Abdul Haleem Butt

Degree Project
E4105D

HÖGSKOLAN
Dalarna

Speech Assessment for the Classification of Hypokinetic Dysarthria in Parkinson's Disease

Abdul Haleem Butt

2012

Master
Thesis
Computer
Engineering
Nr: E4105D

DEGREEPROJECT

Computer Engineering

Program **Master of Science in Computer Engineering**	Reg Number **E4105D**	Extent **15 ECTS**
Name of student **Abdul Haleem Butt**	Year-Month-Day **2012-03-16**	
Supervisor **Taha Khan**	Examiner **Mark Dougherty**	
Company/Department **Department of Computer Engineering, Dalarna University**	Supervisor at the Company/Department **Taha Khan**	
Title **Speech Assessment for the Classification of Hypo kinetic Dysarthria in Parkinson's Disease**		
Keywords **Parkinson's disease, Hypokinetic dysarthria, Speech segmentation, Levodopa, Acoustic analysis**		

ABSTRACT

The aim of this thesis is to investigate computerized voice assessment methods to classify between the normal and Dysarthric speech signals. In this proposed system, computerized assessment methods equipped with signal processing and artificial intelligence techniques have been introduced. The sentences used for the measurement of inter-stress intervals (ISI) were read by each subject. These sentences were computed for comparisons between normal and impaired voice. Band pass filter has been used for the preprocessing of speech samples. Speech segmentation is performed using signal energy and spectral centroid to separate voiced and unvoiced areas in speech signal. Acoustic features are extracted from the LPC model and speech segments from each audio signal to find the anomalies. The speech features which have been assessed for classification are Energy Entropy, Zero crossing rate (ZCR), Spectral-Centroid, Mean Fundamental-Frequency (Meanf0), Jitter (RAP), Jitter (PPQ), and Shimmer (APQ). Naïve Bayes (NB) has been used for speech classification. For speech test-1 and test-2, 72% and 80% accuracies of classification between healthy and impaired speech samples have been achieved respectively using the NB. For speech test-3, 64% correct classification is achieved using the NB. The results direct the possibility of speech impairment classification in PD patients based on the clinical rating scale.

List of Figures

List of Tables

ACKNOWLEDGMENT

I wish to express my appreciation for my dedicated supervisor Taha Khan for his advice, support and guidance in my thesis work. I am able to complete this thesis because of their guidance and long-term, laudable support. He is always willing to discuss my progress with me any time he is available. I greatly appreciate their patients and kindness.

At the same time, I wish to acknowledge all of my teachers in Dalarna University specially Jerker Westin, Hassan Fleyeh, Siril Yella and Mark Dougherty for their guidance during my study in Dalarna University. I am also very thank full to my family for their kind support.

Chapter 1 Introduction

1.1 Hypokinetic Dysarthria

Parkinson's disease (PD) is a degenerative disorder of the central nervous system. PD occurs when a group of cells in an area of brain called substantia-nigra begin to malfunction. These cells in the substantia-nigra produce a chemical called dopamine. Dopamine is a chemical messenger which sends the information to part of the brain that controls the body movement and coordination [1].

PD is a progressive disease that increases with the time. It directly affects the muscles that are used for speech production. This phenomenon is known as Hypokinetic Dysarthria (HKD). Hypokinetic means reduced movement and Dysarthria means anomaly due to uncontrollable movement of the muscles that are used for speech production [1]. HKD is a speech anomaly due to uncontrollable movement of the muscles that are used for speech production (face and jaw). HKD can affect respiration (breathing), phonation (voice production), resonation (richness of voice), and articulation (clarity of speech).

To maintain the adequate amplitude (loudness) of the speech, the air flows periodically through the lungs. In PD, the flow of air is affected which directly affect the loudness of speech [2]. Due to the flow of air through lungs vocal folds vibrate and in high pitched sound, vibration of vocal folds is fast. Similarly for low pitched sound, the vibration is slow. Change in pitch is most common complaint in the voice of PWP [2]. Some male reports higher pitch sound while some females report lower pitch sound. Richness of voice is determined by resonating system. Due to an abnormal resonation nasal sounds are also very common in PWP. Articulatory system is affected in HKD because of uncontrollable movement of the muscles.

For HKD evaluation, conversational speech, articulation errors and vowel prolongation may be analyzed to assess harshness, breathiness, loudness pitch and repeated phonemes. There is an evidence of improvement in speech production with the Levodopa treatment [1]. Unfortunately patients have physical limitations to reach the clinicians and speech therapists. Mobile device assessment tools can be used to monitor the speech impairment when patient have PD.

1.2 Aim of this work

The goal of this thesis is to investigate speech processing methods to detect healthy and impaired voice (in case of HKD) based on speech recordings. Proposed technique is based on four steps i.e. speech preprocessing, speech segmentation, feature extraction and feature classification.

1.3 Challenges

Characterization of the voice in real time environment is a big challenge [3]. Patient's speech can be collected from different sources. It can be acquired from a phone call or from mobile device assessment in which background noise may add to the human speech. To separate noise from speech is a difficult task. Male and female voice pitches are different from each other. The distance between the mouth and the phone during the collection of speech data will also contribute to the quality of the speech. Especially in case of HKD finding the exact boundary position of successive vowels is a difficult task. All these issues may result in incorrect speech detection [2].

Chapter 2 Literature Review

2.1 Previous Research

Nowadays researchers are investigating the relationship between the pathological acoustic parameters and HKD. In previous research measurements of acoustic features i.e articulator rate, segment durations, vowel format frequencies and first moment coefficients have been used. Experiments showed that HKD can be distinguished from healthy voice based on acoustic features. In another experiment, linear predictive coding has been used to distinguish the normal and HKD voice. LPC model has been used to monitor the resonance system. Problems in resonance system affect tone, quality and resonance. Patient suffering from Parkinson's disease opens his mouth much wider, which can increase the loudness of the voice. Voice segmentation in HKD is difficult task because of variations in the audio speech signal. Previous work shows that zero crossing rate (ZCR) provides essential information about voiced and unvoiced speech segments [2]. Unvoiced speech crosses horizontal axis more than voice speech. In another experiment, jitter, shimmer and fundamental frequency were used as acoustic features for classification of impaired and normal speech. The classifier used for this purpose was Multilayer Neural Network. Results showed that MLP can be used to distinguish normal and impaired voices in case of HKD [4].

2.2 Speech Segmentation

In order to analyze the speech impairment in source filter model speech segmentation can be performed on the base of harmonic frequencies and resonance frequencies. The idea of source is that, the air is produced through the lungs and the vocal tract works as the filter to produce voice. In speech impairment source is produced through excitation does not work properly. Air flow is not periodic through lungs in speech pathology. The vocal fold due to un-periodic flow of air produces irregular vibration. To find out fluctuation the harmonic frequencies can be estimated. Characteristics of harmonic frequencies can be analyzed using acoustic features in order to classify between healthy and impaired voice. Peak to peak variation in fundamental frequency or residual frequency can be analyzed using acoustic features in order to classify between healthy and impaired voice. The peaks are residual frequencies also known as formant frequencies. The vocal tracts produce resonance. Vocal tract region begins at the opening between the vocal cords and ends at the lips. It changes shape after every 10ms. In order to distinguish between two sounds we need to analyze the resonance frequencies.

Filters are required to limit one speech frequency range or range of a mixture of different frequencies. Filters have the function to remove the unwanted frequency from the signal. Low-pass filter is used to allow only frequency under the cutoff frequency and attenuate all other frequencies. High pass filter only allows the frequency above the cutoff frequency and blocks the frequency under the cutoff frequency. Band Pass filter is the combination of high pass and low pass filter which can be used for this purpose [5].

Main issue in the HKD assessment is speech segmentation in noisy environment. Acoustic features can be used for speech segmentation in order to find out the exact boundaries of uncontrollable speech signal and noisy signal. Signal energy is favorable to detect the high variation in speech signal. In noisy signal energy is low and fluctuations or variation in the speech can be observed in the energy values. Spectral centroid is the center of gravity of spectrum and it can be used for speech segmentation. If unvoiced segments only contains environmental sounds then spectral centroid values will be low because of low frequency, similarly the spectral centroid for voiced segment will be high because of high frequency [5].

Linear predictive coding (LPC) can be used for speech segmentation on the basing of autocorrelation of residual frequency. After speech segmentation residual frequency can be used to find out variations in pronounced pulses. Source filter model is a basic model for speech production. Two steps have been used in LPC model in order to separate the voice and unvoiced segments. First step is to calculate amplitude of the signal. If the amplitude is large then segment will be considered as a voice segment. Of course we need to pre-determine the range of the amplitude levels associated with voiced and unvoiced sounds. On the basis of this range we can determine the voice and unvoiced speech [6]. Second step is to determine voiced and unvoiced segments most accurately using zero crossing rate (ZCR).

2.3. Feature Extraction

Criteria used by clinicians to rate hypokinetic dysarthria are often difficult to quantify. Now we sidestep the difficult task of quantifying using acoustic features. Source filter model is a basic model to produce voice. Air being pushed from the lungs through vocal tract and out through the mouth to generate the speech. Lungs can be thought of source of the sound and the vocal tract can be thought of filter that produces various types of sounds. Variation in the vocal folds vibration and fluctuation in the resonance frequencies in the vocal tract can be analyzed with different acoustic features. Information that can be extracted from the speech signal can be grouped into the frequency domain (e.g., pitch and spectral properties), the time domain (e.g., energy and duration), and the cepstral domain. In this task those acoustic features must be used which are medically correlated with pathological voices.

