American Journal of Biomedical Science and Engineering
2015; 1(5): 71-81
Published online September 30, 2015 (http://www.aascit.org/journal/ajbse)

American Association for Science and Technology

Keywords
Neurological Disorder,
Voice,
MFCC,
SVM

Received: August 25, 2015
Revised: September 6, 2015
Accepted: September7, 2015

Neurological Disorder Detection Using Acoustic Features and SVM Classifier

Uma Rani K.[1], Mallikarjun S. Holi[2]

[1]Department of Biomedical Engineering and Research Centre, Bapuji Institute of Engineering and Technology, Davangere, Karnataka, India
[2]Department of Electronics and Instrumentation Engineering, University B. D. T. College of Engineering, Visvesvaraya Technological University, Davangere, Karnataka, India

Email address
uma_devoor@yahoo.com (Uma Rani K.), msholi@yahoo.com (Mallikarjun S. Holi)

Citation
Uma Rani K., Mallikarjun S. Holi. Neurological Disorder Detection Using Acoustic Features and SVM Classifier. *American Journal of Biomedical Science and Engineering*. Vol. 1, No. 5, 2015, pp. 71-81.

Abstract
In neurological disordered patients, the physiological substrates necessary for the speech production may be altered and hence the acoustic properties may also change. The measurable information in the acoustic output of individual patients may provide valuable clues for diagnosing certain neurological diseases, course of disease progression, assessing response to medical treatment, or a combination of these. The various acoustic features can be extracted in time domain, frequency domain, time-frequency domain and by non linear feature methods and these features can be used for disordered voice detection. In the present work time domain features like pitch variation, jitter, shimmer, harmonic to noise ratio (HNR) and frequency domain features like Mel Frequency Cepstral Coefficients (MFCCs) are extracted from normal and neurological disordered subject's voice signals. Both time domain and frequency domain features are given to a Support Vector Machine (SVM) classifier and the results are compared in detecting normal and neurological disordered subjects. It is observed that SVM classifier perform well for time domain features with a classification accuracy of 81.43% compared to the frequency domain features with classification accuracy of 71.43%.

1. Introduction

The speech production process is a complex system which involves coordination of numerous individual muscles, cranial and spinal nerves, cortical and subcortical neural areas. Generation of appropriate sounds is necessary to convey a message spoken by a speaker. When a speaker's respiration, phonation, articulation, resonance, and prosody are combined in a well-executed manner, then a meaningful speech message is obtained. However, there will be measurable changes in the acoustic output if there is any problem in these interdependent physiological systems, starting from the diaphragm to the cortex and to the outermost border of the lips. In neurological disordered patients, the physiological substrates necessary for the speech production may be altered and hence the acoustic properties may also change [1]. Measurable information in the acoustic output of individual patients may provide valuable clues for diagnosing certain diseases, course of disease progression, assessing response to medical treatment, or a combination of these. In previous studies[2], [3] it has been reported that in neurological disorders, such as Parkinson Disease (PD) approximately 70% - 90% of patient show some form of vocal impairment [3], [4] and this deficiency may also be one of the earliest indicators of the disease. Hence acoustical voice analyses and measurement methods might provide

useful biomarkers [4] for the diagnosis of such diseases in the early stage, possible remote monitoring of patients, and providing important feedback in voice treatment for clinicians or patients themselves [5]. Acoustic measurements can also improve the individual treatment and avoid inconvenience and cost of physical visits by the patient to the clinic. Moreover, voice recording and analysis is non-invasive, cost effective, and simple to perform [6].

The time domain, frequency domain, time-frequency domain and non linear feature extraction methods are becoming very popular in disordered voice detection. The time domain features like pitch variation, jitter, shimmer, harmonic to noise ratio (HNR) are widely used features in speech analysis and speech detection systems [7],[8],[9],[10].From past decade the frequency domain features Mel Frequency Cepstral Coefficients (MFCCs) are widely used in disordered voice detection systems [11],[12],[13],[14]. Hence a comparative study of both time domain features and the frequency domain features given to a Support Vector Machine (SVM) for identification of normal subjects and subjects with disordered voice affected by neurological disease has been considered in the present work.

2. Materials and Methods

2.1. Data Collection

The present work consists of 281 phonations of sustained vowel /ah/. Among them 175 phonations were collected from 49 male subjects (62.72 ± 8.0 yrs) and 25 female subjects(65.19 ± 8.8 yrs), who were found to be suffering from one or the other neurological disorder like PD, cerebellar demyelination and stroke. Remaining 106 phonations were from 56 normal subjects, who were selected among the age and gender-matched healthy persons who were not complaining of any voice problems. The data were collected from Outpatient Wing, Department of Neurology, J.S.S.Hospital, Mysuru after getting the consent from local ethical committee. Voice signals are recorded as per the standards through a microphone at a sampling frequency of 44,100 Hz using a 16-bit sound card in a laptop computer with a Pentium processor [15], [16]. The microphone to mouth distance was at 5 cm and the subjects were asked to phonate the vowels /ah/ for at least 3 sec at a comfortable level. Further, a steady portion of the signal of 2 sec duration was selected for the acoustic analysis. Figure 1 shows the typical recording for sustained phonation of normal and neurologically disordered subject (PD). All the recordings were done using the PRAAT software, in mono-channel mode and saved in WAVE format on the hard disk and acoustic analysis were done on these recordings [17].

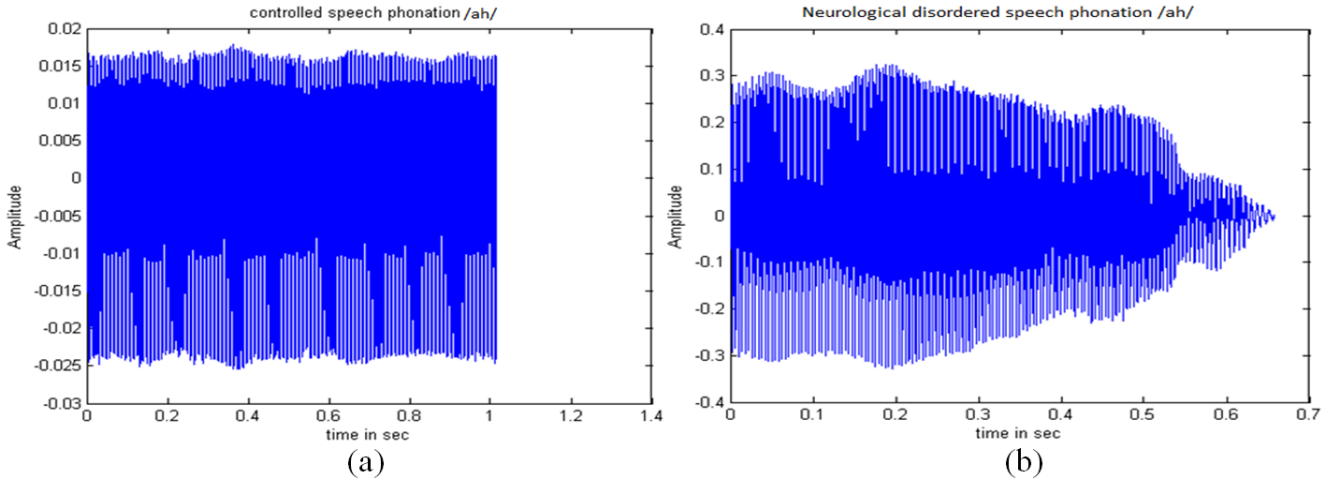

Fig. 1. *Sustained phonation /ah/ of (a) controlled subject (normal) and (b) neurological disordered subject (PD).*

