in terms of classification accuracy. Additionally, when static, delta, delta-delta features are combined for training the system performance of proposed feature for PD detection system is above 10% in terms of classification accuracy.

It can be concluded the cepstral coefficients obtained from Wavelet Transform method showed better performance in terms of classification accuracy, AUC and F1-score for both static and static along with its dynamic coefficients. While the best classification accuracy of 70.7% is obtained using WTCCs, its delta and delta-delta coefficients, it is not a significant improvement over WTCCs. The reason might that, during the wavelet transformation, both approximations and details are derived and used to obtain cepstral coefficients, which in turn reduces the need of monitoring the dynamic changes which are measured by the delta and delta-delta of cepstral coefficients.

Table 3.1 Performance of Parkinson's detection systems in terms of classification accuracy (in %) for diadochokinetic word using baseline feature and WTCCs features on PC-GITA dataset for Support Vector Machine.

Feature	Accuracy (%)	AUC	F1-score
MFCCS	54.3	0.54	0.52
WTCCs	64.3	0.7	0.64
MFCCs+Δ+$\Delta\Delta$	**70.7**	**0.8**	**0.73**
WTCCs+Δ+$\Delta\Delta$	65	0.74	0.63

3.5 Conclusions

In this experiment, we used PC-GITA database of uttering of the diadokinetic word/pa-ta-ka/ to train a Parkinson's disease detection system. The system is based on a cepstral analysis after using a signal transformation by the time frequency treatment by the discrete wavelet transform(DWT). Daubechies wavelet was used in order to transform the vocal signals by the third-scale approximations. The extraction of the Cepstral Coefficients was realized and are employed in the classification by using SVM. The results show a 65% accurate classification rate.

Chapter 4

AUTOMATIC DETECTION OF PARKINSON'S DISEASE USING ZERO-TIME WINDOW BASED CEPSTRAL FEATURE

Parkinson's diseases (PD) is one of the neurological condition affecting both motor and non-motor functions. It is second most common age related neurodegenerative disorder after Dementia impacting life of millions of individual. Dysarthria is one of the prominent symptoms observed for an individual suffering with PD. It is a speech disorder resulting from damage to the central or peripheral nervous system which in turn affect the respiratory, resonatory, phonatory, and articulatory subsystem involved in speech production. Due to non-invasive nature and allowance for remote screening and monitoring [8], detection of PD from the speech is considered as reliable method compared to the traditional approach.

Recently authors in the study [55] explored the importance of single frequency filtering cepstral coefficients (SFCCs) for the discrimination of PD subjects from healthy subjects. The results of this study highlighted the importance of higher spectro-temporal resolution compared to the STFT for automatic detection of PD subjects. ZTW is a time-frequency analysis method which is used to estimate the spectral characteristic of vocal tract system [56]. ZTW allows the possibility of deriving the instantaneous spectral characteristics at every sampling instant which in turn allows the result to be interpreted as instantaneous spectral features. It subsequently provides better resolution of the speech signal in both time as well as frequency domain [57], [58]. This chapter explores cepstral features derived from ZTW method to distinguish healthy subjects from the subjects suffering with PD.

4.1 Time-frequency based feature analysis for automatic detection of PD

4.2 Zero-time windowing method

Zero-time windowing (ZTW) method captures the spectral characteristics of the time varying vocal tract system [56] at every sampling instant. In this method, a short duration (5 ms) of speech signal is multiplied with impulse-like window in order to capture the characteristics of vocal tract system at the instant of glottal closure. Impulse-like window has most of the energy at beginning of window i.e. at zero time which emphasises the speech signal at the beginning of window. This windowing operation in the time-domain is analogous to integration operation in the frequency domain. Loss in spectral resolution (due to impulse-like window) is restored by applying the differencing operation in the frequency domain. Figure 4.1 shows us the Block Diagram of the PD detection system using ZTW

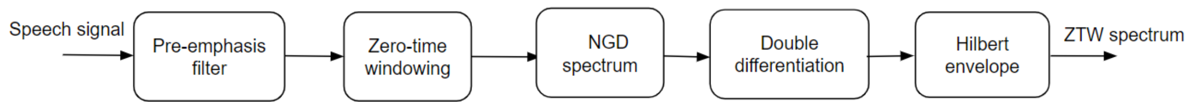

Figure 4.1 Block diagram of ZTW method.

method where the cepstral coefficients are extracted from the zero-time windowed signal.

4.2.1 Extraction of cepstral coefficient from ZTW method

In the ZTW based method, the speech signal is windowed with a heavily decaying window at each instant of time which in turn highlights the samples at the beginning of the window (near the 0th instant). This spectrum which provides higher temporal resolution can also obtain higher spectral resolution when estimated using group delay function [56]. The procedure to derive ZTW spectrum is:

1. To remove the effects of low frequency trend in the signal, the input speech signal,s[n] is pre-emphasised.

2. Speech segment of length L ms at each instant is considered. i.e, $s[n]$ is defined for $n = 0, 1, ..., M-1$, where the number of samples $M = \dfrac{L * fs}{1000}$ and fs is the sampling frequency.