Voiced fundamental frequency or pitch as well as measures like sequential cycle-to-cycle frequency (jitter) and amplitude (shimmer) variations are peculiarly powerful parameters in assessing the variations in fundamental frequency [8]. Speech is produced because of excitation of vocal tract by the periodic flow of air. ZCR is a powerful parameter to assess the periodic and un-periodic flow of the air [9]. Muscles weakness or rigidity affects the f0 abilities [10].The acoustic features which can be used to analyze the behavior of f0 are meanf0, energy entropy and spectral centroid. Meanf0 is the mean of fundamental frequency present in the signal. Uncontrollable movement of muscles causes abnormal f0. Abnormal f0 values have been reported for right brain damaged patients and various type of dysarthria. It cannot be excluded that brain damage in general is associated with f0 mean raised above normal values and altered f0 variability, resulting from a global increase in neurological tone [10]. One way to describe the characteristics of a spectrum is with statistical measures of the energy distribution.

These spectral moments reflect the central tendency and shape of the spectrum. Recent articulatory acoustic studies of dysarthria have shown that the spectral distribution of noise energy in fricatives can be used to quantify articulatory deficits [30]. Central frequency sometimes called spectral centroid is defined as: average frequency weighted by amplitudes, divided by the sum of the amplitudes. Building the concept of irregular vibration of vocal folds, earlier studies have proposed entropy measures [11]. Energy entropy is used to find out the sudden changes or micro level changes in the fundamental frequency [12]. In Parkinson disease irregular vocal fold vibration can be analyzed through energy entropy. Furthermore, these features are used to classify between the two classes marked with 0, and 1 respectively. 0 represents healthy voice and 1 represents impaired voice as marked by the clinicians.

2.4. Feature Classification

For acoustic features classification Naive Bayes can be used. The detail description of classifier is given below.

2.4.1 Naïve Bayes

Naïve Bayes classifier is a simple probabilistic classifier based on Bayes theorem with strong independence assumptions. It assumes that the presence or absence of particular features of a Class is unrelated to the presence or absence of any other feature [13]. Bayes' theorem can be stated as follows

$$Prob(B|A) = \frac{Prob(AandB)}{Prob(A)} \qquad (2.1)$$

Bayes theorem calculates the probability of both parameters A and B. For B given A it will count the number of cases where A and B occurred together and divided the number of cases Where A occurs alone [13]. Naïve Bayes classifier required small amount of training set for classification. In real time data Naïve Bayes perform better than J48 classifier.

In J48 classifier provide decision trees which are difficult to understand in real time data. Naïve Bayes calculate probability which is very simple to understand and implement.

2.4.2 Cross validation

Classifier must be trained to check the reliability of classifier for new data. In this way performance of classifier can be checked in training phase. After that testing is performed to check the progress of classifier. For this purpose we need unseen instances which will be pre-classified. Cross validation is a good technique to do this task. It works as follows:

1. Separate the data in fixed number of partitions (or folds)
2. Select the first fold for testing, whilst the remaining folds are used for training.
3. Perform classification and obtain performance metrics.
4. Select the next partition as testing and use the rest as training data.
5. Repeat classification until each partition has been used as the test set.
6. Calculate an average performance from the individual experiments.

2.5 Performance Evaluation Parameters

Two criteria are discussed here in order to evaluate the performance of the statistics model. These parameters are helpful to provide the system efficiency and validation.

2.5.1 Sensitivity and Specificity

Sensitivity and specificity are statistical measures of the performance of a binary classification test, also known in statistics as classification function. Sensitivity (also called recall rate in some fields) measures the proportion of actual positives which are correctly identified as such (e.g. the percentage of sick people who are correctly identified as having the condition).

Specificity measures the proportion of negatives which are correctly identified (e.g. the percentage of healthy people who are correctly identified as not having the Condition).

This can also be written as:

$$Sensitivity = \frac{number\ of\ true\ positives}{number\ of\ true\ positives + number\ of\ false\ negatives} \qquad (2.2)$$

Specificity is proportion of the patients who have no disease. This can be written as

$$Specificity = \frac{numberoftruenegatives}{numberoftruenegatives + numberoffalsepositives} \qquad (2.3)$$

2.5.2 ROC Curve

In medical diagnostic ROC graph is very widely used. ROC is a graph between True Positive (TP) rate (plotted on Y axis) and False Positive (FP) rate (plotted on X axis).

True positive rate is also known as sensitivity. Sensitivity is the measure of the proportion of actual positives which are correctly identified as such (e.g. the percentage of sick people who are correctly identified as having the condition). Similarly false positive rate (also known as 1-specificity or true negative rate). An ROC graph depicts the performance of the classification. The point (0, 1) represents perfect classification. [16].

2.5.3 Chi-squared attribute evaluation

To assess the acoustic features performance chi-squared test has been performed. It evaluates the worth of a feature by computing the value of the chi-squared statistic with respect to the class. The initial hypothesis is the assumption that the two features are unrelated, and it is tested by chi-squared formula:

$$X^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(O_{IJ} - E_{ij})2}{E_{ij}} \qquad (2.4)$$

Where $O_{ij}$is an observed frequency and E_{ij} is an expected (theoretical) frequency, asserted by the null hypothesis. The greater the value of evidence against the hypothesis [17].

The range of each feature is subdivided in a number of intervals, and then for each interval the number of expected instances for each class is compared with the actual number of instances. This difference is squared, and the sum of these differences for all intervals, divided by the total number of instances is the x^2 value of that feature.

2.5.4 Information Gain

Information gain is the change in information entropy from prior state to a state that takes some information in (equation 2.5).

$$IG(Class, Attribute) = H(Class) - H(Class|Attribute)/H(Attribute) \qquad (2.5)$$

H specifies the entropy. It evaluates the worth of an attribute by measuring the information gain with respect to the class. A weakness of the IG criterion is that it is biased in favor of features with more values even when they are not more informative [18].

2.5.5 Gain Ratio

The Gain Ratio evaluates the worth of an attribute by measuring gain ration with respect to the class. This is non-symmetrical measure that is introduced to compensate for the bias of the IG. GR is given by

$$\mathrm{GR} = \frac{IG}{H\ (X)} \qquad (2.6)$$

As equation (2.6) presents, when an attribute has to be predicted we normalize the IG by dividing by the entropy of X and vice-versa. Where X is child nodes of attribute. Due to this normalization, the GR values fall always in the range [0, 1].

2.5.6 Correlation Coefficient

Correlation coefficient has been used to find out the statistical relationship between two random variables or two sets of data. Correlation coefficient (CC) is a numerical value between -1 and 1 that expresses the strength of the relationship between two variables. Jacob Cohen suggested that a correlation between 0.9 to 1 is almost prefect, correlation 0.7 to 0.9 is very high correlation, 0.5 to 0.7 is high correlation, 0.3 to 0.5 is moderate correlation, 0.1 to 0.3 is low and Correlation between 0 to 0.1 is very small correlation [19]. Positive correlation exists when one variable decreases and the other variable also decreases. Similarly negative relationship between two variables is one in which one variable increases as the other variable decreases.

Guttman scale is a procedure to determine whether a set of items can be ranked in an order on a one-dimensional scale. It utilizes the intensity structure among several indicators of a given variable. The function used for the calculation of intensity structure is MU2 function. On the basis of this function Guttman ranks the data in an order of one dimensional scale [20].

2.5.7 Features Analysis using Student's t-Tests

A t-test is a statistical hypothesis test in which the tests Statistics follow a Student's t distribution where the null hypothesis is supported. The unpaired, or "independent samples" t-test is used when two separate sets of independent and identically distributed samples are obtained. Error bars are graphical elements included in a statistical plot to represent the uncertainty in a sample statistic. Overlapping of error bars shows the significant difference between populations. If the error bars do not overlap, it is presumed that there is a statistically significant difference between them. The test determines whether the data has come from the same population or not. In general, observation if the two error bars from two populations overlap, there is chance the true mean of two populations falls somewhere in the region of overlap. So the true population mean could be the same. In this case, we conclude that the Samples do not support the hypothesis of difference between two populations. There is another possibility that true population mean may not fall in the region of overlap.

In this case, we conclude that populations are different. In t test comparison of mean shows the stronger or weaker relation between two groups. Big difference in mean values indicates the big difference between two populations. Similarly, small differences between two means values indicate the small difference between two populations. Standard deviation is also used to find out the variability between the classes. STD is calculating the variability that is noise which may make difficult to see group difference. Through STD we are able to see variation individually in each class.

Chapter 3 Methodology

3.1 Data Acquisition

We used data collected in the study of Goetz et al. (2009), recently summarized in Tsanas et al.(2010a). The data set that is used in this work is collected through the Quantitative Motor Assessment Tool (QMAT) system. All the data has been de-identified. Each speech test is paired with UPDRs test. The data consisted of both normal and pathological voices. For spoken passage tests. Speech audio samples are rated on the performance of subjects in spoken sentences. Three types of sentences spoken by each subject. Each sentence spoken by 120 subjects. Sentence for speech test-1 is "The north wind and the sun was disputed which was stronger……". Sentence for speech test-2 is "When the sun light strike rain drops in the air……". Sentence for speech test-3 "you wish to know all about my grandfather……..". Two hundred twenty (220) audio samples have been assessed. Further aim to use this data set is to discriminate healthy and impaired voice in the case of hypokinetic dysthria. Many of these data points come from clinical visits where the subject's took QMAT tests are included in this data set.

3.2 Methodology

This section explains the methodology which is based on finding out voice impairment through source filter model during speech production. Flow chart in figure 3.1 shows the procedure to classify between the healthy and impaired voice (i.e 0 and 1). Where 0 is for healthy voice and 1 is for impaired voice. Following steps are taken to complete the process of classification.

1. Band pass filter has been used to separate the human speech from noisy frequencies.
2. Speech segmentation is performed to separate between the voiced and unvoiced segments.

3. Feature extraction is performed to analyze the voiced segments. EE, ZCR, spectral centroid, jitter (RPQ), jitter (PPQ), shimmer and f0 have been extracted from voiced segments.
4. Features are trained using NB classifier in WEKA tool to classify between 0 and 1.