2.2. Acoustic Parameter

2.2.1. Time Domain Features

The time domain features in this study include three measures of fundamental frequency, five measures on *jitter*, six measures on *shimmer*, and two measures on signal to noise ratios (*harmonics to noise ratio*) [7],[8],[9],[10]. All these measures were calculated using the PRAAT software after selecting a steady portion of 2 sec duration from the acquired voice sample. The voice/speech oscillation interval is called pitch period, which is the physiological determination of the number of cycles that the vocal folds vibrate in a second. Change in this pitch period is a common manifestation of vocal impairment due to incomplete vocal fold closure and also imbalanced vocal fold movement resulting in excessive breathiness (noise) and affecting the signal pattern severely. This imbalanced vocal fold movement also results in turbulent noise and the appearance of vortices in the airflow from the lungs as shown in Fig. 1. In general, people with voice disorders cannot elicit steady phonations [9].

Jitter and Shimmer Measures:

Jitter and shimmer are the common measures of prolonged sustained vowels. The values of these measures above a certain threshold are related to voice pathology, usually

perceived as breathy, rough or hoarse voices. Jitter refers to the variability of F0 the fundamental frequency, and it is affected due to the lack of control of the vocal fold vibration [7],[8],[9],[18],[19]. On the other hand, the air column pressure on sub-glottis is related as vocal intensity (shimmer), which in turn depends on factors like amplitude of vibration and tension of vocal folds[18]. Shimmer is affected mainly due to the reduction in tension or mass lesions in the vocal folds. These measures are also said to change with gender; for instance, F0 and amplitude instability increases in aged voice, resulting in greater jitter and shimmer values, leading to tremor and increased hoarseness [19].The jitter and Shimmer values are calculated as shown below:

Jitter (relative): Average absolute difference between consecutive periods, divided by the average period

$$jitt(\%) = \frac{\frac{1}{n-1}\sum_{i=n-1}^{1}|T_{i+1}-T_i|}{\frac{1}{n}\sum_{i=1}^{n}T_i} \quad (1)$$

Jitter (absolute): It is the cycle-to-cycle variation of fundamental frequency, that is, the average absolute difference between consecutive periods, given as

$$jitt(ab) = \frac{1}{n-1}\sum_{i=n-1}^{1}|T_{i+1} - T_i| \quad (2)$$

where T_i are the extracted F0 period lengths and n is the number of extracted F0 periods, as shown in Fig. 2.

Similarly the other Jitter measures, relative average perturbation (RAP) and the Jitter Five-point Period Perturbation Quotient(ppq5) are calculated as shown in Table I.

Shimmer (absolute): Variability of the peak-to-peak amplitude in decibels, that is, the average absolute base-10 logarithm of the difference between the amplitudes of consecutive periods, multiplied by 20

$$shimm(dB) = \frac{1}{n-1}\sum_{i=n-1}^{n-1} 20 \times log\frac{A_i}{A_{i+1}} \quad (3)$$

where A_i are the extracted peak-to-peak amplitude data and n is the number of extracted fundamental frequency periods, as shown in Fig. 3.

Shimmer (relative): average absolute difference between the amplitudes of consecutive periods, divided by the average amplitude

$$shimm = \frac{\frac{1}{n-1}\sum_{i=n-1}^{1}|A_{i+1}-A_i|}{\frac{1}{n}\sum_{i=1}^{n}A_i} \quad (4)$$

The other Shimmer calculations along with the ratios of harmonics and noise are summarized in Table I.

A total of 16 acoustic features were extracted from the voice samples and are summarized in Table I.

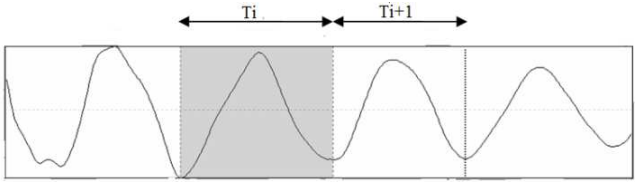

Fig. 2. *Jitter measurement for four F0 Periods.*

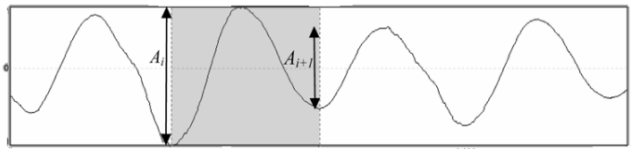

Fig. 3. *Shimmer measurement for four F0 Periods.*

Table I. *Time Domain Features description with formulae.*

Sl.No.	Feature	Description	Formulae
1.	F0(Hz)	Mean pitch	$MeanF_0 = \frac{1}{n}\sum_{i=1}^{n} F_i$
2.	Flo(Hz)	Minimum pitch	$MinF_0 detected = Min(F_i)$
3.	Fhi(Hz)	Maximum pitch	$MaxF_0 detected = Max(F_i)$
4.	Jitter (%)	Fundamental frequency perturbation (%)	$jitt(\%) = \frac{\frac{1}{n-1}\sum_{i=n-1}^{1}\lvert T_{i+1} - T_i\rvert}{\frac{1}{n}\sum_{i=1}^{n} T_i}$
5.	Jitter (Abs)	Fundamental frequency perturbation (absolute)	$jitt(ab) = \frac{1}{n-1}\sum_{i=n-1}^{1}\lvert T_{i+1} - T_i\rvert$
6.	RAP	Relative Average Perturbation	$RAP = \frac{\frac{1}{n-2}\sum_{i=2}^{n-1}\lvert\frac{T_{i+1}+T_i+T_{i-1}}{3} - T_i\rvert}{\frac{1}{n}\sum_{i=1}^{n} T_i} \times 100$
7.	PPQ	Five-point Period Perturbation Quotient	$PPQ = \frac{\frac{1}{n-4}\sum_{i=3}^{n-2}\lvert\frac{\sum_{j=-2}^{2} T_{i+j}}{5} - T_i\rvert}{\frac{1}{n}\sum_{i=1}^{n} T_i} \times 100$
8.	DDP	Average absolute difference of differences between cycles, divided by the average period	$DDP = \frac{1}{n-1}\sum_{i=n-1}^{1}\lvert A_{i+1} - A_i\rvert$
9.	Shimmer	Shimmer Local amplitude perturbation	$shimm = \frac{\frac{1}{n-1}\sum_{i=n-1}^{1}\lvert A_{i+1} - A_i\rvert}{\frac{1}{n}\sum_{i=1}^{n} A_i}$
10.	Shimmer (dB)	Local amplitude perturbation (decibels)	$shimm(dB) = \frac{1}{n-1}\sum_{i=n-1}^{n-1} 20 \times log\frac{A_i}{A_{i+1}}$