3. The segment is multiplied with a heavily decaying window $w_1[n]$, where:

$$w_1[n] = \begin{cases} 0, & n = 0 \\ \dfrac{1}{4sin^2(\pi n/2N)}, & n = 1, ...N-1 \end{cases} \tag{4.1}$$

   where N is the number of samples used for Discete Fourier Tansform (DFT) and N ¿¿ M. The result of the multiplication of the signals s[n] and the window $w_{12}[n]$ is equivalent to four times integration in the frequency domain. The resulting ripple effect in the frequency domain when the signal is terminated at the instant *n=M-1* is reduced by using window $w_2[n]$, which is square of half cosine window.

$$w_2[n] = 4cos^2(\pi n/2M), \quad n = 0, ...M-1. \tag{4.2}$$

4. The spectrum is estimated using the numerator of the group delay (NGD) function (g[k]) for the windowed signal (i.e., $x[n] = w_1^2[n]w_2[n]s[n]$) and is given by:

$$g[k] = X_R[k]Y_R[k] + X_I[k]Y_I[k], \quad k = 0, ..., N-1, \tag{4.3}$$

where $X_R[k]$ and $X_I[k]$ are the real and imaginary parts of the N-point DFT $X[k]$ of $x[n]$. Likewise, $Y_R[k]$ and $Y_I[k]$ are the real and imaginary parts of the N-point DFT $Y[k]$ of $y[n] = nx[n]$.

5. To obtain the peaks in the spectrum corresponding to formants of the vocal tract system, the numerator of the group delay function is double differentiated and its Hilbert envelope is denoted by $X[n, k]$.

4.3 Parkinson's diseases detection system using ZTWCCs

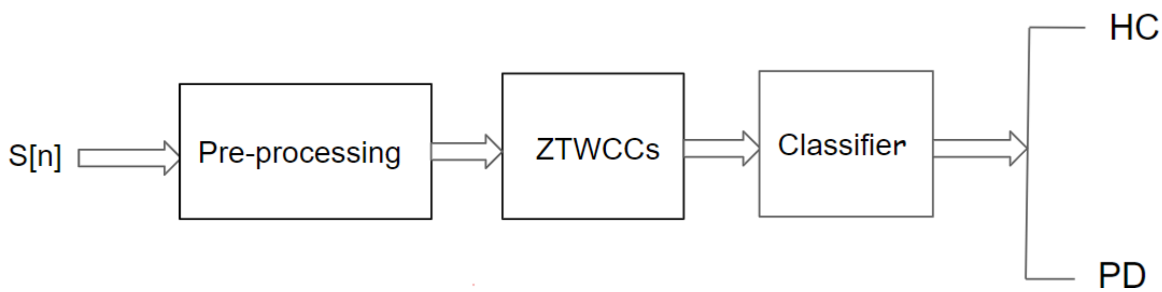

Figure 4.2 Block diagram of PD detection system.

Figure 4.2 shows the block diagram of PD detection system used in this paper. The pre-emphasised speech signal is segmented and frame shifted as will be discussed in the following section. The resultant signal is used to obtain ZTWCC (zero time windowing cepstral coefficients), ZTWCC+Δ and ZTWCC+$\Delta\Delta$ features. The statistics based on these features are used as training data to a classifier which gives a binary output label. The labels used are Healthy Cases (HC) and Parkinson's Disease (PD).

Figure 4.3 illustrate the speech signal, corresponding STFT, ZTW based spectrogram for healthy control (HC) subject and subject suffering from Parkinson's diseases (PD). It can be observed from the figure that STFT based spectrogram (for the window size of 20 ms) provides the good frequency resolution but smear the information in time domain, while ZTW-based spectrogram provides more better resolution in both time and frequency domain. Moreover, ZTW-based spectrogram shows the significant discrimination between subject with HC and and subject suffering from PD.

4.4 Experimental set up

This section explains database used in this study along with features used for performing the experiments. Further it also discusses the classifier we used to build the PD detection system.