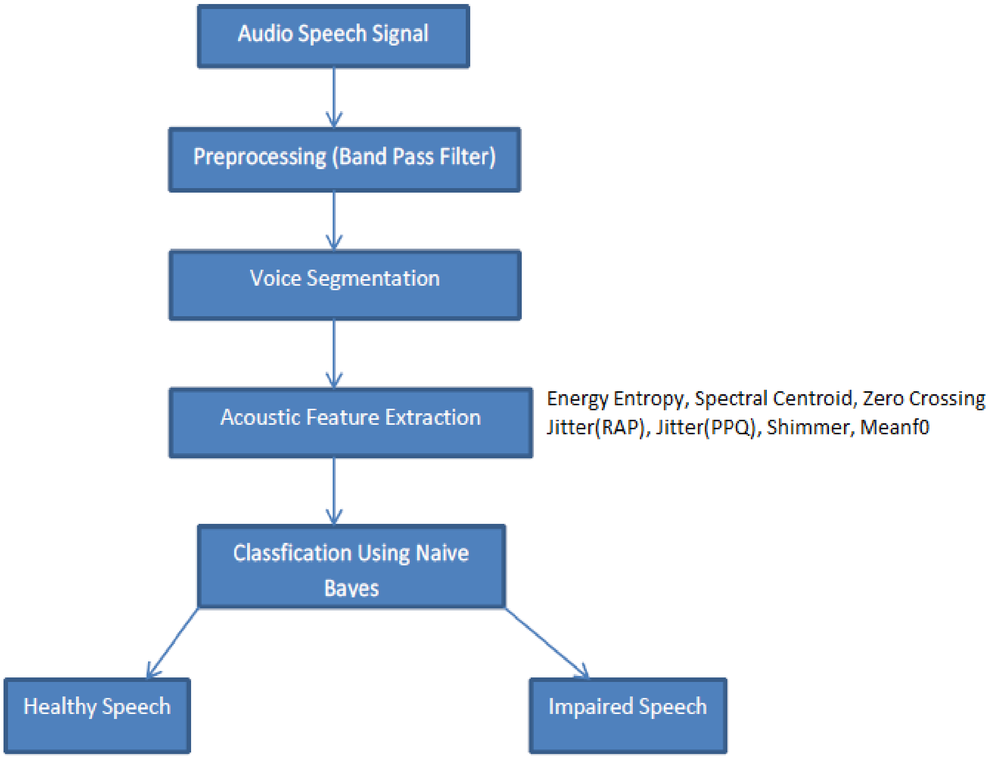

Audio Speech Signal
Preprocessing (Band Pass Filter)
Voice Segmentation
Acoustic Feature Extraction
Energy Entropy, Spectral Centroid, Zero Crossing
Jitter(RAP), Jitter(PPQ), Shimmer, Meanf0
Classfication Using Naive Bayes
Healthy Speech
Impaired Speech

Figure 3.1: Propose Methodology Flow Chart.

3.3 Band Pass Filter

Males and females differ by 88 Hz in their high tones. Women produce a fundamental frequency of 358 Hz and men 270 Hz on average. For female the higher range of the frequency is 140 to 400 Hz and for males 70 to 200 Hz . For low tones, the baseline values for females and males are 289 Hz and 201 Hz respectively. So the most recommended frequency for human speech segmentation is between 70to 400 Hz [18]. Band-pass filter has been used to segment the speech between 70 to 400Hz. Band-pass filter allowed to pass frequency within a certain range and rejects frequencies outside that range.

Band Pass filter use Fourier transform to convert signal from time domain to frequency domain. Before giving the output signal it is converted back in to the time domain using the inverse Fourier transform. An ideal band pass filter would completely attenuate all frequencies outside the pass band.

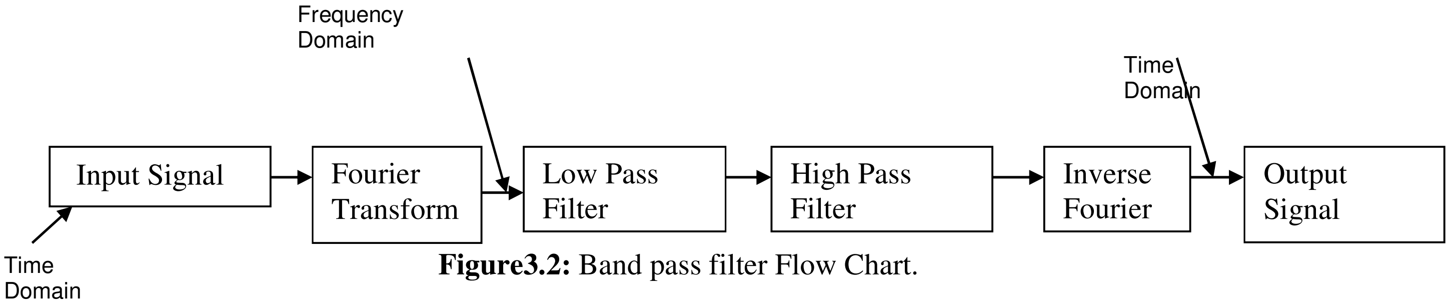


Figure3.2: Band pass filter Flow Chart.

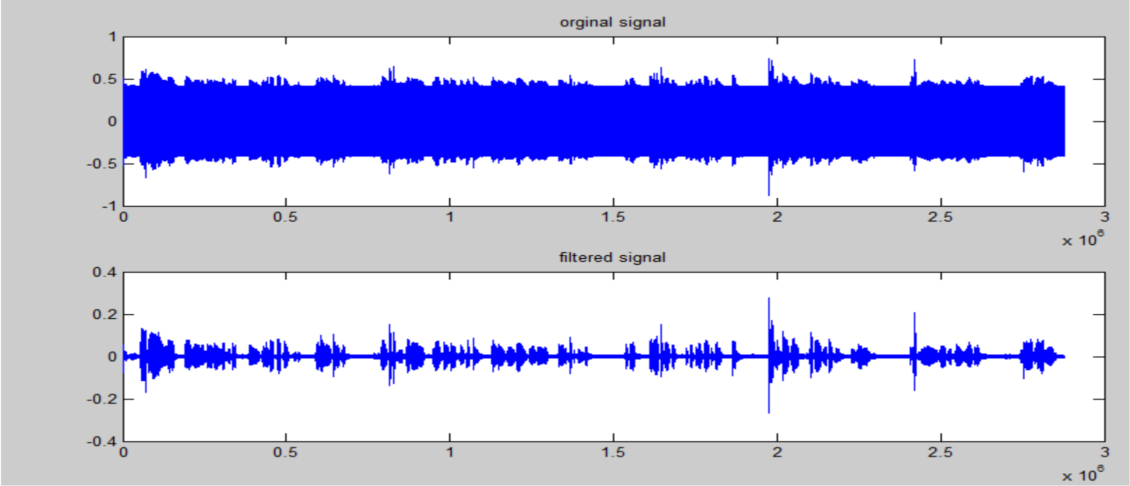


Figure3.3: Original and filtered signal wave.

3.4 Voice Detection

Speech signal consists areas of silence and noise. Therefore, in speech analysis it is needed to first apply a silence removal method in order to detect the clean speech segments [9]. Further these speech segments have been used to find out the variations in harmonic frequencies using different acoustic features. Method of speech segments detection is given below [22].

1. Compute signal energy and spectral centroid from audio signal
2. Threshold is estimated for signal energy and spectral centroid
3. This threshold is used to find the speech segment from audio signal
4. Finally post processing is done to merge the speech segments

3.4.1 Short-term Processing For Audio Feature Extraction

Audio signal may be divided into overlapping or non-overlapping short-terms frames or windows. The reason to use this technique is that audio signal varies with time, close analysis for each frame is very necessary [23].

Let's suppose we have a rectangular window w (n) which has N samples

$$w(n) = \begin{cases} 1, & 0 \leq n \leq N\text{-}1 \\ 0, & \text{elsewhere} \end{cases} \tag{3.1}$$

Frame of the original signal is related to the shifting process of the window with the time. The sample of any frame is computed using equation (3.2)

$$xi(n') \equiv x(n)w(n - m_i) \tag{3.2}$$

m_i is the shift of the window with the i'th frame, and its value depends on the size of the window and step. Window size must be large enough to calculate the data and will be short enough for the validity. Commonly, window size varies from 10 to 50 msec and step size depends on the level of overlap [6]. To extract the features, signal is divided into non-overlapping frames size of 0.05 seconds length in order to extract the features from each frame. As long as window size and step is selected the feature value F is calculated for each frame. Therefore, an M- element array of feature values F = fj , j = 1…..M, for the whole audio signal is calculated. Obviously, the length of that array is equal to the number of frames. Number of frames in each audio signal will be calculated using $M = \left[\frac{L-S}{N}\right] + 1.$ Where N is the window length, S is the window step, and L is the total number of samples of the signal [7].

3.4.2 Threshold Based Segmentation

In order to find out the threshold from acoustic feature, the signal is broken in to non-overlapping short-term frames. For each frame spectral centroid and short time energy are calculated. The sequence values of audio signal are compare with the calculated threshold from both acoustic features in order to separate the voiced and unvoiced area.

Signal energy (STE) is a time domain audio feature. Speech signals consist of many silence area between high energy values. In general observation the energy of the voiced segment is larger than the energy of the silent segments. Let $\boldsymbol{xi}$ (n); (for n = 1……N) the audio samples of the ith frame, of length N. Then, for each frame i the energy is calculated according to the equation (3.3):

$$E(i) = \frac{1}{N}\sum_{n=1}^{N}|xi(n)|^2 \tag{3.3}$$

To extract the short time energy from audio signal it is divided in to the short term frames. In order to calculate threshold, normalization of value is required. Normalization of audio signal value between 0 and 1 is performed. Number of frames in the audio signal is calculated using the formula $M = \left[\frac{L-S}{N}\right] + 1$. N is the window length, S is the window step and L is the total number of samples of the signal. Window and step size is in second which is 0.050 second. After calculation of number of frames from audio signal,energy is calculated for each frame.The energy for each frame is used to find the silent period in each frame on the basis of the feature thershold computed from the sequence values.