Sl.No.	Feature	Description	Formulae
11.	Shimmer:APQ3	Three point Amplitude Perturbation Quotient	$APQ3 = \frac{\frac{1}{n-2}\sum_{i=2}^{n-1}\lvert\frac{A_{i+1}+A_i+A_{i-1}}{3} - A_i\rvert}{\frac{1}{n}\sum_{i=1}^{n} A_i} \times 100$
12.	Shimmer: APQ5	Five point Amplitude Perturbation Quotient	$APQ5 = \frac{\frac{1}{n-4}\sum_{i=3}^{n-2}\lvert\frac{\sum_{j=-2}^{2} A_{i+j}}{5} - A_i\rvert}{\frac{1}{n}\sum_{i=1}^{n} A_i} \times 100$
13.	Shimmer: APQ11	11-point Amplitude Perturbation Quotient	$APQ11 = \frac{\frac{1}{n-10}\sum_{i=6}^{n-5}\lvert\frac{\sum_{j=-5}^{5} A_{i+j}}{11} - A_i\rvert}{\frac{1}{n}\sum_{i=1}^{n} A_i} \times 100$
14.	Shimmer: DDA	Average absolute difference between consecutive differences between the amplitudes of consecutive periods	$DDA = 3\left\{\frac{\frac{1}{n-2}\sum_{i=2}^{n-1}\lvert\frac{A_{i+1}+A_i+A_{i-1}}{3} - A_i\rvert}{\frac{1}{n}\sum_{i=1}^{n} A_i}\right\} \times 100$
16.	HNR	Harmonics-to-Noise Ratio	$HNR = 10log10\left\{\frac{\sum_{i}^{N/2}\lvert S_i\rvert^2}{\sum_{i=1}^{N/2}\lvert N_i\rvert^2}\right\}$
15.	NHR	Noise-to-Harmonics Ratio	$NHR = \frac{\sum_{i}^{N/2}\lvert S_i\rvert^2}{\sum_{i=1}^{N/2}\lvert N_i\rvert^2}$

2.2.2. Frequency Domain Features

Mel Frequency Cepstral Coefficients (MFCCs)

Figure 4 shows the method involved in the calculation of MFCCs. MFCC is based on human hearing perceptions, the term mel refers to a kind of estimate related to the perceived frequency. The mapping between the real frequency scale (Hz) and the perceived frequency scale (mels) is approximately linear below 1 kHz and logarithmic for higher frequencies. The method involves two types of filter; which are spaced linearly at low frequency below 1000 Hz and logarithmic spacing above 1000Hz. A subjective pitch is present on Mel Frequency Scale to capture important characteristic of speech / voice signal.

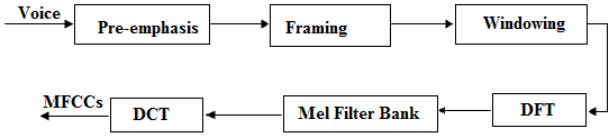

Fig. 4. *Calculation of MFCCs.*

a. *Pre-emphasis:* The voice signal is first pre emphasized, that is, the signal is first passed through a high pass filter. The filter enhances the high frequency components of the spectrum, which are usually reduced during the speech production process. The pre emphasized signal is obtained by applying the following 1st order high pass FIR filter of the form given in eq. 5.

$$H(z) = 1 - az^{-1} \tag{5}$$

It is clear from the equation that there will be a 'Zero' when z = a. By setting 'a' to 0.97 puts the 'zero' at 0.97, which will attenuate the low frequencies that are close to ω = 0. Hence eq. 5 can now be represented as

$$y(t) = x(t) - 0.97x(t-1) \tag{6}$$

where $x(t)$ is theinput voice signal and $y(t)$ is the output.

b. *Framing:* The time-domain waveform is divided into overlapping fixed duration segments called *frames*. Here frames of 20 ms with 10 ms overlapping are considered as shown in Fig. 5.This reduces the amplitude of the discontinuities at the boundaries of each finite sequence acquired by the digitized signal.

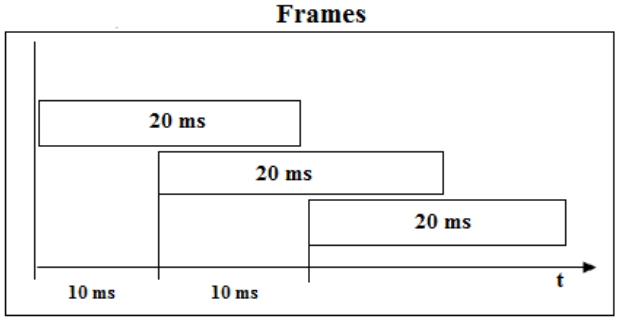

Fig. 5. *Frames of the voice signal.*

The voice signal is locally analyzed by applying window whose duration in time is shorter than the signal. The window is first applied at the beginning of the signal, then, moved further until the end of the signal is reached. The length of the window chosen is 20ms; this window is further moved with an overlap period of 10ms. This is continued till the end of the signal [20], [21].

c. *Windowing:* The framing operation has a rectangular window effect which will generate undesirable spectral artifacts. Thereby each frame is multiplied by a window function to smooth the effect by tapering each frame at the beginning and end edges. The Hamming and the Hanning windows are the commonly used in speech analysis. Here a Hamming window of 20ms is used to reduce the side effects. This tapered window function creates a smoother and less distorted spectrum.

d. *Discrete Fourier Transform(DFT):*A Fast *Fourier* Transform (*FFT*) operation is applied to each frame to the pre-emphasized, windowed voice signal which will give complex spectral values. The only parameter to be fixed for the FFT calculation is the number of points N, which is usually a power of 2, and greater than the number of points in the window. Here, a 512-point FFT

is applied, then 256 complex spectral values uniformly spaced from 0 to *Fs* /2 (where *Fs* is the sampling frequency) are produced (ignoring the mirror values). In speech processing the phase information is ignored and only the FFT magnitude is considered.

e. *Mel-filter bank:* The available spectrum after DFT presents a lot of fluctuations and too much detailed information. Only the envelope of the spectrum is of interest, hence the smoothing of the spectrum is done, which will also reduce the size of the spectral vectors, for this the available '*N*' FFT magnitude co-efficient are converted to *K* filter bank values. The filters are triangle in shape as shown in Fig 6. This is necessary because N=256 represents too much spectral detailed information and by smoothing the spectrum to K = 20 values per frame; a more efficient representation is achieved. The filter bank values are derived by cross-wise multiplying the *'N'* FFT magnitude co-efficient by the *K* triangular filter bank weighting function and then accumulating or *binning* the results from each filter triangle. The centers of the triangle filter banks are spaced according to the Mel scale as in eq.7.

$$f_{MEL} = 2595\, log_{10}\left(1 + \frac{f_{LIN}}{700}\right) \quad (7)$$

If the accumulated output from the k[th] filter bank is denoted as S_k, then log of the filter bank output, log $(S_k)$is taken to reflect the logarithmic compression in the dynamic range exhibited by the human hearing system. Taking the logarithm, also transforms multiplicative frequency filtering channel distortions into additive effect, hence, making it easier for compensation if required.