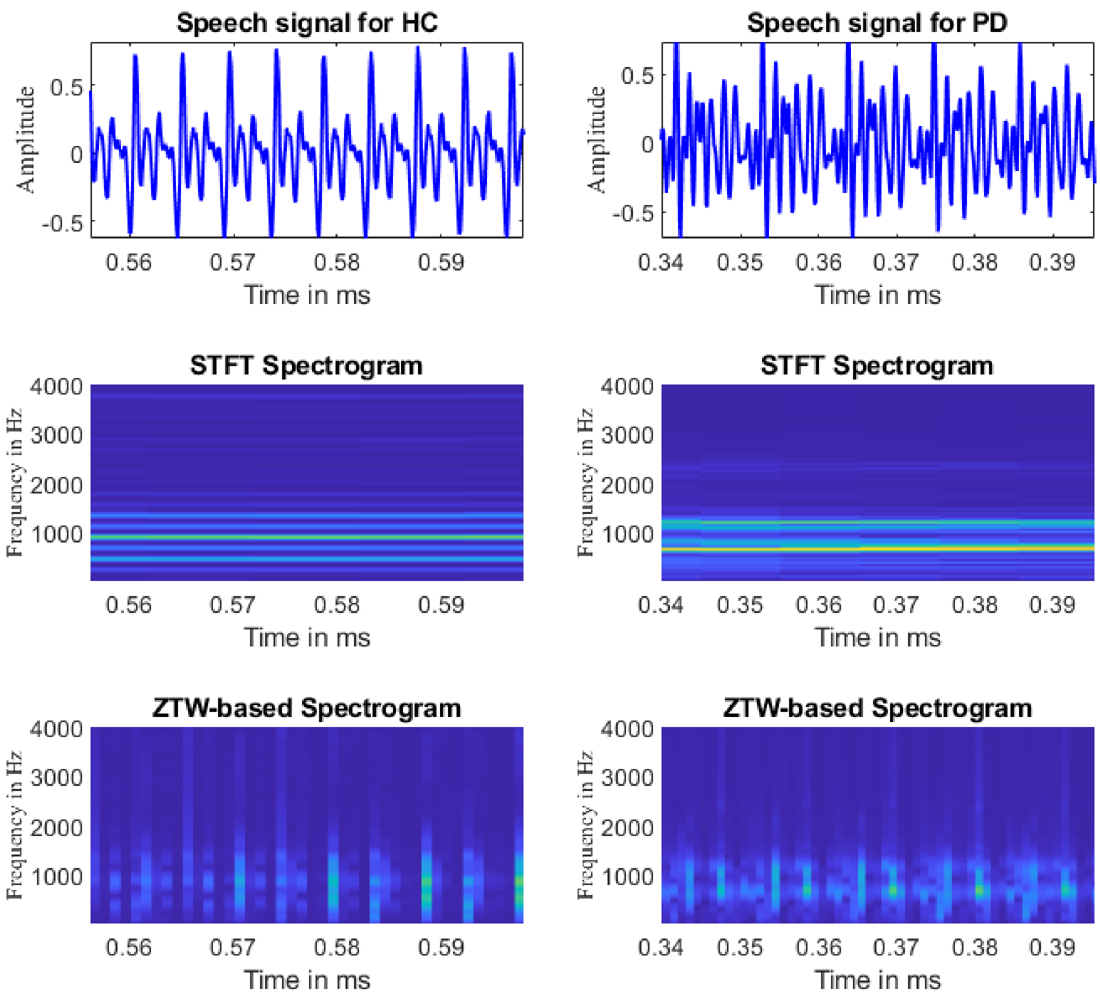

Figure 4.3 Illustration of speech signal, spectrograms obtained from STFT, and ZTW methods for HC and PD.(a) and (d) Speech signal, (b) and (e) STFT spectrogram, (c) and (f) ZTW spectrogram for healthy subject and subject suffering with PD, respectively.

4.4.1 Database

PC-GITA database [52] is used in this paper for performing the experiments. It is a well balanced dataset in terms gender and number of speaker for both the classes. The dataset consists of vowels, isolated words, diadochokinetic words, sentences, and reading text for 70 healthy and 70 subjects suffering with PD. All the speech samples were recorded in the sound proof room. In this study we used diadochokinetic words and vowel /A/ for performing the experiments.

Table 4.1 Details of PC-GITA database and number of samples used in this paper for PD detection task.

Class	No. of Speakers
HC	70 (35 male and 35 female)
PD	70 (35 male and 35 female)

4.4.2 Features

- MFCCs features derived from speech signal are used as baseline feature in this study. It is computed using 20 ms frame size and with the frame shift of 5 ms. 13 static, and 13 delta (first derivative) and 13 delta-delta (second derivative) coefficients make it as 39 dimension feature vector. Four statistics of 39 dimension feature vector are computed which in turn resulted in 156 dimension feature vector. The statistics considered are mean, standard deviation, kurtosis, and skewness.

- For the computation of proposed feature, first the speech signal is divided into segments of 5 ms, with frame shift of 1 ms. Then feature vector is computed from the ZTW spectrum of speech signal. It consists of static, delta and delta-delta cepstral coefficient each of size 13, which results in 39 dimension feature vector. Additionally, four statistics namely, mean, standard deviation, kurtosis, and skewness are computed on these 39 dimension feature vector which resulted in 156 dimension feature vector.

