

ORIGINAL RESEARCH

A hybrid model for neurological disordered voice classification using time and frequency domain features

Uma Rani. K^{*1}, Mallikarjun S. Holi²

¹Department of Biomedical Engineering and Research Centre, Bapuji Institute of Engineering and Technology, Davangere, Karnataka, India

²Department of Electronics and Instrumentation Engineering, University B. D. T. College of Engineering, Visvesvaraya Technological University, Davangere, Karnataka, India

Received: August 31, 2015

Accepted: October 21, 2015

Online Published: November 16, 2015

DOI: 10.5430/air.v5n1p87

URL: <http://dx.doi.org/10.5430/air.v5n1p87>

ABSTRACT

Different neurological disorders may lead to speech related problems, due to paralysis in vocal fold or weakness of the related muscles. This may modify the acoustic characteristics of the subject's voice which may provide important information for detecting certain neurological diseases. The vowel phonation which is acoustically informative and uttered by the patient with not much difficulty is collected and various acoustic features are extracted by time domain and frequency domain techniques. The use of all these features for classification may lead to a large feature space, which may lead to complexity. Hence to avoid this, in the present work experimentation is done by fusing different classifiers which are fed with features extracted from different domains. The time domain and frequency domain features are given to Support Vector Machine (SVM) and Gaussian Mixture Model (GMM) respectively, and the intermediate decision of these classifiers is given to another SVM to identify the voice signal as normal or diseased. It is observed that this hybrid classifier model has shown some improvement with a classification accuracy of 91.43% compared to single GMM classifier with an accuracy of classification of 90% with frequency domain features as input.

Key Words: Neurological disorder, Voice, Support Vector Machine, Gaussian Mixture Model, Classifiers

1. INTRODUCTION

Neurological disorder in adults such as apraxia, dysarthria and brain injury may result due to stroke, progressive neurological disease like Parkinson Disease (PD) and head injury respectively.^[1] After the occurrence of stroke there can be paralysis of the vocal folds and weakness of the muscles which are related to phonation. During phonation air from the lungs passes through the larynx containing the vocal folds which vibrate and produce voice. Death or damage to nerve

cells in case of progressive diseases like PD and cerebellar demyelination results in hoarseness, weakness, breathy or roughness of voice known as dysphonia. This is often observed during production of vowel sounds. Dysphonia is measured using a variety of examination tools that allow the clinician to see the pattern of vibration of the vocal folds, principally by laryngeal videostroboscopy, which may cause discomfort to the patient.^[2] Hence acoustic analysis of voice to detect voice disorders is gaining more interest in present

*Correspondence: Uma Rani. K; Email: uma_devoor@yahoo.com; Address: Department of Biomedical Engineering and Research Centre, Bapuji Institute of Engineering and Technology, Davangere, Karnataka, India.

days.^[2-4] The acoustic analysis of voice is non invasive, cost effective, comfortable to patients and gives information about the hidden voice disorders. Therefore, the voice measurements are well suited for acoustic analysis with no profound effects of the articulators used during the speech.^[2,5]

The voice changes may include changes in the pitch, cycle-to-cycle variations; jitter, decrease in amplitude; shimmer, loss of power, addition of noises; harmonic to noise ratios, constriction of voice range; *i.e.*, displacement towards lower frequency.^[2,3,6,7] The normal acoustic features lie in the frequency range 20 Hz - 4,000 Hz, but the pathological voice features are spread over 20 Hz - 20 kHz. The changes in the acoustic features may not be noticeable by normal hearing. The feature extraction methods and design of classifier is the most challenging task in the pathological voice diagnosis. From past decade mel frequency cepstral coefficients (MFCCs) have been widely used in pathological voice detection systems with good results.^[8-10] The main advantage of MFCCs is that, its calculation does not require any previous pitch information, which is a difficult task in the presence of pathology. The basic time domain features and MFCCs can be used in developing classifiers models for detection of pathological voices. In the earlier work^[11] the neurological disease; PD is considered in which combined features; the time domain and nonlinear dynamic features are given to different classifiers considered independently like SVM, K-Nearest Neighbor (KNN) and discrimination-function-based (DBF) shows classification accuracy of 91.04%, 93.82% and 82.2% respectively. From the literature it is found that single or elementary classifiers with a single domain or combined features show a lesser classification rate compared to combined features to combined/hybrid classifiers.