Spectral centroid (SC) it is a frequency domain audio feature. This feature is use to find the high values in the spectral position of the brighter sounds [22]. Spectral centroid is basically used to calculate the center of gravity of each spectrum by using the equation 3.9 given below.

Threshold Estimation: For Threshold from both features following step are followed

1. Compute Histogram of the feature sequence's values for SC and STE
2. Apply Smoothing filter on histogram
3. Find local maxima of histogram of feature sequence values

If there are two local maxima in the histogram it is calculated using formula.

$$T = \frac{W.M_1 + M_2}{W+1} \tag{3.4}$$

Where M_1 is first local maxima and M_2 is second local maxima and W is user defined parameters [21]. Large value of W increases the threshold value.

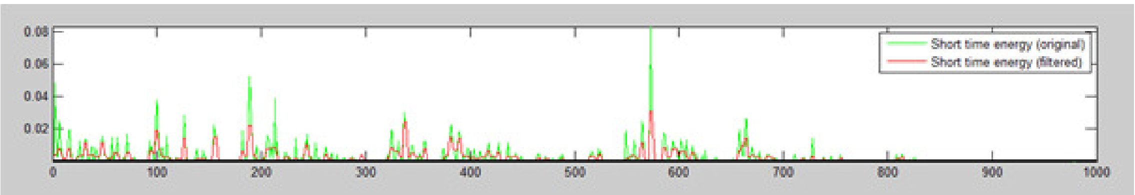


Figure3.4: Original (green lines) and filtered (red lines) short time energy of audio signal.

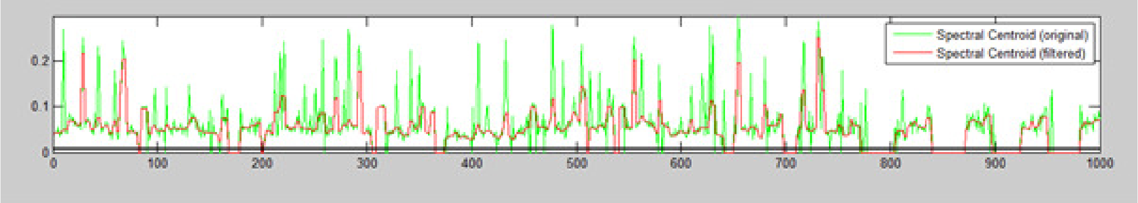


Figure3.5: original (green lines) and filtered (red lines) spectral centroid of audio signal.

The above figures 3.2 and 3.3 shows the process computed for two thresholds T1 and T2 for the signal energy and spectral centroid sequence respectively. The green line shows the original sequence values of the signal energy and spectral centroid of each frame. Red line shows the filtered sequence values of the signal energy and spectral centroid of each frame.

Filtering is performed using median filter to remove the random noise. Repetition of values is also a noise. Main idea of median filter is run through signal entry by entry replacing each entry with median of neighboring entries [22].

The segment is considered speech segment if the feature sequence values are greater than the computed threshold T1 &T2 [9]. After defining the limits or threshold, speech segments are found on the base of these limits and these speech segments are put in to one array in order to merge the all overlapping speech segments. Red color in the figure 3.4 shows the detected speech area from the audio signal.

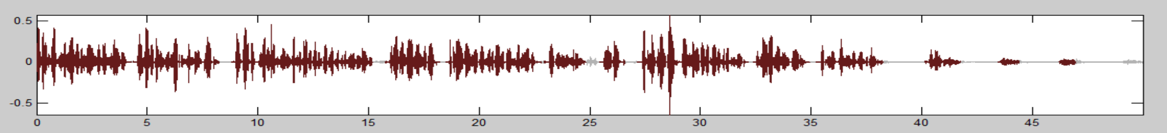


Figure3.6: Detected Voice Segments.

3.4.3 Speech Segmentation Using Linear Predictive Coding

LPC model is a vocal tract model. The basic speech parameters that can be estimated using this technique are pitch, and formant. The basic idea of LPC model is that one speech sample can be predicted from the past samples which is why known as linearly predictable. Sometimes we need to analyze overall formant pattern without interference of harmonic frequencies. LPC model remove the harmonic frequencies and produces the peaks of pulses also known as resonance frequency. The spectrum has been calculated according to equation (3.5).

$$S[n] = \sum_{k=1}^{p} a_k \, s\,[n-k] - G_u[n] \qquad (3.5)$$

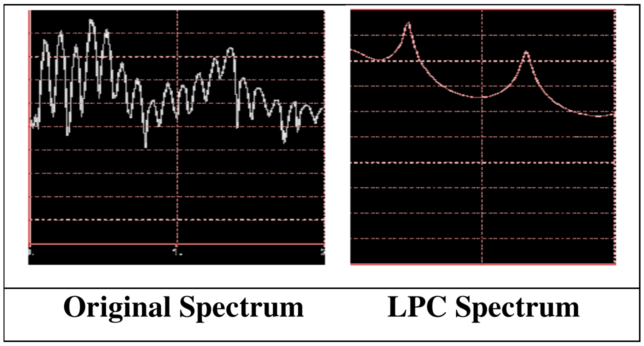
Where s[n] is speech signal. s [n-k] are past samples multiplied with a constant a_k also known as vocal tract. Where $G_u[n]$ is excitation. In LPC we are only looking at the vocal tract resonance. Pulse excitation $G_u[n]$ must be equal to zero in order to remove the interference of excitation between the pulses to predict the vocal tract resonance signal. Constant a_k is changed randomly until excitation $G_u[n]$ becomes equal to zero. In order to find out the excitation in the signal, past samples has been used.

Past samples depend on the LPC model. 12 fold LPC model has been used. It means we need 12 past samples in order to find out the excitation (error). Current sample is subtracted from past samples to get an error. Error is basically equal to G_u [u] or excitation. To minimize error coefficient a_k is altered until error equals to zero. After getting LPC residual signals autocorrelation function has been used in order to get smoother signal. LP residual signal has been used to estimate the pitch period through autocorrelation.

$$\emptyset(k) = \frac{1}{N} \sum_{n=0}^{N-1} s[n]s[n+k] \qquad (3.6)$$

Where $\emptyset(k)$ is pitch period. S[n] is current signal and S [n+k] is shifted signal. If we multiply two same values and add them, they give the large value.

Similarly in auto correlation it estimates the two correlated highest peaks in the pitch period using given equation and remove interference of all other harmonic frequencies or peaks. After successful implementation of this equation we will get smooth spectrum which is shown in figure 3.7.



Original Spectrum **LPC Spectrum**

Figure3.7: Comparison of Original and LPC Speech spectrum.

After getting smooth spectrum speech segmentation has been performed. Approach which is use for voicing in LPC model consists of two steps. Firstly, amplitude of the signal is calculated (also known as signal energy). If the amplitude is large then signal is determined as speech signal. For the classification between voiced and unvoiced segments threshold values are predefined for both types of sound.

Final determination of voice and unvoiced signal is based on counting the number of times waveform crosses from horizontal axis. These values are compared with the normal range of voiced and unvoiced sounds. This counting is also known as zero crossing rates (ZCR).

3.5 Acoustic Feature Extraction

Feature extraction has been performed to analyze the un-stationary behavior of the audio samples of speech signal in order to detect the healthy and impaired speech voice. It is also helpful to find out the relevance of the acoustic features with the pathological voices in case of HKD in Parkinson disease. For this purpose different acoustic features have been extracted. Three acoustic features have been extracted after applying threshold based speech segmentation which are Signal energy, Spectral centroid, and Zero crossing rate. Other four features have been extracted using LPC based model using residual frequency.

3.5.1 Feature Extraction from Speech Segments

Three acoustic features extracted from speech segments are energy entropy, spectral centroid, and zero crossing. These features are basically used to analyze the variation in harmonic frequencies. Variation in harmonic frequencies is basically irregular vibration of vocal folds due to un-periodic flow of air through lungs.

3.5.2 Energy Entropy

This is a time domain audio feature. To compute sudden changes in the energy, frames are divided in K sub-frames of fixed duration. For each sub-frame normalized energy is calculated and divided with the total frame energy using equation (3.7) [23].

$$e_j^2 = \frac{E_{subFrame_j}}{E_{shortFrame_i}} \tag{3.7}$$

Sum of all sub-frames normalized energy is energy entropy of that frame. EE for particular frame is computed using given equation (3.8).

$$H(i) = -\sum_{j=1}^{K} e_j^2 . \log(e_j^2) \tag{3.8}$$

$H(i)$ is EE of ith frame. Where $e_j^2$is normalized energy of sub frames. In the vocal fold related voice impairment, irregular distribution in the spectrum of speech signal, un-periodic flow of air through lungs, decrease the intensity of speech waveform.

Energy distribution in sub-bands in pathological speech shows high fluctuations compare to normal speech. Energy entropy has been used to evaluate irregularities in these sub-bands. Value of energy entropy is low if there are more irregularities in the energy distribution in these sub bands.