4.4.3 Classifier

SVM classifier is a most widely used classifier for pathological speech processing application due to its reliable performance even on small dataset. In this study, SVM classifier is used for training the Parkinson's Disease detection system. Additionally, various kernel functions such as linear, polynomial, and RBF are also used for training the system. Among them, the RBF kernel exhibited the best performance. 10-fold cross validation is used to conduct the experiments and the average classification accuracy of all folds is referred to as classification accuracy and is measured in percentage. The performance of the PD detection system is also measured in terms of AUC, and F1-score. The values of AUC and F1-score lies between 0 to 1. A value of 1 indicates that classifier can correctly distinguish between the two classes, while 0 indicates that classifier gets confusion between the classes.

Additional classifiers which we used in this experiment for training the PD system are Narrow Neural Network (NNN) and medium Neural Network (NN) as they allow for greater generality in unseen data, faster training time and greater interpretability due to fewer parameters. In this experiment, the neural networks with a single layer are used for training the system utilizing Rectified Linear Function (ReLU) function due to its effectiveness and simplicity in computations as it uses max function compared to those that use computationally difficult functions like exponential functions.

4.5 Result and Discussion

This study explored ZTWCCs features for automatic detection of Parkinson's diseases on PC-GITA dataset. The performance of the PD detection system is also compared with baseline MFCCs feature. The experiments are carried out using four features, namely static MFCCs (52-dimension feature vector), MFCCs along with its delta and delta-delta coefficients (156-dimension feature vector), static Zero-Time Windowing Cepstral Coefficients (ZTWCCs) (52-dimension feature vector), and ZTWCCs along with its delta and delta-delta coefficients (156-dimension feature vector).

For building the detection system, we used SVM classifier with an RBF kernel and narrow neural network were used. The results of this study is reported in terms of classification accuracy (in %), AUC, and F1-score in Table 4.2, Table 4.3, Table 4.4 and Table 4.5 for diadochokinetic word and isolated vowel.

From the tabulated results, it can be observed that proposed features outperformed the baseline MFCCs feature. This improvement in the performance might be due to better spectro-temporal resolution property provided by ZTW method as compared to STFT which is used for calculation of MFCCs.

From the Table 4.3, which shows the results using SVM classifier, static ZTWCCs show better performance in terms of classification accuracy, AUC and F1-score of 63.6%, 0.74, and 0.64, respectively as compared to static MFCCs feature. It can be seen that performance is improved up to 9% improvement in terms of classification accuracy. Additionally, when static, delta, delta-delta features are combined for training the system performance of proposed feature for PD detection system is above 1% in terms of classification accuracy.

It can be concluded from the Table 4.3, cepstral coefficients obtained from ZTW method showed better performance in terms of classification accuracy, AUC and F1-score for both static and static along with its dynamic coefficients. The best classification accuracy of 70% is obtained using ZTWCCs, its delta and delta-delta coefficients.

Table 4.4 and Table 4.5 shows the results when the Parkinson's detection systems is built using narrow neural network for diadochokinetic word for PC-GITA database in terms of classification accuracy. It can be observed from the tables that the performance of the system is improved by including the delta and double delta features along with static cepstral coefficients. The performance is improved up to 8%

Table 4.2 Performance of Parkinson's detection systems in terms of classification accuracy (in %) for diadochokinetic word using baseline feature and proposed features on PC-GITA dataset for Support Vector Machine.

Feature	Accuracy (%)	AUC	F1-score
MFCCS	54.3	0.54	0.52
ZTWCCs	63.6	0.74	0.64
MFCCs+Δ+$\Delta\Delta$	70.7	**0.8**	0.73
ZTWCCs+Δ+$\Delta\Delta$	**72.1**	0.79	**0.74**

Table 4.3 Performance of Parkinson's detection systems in terms of classification accuracy (in %) for isolated vowel /a/using baseline feature and proposed features on PC-GITA dataset for Support Vector Machine.

Feature	Accuracy	AUC	F1-score
MFCCS	55	0.45	0.54
ZTWCCs	66.4	0.7	0.68
MFCCss+Δ+$\Delta\Delta$	68.6	0.72	0.69
ZTWCCs+Δ+$\Delta\Delta$	**70**	**0.76**	**0.72**

Table 4.4 Performance of Parkinson's detection systems in terms of classification accuracy (in %) for diadochokinetic word using baseline feature and proposed features on PC-GITA dataset for narrow neural network.

Features	Accuracy	AUC	F1-Score
MFCCs	48.6	0.48	0.52
ZTWCCS	63.6	0.72	0.62
MFCCs+Δ+$\Delta\Delta$	65.7	0.73	0.65
ZTWCCs+Δ+$\Delta\Delta$	**73.6**	**0.81**	**0.72**

for ZTWCCs features as compared to the baseline MFCCs features. The best performance of 73.6%, 0.81 and 0.76 is observed in terms of classification accuracy, AUC and F1-score, respectively when the system is trained with narrow neural network.