The success of any classification model depends on the input features, structure of the model and low errors which provide high accuracy. If the elementary (single) classifier models do not give low errors or they do not have high qualities in providing accurate results, then experimentation can be carried out by combining the outputs of these elementary classifiers to build a hybrid model which can give better results. The present work focuses on developing a hybrid model which is based on elementary classifiers with time domain features (SVM) and frequency domain features (GMM) to enhance the accuracy of classification.

The identification of neurological disease in subjects is important and a challenging task. These diseases are progressive in nature, if not diagnosed and treated early may lead to critical disabilities, and affecting the subject's employment, family life, and social activities. These subjects will also not be able to communicate properly due to disability in their

vocal system and hence have a tough time in family and in their work place forcing them to early retirement, sometimes leading to depression. Therefore, it would be very beneficial to medical field if a technique is developed which would automatically identify a subject suffering from neurological disease with ease and less discomfort. The present work proposes a technique to address this issue based on voice signal analysis and modelling.

2. MATERIALS AND METHODS

2.1 Database

The database consists of 281 sustained vowel /ah/ phonations acquired from 49 male (62.72 ± 8.0 years) and 25 female subjects (65.19 ± 8.8 years) suffering from different neurological diseases. The PD, cerebellar demyelination and stroke are the diseases considered in this work. 175 phonations were acquired from the diseased group and 106 phonations from normal subjects. The normal subjects were 56 in number in the same age group and healthy with no complaint of voice/speech related problems. The data collection was done using a unidirectional microphone connected to a laptop computer with Pentium processor at the Outpatient Wing, Neurology Department, J.S.S hospital, Mysuru, after getting the consent and clearance from local ethical committee. The subjects were instructed to phonate the vowel/ah/ with comfort into the microphone which was at a distance of 5 cm from the mouth. The signal acquisition was carried out using PRAAT software by selecting the sampling frequency as 44,100 Hz and channel mode as mono.^[12-14] The voice signals were saved on the hard disk in WAVE format and acoustic analysis was carried out on selected 2 sec steady portion of the acquired voice signal.

2.1.1 Time domain features

The classical dysphonia features in the present study include fundamental frequency—three measures, fundamental frequency perturbations—five measures (jitter), amplitude perturbations—six measures (shimmer), and signal to noise ratio—two measures.^[6,15] Table 1 gives the description of the 16 voice features used in this study. In a healthy person the vocal fold opening and closing is considered typically periodic and the time interval between successive cycles is said to be equal and is called pitch period or fundamental frequency. In subjects with voice disorder this pattern may alter due to the incomplete closure of the vocal folds, which leads to an unsteady phonation. Due to uncontrolled vibration of the vocal fold there is variability in F0, this is referred as Jitter. The amplitude of vibration called the Shimmer depends on the tension of vocal folds and subglottic pressure of the air column.^[7,10,15-17] This imbalanced movement of the vocal fold results in excessive breathiness (turbulent noise),

which can be measured by Harmonics to Noise Ratio (HNR) and Noise to Harmonics Ratio (NHR).^[18] A free shareware program “PRAAT” was used to quantify these time domain features.^[14]

Table 1. Description of Voice Features

GROUP	FEATURE	DESCRIPTION	
Fundamental frequencies	F ₀ (Hz)	Mean pitch	
	F _{lo} (Hz)	Minimum pitch	
	F _{hi} (Hz)	Maximum pitch	
Fundamental frequency variations	Jitter (%)	Fundamental frequency in %	
	Jitter (Abs)	Absolute Fundamental frequency	
	RAP	Relative Amplitude variations	
	PPQ	Five-point Period variation Quotient	
	DDP	Ratio of average absolute difference of differences between cycles to average period	
Amplitude variations	Shimmer	Shimmer Local amplitude variations	
	Shimmer (dB)	Local amplitude variations in decibels	
	Shimmer: APQ3	Three point Amplitude variation Quotient	
	Shimmer: APQ5	Five point Amplitude variation Quotient	
	Shimmer: APQ11	11-point Amplitude variation Quotient	
	Shimmer: DDA	Average absolute amplitude difference between consecutive differences between the amplitudes of consecutive periods	
	Signal to Noise ratios	NHR	Ratio of Noise-to-Harmonics
		HNR	Ratio of Harmonics-to-Noise

2.1.2 Frequency domain features: Linear Predictive Coding (LPC) – MFCCs

Figure 1 show the different steps involved in the calculation of LPC-MFCCs. The human speech production mechanism has the tendency to attenuate the components present in high frequency range, therefore initially the voice signal is pre-emphasized using a 1st order high pass finite impulse response (FIR) filter using Eq.(1). FIR filter has only zero with no poles and if a point on the unit circle trace is close to zero attenuating the corresponding frequencies. Hence the coefficient is taken as 0.97 which puts a zero at 0.97 resulting in attenuation of frequencies close to $\omega=0$.