3.5.3 Spectral Centroid

Auditory feature related to shape of spectrum is brightness, which is often measured by the spectral centroid [6]. For example vowel sounds 'ee' is brighter than 'oo'. In HKD brightness of voice is affected because of irregular vibration of vocal folds. Spectral centroid is the average frequency for a given sub bands or harmonic frequencies. Harmonic frequencies produced by vocal source after the excitation of un-periodic flow of air through lungs affect the center of gravity of spectrum. Centroid is computed using equation 3.9.

$$C_i = \frac{\sum_{k=1}^{N}(k+1)Xi(k)}{\sum_{k=1}^{N}Xi(k)} \tag{3.9}$$

Spectral centroid base on the harmonic frequencies where k is the number of harmonics frequencies where C_i, is the average of harmonic frequencies at the time t (i). x_i **(k)** K=1……N is the Fourier Transform (DFT) of harmonic frequency on ith time.

3.5.4 Zero Crossing Rate (ZCR)

This is also a time domain audio feature. ZCR is basically the rate of change in the signal where the signal changes from positive to negative or back to its position; at that time signal have zero value. A ZCR is said to have occurred in a signal when its waveform crosses the time axis or changes its algebraic sign.

$$Z_j = \frac{1}{2S}\sum_{i=1\ldots S}\|sgn(x_i) - sgn(x_{i-1})\| \tag{3.10}$$

Where x_i is the discrete point value of the ith frame. Where Sgn(.) is the sign function:

$$Sgn[x_i(n)] = \begin{cases} 1 & x_i(n) \geq 0 \\ -1 & x_i(n) < 0 \end{cases} \qquad (3.11)$$

Voice speech is produced because of excitation of vocal tract by the periodic flow of air at the glottis. Usually it shows a low zcr count in the case of healthy voice and high zcr in the case of the impaired voice. Because in the healthy voice excitation of vocal tract produces periodic flow of the air but in case of impaired voice, un-periodic flow of air causes the high zcr.

3.5.5 Pitch Period Estimation using LPC Model

Pitch period is time required to one wave cycle to completely pass a fix portion. For speech signals, the pitch period is a thought of as the period of the vocal cord vibration that occurs during the production of voiced speech. Pitch period estimation has been performed using autocorrelation of residual signal. In order to achieve the highest level of performance, only positive going peaks have been estimated. To estimate the positive peaks in pitch period Peak Picker (PP) algorithm has been used. PP operates on period by period. If the algorithm has succeeded then there is only one peak left in one period [24]. Positive peaks are basically Time of occurrence of vowels (Tx). Tx has been used to calculate the standard statistics of voice quality that are mean fundamental frequency (meanf0), Jitter relative average perturbation (RAP), Jitter pitch perturbation quotient (PPQ) and shimmer.

3.5.5.1 Jitter Measurement

Fundamental frequency is determined physiologically by the number of cycles the vocal fold vibrates in one second. Jitter is used to find out the cycle to cycle variation in fundamental frequency. Cycle-to-cycle jitter is the change in a clock’s output transition from its corresponding position in the previous cycle shown in Figure 3.8. Jitter is variability in f0 and it is affected because of lack of control of vocal fold vibration in HKD [8].

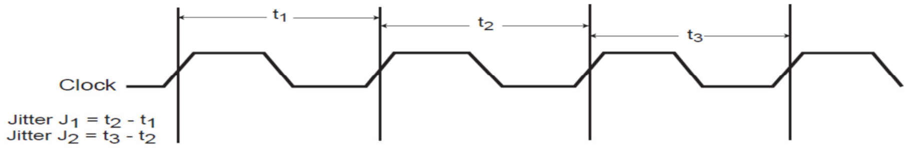


Figure3.8: Cycle-to-Cycle Jitter.

Jitter (RAP): RAP stand for relative average perturbation. Perturbation mean disturbance of motion. It is the absolute average difference between a period and its neighbors divided by the average difference between periods [8].

$$Jitter(RAP) = \frac{\left(\frac{1}{N}-2\right)\sum_{i=2}^{N-1} |T_i-\left(\frac{(T_i+T_{i-1}+T_{i+1})}{3}\right)}{\frac{1}{N}\sum_{i=1}^{N} T_i} \quad (3.12)$$

Where T_i is time period value of the i_{th} window and N is the number of voiced frames. Variation in the fundamental frequency in the healthy voice is less than as compare to the impaired voice in case of HKD.

Jitter (PPQ): stands for point period perturbation. It is the average absolute difference between a period and the average of it and its five neighbors divided by the average period [8].

$$jitter(PPQ) = \frac{\left(\frac{1}{N}-4\right)\sum_{i=3}^{N-2}\left|T_i-\left(\frac{(T_i+T_{i-2}+T_{i-1}+T_{i+1}+T_{i+2})}{N}\right)\right|}{\frac{1}{N}\sum_{i=1}^{N} T_i} \quad (3.13)$$

3.5.5.2 Shimmer Measurement

Shimmer is amplitude variation in fundamental frequency. Vocal intensity is related to sub glottis pressure of the air column, which, in turn, depends on other factors such as amplitude of vibration and tension of the vocal folds.

Shimmer is affected mainly because of the reduction in this tension and mass lesions in the vocal folds [8]. For healthy voice, the variation in the amplitude and frequency will be low as compare to impaired voice.

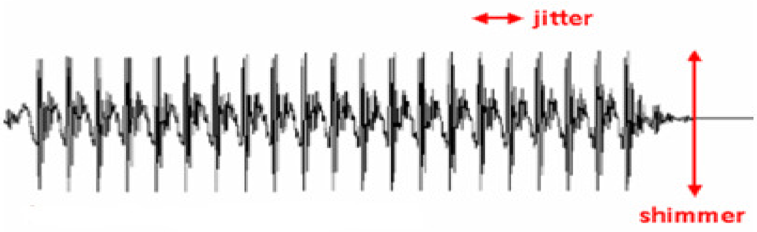


Figure3.9: Shimmer.

Shimmer (APQ): This is the five-point Amplitude Perturbation Quotient, the average absolute difference between the amplitude of a period and the average of the amplitudes of it and its five closest neighbors, divided by the average amplitude [8].

$$Shimmer(APQ5) = \frac{\frac{1}{N}-4\sum_{i=3}^{N-2}\left|A_i-\frac{(A_{i-2}+A_{i-2}+A_{i+1}+A_{i+2})}{5}\right|}{\frac{1}{N}\sum_{i=1}^{N}A_i} \qquad (3.14)$$

Where A_i is the peak amplitude value of the i_{th} window and N is the number of voiced frames.

3.5.5.3 Mean Fundamental Frequency (Meanf0)

Fundamental frequency is consisting of cycles that which vocal folds produce in one second. Meanf0 is basically mean of these cycles. PD is a progressive disease. F0 instability increases with the disease. In many people PWP opening the mouth wider can increase the loudness of the voice. This directly affects the f0. Similarly uncontrollable movement of vocal fold vibration cause the non-periodic cycles of fundamental frequency. F0 is basically number of cycles produce in one second. After pitch period estimation using peak picker algorithm peaks of each cycle are achieved. Meanf0 is the mean of all the peaks of these cycles. In impaired voice especially voice of PWP meanf0 will be high because of high fluctuation in the peak of the cycles.

3.6 Feature selection and Classification

Finally classification has been performed using open source tool known as weka. Naïve Bayes Classifier has been used for classification. Before the classification, features selection has been performed using different attributes evaluation techniques in order to select the important features. Chi-square, info-gain and Gain info ratio has been performed in order to evaluate the important features. For this purpose weka tool is used where all these methods are built-in.

3.6.1 Chi-squared Attribute Evaluation

Chi-square is non-parametric technique used to check the difference between the theoretical expected values and actual values. Because it is non-parametric test it use ordinal data for evaluation instead of mean and variances.

Feature	Speech test-1	Speech test-2	Speech test-3
Jitter(PPQ)	2.29	0.6	0.9
Spectral Centroid	2	1.2	1.5
Mean F0	1.04	1.5	0.3
Energy Entropy	0.3	0.3	0.6
Zero Crossing	0.3	0.3	0.3
Shimmer(APQ)	0.3	0.6	0.9
Jitter(RAP)	0.3	0.3	0.6

Table3.1: Chi-square evaluation for all Speech tests.

The above result Table 3.1 shows the ranking of each feature through Chi-square attribute evaluation. Results shows all acoustic features pass attribute evaluation test in all speech tests.

3.6.2 Info Gain Attribute Evaluation

Similarly info gain is similar to chi-square as mention before. The result shows the evaluation of features.

Feature	Speech test-1	Speech test-2	Speech test-3
Jitter(PPQ)	2.29	0.6	0.6
Spectral Centroid	2	1.2	1.5
Mean F0	1.04	1.5	0.3
Energy Entropy	0.3	0.3	0.6
Zero Crossing	0.3	0.3	0.3
Shimmer(APQ)	0.3	0.6	1.2
Jitter(RAP)	0.3	0.3	0.3

Table3.2: Info Gain Attribute Evaluation for speech test1.

Info gain attribute evaluation is working same like chi-square attribute evaluation. It also shows almost same result as in chi-square. Highest ranked feature are Jitter (PPQ), same like chi-square. All acoustic features pass info gain attribute evaluation test.

3.6.3 Gain Ratio Attribute Evaluation

Feature	Speech test-1	Speech test-2	Speech test-3
Jitter(PPQ)	2.29	0.6	0.6
Spectral Centroid	2	1.2	1.5
Mean F0	1.04	1.5	0.3
Energy Entropy	0.3	0.3	0.6
Zero Crossing	0.3	0.3	0.3
Shimmer(APQ)	0.3	0.6	1.2
Jitter(RAP)	0.3	0.3	0.6

Table3.3: Gain Ratio Attribute Evaluation for speech test1.

All the above attribute evaluation methods shows almost same ranking of acoustic features. All the acoustic features pass the attribute evaluation test which depicts that all features are correlated with pathological voices and all are important for classification.

Finally classification has been performed using with open source tool known as weka. There are many classifiers which we can use for different problems. Here NB classifier with the 10 fold cross validation has been used to classify the extracted features data with the clinical rated data. Results are discussed in the next section.