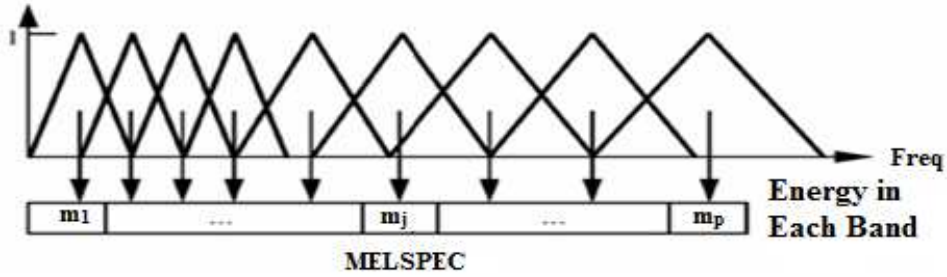

Fig. 6. *Triangle filter bank.*

f. *Discrete Cosine Transform (DCT):* The final step is to convert the *K* log filter bank spectral values, $\{log(S_k)\}_{k=1}^{K}$,into *L* cepstral coefficients using the DCT is given by eq. 8.

$$c_n = \sum_{k=1}^{K} \log(S_k) \cos\left[n(k-0.5)\frac{\pi}{K}\right] n = 1,2,\ldots\ldots,L. \quad (8)$$

Unlike spectral features which are highly correlated, cepstral features yield a more de-correlated and compact representation. Here *L* = 13 MFCC coefficients are extracted per frame which forms the feature vector for that frame[11],[12],[13],[22].

2.2.3. Classifier

Support Vector Machine (SVM)

The foundation of SVM developed by Vapnik [23] has gained popularity due to many attractive features and good performance. The Structural Risk Minimization (SRM) principle employed in SVM has shown to be superior to the traditional Empirical Risk Minimization (ERM) principle, employed in the conventional Neural Networks (NN). In ERM, (NN) choosing an appropriate structure, i.e. order of polynomials, number of hidden layer, and keeping the confidence interval fixed, minimization of the training error is done. Whereas in SRM (SVM) keeping the value of the training error fixed (equal to zero or equal to some acceptable level) and minimization of the confidence interval is done. SVMs were developed to solve the classification problem, but recently they have also been extended to the domain of regression problems. The structure of the SVM used for both time domain and frequency features is as shown in Fig.7. For time domain features the inputs are Xn where n=16. In the case of MFCC features n=13. It can be seen that the structure is similar to a NN, but the only difference between NN and SVM is the learning algorithms. The NN usually uses the error back propagation algorithm or a more sophisticated gradient descent algorithm or some other linear algebra based approach, whereas the SVMs learn to select an optimal subset by Learning Programming (LP) or solving the Quadratic programming (QP) [23],[24].

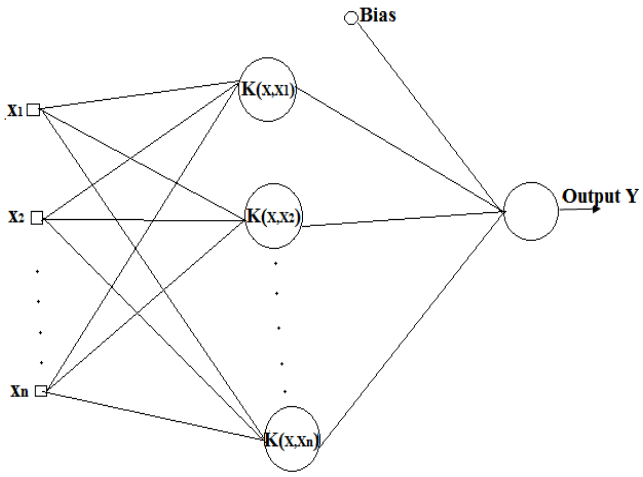

Fig. 7. *Structure of SVM.*

The goal of SVM is to produce a model which predicts target value of data instances in the testing set which are

given with the attributes. The classification in SVM is an example of supervised learning. A step in SVM classification involves identification of features which are intimately connected to the known classes. SVM models were initially defined to classify linearly separable classes with no sample overlap, and then an infinite number of hyper-planes can separate the data. Hence an optimum separating hyper-plane with a maximum margin has to be calculated. This hyper-plane is uniquely determined by the vectors on the margin, called as support vectors. The separating hyper-plane is chosen to maximize separation distance between the closest training samples. An example of two linearly separable classes is shown in Fig. 8. In the classification mode the equation of the hyper-plane separating two different classes is given by the relation

$$y(x) = w^T \emptyset(x) = \sum_{j=1}^{k} w_j \emptyset_j(x) w_0 = 0 \quad (9)$$

Where the vector $\emptyset(x) = \emptyset_0(x), \emptyset_1(x), \ldots\ldots\ldots \emptyset_k(x)$ is composed of activation function of hidden units with $\emptyset_0(x) = 1$ and $w = [w_0, w_1, \ldots . w_k]^T$ is the weight vector of the network.

The most distinctive fact about SVM is that the learning task is reduced to quadratic programming by introducing α_i the Lagrange multipliers. All operations in learning and testing modes are done in SVM using kernel functions satisfying Mercer conditions [23]. The kernel is defined as

$$K(x, x_i) = \emptyset^T(x_i)\, \emptyset(x) \quad (10)$$

The well known kernels include polynomial, radial Gaussian, or tanh activation function.

i. Polynomial kernel of degree d:

$$K(x, x_i) = (\langle K(x, x_i) \rangle + 1)^d \quad (11)$$

ii. Radial basis function with Gaussian kernel of width C > 0:

$$K(x, x_i) = exp^{\frac{-|x - x_i|^2}{c}} \quad (12)$$

iii. Neural networks with tanh activation function:

$$K(x, x_i) = \tan h(K\langle x, x_i \rangle + \mu) \quad (13)$$

Where the parameters K and μ are the gain and shift.

The final problem of learning SVM, formulated as the task of separating learning vectors x_i into two classes ofthe destination values either $d_i = 1$ or $d_i = -1$, with maximal separation margin, is reduced to the dual maximization problem of the quadratic function [23],[24].

$$\max Q(\alpha) = \sum_{i=1}^{P} \alpha_i - \frac{1}{2} \sum_{i=1}^{P} \sum_{j=1}^{P} \alpha_i \alpha_j d_i d_j K(x_{i,} x_j) \quad (14)$$

with the constraints $\sum_{i=1}^{P} \alpha_i d_i = 0, 0 \le \alpha_i \le C$, where C is a user-defined constant and p is the number of learning data pairs (x_i, d_i). C represents the regularizing parameter and determines the balance between the complexity of the network, characterized by the weight vector $\boldsymbol{w}$and the error of classification of data. For the normalized input signals the value of C is usually much higher than 1 and adjusted by cross validation.