$$y(t) = x(t) - 0.97x(t - 1) \quad (1)$$

Further, as the speech signal is non-stationary the signal is divided into frames of 20 ms duration with 10 ms overlapping and hence are said to have stationary behavior within the frames. The framing operation is equivalent to rectangular window effect which gives rise to unwanted noise in spectral domain. Thereby to remove this spurious noise each frame is multiplied with a hamming window which is tapered at

the beginning and the end edges of the frame to give a less distorted and smoother spectrum.^[19,20]

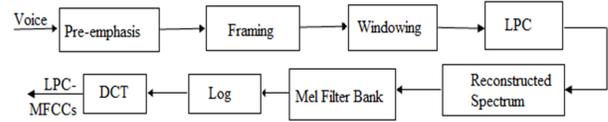


Figure 1. Calculation of LPC-MFCCs

The Fast Fourier Transform- Mel Frequency Cepstral Coefficient (FFT-MFCC) method gives a nonparametric description of the spectra describing the peaks and valleys of the spectrum. Figure 2 shows FFT spectrum containing lot of detailed information resulted due to the deep fluctuations in the spectrum. Only the envelope of the spectrum is of interest, hence to get a smoothing spectrum the LPC method is employed. A detail description of calculation of the method is given in our earlier work.^[21] Here the LPC is used to estimate the main parameters of the signal. According to^[22] the speech production model can be often called as linear production model or autoregressive model. The LPC coefficients are obtained from each frame and the prediction error for each frame can be given as

$$E_m = \sum_n e_m^2[n] = \sum_n \left(x_m[n] - \sum_{j=1}^p a_j x_m[n-j] \right)^2 \quad (2)$$

Where $x[n]$ is a frame of the speech signal and p the order of the LPC analysis. While estimating LPC coefficients from the speech frame, the orthogonality principle is assumed and the Yule Walker equations are obtained as

$$\sum_{j=1}^p a_j \phi_m[i, j] = \phi_m[i, 0], \quad i = 1, 2, \dots, p \quad (3)$$

Where ϕ_m is the correlation coefficients defined as

$$\phi_m[i, j] = \sum_n x_m[n-i] x_m[n-j] \quad (4)$$

Solution of the p linear equations gives p LPC coefficients that minimize the prediction error, and the solution for the Yule Walker equations is obtained using the Levinson-Durbin recursion algorithm.^[19] The recursion finds the solution of all prediction coefficients of order less than p . After the LPC analysis, the power spectrum of the speech frame can be calculated from its LPC parameters. LPC analysis produces an estimate smoothed spectrum. The higher order cepstral coefficients give rise to distinguished ripples in the spectrum, which is a representation of pitch harmonics. Thus the coefficients are truncated at order 10-24 usually such that a

relatively smooth spectrum of the voice signal can be used for reconstruction. In this work, the spectral estimation is carried out using an Auto Regressive (AR) filter of order ten designed using the LPC analysis technique. With an increase in the order more smoothing was observed. Reconstruction of spectral envelope was done to get back the voice signal and to scale back the signal energy level back to the original level using the residual signal energy. Thirteen MFCCs are derived from the log of the mel bank outputs using the Discrete Cosine Transform (DCT). Logarithm is taken to reflect the logarithmic compression in the dynamic range exhibited by the human hearing system and also to transform multiplicative frequency filtering channel distortions into additive effect.^[19]

2.2 Classifiers

2.2.1 Support Vector Machine (SVM)

A SVM is a two-class classifier; the support vector algorithm is a promising algorithm for supervised learning problems. The algorithm looks for the hyperplane that separates the two classes with a margin. The hyperplane should have a maximum margin for a good separation between the classes. The instances that lie around very close to the hyperplane are termed as support vectors. The SVM may be equally used to separate non-linearly separable patterns. For this, the input space is transformed into a new feature space, where a linear model is constructed to represent the nonlinear decision boundary.