Chapter 4 Results and Analysis

4.1 Classification Results

The 10 fold cross validation using Naïve Bayes algorithm produces results from all speech tests which are given below. True positive rate is number of those people who are correctly diagnosed as sick. True Negative rate is number of those people who are correctly identified as healthy. False positive rate is the number of healthy people diagnosed as sick. False negative rate is the sick people identify as healthy. Sensitivity is the percentage of sick people who are correctly identified as having the HKD. Specificity the percentage of healthy people who are correctly identified as not having the HKD. These two parameters have been used to estimate the performance of classifier. ROC is a graphical plot of true positive and false positive rates.

	Positive	**Negative**
Positive	8(TP)	12(FN)
Negative	7(FP)	43(TN)
Sensitivity	TP/(TP+FN) = 8/(8+12)	40%
Specificity	TN/(TN+FP)=43/(43+7)	86%

Table4.1: Results obtained from NB classifier for speech test-1 (70 audio samples).

	Positive	**Negative**
Positive	16(TP)	6(FN)
Negative	8(FP)	42(TN)
Sensitivity	TP/(TP+FN) = 16/(16+6)	72%
Specificity	TN/(TN+FP)=42/(42+8)	84%

Table4.2: Results obtained from NB Classifier for speech test-2 (72 audio samples).

	Positive	**Negative**
Positive	2(TP)	17(FN)
Negative	8(FP)	43(TN)
Sensitivity	TP/(TP+FN) = 2/(2+17)	10.5%
Specificity	TN/(TN+FP)=43/(43+8)	84%

Table4.3: Results obtained from NB classifiers for speech test-3(70 audio samples).

	Sensitivity	**Specificity**	**Overall Accuracy**	**ROC Area**
Specchtest1(70 audio samples)	40%	86%	72%	0.74
Speechtest2 (72 audio samples)	72%	84%	80%	0.74
Speechtest3 (70 audio samples)	10.5%	84%	64%	0.45

Table4.4: NB Classifier performance parameters.

We can see the overall classification results in all speech test are good enough for practical implementation. Sensitivity refers to the measure of proportion of dyarsthric audio samples which are correctly identified. Sensitivity is low as compare to specificity. Low sensitivity directs to need the improvement in this method to make it high sensitive to diagnose the HKD. Specificity is measure of proportion of healthy audio samples which are correctly identified. Fluctuation in the overall results in all speech tests is due to non-stationary behavior of the signal. In real time environment speech signal is not standard. No matter what the speaking task is used. Speech properties will be different because of environmental conditions and speaking ability. ROC values in speech test1 and speech test2 are good which depicts that this methodology is feasible for practical implementation. Speech test3 has less ROC value because of more speech impairment in speech test3 audio samples as compare to the other speech tests which is discussed in detail in next section.

4.2 ROC Graph

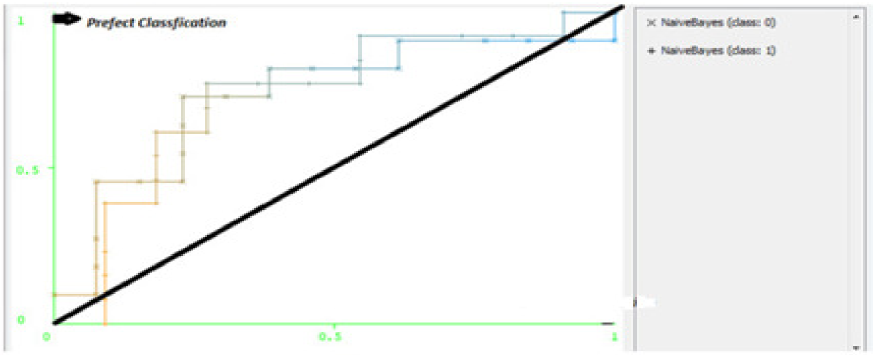


Figure4.1: Roc Graph; represent the classification performance of speech test-1.

ROC is graphical plot of true positive and false positive. True Positive also called Sensitivity and false positives also known as 1-specificity. The number of instances lies in the region of true positives near the value 0 and 1 on y-axis is considered good classification. The above graph of speech test-1 is near to region of correct classification with ROC value 0.74. This value depicts that this method is feasible for practical implementation with some improvement.

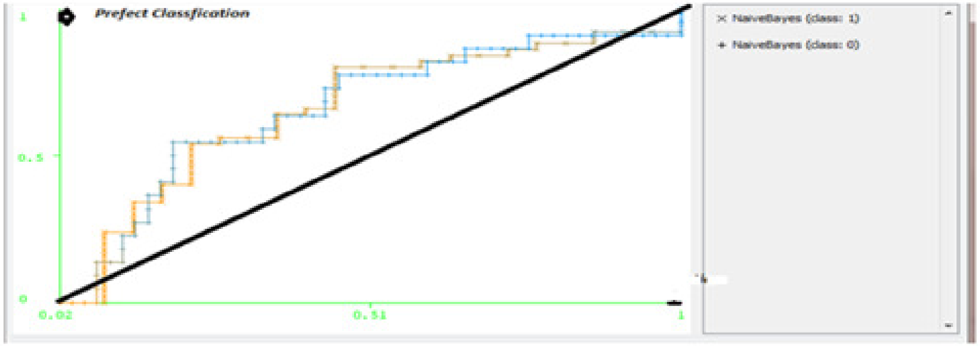


Figure4.2: Roc Graph; represent the classification performance for speech test-2.

ROC graph for speech test-2 is towards the prefect classification region with ROC value 0.74.

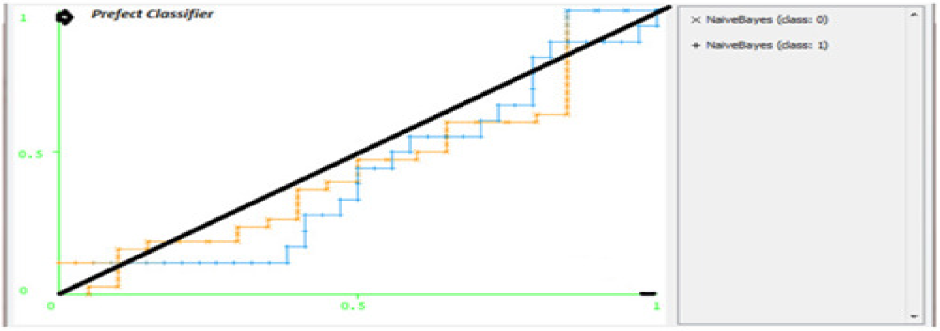


Figure4.3: ROC graph; represent the classification performance for speech test3.

ROC graph for speech test-3 is not in the prefect classification region with the value 0.47. Variation in the classification results in all speech tests has many reasons which are discussed in next section. One reason is un-stationary behavior of the signal. Secondly large number of fricatives, consonants and vowels present in speech test 3. This indicates the large impairment in speech test3 audio signals.

4.3 Acoustic features Correlation with voice pathology

Correlation between extracted values and the targets values has been found using MYSTAT statistical tool. There are two types or directions of correlation, i.e. positive correlation and negative correlation. The negative correlation shows that one variable increases and other variable decreases. Similarly positive correlation shows that both variables decreases or increases. Negative sign does not indicate anything about strength. It only indicates that the correlation is negative in direction.

	Energy Entropy	**Zero Crossing Rate**	**Spectral Centroid**	**Mean F0**	**Jitter(RAP)**	**Jitter(PPQ)**	**Shimmer**
Correlation	0.592	0.565	0.551	-.130	0.371	0.420	0.085

Table4.5: Acoustic Features Correlation with targets values for speech test1.

The above result shows that Energy entropy, Zero Crossing rate, and Spectral Centroid, are highly correlated with speech pathology in speech-test1. Similarly Jitter (PPQ) and Jitter (RAP) are in the region of moderate correlation with positive direction. Meanf0 is in the region of low correlation with negative direction. Overall correlation result shows that all features are correlated with voice pathology in speech test1.

	Energy Entropy	**Zero Crossing Rate**	**Spectral Centroid**	**Mean F0**	**Jitter(RAP)**	**Jitter(PPQ)**	**Shimmer**
Correlation	0.780	0.179	0.032	-.385	0.410	0.470	0.179

Table4.6: Acoustic Features Correlation with targets values for speech test2.

Energy entropy shows very high correlation with voice pathology. Similarly meanf0, jitter (RAP), and Jitter (PPQ) are in the region of moderate correlation. Spectral centroid and shimmer shows low correlation with speech pathology in speech test2.

	Energy Entropy	**Zero Crossing Rate**	**Spectral Centroid**	**Mean F0**	**Jitter(RAP)**	**Jitter(PPQ)**	**Shimmer**
Correlation	0.659	0.761	0.646	-.284	-0.420	-0.462	-0.540

Table4.7: Acoustic Features Correlation with targets values for speech test3.

Zero crossing rates shows very high correlation with voice pathology in speech test3. Similarly energy entropy, spectral centroid shows the high correlation with speech pathology. Jitter (RAP), Jitter (PPQ) and Shimmer are in the region of moderate correlation.

4.3 Student t test for speech test1 acoustic features

The acoustic features were tested using a one-tailed dependent sample's t-test with the type one error rate. An unpaired sample's t-test was used for these acoustic features to analyze the significant difference between two independent classes 0 and 1. The more the error bars are separated, the greater is the chance that the population is different [17]. In t test mean and standard deviation values has been estimated in order to find out the difference and variance in the population. As mentioned before if the two error bars are from the different population then the mean may not fall in the region of overlap. In this case, we conclude that the population is different. The most common way to describe the range of variation is standard deviation. The standard deviation is simply the square root of the variance. STD is calculating the variability, i.e. noise which may make it difficult to see group differences.