The solution of eq. 14 with respect to the Lagrange multipliers produces the optimal weight vector w_{opt} as $w_{opt} = \sum_{i=1}^{N_s} \alpha_{oi} d_{oi} \emptyset(x_{oi})$ .In this equation N_s means thenumber of support vectors, i.e. the learning vectors x_i, for which the relation is

$$d_i\left(\sum_{j=1}^{k} w_j \emptyset_j x_i\right) + w_o \ge 1 - \xi_i \quad (15)$$

$\xi_i \ge 0$, the nonnegative slack variables of the smallest possible values are fulfilled with the equality sign [23],[24]. The output signal y(**x**) of the SVM network in the retrieval mode (after learning) is determined as the function of kernels.

$$y(x) = \sum_{i=1}^{N_s} \alpha_{oi} d_i K(x_{oi}, x) + w_0 \quad (16)$$

and the explicit form of the nonlinear function $\emptyset(x)$ need not be known

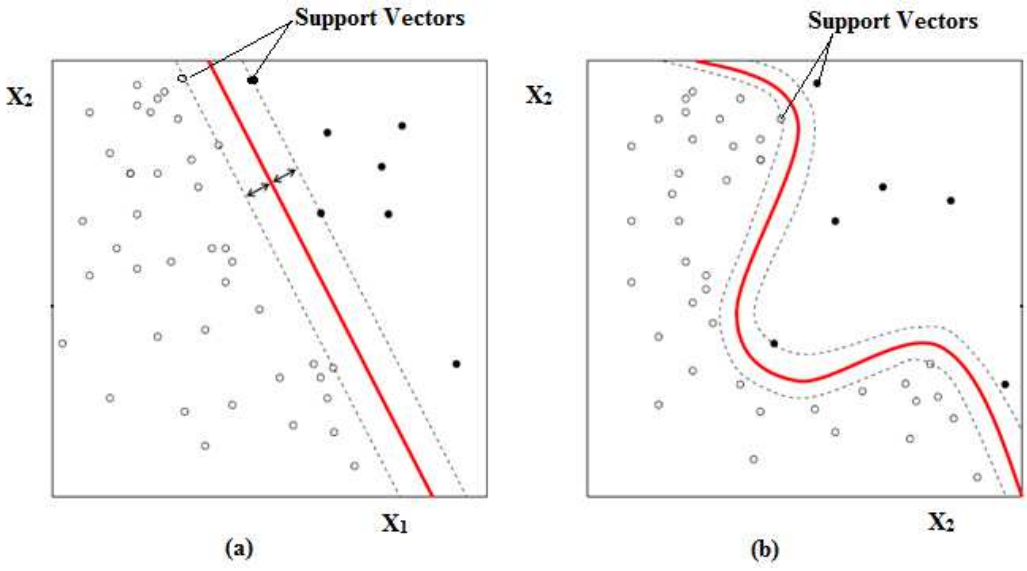

Fig. 8. *Basic Principle of SVM with (a) Linearly separable data (b) Nonlinearly separable data.*

3. Experimentation and Results

3.1. Time Domain Feature Analysis

Sixteen time domain features shown in Table I were extracted from normal and neurological disordered subjects voice signal. The distribution of the 16 features of neurological disordered subject voices is shown in Fig. 9 as arranged in Table 1. It can be seen that the notches representing the range of values of the features do not overlap to a great extent and hence can be considered as significant features, which can be given as input to the classifier. Figure 10 shows the distribution of the Pitch, Jitter, Shimmer, NHR and HNR measurements in box plots of normal and neurological disordered subject voices. The boxes have lines at the lower quartile, median, and upper quartile values. The whiskers are lines extending from each end of the boxes to show the extent of the rest of data and "+" symbols mark the outlying points. If the median line in the box plot does not overlap, it can be concluded with 95% confidence that the true medians do differ, so medians are statistically different for normal and neurological disorder voices and hence can be used as features for identification of neurological disordered subjects. The data is also analyzed statistically by student *t-test* and found that the normal and the pathological values significantly differ ($p< 0.05$) for all features except for F0, Flo, Fhi, as per the findings from our earlier study. Four jitter measurements have values of $p< 0.01$ whereas local jitter has $p<0.005$. All shimmer measurements have values of $p< 0.001$, whereas F0 has $p =0.5845$, Flo; $p = 0.7599$, Fho ; $p = 0.4795$ [9].

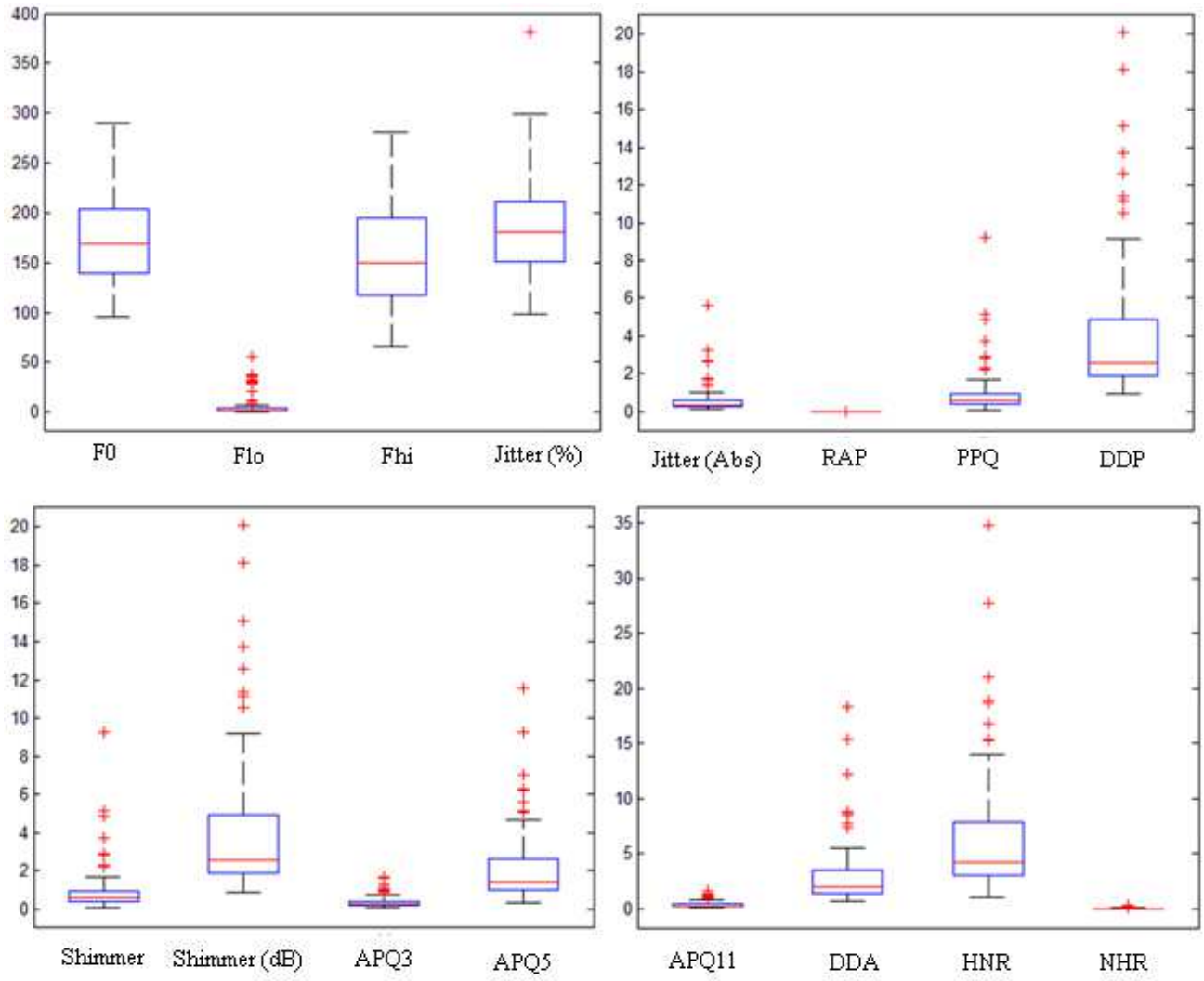

Fig. 9. *Box Plots showing the distribution of values of the Time-domain features of Neurologically disordered subjects voice tabulated in Table I.*

3.2. Frequency Domain Feature Analysis

The MFCCs parameters were calculated for both normal and neurological subjects for a dimension of 13.The variation of MFCC of normal and neurological disordered voices is shown in Fig.11 (a). It can be observed that the variation of the coefficients from frame to frame is static whereas in case

of neurological disordered voice the variation is dynamic. This may be due to the fact that the impulses from the brain neurons of the neurologically disordered subjects are randomly varying. Figure 11(b) show the power spectrum of the normal and neurological disordered voice signals, where the energy of the neurological disordered voice is more than the normal voice.