Table 4.5 Performance of Parkinson's detection systems in terms of classification accuracy (in %) for diadochokinetic word using baseline feature and proposed features on PC-GITA dataset for medium neural network.

Features	Accuracy	AUC	F1-Score
MFCCs	49.3	0.48	0.50
ZTWCCS	70.7	0.78	0.71
MFCCs+Δ+$\Delta\Delta$	69.3	0.73	0.68
ZTWCCs+Δ+$\Delta\Delta$	**72.1**	**0.83**	**0.76**

4.6 Conclusion

This study used ZTWCCs for building a PD system which can distinguish a subject suffering with Parkinson's disease from healthy control subjects. The experiments were carried out on PC-GITA database for the diadochokinetic word and vowel /a/. The PD system was built using SVM classifier with RBF kernel and neural networks Narrow NN and medium NN with one layer of size 10 and 25 respectively. The performance of the system is also compared with the baseline MFCCs feature. The best classification accuracy of 72.1% and 70% is achieved using SVM classifier for diadochokinetic word and vowel, respectively on PC-GITA database. Moreover the system was also trained using the neural network architectures. In this case the best performance in terms of classification accuracy of 73.6% is obtained when the PD detection system is trained using narrow neural network using the 156 dimension ZTWCCs features. Moreover, it was observed that by increasing the number of neurons from 10 (for narrow neural network) to 25 (medium neural network) performance does not improved significantly which may be due to limited amount of data.

From the result of all the experiments, we conclude that the performance of the system is improved in terms of classification accuracy, AUC and F1-score using ZTWCCs features as compared to baseline features. The improvement in the performance may be due to good spectro-temporal resolution provided by ZTW method.

Chapter 5

Conclusion and Future Scope

5.1 Conclusion

In this thesis, we studied the available systems for the detection of Parkinson's Disease using the speech signal of the person of interest. We have also reviewed the various approaches to the systems based on the different models with varied computational demands. We have also found that the performance of the detection system of Parkinson's Disease based on cepstral features is high and more effective when compared with systems that have similar computational requirements. On this basis, we used Mel Frequency Cepstral Coefficients and their derivatives as the classification feature of the proposed system.We found that the different datasets of vowel and diadokinetic words perform differently when used with different classifiers. Also,Δ+$\Delta\Delta$ consistently outperforms the various other features and medium neural networks and narrow neural networks work as the best classifiers.

5.2 Future Scope

The proposed systems showed a significant improvement of at least 9% over the baseline systems. The system can be further improved by applying a greater variety of wavelets for comparison. It would also allow for a better match for the type of datasets based on monologues or repeated uttering of an vowel as they capture different characteristic information of the subject. AS the accuracy improves, the system can be developed to be easier to access and use with a variety of languages. As the Parkinson's Disease is a very prevalent and progressive disease, repeated and remote testing allows for earlier diagnosis for the subject and allows for higher prognosis.

List of Related Publications

- Monica Ponnam, Purva Barche, and Anil Kumar Vuppala, **"Automatic Detection of Parkinson's Disease using Zero-Time Window based Cepstral Coefficients"**, in proceedings of *IEEE - INDICON*, 2023.
- Monica Ponnam, Purva Barche, and Anil Kumar Vuppala, **"Automatic Detection of Parkinson's Disease based on Cepstral Coefficients derived from Wavelet Transformations "**, (to be published).

Bibliography

[1] T. B. Stoker and J. C. Greenland, "Parkinson's disease: pathogenesis and clinical aspects [internet]," 2018.

[2] T. Taguchi, H. Tachikawa, K. Nemoto, M. Suzuki, T. Nagano, R. Tachibana, M. Nishimura, and T. Arai, "Major depressive disorder discrimination using vocal acoustic features," *Journal of affective disorders*, vol. 225, pp. 214–220, 2018.

[3] S. Besbes and Z. Lachiri, "Multitaper mfcc features for acoustic stress recognition from speech," *International Journal of Advanced Computer Science and Applications*, vol. 8, no. 3, 2017.

[4] F. Chenchah and Z. Lachiri, "Acoustic emotion recognition using linear and nonlinear cepstral coefficients," *Int. J. Adv. Comput. Sci. Appl*, vol. 6, no. 11, pp. 1–4, 2015.

[5] M. Asgari and I. Shafran, "Predicting severity of parkinson's disease from speech," in *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology*. IEEE, 2010, pp. 5201–5204.

[6] J. R. Orozco-Arroyave, J. D. Arias-Londoño, J. F. Vargas-Bonilla, and E. Nöth, "Perceptual analysis of speech signals from people with parkinson's disease," in *Natural and Artificial Models in Computation and Biology: 5th International Work-Conference on the Interplay Between Natural and Artificial Computation, IWINAC 2013, Mallorca, Spain, June 10-14, 2013. Proceedings, Part I 5*. Springer, 2013, pp. 201–211.