	Mean Class 0	**Mean Class 1**	**SD Class0**	**SD Class1**
Energy Entropy	4.15	3.02	5.63	0.922
Spectral Centroid	9.002	6.708	7.769	2.221
Zero Crossing Rate	3.790	2.592	4.576	1.115
Meanf0	171	164	56.4	39.0
Jitter(RAP)	12.3	15.0	4.75	7.29
Jitter(PPQ)	10.8	13.9	5.00	7.05
Shimmer	17.5	17.9	4.68	5.19

Table4.8: Mean and STD of Acoustic features for speech test1.

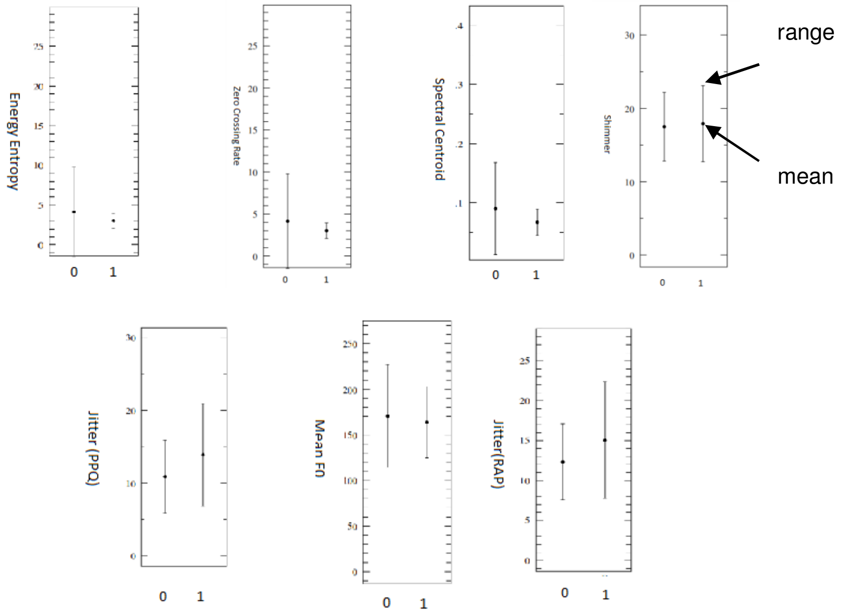


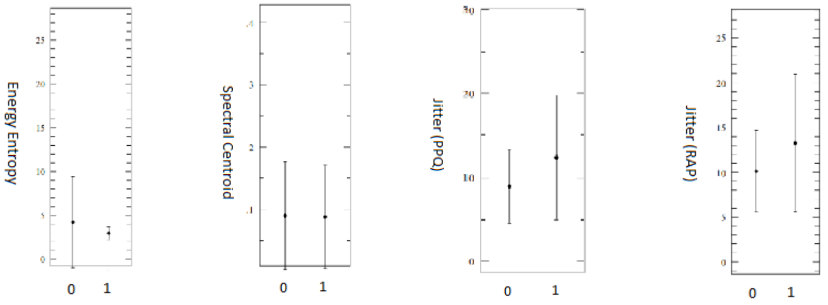
Figure4.4: Error Bars for speech test-1; overlapping b/w bars shows the significant difference between class 0 and 1 samples for each feature.

Overlapping between the error bars shows the significant difference between two populations. In the above given t-test for acoustic features overlapping is observed in all features. Energy Entropy (EE), Spectral centroid (SC), and Zero Crossing Rate (ZCR) shows that mean values are not in the region of overlap which also depicts that these features are highly correlated with voice pathology. Standard Deviation shows variance across each group. We can clearly see that EE, SC and ZCR have low variance in both groups as compared to other acoustic features.

4.4 Student's t test for speech test-2acoustic features

	Mean Class 0	Mean Class 1	SD Class0	SD Class1
Energy Entropy	4.23	2.95	5.19	0.743
Spectral Centroid	9.00	8.788	8.691	8.295
Zero Crossing Rate	6.572	8.221	0.126	0.168
Meanf0	166	143	56.5	58.0
Jitter(RAP)	10.1	13.3	4.56	7.68
Jitter(PPQ)	8.92	12.4	4.41	7.39
Shimmer	15.9	17.1	6.15	6.96

Table4.9: Mean and STD of Acoustic features for speech test2.



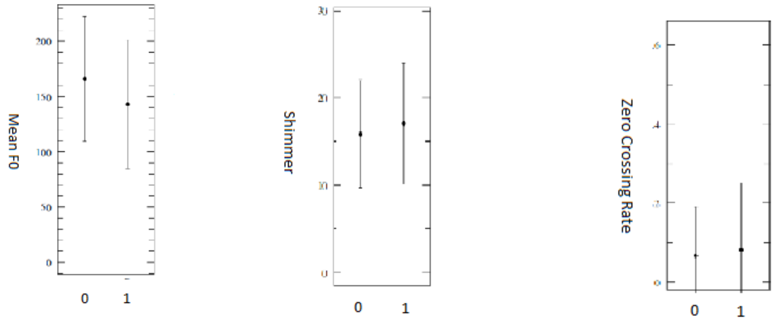


Figure4.5: Error Bars for speech test-2; overlapping b/w bars shows the significant difference between class 0 and 1 samples for each feature.

In the above given t-test for acoustic features overlapping is observed in all features. We can clearly observe that energy entropy has small overlapping between two groups, which shows that there is large significant difference between two groups as compared to other features.

4.5 Student’s t test for Speech test-3 acoustic features

	Mean Class 0	Mean Class 1	SD Class0	SD Class1
Energy Entropy	3.36	5.73	1.43	8.76
Spectral Centroid	5.344	9.016	2.638	0.103
Zero Crossing Rate	2.656	7.186	1.492	0.149
Meanf0	167	151	37.5	64.8
Jitter(RAP)	16.1	12.4	7.02	7.07
Jitter(PPQ)	14.8	11.1	6.66	6.39
Shimmer	18.5	15.4	4.29	6.43

Table4.10: Mean and STD of Acoustic features for speech test3.

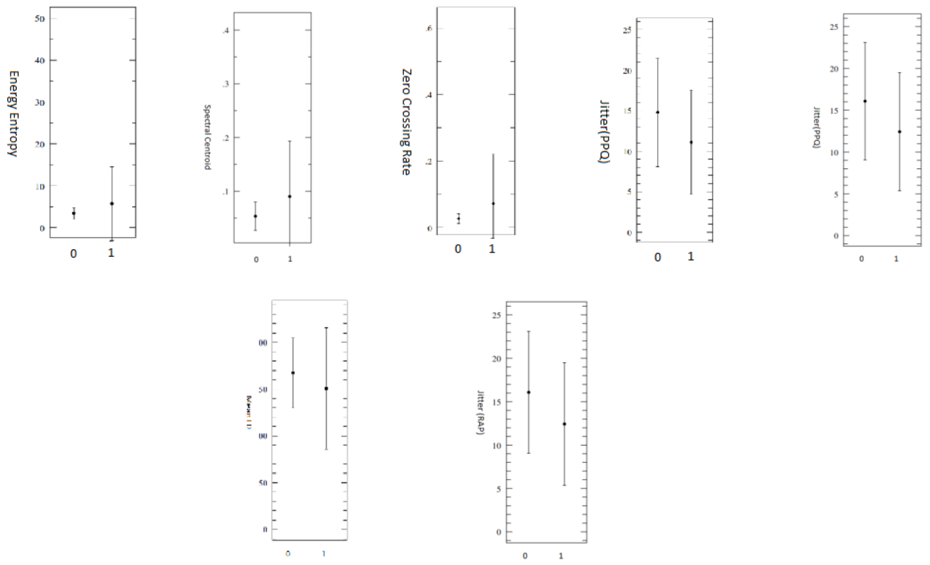


Figure4.6: Error Bars for speech test-3; overlapping b/w bars shows the significant difference between class 0 and 1 for each feature.

Overlapping has been observed in t test for all acoustic features. We can clearly observe that zero crossing has small overlapping between two groups which shows that there is large significant difference between two groups as compared to other features. Minimum overlapping has been observed across ZCR. Standard deviation also shows minimum variance in both populations as compared to other features.

4.6 Discussion

Above results shows that overall classification results direct the relevance of all acoustic features with pathological voices in Hypokinetic Dysthria. Zero Crossing Rate, Spectral Centroid and Energy Entropy showed highly correlated features with voice pathology. Jitter (RAP), Jitter (PPQ), shimmer and meanf0 showed low correlation with voice pathology. Similarly low sensitivity has been observed in all speech tests results. It indicated the sensitivity affected due to low correlation of the acoustic features (Jitter (RAP), Jitter (PPQ), shimmer and meanf0).

Zero crossing rate is highly correlated in speech test1 (p= 0.565) and in speech test3 (p= 0.761) with voice pathology. Due to voiced spectrums in speech signal the ZCR is low in normal speech. In contrast, ZCR is higher in abnormal speech due to unvoiced spectrums. The speech signal holds most of its energy in voiced signal at low frequencies. For unvoiced sound, noise excitation takes place at higher frequencies due to short length of vocal tract. Consequently a high and a low ZCR transmits to unvoiced and voiced speech respectively. It also directed the stress induced in speech test1 and speech test3 due to presence of fricatives, which causes difficulties to pronounce.

Spectral Centroid is highly correlated in speech test1 (p= 0.551) and in speech test3 (p= 0.646) with voice pathology. Brightness is also called spectral centroid. It has been observed in normal speech spectral centroid is high because of high dominant frequency in each spectrum. Speech spectrum has strong spectral peaks because these peaks are generally not affected by noise. In contrast, spectral centroid is lower in abnormal speech; due to noisy sounds. Noise spectrum does not have strong spectral peaks. It also indicates the presence of a large number of fricatives in speech test1 and speech test3. Large number of fricatives causes' speech impairments resulted in weak spectral peaks.