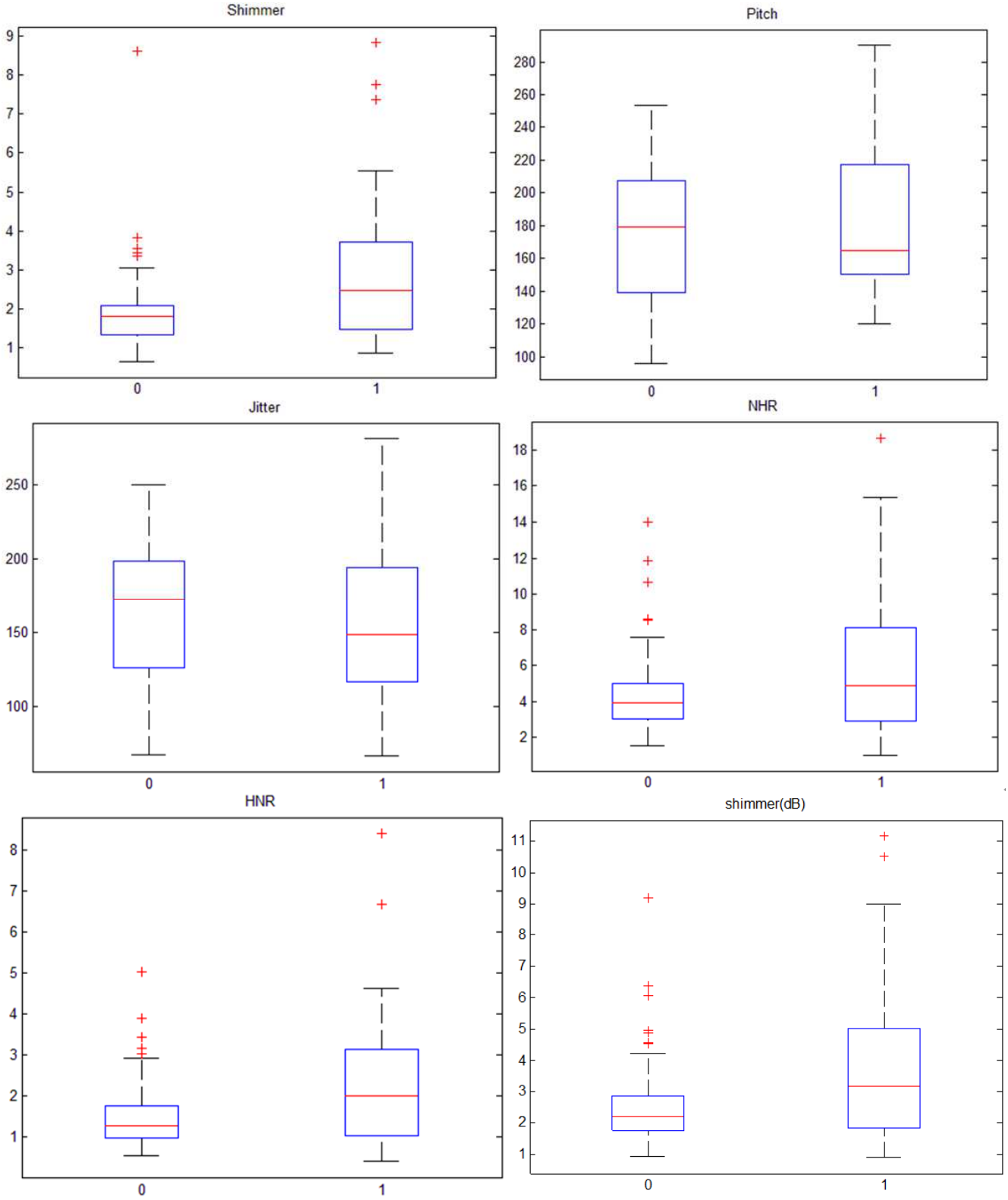

Fig. 10. *Box plots showing the distribution of the five features Pitch, Jitter, Shimmer, NHR,HNR of Normal (0) and Neurological disordered (1) Voices.*

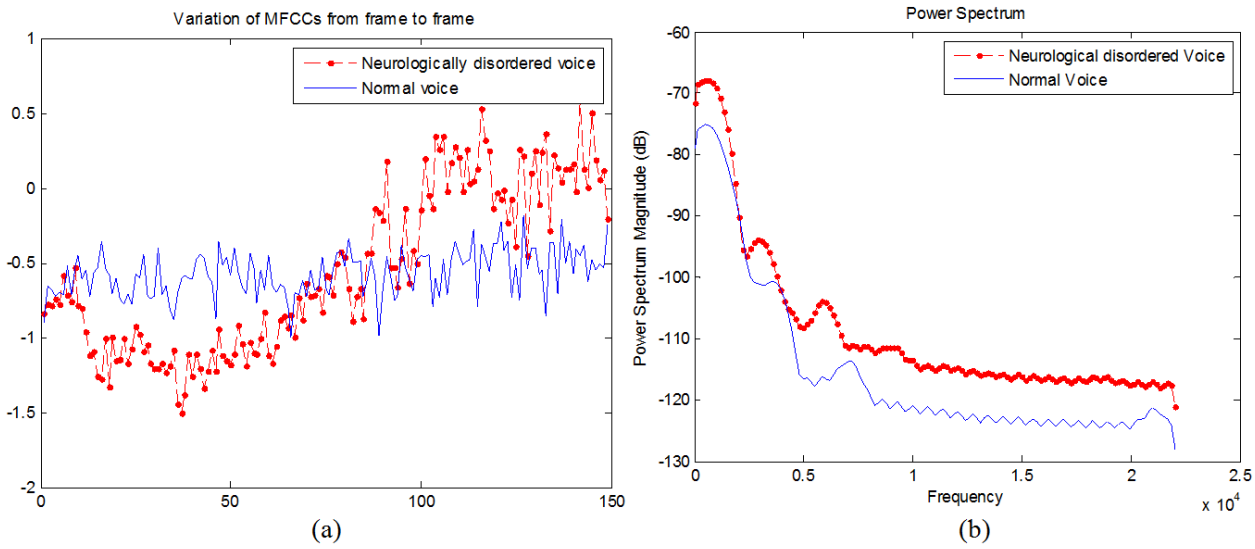

Fig. 11. *(a) Variation of MFCCs from frame to frame (b) Power spectrum of normal and neurolgicaldisordered voice signal.*