[7] L. W. Lopes, L. B. Simões, J. D. da Silva, D. da Silva Evangelista, A. C. d. N. e Ugulino, P. O. C. Silva, and V. J. D. Vieira, "Accuracy of acoustic analysis measurements in the evaluation of patients with different laryngeal diagnoses," *Journal of Voice*, vol. 31, no. 3, pp. 382–e15, 2017.

[8] S. Arora, L. Baghai-Ravary, and A. Tsanas, "Developing a large scale population screening tool for the assessment of parkinson's disease using telephone-quality voice," *The Journal of the Acoustical Society of America*, vol. 145, no. 5, pp. 2871–2884, 2019.

[9] D. Gupta, A. Julka, S. Jain, T. Aggarwal, A. Khanna, N. Arunkumar, and V. H. C. de Albuquerque, "Optimized cuttlefish algorithm for diagnosis of parkinson's disease," *Cognitive systems research*, vol. 52, pp. 36–48, 2018.

[10] M. B. Er, E. Isik, and I. Isik, "Parkinson's detection based on combined cnn and lstm using enhanced speech signals with variational mode decomposition," *Biomedical Signal Processing and Control*, vol. 70, p. 103006, 2021.

[11] N. Souissi and A. Cherif, "Artificial neural networks and support vector machine for voice disorders identification," *International Journal of Advanced Computer Science and Applications*, vol. 7, no. 5, 2016.

[12] M. Cernak, J. R. Orozco-Arroyave, F. Rudzicz, H. Christensen, J. C. Vásquez-Correa, and E. Nöth, "Characterisation of voice quality of parkinson's disease using differential phonological posterior features," *Computer Speech & Language*, vol. 46, pp. 196–208, 2017.

[13] A. Tsanas, M. A. Little, P. E. McSharry, and L. O. Ramig, "Nonlinear speech analysis algorithms mapped to a standard metric achieve clinically useful quantification of average parkinson's disease symptom severity," *Journal of the royal society interface*, vol. 8, no. 59, pp. 842–855, 2011.

[14] J. R. Orozco-Arroyave, E. A. Belalcazar-Bolanos, J. D. Arias-Londoño, J. F. Vargas-Bonilla, S. Skodda, J. Rusz, K. Daqrouq, F. Hönig, and E. Nöth, "Characterization methods for the detection of multiple voice disorders: neurological, functional, and laryngeal diseases," *IEEE journal of biomedical and health informatics*, vol. 19, no. 6, pp. 1820–1828, 2015.

[15] M. Little, P. McSharry, E. Hunter, J. Spielman, and L. Ramig, "Suitability of dysphonia measurements for telemonitoring of parkinson's disease," *Nature Precedings*, pp. 1–1, 2008.

[16] S. Hawi, J. Alhozami, R. AlQahtani, D. AlSafran, M. Alqarni, and L. El Sahmarany, "Automatic parkinson's disease detection based on the combination of long-term acoustic features and mel frequency cepstral coefficients (mfcc)," *Biomedical Signal Processing and Control*, vol. 78, p. 104013, 2022.

[17] R. Fraile, N. Sáenz-Lechón, J. I. Godino-Llorente, V. Osma-Ruiz, and C. Fredouille, "Mfcc-based remote pathology detection on speech transmitted through the telephone channel-impact of linear distortions: Band limitation, frequency response and noise," in *International Conference on Bio-inspired Systems and Signal Processing*, vol. 1. SCITEPRESS, 2009, pp. 41–48.

[18] S. S. Upadhya, A. Cheeran, and J. H. Nirmal, "Thomson multitaper mfcc and plp voice features for early detection of parkinson disease," *Biomedical Signal Processing and Control*, vol. 46, pp. 293–301, 2018.

[19] M. Novotnỳ, J. Rusz, R. Čmejla, and E. Ržička, "Automatic evaluation of articulatory disorders in parkinson's disease," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 22, no. 9, pp. 1366–1378, 2014.

[20] T. Tykalova, J. Rusz, J. Klempir, R. Cmejla, and E. Ruzicka, “Distinct patterns of imprecise consonant articulation among parkinson’s disease, progressive supranuclear palsy and multiple system atrophy,” *Brain and language*, vol. 165, pp. 1–9, 2017.

[21] B. Harel, M. Cannizzaro, and P. J. Snyder, “Variability in fundamental frequency during speech in prodromal and incipient parkinson’s disease: A longitudinal case study,” *Brain and cognition*, vol. 56, no. 1, pp. 24–29, 2004.

[22] A. M. García, F. Carrillo, J. R. Orozco-Arroyave, N. Trujillo, J. F. V. Bonilla, S. Fittipaldi, F. Adolfi, E. Nöth, M. Sigman, D. F. Slezak *et al.*, “How language flows when movements don’t: an automated analysis of spontaneous discourse in parkinson’s disease,” *Brain and language*, vol. 162, pp. 19–28, 2016.