Energy Entropy also showed high correlation with voice pathology in speech test1 (p= 0.551) and in speech test3 (p= 0.646). Energy entropy has been used to measure the sudden changes in the energy level of an audio signal. It has been found in normal speech spectrum's energy distribution is constant because of constant positive stress factors in the speech spectrum. No sudden change in energy level has been observed in normal speech. In pathological voices, energy distribution is not constant throughout the speech spectrum due to negative stress factors in the speech spectrum. One other reason is that the fluctuation in the energy is distanced between the mouth and the phone during the collection of speech data. Large distance of mouth from phone also decreases the energy of the speech spectrum.

Less sensitivity (speech test1 40%, speech test3 10.5%) has been noticed in speech test 1&3. Similarly low correlation has been noticed in the acoustic features, jitter, shimmer and meanf0 as compare to the other acoustical features with voice pathology. It indicates low correlation of these acoustic features cause the less sensitivity. Ignorance of nasal sounds in LPC model causes the less correlations of these acoustical features with voice pathology. LPC model is based on the source filter model, but sound will not always produce according to LPC model. Nasal cavity which is part of a source filter model has been ignored in LPC model. Nasal sounds may lead the incorrect speech segmentation. Reason to ignore nasal cavity in LPC model is that for nasal sounds, the nose cavity forms a side branch. Nasal sounds require more complex and different algorithms, which was impossible to embed in LPC model.

Chapter 5 Conclusions and Future Work

The main aim of this thesis was to categorize healthy and impaired speech recordings to assess speech impairment in case of HKD using the acoustic features. For this purpose, three types of running speech test have been used, which consisted of both normal and pathological speech samples. Preprocessing on audio speech signal was performed using Band pass filter. Speech segmentation was performed using linear predictive coding and silent removal algorithm. Feature extraction was performed on the basis of short time energy (STE) and spectral centroid (SC). Some features were used to analyze peak to peak variation in fundamental frequency using linear predictive coding (LPC). Features which were used for this purpose was Jitter (RAP), Jitter (PPQ), Shimmer and meanf0. Other features Energy entropy, Zero crossing rate, and Spectral centroid are used to analyze the fluctuation in the harmonic frequencies. For this purpose speech segments are used after speech segmentation using silent removal algorithm.

Chi-square, Info-gain and Gain-ratio attribute evaluation test was performed to select the important features for classification. All the acoustic features passed attribute selection tests. Discrete values of these features were classified using 10 fold cross validation and Naïve Bayes classifier. Results achieved from the classification directs the correlation of the acoustic features with speech pathology in Parkinson disease.

T-test was performed between healthy and impaired groups for each of the features. In all the features, overlapping in both classes are observed for 95% confidence interval. Correlation coefficient was calculated to find out the correlation of features with speech pathology. All acoustic features somehow less or more shows correlation with the clinical ratings in running speech tests. Correlation coefficient analysis showed Energy Entropy (EE), Spectral Centroid and Zero Crossing Rate (ZCR) to be significantly correlated as compared to other acoustic features.

The ROC graph for speech test1& speech test2 was near to the region of prefect classification. This depicts that this method is feasible for practical implementation with some improvement. To make it more effective method some future work has been proposed.

Nasal sounds have been ignored in LPC model which are very common in HKD affected speech. Nasal sounds may lead to incorrect speech segmentation. In future cepstrum base methods can be used to overcome this issue. Cepstrum based methods can be used to deal with nasalic sounds. Formant analysis may be performed to investigate speech impairment in HKD. Future, research will focused on the classification of speech impairment using UNIFIED PARKINSON'S DISEASE RATING SCALE (UPDRS). This will be helpful for the self-assessment of Patient with Parkinson (PWP) using the mobile device assessment tool.

References

[1] Jennifer Cambur, Philadelphia, Pennsylvania, Stefanie Countryman, Janet Schwantz, MS, **"Parkinson's Disease: Speaking Out"** The National Parkinson Foundation 2000

[2] Taha khan and jerker westin **"Methods for detection of speech impairment using mobile devices"** Academy of industry and society, Computer Engineering, Dalarna University Received: 15 February 2011; Revised: 22 March, 2011; Accepted: 10 April, 2011

[3] Eric J. Hunter, Jennifer Spielman, Lorraine O. Ramig **"Suitability of dysphonia measurements for telemonitoring of Parkinson's disease"** Biomedical Engineering IEEE Transactions on Volume:56, Issue:4

[4] LotfiSalhi, TalbiMourad, and AdneneCherif, **"Voice Disorders Identification Using Multilayer Neural Network",**Signal Processing Laboratory Sciences Faculty of Tunis, University Tunis ElManar, Tunisia; Accepted December 28, 2008

[5] "Band Pass Filter" Lessons in the electrical circuits Volume 2-AC Sixth Edition, last update July 25, 2007

[6] "Linear Predictive Coding" Jeremy Bradbury Linear Predictive Coding December 5, 2000

[7] Meysam Asgari and Izhak Shafran **"Predicting Severity of Parkinson's Disease from Speech"** 32nd Annual International Conference of the IEEE EMBS Buenos Aires, Argentina, August 31 - September 4, 2010

[8] Farrus, M.; Hernando, J **"Using jitter and shimmer in speaker verification Signal Processing", Publish in IET** Volume: 3, Issue: 4 Received on 7th Agust 2008, Revised on 28th November 2008

[9] Eric J. Hunter, Jennifer Spielman, Lorraine O. Ramig **"Suitability of dysphonia Measurements for Telemonitoring of Parkinson's disease"** Biomedical Engineering IEEE Transactions on Volume: 56, Issue: 4

[10] Diana Van LanckerSidtis, Ph.D., "**Fundamental Frequency (F0) Measures Comparing Speech Tasks in Aphasia and Parkinson Disease"** Journal of Medical Speech Language Pathology, Volume 12, Number 4, pp. 207-212

[11] AthanasiosTsanas, Max A. Little, Patrick E. McSharry, Senior Member, **"Novel Speech signal processing algorithms for high-accuracy classification of Parkinson's Disease"** IEEE, TBME-00887-2011.R1

[12] Mattias Nilsson **"Entropy and Speech"** Thesis for the degree of Doctor of Philosophy Sound and Image Processing Laboratory School of Electrical Engineering KTH (Royal Institute of Technology) Stockholm 2006

[13] G. Parthiban, A. Rajesh, S.K.Srivatsa **" Diagnosis of Heart Disease for Diabetic Patients using Naive Bayes Method"** International Journal of Computer Applications (0975-8887) Volume 24- No.3, June 2011

[14] JOHN R. LANZANTE**, "A Cautionary Note on the Use of Error Bars Manuscript"** Received 18 May 2004, in final form 3 March 2005)

[15] "Sensitivity and specificity" http://en.wikipedia.org/wiki/Sensitivity_and_specificity

[16] Tom Fawcett **"ROC Graphs: Notes and Practical Considerations for Researchers",** HP Laboratories, MS 1143, 1501 Page Mill Road, Palo Alto, CA 94304 March 16, 2004

[17] "Student t test" http://www.fgse.nova.edu/edl/secure/stats/lesson7.htm

[18] Jasmina Novakovic **"The Impact of Feature Selection on the Accuracy of Bayes Classifier"** 18th Telecommunications forum TELFOR 201, Serbia, Belgrade, November 23-25, 2010.

[19] "Jacob Cohen Correlation**"** http://www.sportsci.org/resource/stats/effectmag.html

[20] Louis Guttman, 1944**, "A Basis for Scaling Qualitative Data",** American Sociological Review 9:139-150

[21] "Chi Square" http://www.fgse.nova.edu/edl/secure/stats/lesson7.htm

[22] Theodoros Giannakopoulos **"A method for silence removal and segmentation of speech signals"** Computational Intelligence Laboratory (CIL), Insititute of Informatics and Telecommunactions, NSCR Demokritos, Greece 2009

[23] Theodoros D. Giannakopoulos **"Study and application of acoustic information for the detection of harmful content, and fusion with visual information"** Department of Informatics and Telecommunications University of Athens, Greece 2009

[24] Ian Howard **"Speech Fundamental Period Estimation Using Pattern Classification"** Department of Phonetics and Linguistics University College London Post-Viva Edition 8/2/92

[25] Theodoros Giannakopoulos 1, DimitriosKosmopoulos 2,AndreasAristidou 1, and SergiosTheodoridis 1**"Violence Content Classification Using Audio Feature"**.Department of Informatics and Telecommunications University of Athens, Greece G. Antoniou et al. (Eds.): SETN 2006, LNAI 3955, pp. 502 – 507, 2006

[26] Carlos Matos, MS Lorraine Ramig, PhD Jennifer Spielman, MA CF-SLP John Bennett, PhD Angela Halpern MS, **"Speech Treatment for Parkinson's Disease",** University of Colorado Boulder, Campus Box 409, *Expert Rev.* Neurotherapeutics 8(2), 299–311 (2008)

[27] Jonathan P.Evans, Man-ni Chu, John A. D. Aston, Chao-yu Su, **"Linguistic and human effects on F0 in a tonal dialect of Qiang", a.Institute of Linguistics"**, This is a pre-publication version of: Evans, J. P., Chu M., Aston J. A. D., & Su C. (2010). Linguistic and human effects on F0 in a tonal dialect of Qiang. Phonetica, 67, 1-2.

[28] Richard B. Reilly, Rosalyn Moran, Peter Lacy, **"Voice Pathology Assessment based on a Dialogue System and Speech Analysis"**, Department of Electronic and Electrical Engineering, University College Dublin, Ireland St James's Hospital, Dublin 8, Ireland

[29] JasminaNovakovic **"The Impact of Feature Selection on the Accuracy of Bayes Classifier"** 18th Telecommunications forum TELFOR 201, Serbia, Belgrade, November 23-25, 2010.

[30] http://www.wisegeek.com/what-is-statistical-significance.htm