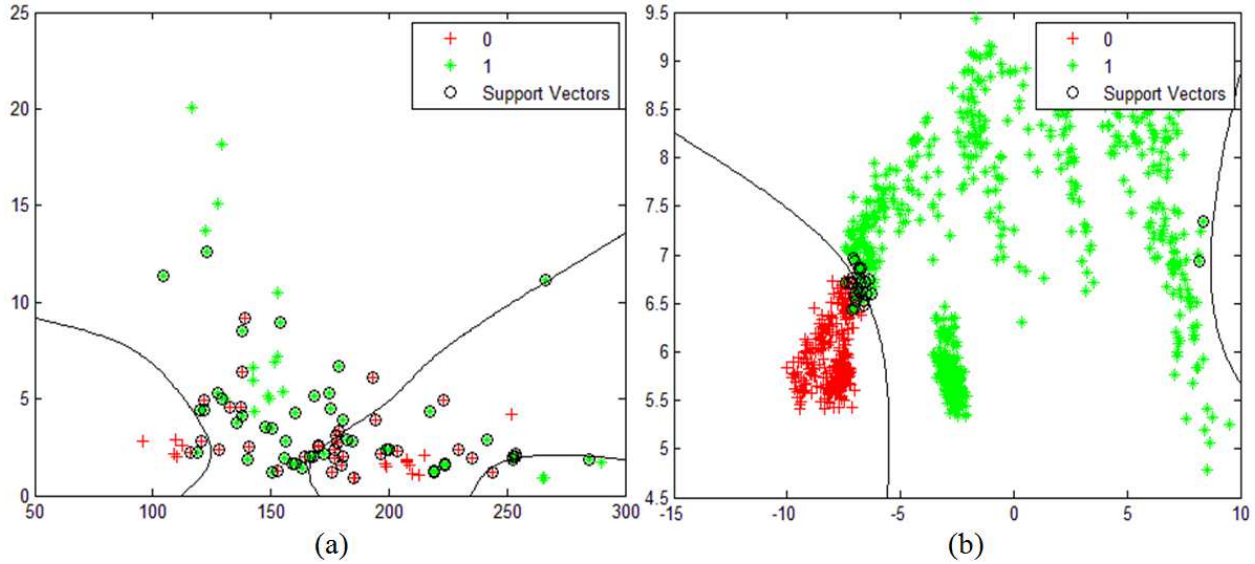

Fig. 12. *(a) Separating Normal (0) from Disordered (1)Voices using polynomial SVM using Time Domain features; (b) Separating Normal (0) from Disordered (1) Voices using polynomial SVM using Spectral (MFCCs) features.*

3.3. Classifier

The structure of the SVM network is shown in Fig.7.To train the network a polynomial kernel with order 3is chosen, setting the maximum iteration to 2000 setting the error to zero. A Sequential Minimal Optimization method is used to find the separating hyper-plane between the classes. In order to evaluate the performance of the classifier and to make comparisons, several measurements (TP, TN, FN, FP) and ratios (SE, SP, and Acc) were taken into account [25].

1. True negative (TN): The detector found no event (normal voice) when indeed none was present.
2. True positive (TP): The detector found an event (pathological voice) when one was present.
3. False negative (FN): The classifier missed an event, also called false rejection
4. False positive (FP): The detector found an event when none was present, also called as false acceptance.
5. Sensitivity (SE): Likelihood that an event will be detected given that it is present

$$SE = \frac{TP}{TP+FN} \cdot 100 \quad (17)$$

6. Specificity *(SP)*: Likelihood that the absence of an event will be detected given that it is absent

$$SP = \frac{TN}{TN+FP} \cdot 100 \quad (18)$$

7. Accuracy (Acc): Likelihood that the classification is correct

$$Acc = \frac{TP+TN}{TP+TN+FP+FN} \cdot 100 \quad (19)$$

A comparative study to classify the normal voice from the neurological disordered voice is presented in Table II. In our earlier work the experimentation was done using the 16

time domain features as input to multilayer perceptron neural network (MLPNN). In the first trial, MPLNN with 20 hidden nodes is trained and tested, and achieved a classification accuracy of 75.7%. Later in the second trial the hidden layer neurons was increased to 40 and the classification accuracy achieved was 78.57%. The experimentation was also carried out using the spectral domain features with 13 MFCCs as input to MLPNN. In a similar manner, in the first trial, MPLNN with 20 hidden nodes is trained and tested, which resulted in a classification accuracy of 77%. In the second trial the hidden layer neurons is increased to 40 and the classification accuracy achieved was 80%. From the above experimentation MLPNN with 13 MFCCs as input with 40 hidden layer neurons was found to be an optimized classifier.

From the present work it is observed from Table II that the rate of identification of neurological disordered voice with SVM is more with 83.3%, with time domain features. Whereas the identification of disordered voice with MFCC features is only 42.86%. Confusion matrix of train and test dataset shows that the system is able to identify both normal and disordered voice 100% with the MFCC features in the train dataset. The identification rate in case of test dataset for normal voice is 100% but for disordered voice is only 42.86%. This reason for the drop in the overall accuracy of the classifier performance may be because, SVM uses supervised training algorithm and requires less training pattern to estimate a good model of the class under analysis and generally will not perform well with large training attributes. Figure 12 (a) shows the plot of the time domain features using the polynomial kernel of order 3. The support vectors can be seen around the nonlinear boundary, are quite less, well separated and are not overlapping. Figure 12(b) shows the plot of the MFCC features using the same polynomial kernel, but here the support vectors crowed and overlapping around the boundary, which may be one of the reason for non-identification of the features and hence resulting in misclassification.

Hence any other classifier which is able to generate class specific model (i.e. normal and disordered models) and can handle large training data with unsupervised learning algorithm may be used to check and see if the misclassification is reduced.

4. Conclusion

Time domain parameters used for classification of normal voice from neurological disorder voice show significant differences in their *p* value in all types of shimmers, jitters, NHR, and HNR except in pitch features. Both time domain and spectral based parameters were used to train the SVM network separately and later used for classification of normal and neurological disordered subject voices for a comparative study. The time domain features with SVM classifier gives better classification of normal and pathological voices. Though frequency domain features are not giving good results using SVM, for analysis we require only short duration data with more information compared to long duration data for time domain features. In future work, to improve the classification accuracy, experimentation could be done with spectral features as inputs for some type of generative classifiers with unsupervised learning algorithm, and also combine different classifiers to see whether there is an improvement in accuracy of classification.

Table II. *The classification accuracy of SVM for time domain features and Spectral domain features (MFCCs).*

Classifier	Features to classifier	Classifier's Parameter	Subset	Confusion Matrix		Sensitivity	Specificity	Accuracy (%)
ANN	Time Domain Classical Features	Hidden neurons 20	Train	60 20	11 120	84.5	85.7	85.3
			Test	32 14	3 21	91.4	60	75.7
	Time Domain Classical Features	Hidden neurons 40	Train	68 18	3 122	95.77	87.14	90
			Test	24 4	11 31	68.57	88.57	78.57
	Spectral Domain Features MFCCs	Hidden neurons 20	Train	63 17	7 123	90	87.85	88
			Test	23 4	12 31	65.7	88.5	77
	Spectral Domain Features MFCCs	Hidden neurons 40	Train	69 16	2 124	97	88.57	91.47
			Test	24 3	11 32	68.57	91.42	80
SVM	Time Domain Classical Features	Polynomial kernel of order 3	Train	70 14	1 126	98.59	90	92.89
			Test	29 7	6 28	82.86	80	81.43
	Spectral Domain Features MFCCs	Polynomial kernel of order 3	Train	71 0	0 140	100	100	100
			Test	15 0	20 35	42.86	100	71.43

Acknowledgment

The authors are grateful to Dr. Harsha and Dr.Keshav, Neurological Department, J.S.S., Hospital, Mysuru, for helping us to collect the voice data of neurological disordered patients.