[23] K. C. Fraser, J. A. Meltzer, and F. Rudzicz, “Linguistic features identify alzheimer’s disease in narrative speech,” *Journal of Alzheimer’s Disease*, vol. 49, no. 2, pp. 407–422, 2016.

[24] P. A. Pérez-Toro, J. C. Vásquez-Correa, M. Strauss, J. R. Orozco-Arroyave, and E. Nöth, “Natural language analysis to detect parkinson’s disease,” in *Text, Speech, and Dialogue: 22nd International Conference, TSD 2019, Ljubljana, Slovenia, September 11–13, 2019, Proceedings 22*. Springer, 2019, pp. 82–90.

[25] M. Novotnỳ, P. Dušek, I. Daly, E. Ržička, and J. Rusz, “Glottal source analysis of voice deficits in newly diagnosed drug-naïve patients with parkinson’s disease: correlation between acoustic speech characteristics and non-speech motor performance,” *Biomedical Signal Processing and Control*, vol. 57, p. 101818, 2020.

[26] Z. Soumaya, B. D. Taoufiq, B. Nsiri, and A. Abdelkrim, “Diagnosis of parkinson disease using the wavelet transform and mfcc and svm classifier,” in *2019 4th World Conference on Complex Systems (WCCS)*. IEEE, 2019, pp. 1–6.

[27] A. Tsanas, M. A. Little, C. Fox, and L. O. Ramig, “Objective automatic assessment of rehabilitative speech treatment in parkinson’s disease,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 22, no. 1, pp. 181–190, 2013.

[28] B. E. Sakar, M. E. Isenkul, C. O. Sakar, A. Sertbas, F. Gurgen, S. Delil, H. Apaydin, and O. Kursun, “Collection and analysis of a parkinson speech dataset with multiple types of sound recordings,” *IEEE Journal of Biomedical and Health Informatics*, vol. 17, no. 4, pp. 828–834, 2013.

[29] C. M. Travieso, J. B. Alonso, J. R. Orozco-Arroyave, J. F. Vargas-Bonilla, E. Nöth, and A. G. Ravelo-García, “Detection of different voice diseases based on the nonlinear characterization of speech signals,” *Expert Systems with Applications*, vol. 82, pp. 184–195, 2017.

[30] S. Davis and P. Mermelstein, "Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences," *IEEE transactions on acoustics, speech, and signal processing*, vol. 28, no. 4, pp. 357–366, 1980.

[31] A. Tsanas, M. Little, P. McSharry, and L. Ramig, "Accurate telemonitoring of parkinson's disease progression by non-invasive speech tests," *Nature Precedings*, pp. 1–1, 2009.

[32] P. Henríquez, J. B. Alonso, M. A. Ferrer, C. M. Travieso, J. I. Godino-Llorente, and F. Díaz-de María, "Characterization of healthy and pathological voice through measures based on nonlinear dynamics," *IEEE transactions on audio, speech, and language processing*, vol. 17, no. 6, pp. 1186–1195, 2009.

[33] J. I. Godino-Llorente, V. Osma-Ruiz, N. Sáenz-Lechón, P. Gómez-Vilda, M. Blanco-Velasco, and F. Cruz-Roldán, "The effectiveness of the glottal to noise excitation ratio for the screening of voice disorders," *Journal of Voice*, vol. 24, no. 1, pp. 47–56, 2010.

[34] L. Moro-Velazquez, J. A. Gómez-García, J. I. Godino-Llorente, J. Villalba, J. R. Orozco-Arroyave, and N. Dehak, "Analysis of speaker recognition methodologies and the influence of kinetic changes to automatically detect parkinson's disease," *Applied Soft Computing*, vol. 62, pp. 649–666, 2018.

[35] J. R. Orozco-Arroyave, F. Hönig, J. Arias-Londoño, J. Vargas-Bonilla, K. Daqrouq, S. Skodda, J. Rusz, and E. Nöth, "Automatic detection of parkinson's disease in running speech spoken in three different languages," *The Journal of the Acoustical Society of America*, vol. 139, no. 1, pp. 481–500, 2016.

[36] L. Moro-Velazquez, J. A. Gomez-Garcia, J. I. Godino-Llorente, J. Villalba, J. Rusz, S. Shattuck-Hufnagel, and N. Dehak, "A forced gaussians based methodology for the differential evaluation of parkinson's disease by means of speech processing," *Biomedical Signal Processing and Control*, vol. 48, pp. 205–220, 2019.

[37] T. B. Drissi, S. Zayrit, B. Nsiri, and A. Ammoummou, "Diagnosis of parkinson's disease based on wavelet transform and mel frequency cepstral coefficients," *International Journal of Advanced Computer Science and Applications*, vol. 10, no. 3, 2019.