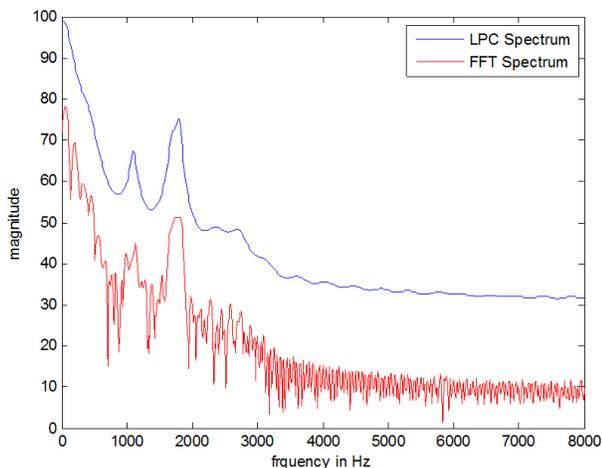


Figure 2. The FFT and LPC based spectrum of the voice signal

The hyperplane function separating the two classes is represented as

$$F(x) = w_0 + w_1 a_1 + w_2 a_2 \quad (5)$$

Where a_1 and a_2 represent the attribute values and w_0, w_1, w_2 the weights

The hyperplane function for maximum margin in terms of support vectors is given as

$$f(x) = b + \sum_{i=1}^l \alpha_i y_i \langle x_i \cdot x \rangle \quad (6)$$

Where i represents the support vector; $y_i = +1$ or -1 gives the class value; x_i is the i^{th} support vector; x represents the test value vector; $\langle x_i \cdot x \rangle$ represents the dot product; b and α 's are parameters of the hyperplane similar to weights w_0, w_1, w_2 .

According to the equation of the maximum margin hyperplane, every time an instance is classified, its dot product considering all support vectors must be calculated. A non-linear transformation applied to the instances will result in a high dimensional space with a large number of attributes. This results in an expensive computation of the dot product and may continue during the training process also. Actually such a high dimensional feature space is not necessary. By using a kernel function, the dot product computation can be done before the non linear transformation, which is called as *kernel trick*. In this process a kernel function is applied to instances in the original input space, which brings the same effect as linear transformation without expanding the feature space by non linear transformation. The major advantage of the kernel is that it operates in the input space by summing the weights of the kernel function obtained by evaluating the support vectors, which can lead to a solution for classification problem. The SVM algorithm can construct a variety of learning machines by use of different kernel functions. The well known kernels may include polynomial, radial Gaussian, or tanh activation function. In our work experimentation was done with all the three kernels and polynomial kernel of degree 3 performed good.^[23,24]

The three kernel functions are given as follows:

- (1) Polynomial kernel of degree d :

$$K(x, x_i) = (\langle K(x, x_i) \rangle + 1)^d \quad (7)$$

- (2) Radial basis function with Gaussian kernel of width $C > 0$:

$$K(x, x_i) = \exp \frac{-|x-x_i|^2}{c} \quad (8)$$

- (3) Neural networks with tanh activation function:

$$K(x, x_i) = \tan h(K \langle x, x_i \rangle + \mu) \quad (9)$$

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

2.2.2 Gaussian Mixture Models (GMM)

A GMM is a parametric modelling technique. Here the probability density function is represented as a weighted sum of Gaussian component densities. A Gaussian model assumes the feature vectors to follow a Gaussian distribution, characterized by a mean and a deviation about the mean. Further, by allowing a mixture of such Gaussians the distribution of the features for the data under consideration can be characterized.^[20,25]

The below equation represents the summation of the weighted values of M component Gaussian densities,

$$P(x|\lambda) = \sum w_i g(x|\mu_i \sum_i), \text{ for } i = 1, 2, \dots, M \quad (10)$$

where x gives the D-dimensional vector of the features, w_i gives the weights of the Gaussian mixtures, and $g(x|\mu_i \sum_i)$ represents the Gaussian densities. For D-variant the function can be expressed as,

$$g(x|\mu_i \sum_i) = \frac{1}{2\pi^{D/2} |\sum_i|} \exp\left\{-\frac{1}{2}(x-\mu_i)'\sum_i^{-1}(x-\mu_i)\right\} \quad (11)$$

μ_i gives the mean vector, \sum_i represents the matrix of covariance. But the weights of the Gaussian mixture should be $\sum_{i=1}^M w_i = 1$.