References

[1] A Wisniecki M Cannizzaro, H.Cohen, P.J.Snyder, "Speech Impairments in Neuro-degenerative Diseases/Psychiatric Illnesses", Elsevier, pp.758-764, 2006.

[2] J. Rusza and R. Cmejla H. Ruzickova and E. Ruzicka, " Quantitative acoustic measurements for characterization of speech and voice disorders in early untreated Parkinson's disease", *J. Acoust. Soc. Amer.,* vol.129, no. 2, pp. 350-366, Jan. 2011.

[3] A. K. Ho, R. Iansek, C. Marigliani, J. Bradshaw, and S. Gates, "Speech impairment in large sample of patients with Parkinson's disease," *J. Behav. Neurol.* vol.11, pp. 131-137, 1998.

[4] J. Rusza and R. Cmejla H. Ruzickova and E. Ruzicka "Objectification of dysarthria in Parkinson's disease using bayes theorem" *in Proc. Recent Researches in Communications, Automation, Signal Processing, Nanotechnology, Astronomy and Nuclear Physics* in (WSEAS), Cambridge, UK, 2011, pp.165-169.

[5] B. T. Harel, M. S. Cannizaro, H. Cohen, N. Reilly, and P. J. Snyder, "Acoustic characteristic of Parkinsonian speech: A potential biomarker of early disease progression and treatment," *J. Neurolinguistics*, vol. 17, pp.439-453, pp.1-19, 2004.

[6] M. A. Little, P. E. McSharry, S. J. Roberts, D. Costello, and I. M. Moroz, "Exploiting nonlinear recurrence and fractal scaling properties for voice disorder detection," *Biomedical Engineering [Online]*. vol. 6, no. 23, 2007.

[7] Athanasios Tsanas., Max A. Little, Patrick E. McSharry, and Lorraine O. Ramig, "Accurate Telemonitoring of Parkinson's Disease Progression by Noninvasive Speech Tests", *IEEE Trans. Biomed. Eng.*, vol. 57, no. 4, pp. 884-893, 2010.

[8] Boyan Boyanov and Stefan Hadjitodorov, "Acoustic Analysis of Pathological Voices- A Voice analysis system for the screening of laryngeal disease" *IEEE Eng Med Biol Mag.,* pp. 74-82, July/Aug. 1997.

[9] Uma Rani. K and Mallikarjun S. Holi, "Analysis of Speech Characteristics of Neurological Diseases and their Classification", *in Proc. of IEEE International conference on Computing Communication & Networking Technologies (ICCCNT), 2012* Coimbatore, India, *pp 1-6.*

[10] M.Hariharan, M. P. Paulraj, SazaliYaacob "Time-Domain Features And Probabilistic Neural Network for the Detection Of Vocal Fold Pathology", *Malaysian Journal of Computer Science*, vol. 23, no. 1, pp. 60-67, 2010.

[11] Julian D. Arias-Londono, Juan I. Godino-Llorente,,Nicolas Saenz-Lechon, Victor Osma-Ruiz, and German Castellanos-Domınguez, "Automatic Detection of Pathological Voices Using Complexity Measures, Noise Parameters, and Mel-Cepstral Coefficients", *IEEE Trans. Biomed. Eng.*, vol. 58, no. 2, Feb. 2011.

[12] J. I. Godino-Llorente and P. Gomez-Vilda, "Automatic detection of voice impairments by means of short-term cepstral parameters and neural network based detectors*" IEEE Trans. Biomed. Eng.*vol. 51,no.2, pp.380–384, 2004.

[13] Ruben Fraile, Juan Ignacio Godino-Llorente, Nicolas Saenz-Lechon, Vıctor Osma-Ruiz, Pedro Gomez-Vilda, " Use Of Cepstrum-Based Parameters For Automatic Pathology Detection On Speech *Analysis of Performance and Theoretical Justification", Proc. 1st. Int. Conf. on Biomed. Elec. and Devices, (BIOSIGNALS 2008)*, Funchal, Madeira, Portugal, Jan. 28-31, vol.1,2008, pp.85-91.

[14] Tripti Kapoor, R.K. Sharma," Parkinson's disease Diagnosis using Mel-frequency Cepstral Coefficients and Vector Quantization", *International Journal of Computer Applications*, vol. 14,no.3, pp.43-46, January 2011.

[15] Youri Maryn, Paul Corthals, Marc De Bodt, Paul Van Cauwenberge, "Perturbation Measures of Voice: A Comparative Study between Multi-Dimensional Voice Program and Praat", *J.Folia phoniatricaetlogodica,* vol. 16,pp.217-226,2009.

[16] Luis M. T. Jesus, Anna Barney, Ricardo Santos, Janine Caetano, Juliana Jorge, Pedro Sa Couto," Universidade de Aveiro Voice Evaluation Protocol", in *Proc. of Interspeech 2009,* Brighton, UK, 7-10 Sept. 2009, pp. 971-974.

[17] P. Boersma, and D. Weenink, "Praat: doing phonetics by computer (Version 5.2.2.1) [Computer program]. Retrieved from http://www.praat.org/, 2011.

[18] Wertzner H.F., Schreiber S., Amaro L., "Analysis of fundamental frequency, jitter, shimmer and vocal intensity in children with phonological disorders", *Revista Brasileira de. Otorrinolaringol*, vol. 71, pp. 582–588, 2005.

[19] M. Farru´ s J. Hernando, " Using Jitter and Shimmer in speaker verification*", J. IET Signal Process.*, vol. 3, no. 4, pp. 247–257,2009.

[20] Febe de Wet, Bert Cranen, Johan De Veth, LoeBoves, A comparison of LPC and FFT-based acoustic features for noise robust ASR, *in Proc. of Eurospeech,* 2001, pp.1-4.

[21] Tomi Kinnunen, Haizhou Li, "An overview of text-independent speaker recognition: From features to supervectors" *J. of Speech Comm*. vol.52,no.1,pp.1-30,2010.

[22] Uma Rani K and Mallikarjun S. Holi, Automatic Detection of Neurological Disordered Voices Using Mel Cepstral Coefficients and Neural Networks*, proc. of IEEE-EMBS Special Topic Conference on Point-Of-Care (POC) Healthcare Technologies*, Bangalore, India, 2013, pp.76-79.

[23] Steve R. Gunn, "Support Vector Machine for Classification and Regression", *Technical Report,* School of Electronics and Computer Science University of Southampton, 1998, pp.1-53.

[24] P. Dhanalakshmi, S. Palanivel, V. Ramalingam, "Classification of audio signals using SVM and RBFNN," *Expert Systems with Applications,* vol.36, no.3, pp. 6069-6075, 2009.

[25] Juan Ignacio Godino-Llorente, Pedro Gomez-Vilda, and Manuel Blanco-Velasco, "Dimensionality Reduction of A Pathological Voice Quality Assessment System Based On Gaussian Mixture Models and Short-Term Cepstral Parameters", *IEEE Trans. Biomed. Eng*., vol.53, no. 10, pp.1943-1953, 2006.