[38] C. O. Sakar and O. Kursun, "Telediagnosis of parkinson's disease using measurements of dysphonia," *Journal of medical systems*, vol. 34, pp. 591–599, 2010.

[39] A. Mansour and Z. Lachiri, "Svm based emotional speaker recognition using mfcc-sdc features," *International Journal of Advanced Computer Science and Applications*, vol. 8, no. 4, 2017.

[40] A. Benba, A. Jilbab, and A. Hammouch, "Voice analysis for detecting persons with parkinson's disease using mfcc and vq," in *The 2014 international conference on circuits, systems and signal processing*, 2014, pp. 23–25.

[41] J. I. Godino-Llorente, S. Aguilera-Navarro, and P. Gómez-Vilda, "Lpc, lpcc and mfcc parameterisation applied to the detection of voice impairments," in *Sixth International Conference on Spoken Language Processing*, 2000.

[42] J. Gómez-García, L. Moro-Velázquez, and J. godino llorente, "On the design of automatic voice condition analysis systems. part ii: Review of speaker recognition techniques and study on the effects of different variability factors," *Biomedical Signal Processing and Control*, vol. 48, pp. 128–143, 02 2019.

[43] J. R. Orozco-Arroyave, F. Hönig, J. D. Arias-Londoño, J. F. Vargas-Bonilla, and E. Nöth, "Spectral and cepstral analyses for parkinson's disease detection in spanish vowels and words," *Expert Systems*, vol. 32, no. 6, pp. 688–697, 2015.

[44] M. K. MacPherson, J. E. Huber, and D. P. Snow, "The intonation–syntax interface in the speech of individuals with parkinson's disease," 2011.

[45] 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 transactions on biomedical engineering*, vol. 59, no. 5, pp. 1264–1271, 2012.

[46] J. Mekyska, E. Janousova, P. Gomez-Vilda, Z. Smekal, I. Rektorova, I. Eliasova, M. Kostalova, M. Mrackova, J. B. Alonso-Hernandez, M. Faundez-Zanuy *et al.*, "Robust and complex approach of pathological speech signal analysis," *Neurocomputing*, vol. 167, pp. 94–111, 2015.

[47] A. Benba, A. Jilbab, A. Hammouch, and S. Sandabad, "Voiceprints analysis using mfcc and svm for detecting patients with parkinson's disease," in *2015 International conference on electrical and information technologies (ICEIT)*. IEEE, 2015, pp. 300–304.

[48] N. Cummins, A. Baird, and B. W. Schuller, "Speech analysis for health: Current state-of-the-art and the increasing impact of deep learning," *Methods*, vol. 151, pp. 41–54, 2018.

[49] B. Karan, S. S. Sahu, J. R. Orozco-Arroyave, and K. Mahto, "Hilbert spectrum analysis for automatic detection and evaluation of parkinson's speech," *Biomedical Signal Processing and Control*, vol. 61, p. 102050, 2020.

[50] A. Benba, A. Jilbab, and A. Hammouch, "Discriminating between patients with parkinson's and neurological diseases using cepstral analysis," *IEEE transactions on neural systems and rehabilitation engineering*, vol. 24, no. 10, pp. 1100–1108, 2016.

[51] A. Jafari, "Classification of parkinson's disease patients using nonlinear phonetic features and mel-frequency cepstral analysis," *Biomedical Engineering: Applications, Basis and Communications*, vol. 25, no. 04, p. 1350001, 2013.

[52] J. R. Orozco-Arroyave, J. D. Arias-Londoño, J. F. Vargas-Bonilla, M. C. Gonzalez-Rátiva, and E. Nöth, "New spanish speech corpus database for the analysis of people suffering from parkinson's disease." in *LREC*, 2014, pp. 342–347.

[53] S. G. Mallat, "A theory for multiresolution signal decomposition: the wavelet representation," *IEEE transactions on pattern analysis and machine intelligence*, vol. 11, no. 7, pp. 674–693, 1989.

[54] I. Daubechies, *Ten lectures on wavelets*. SIAM, 1992.

[55] S. R. Kadiri, R. Kethireddy, P. Alku *et al.*, "Parkinson's disease detection from speech using single frequency filtering cepstral coefficients." in *Interspeech*, 2020, pp. 4971–4975.

[56] Y. Bayya and D. N. Gowda, "Spectro-temporal analysis of speech signals using zero-time windowing and group delay function," *Speech Communication*, vol. 55, no. 6, pp. 782–795, 2013.

[57] R. Kethireddy, S. R. Kadiri, S. Kesiraju, S. V. Gangashetty *et al.*, "Zero-time windowing cepstral coefficients for dialect classification." in *Odyssey*, 2020, pp. 32–38.

[58] S. R. Kadiri and B. Yegnanarayana, "Breathy to tense voice discrimination using zero-time windowing cepstral coefficients (ztwccs)." in *Interspeech*, 2018, pp. 232–236.