The representation of all the parameters can be written as below,

$$\lambda = \{w_i, \mu_i, \sum_i\} \quad i = 1, \dots, M \quad (12)$$

Once the GMM is configured and the training vectors identified, λ can be estimated using the maximum likelihood (ML) technique. Considering $X = \{x_1, \dots, x_t\}$, as training vector the ML maximises the likelihood in the model, given as,

$$p(X|\lambda) = \prod_{t=1}^T p(x_t|\lambda) \quad (13)$$

As the above function is non-linear maximization becomes very complex, hence an iterative algorithm expectation maximization (EM) for the purpose.^[24] To start the algorithm initiates a model λ , then a new model λ' is estimated, such that $(X|\lambda') \geq p(X|\lambda)$. Then in the next step λ' becomes the initial model for which the new model is estimated, the iteration continues until it converges to the threshold value. Now during the testing process for the test vector

$X = \{x_1, x_2, \dots, x_T\}$, the Log likely hood Ratio (LLR) which are also called as scores is given,

$$\Lambda(X) = \log\left[p\left(\frac{x}{\lambda_N}\right)\right] - \log\left[p\left(\frac{x}{\lambda_D}\right)\right] \quad (14)$$

where, λ_N represents the model for normal voice and λ_D for neurological disordered voice. Λ_{TH} which is the threshold value over which the decision is made is set to identify the disordered voices (false rejection) from the normal voices (false acceptance). For this log-likely hood ratio (scores) and Λ_{TH} are compared and if $\Lambda(x) \geq \Lambda_{TH}$ the voice is identified as normal and if $\Lambda(x) < \Lambda_{TH}$ the voice is disordered, as shown in Figure 5(b).^[26,27]

2.2.3 Combining the GMM (Generative) and SVM (Discriminative) Classifiers

The purpose of combining the outputs of different single classifiers rather than fusing features is because the input feature space has to be simple for such classifiers. Also, the performance of an individual classifier may differ, for different feature set. The best classifier may also fail on a feature pattern that the other classifier was successful. Further, it is not necessary that the misclassified outputs be overlapped, hence, making it possible to combine the outputs of the individual classifier and improve the system performance. The main aim of combining classifiers is to utilize the combination in such a way that the qualities of each other are enhanced.

The classification of normal and neurological disordered voices was carried out by combining the GMM (Generative approach) and the SVM (Discriminative approach) as shown in Figure 3. The voice features are calculated in two different domains. In one branch the 16 time domain features are fed to the discriminative classifier; the SVM to discriminate to which class the sample falls into; whether normal/disordered. In the other branch the MFCC are given to the generative model where the likelihood to be normal or the likelihood to be neurological disordered are calculated using GMMs. Further the intermediate decision of these two classifiers about the presence or absence of the disorder is used as a new feature space and given to a SVM based classifier trained to give the presence or absence of the disorder. This method of combining classifiers improves the overall performance due to two reasons. Firstly the time domain and the frequency domain feature are not combined making the feature space very simple and small. Secondly, the voices misclassified would not necessarily overlap, because, the data lying in the overlapping region may belong to either of the class and during classification, only one class is predicted by the system, hence making classification difficult.

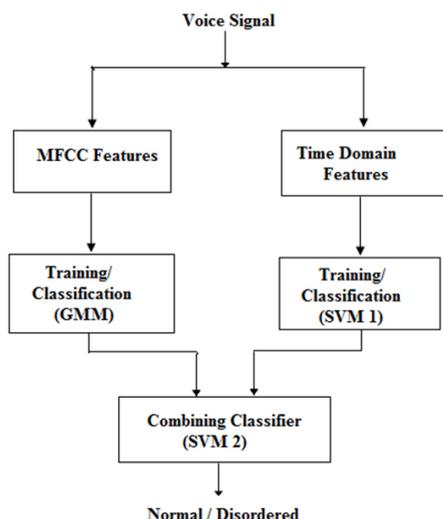


Figure 3. The scheme for automatic detection of neurological disorder using combined classifiers

3. EXPERIMENTATION AND RESULTS

Statistical analysis on the 16 time domain features (see Table 1) was carried out using the students *t*-test and statistical significances of the *p* value ($p < .05$) between the two groups, neurological disordered patients and normal subjects were found in all the jitter, shimmer and signal to noise ratios (NHR and HNR) except in the pitch measures, as discussed in our earlier studies.^[6] All the 16 time domain features were given to train the SVM 1 network with a polynomial kernel of order 3, setting the maximum iteration to 2,000 with error to zero. A Sequential Minimal Optimization method is used to find the separating hyperplane between the classes. The Receiver Operating Characteristic (ROC) curve for the time domain features is shown in Figure 4. The ROC curve is

a graph of sensitivity (*y* axis) with specificity (*x* axis). A large *y* value of the curve indicates maximum sensitivity and a small *x* value of the curve indicates maximum specificity. Thus a good a test cut off value is that value which corresponds to a point on the ROC curve nearest to the upper left corner of the ROC graph. Hence an optimal cut-off point on the ROC curve will move from the vicinity of the upper left corner over toward the upper right corner.

The GMMs were trained using 4, 8, and 16 mixtures. The classification performance for the different mixtures was evaluated and the performance for 8-mixture configuration was found to be the best and as the mixtures were increased, the classifier went into saturation. The experimentation for the study of the performance of different mixtures is discussed in Ref.20. The ROC for GMM with different mixtures is shown in Figure 5(a) indicating that the classifier performs well for 8-mixture. Therefore, the classifier with 8-mixture which is giving an overall accuracy of 90% for LPC-MFCCs input is used for combining the classifiers. Figure 5 (b) shows false acceptance and false rejection plots versus threshold Λ_{TH} . Both lines cross over the equal error rate point (EER).

Table 2 shows the classification accuracy results obtained by the hybrid classifier, input with features of different domain. The SVM 2 provides a feature space with two vectors, one vector as the GMM scores and the other vector as the SVM scores, to take a final decision about the presence or absence of pathology. The table shows the best classification accuracy combining both classifiers with SVM 2, giving an overall accuracy of 91.43%, with an increase in 1.43% accuracy in hybrid model compared to GMM which is significant in such speech studies. Whereas the disordered voice identification is 85.71% and normal voice identification is 97.14%.

Table 2. Performance of Individual Classifier and Hybrid Classifier

Classifier	Features to classifier	Classifier's Parameter	Subset	Confusion Matrix		Sensitivity (%)	Specificity (%)	Accuracy (%)
SVM 1	Time Domain Features	Polynomial kernel of order 3	Train	70	1	98.59	90	92.89
			Test	29	6	82.86	80	81.43
GMM	Frequency Domain LPC-MFCC Features	8 mixtures	Train	68	3	95.77	97.86	97.16
			Test	30	5	85.71	94.28	90
Hybrid	SVM 1 & GMM Scores to SVM 2	Polynomial kernel of order 3 (SVM 1 & 2) and 8 mixture (GMM)	Train	70	1	98.59	97.86	98.1
			Test	30	5	85.71	97.14	91.43
				1	34			

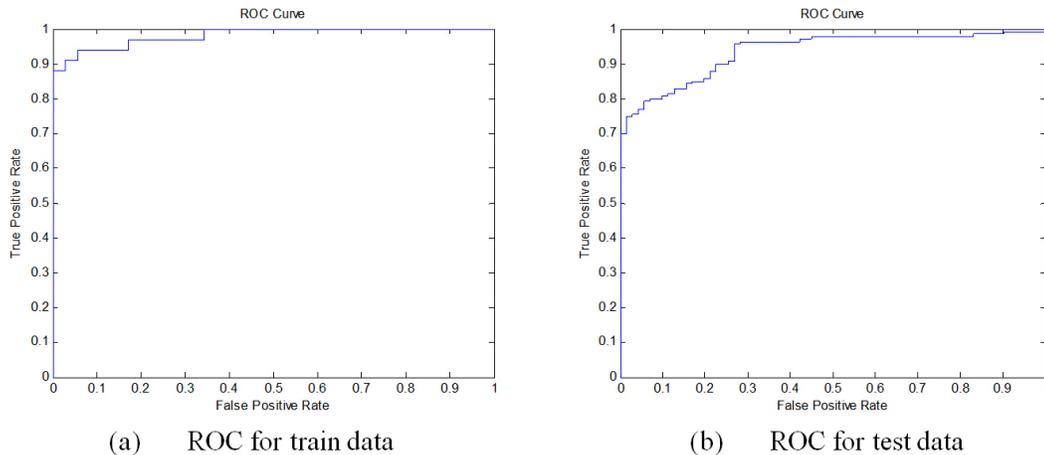


Figure 4. ROC for train data and test data

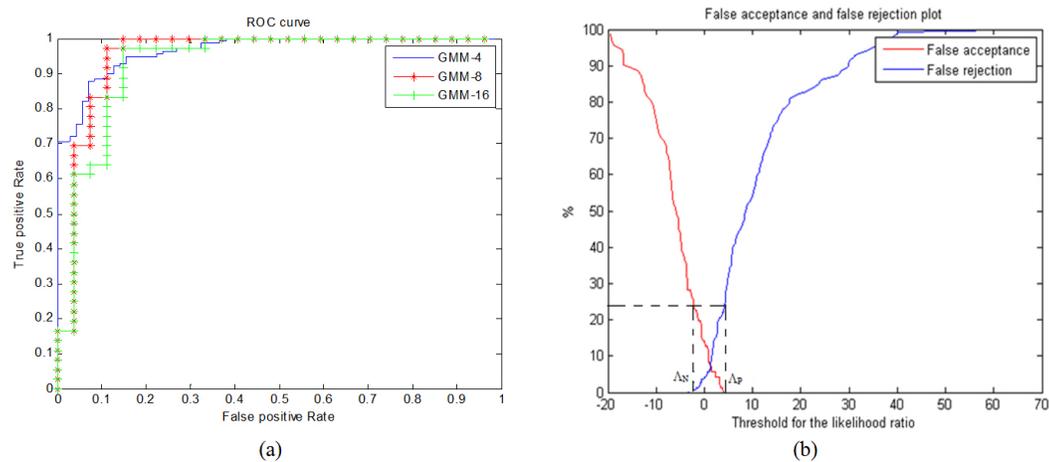


Figure 5. (a) Comparison of ROC for different Gaussian; (b) False rejection and false acceptance indicating mixtures the thresholds of LLR which crossover at EER

4. CONCLUSION

The construction of an accurate model to identify the neurological disease subjects is very essential which would help the neurologist for proper and speedy identification, because neurological disease investigation are time consuming, painful and costly. Therefore, the voice signals which can be collected with ease, without pain and at low cost can be analyzed efficiently and quickly using a microphone and computer. For accurate identification of the diseased subject, design of a good classifier with proper inputs is essential. The use of a hybrid classifier has demonstrated to be a valuable technique alternative to fusion of features and hence avoiding the construction of a huge feature space. The identification rate of disordered voice using time domain features to SVM and frequency domain features to GMM classifiers were good. Hence an experimentation of combining these

classifiers with the time and frequency domain features was carried out. The use of SVM as final detector, allowed an improvement in the overall classification accuracy of 91.53% with no change in the identification rate of disordered voice of 85.71% in both the cases *i.e.* with frequency domain features to GMM classifier and combined features to the hybrid model. In future, the time-frequency domain - the wavelet transform technique may be used as features to the hybrid classifier to observe for any improvement in identification rate of the disordered voice.

ACKNOWLEDGEMENTS

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] Kent RD, Kent JF, Weismer G, *et al.* What dysarthrias can tell us about the neural control of speech. *Journal of Phonetics*. 2000 July; 28(3): 273-302.
- [2] Little MA, McSharry PE, Hunter EJ, *et al.* Suitability of Dysphonia Measurements for Telemonitoring of Parkinson's Disease. *IEEE Trans. Biomed. Eng.* 2009 April; 56(4): 1015-22.
- [3] Rusza J, Cmejla R, Ruzickova H, *et al.* Quantitative acoustic measurements for characterization of speech and voice disorders in early untreated Parkinson's disease. *J. Acoustic Soc. Amer.* 2011 Jan.; 129(2): 350-66.
- [4] Ho K, Ianseck R, Marigliani C, *et al.* Speech impairment in large sample of patients with Parkinson's disease. *J. Behav. Neuro.* 1998; 11(3): 131-7.
- [5] Rusza J, Cmejla R, Ruzickova H, *et al.* Objectification of dysarthria in Parkinson's disease using bayes theorem. *Proceedings of Recent Researches in Communications, Automation, Signal Processing, Nanotechnology, Astronomy and Nuclear Physics in (WSEAS)*; Cambridge, UK; 2011 Feb. p. 20-22, 165-169.
- [6] Uma RK, Mallikarjun SH. Analysis of Speech Characteristics of Neurological Diseases and their Classification. *Proceeding of IEEE International conference on Computing Communication & Networking Technologies (ICCCNT)*; Tamil Nadu, Coimbatore, India; 2012 July 26-28. p. 1-6.
- [7] Farris M, Hernando J. Using Jitter and Shimmer in speaker verification. *J. IET Signal Processing*. 2009 July; 3(4): 247-57. <http://dx.doi.org/10.1049/iet-spr.2008.0147>
- [8] Arias-Londono JD, Godino-Llorente JJ, Saenz-Lechon N, *et al.* Automatic Detection of Pathological Voices Using Complexity Measures, Noise Parameters, and Mel-Cepstral Coefficients. *IEEE Trans. Biomed. Eng.* 2011 Feb.; 58(2): 370-9.
- [9] Godino-Llorente JJ, Gomez-Vilda P, Blanco-Velasco M. Dimensionality. Reduction of a Pathological Voice Quality Assessment System Based on Gaussian Mixture Models and Short-Term Cepstral Parameters. *IEEE trans. Biomed. Eng.* 2006 Oct.; 53(10): 1943-53. PMID:17019858. <http://dx.doi.org/10.1109/TBME.2006.871883>
- [10] Uma Rani K, Mallikarjun SH. Automatic detection of neurological disordered voices using melcepstral coefficients and neural networks. *Proceedings of IEEE International conference on Point-of-Care (PoC) Healthcare Technologies*. Bangalore, India; 2013 Jan. 16-18. p. 76-79.
- [11] Maryn Y, Corthals P, Marc De B, *et al.* Perturbation Measures of Voice: A Comparative Study between Multi-Dimensional Voice Program and Praat. *J. Folia phoniatricaetlogodica*. 2009 July; 61: 217-26.
- [12] Jesus LMT, Barney A, Santos R, *et al.* Universidade de Aveiro's Voice Evaluation Protocol. *Proceedings. of Interspeech*, Brighton, UK. 2009 Sept. 7-10. p. 971-4.
- [13] Weenink BP. Praat: doing phonetics by computer (Version 5.2.2.1) [Computer program]. Available from: <http://www.praat.org/>
- [14] Wertzner HF, Schreiber S, Amaro L. Analysis of fundamental frequency, jitter, shimmer and vocal intensity in children with phonological disorders. *Revista Brasileira de Otorrinolaringol.* 2005 Sep-Oct.; 71(15): 582-8.
- [15] Tsanas A, Little MA, Mc Sharry PE, *et al.* Accurate Telemonitoring of Parkinson's Disease Progression by Noninvasive Speech Test. *IEEE Trans. Biomed. Eng.* 2010 April; 57(4): 884-93.
- [16] Boyanov B, Hadjitodorov S. Acoustic Analysis of Pathological Voices—A Voice analysis system for the screening of laryngeal disease. *IEEE Eng Med Biol Mag.* 1997 July/Aug. p. 74-82.
- [17] Yumoto E, Gould WJ. Harmonics-to-noise ratio as an index of the degree of hoarseness. *J. Acoust. Soc. Am.* 1982 June; 71(6): 1544-50. PMID:7108029. <http://dx.doi.org/10.1121/1.387808>
- [18] de Wet F, Cranen B, De Veth J, *et al.* A comparison of LPC and FFT-based acoustic features for noise robust ASR. *Proceedings of Eurospeech 2001, 7th European Conference on Speech Communication and Technology*. 2001 Sept-7: 1-4.
- [19] Kinnunen T, Li HZ. An overview of text-independent speaker recognition: From features to supervectors. *J. of Speech Comm.* 2010; 52(1): 1-30.
- [20] Uma Rani K, Mallikarjun SH. GMM Classifier for Identification of Neurological Disordered Voices Using MFCC Features. *IOSR J. of VLSI and Sig. processing*. 2015 Mar-Apr.; 5(2): 44-51.
- [21] Mammone RJ, Zhang X, Ramachandran RP. Robust Speaker Recognition - A Feature-based Approach. *IEEE Signal Processing Magazine*. 1996 Sept.: 58-71.
- [22] Gunn SR. Support Vector Machine for Classification and Regression. *Technical Report*. School of Electronics and Computer Science University of Southampton. 1998 May: 53.
- [23] Dhanalakshmi P, Palanivel S, Ramalingam V. Classification of audio signals using SVM and RBFNN. *Expert Systems with Applications*. 2009 April; 36(3): 6069-75. <http://dx.doi.org/10.1016/j.eswa.2008.06.126>
- [24] Tongeri R, Pallella D. An Overview of Speaker Identification: Accuracy and Robustness Issues. *IEEE Circuits and Sys. Mag.* 2011: 23-61.
- [25] Reynolds D. Speaker identification and verification using Gaussian mixture speaker models. *J. of Speech Comm.* 1995; 17: 91-108. [http://dx.doi.org/10.1016/0167-6393\(95\)00009-D](http://dx.doi.org/10.1016/0167-6393(95)00009-D)
- [26] Reynolds DA, Quatieri TF, Dunn RB. Speaker Verification Using Adapted Gaussian Mixture Models. *J. of Digital Signal Processing*. 2000 Jan; 10(1): 19-41.
- [27] Rouzbahani HK, Daliri MR. Diagnosis of Parkinson's Disease in Human Using Voice Signals. *J. of Basic and Clinical Neuroscience*. 2011; 2(3): 12-20.