IOSR Journal of VLSI and Signal Processing (IOSR-JVSP)
Volume 5, Issue 2, Ver. I (Mar. - Apr. 2015), PP 44-51
e-ISSN: 2319 – 4200, p-ISSN No. : 2319 – 4197
www.iosrjournals.org

GMM Classifier for Identification of Neurological Disordered Voices Using MFCC Features

K. Uma Rani[1], Mallikarjun S Holi[2]
[1]*Department of Biomedical Engineering and Research Center, B.I.E.T, Davangere, India.*
[2]*Department of Electronics and Instrumentation, University B.D.T College of Engineering Davangere,India.*

Abstract : *Automatic detection of neurological disordered subjects voice mostly relies on parameters extracted from time-domain processing. The calculation of these parameters often requires prior pitch period estimation; which in turn depends heavily on the robustness of pitch detection algorithm. In the present work cepstral-domain processing technique which does not require pitch estimation has been adopted to extract the features of voice signal. The Mel frequency cepstral coefficients (MFCCs) are computed using two methods; the fast Fourier transform (FFT) and the linear predictive coding (LPC) method. The cepstral parameters estimated from these methods are used as features to classify normal subject voice from neurologically disordered subject's voice using Gaussian mixture model (GMM). The results of the two methods are compared, and it is found that the accuracy of LPC-MFCC based GMM classifier is 89.55% compared to FFT-MFCC based GMM classifier which is giving an accuracy of classification of 88.5%.*
Keywords *- Fast Fourier Transform, Gaussian mixture model, Linear prediction coefficient, Mel frequency cepstral coefficient.*

I. Introduction

The voice problems may be caused by abnormal control, coordination, or strength of voice box muscles due to an underlying neurological disease such as; stroke, Parkinson's disease (PD), cerebral demyelination, amyotrophic lateral sclerosis. The change in voice is one of the symptoms for neurological disorder and dysphonia can be the first sign of a neurological disorder. The voices of these neurologic disorder tend towards either constancy or variability of phonation (dysphonia) depending upon whether or not the pathophysiology of the disease produces relatively steady or relatively fluctuating abnormal laryngeal or respiratory muscle movements. Neurological voice problems are primarily diagnosed via patient history and physical examination [1]. These physical examinations are invasive and cause significant discomfort to the patient. Hence numerous laboratories worldwide concentrate on diagnostic support methods based on acoustic voice analysis which, combined with classification methods provide the development of an expert aided system for the detection of speech system pathologies and also serves as a good automatic screening system [2], [3], [4], [6], [7], [8], [9], [10]. The traditional measurement methods to characterize the speech signal includes F0 (the fundamental frequency or pitch of vocal oscillation), absolute sound pressure level (indicating the relative loudness of speech), jitter(the extent of variation in speech F0 from one vocal cycle to other), shimmer (the extent of variation in speech amplitude from cycle to cycle), and noise-to-harmonics ratios (the amplitude of noise relative to tonal components in the speech). The earlier studies have shown variations in all these measurements for comparison of healthy controls to PD patients, indicating that these could be useful measures in assessing the extent of disorder in voice [1], [4], [9]-[23] . Studies have shown variations in all these measurements for comparison of healthy controls to PD patients, indicating that these could be useful measures in assessing the extent of dysphonia [2], [7], [8]-[10], [12], [14], [15], [23]. Also for these measurements long duration of signal is required, which is sometimes very difficult to collect them from voice affected patients. An alternate to this is the frequency domain analysis which requires less data and gives more information [24], [25]. In the earlier study [12], time domain features and the classification accuracy may be due to the fact that the measures considered are depending on the pitch measurements. Hence in the present work the well-known MFCCs; a cepstral domain measure is used as an alternative, as they do not show the same dependency on pitch of the signal. Here two methods has been used to extracts the MFCCs; the FFT based MFCC and LPC based MFCC and an attempt of evaluation of these two methods has been exercised on the voice samples [26].

II. Materials And Methods

2.1 Data Collection

The database containing total 182 phonations of sustained vowel /ah/ was collected from both male (62.72 ± 8.0 yrs) and female (65.19 ± 8.8 yrs) subjects suffering from different neurological disorders which include PD, cerebellar demyelination, and stroke. The disordered patient data were collected from outpatient wing of Neurology Department, J.S.S. Hospital, Mysore. The controlled group consist of 98 phonations of sustained vowel /ah/ from both male and female subjects who were not complaining any voice problems and their age matching with that of patient group. For this study, approval is obtained from the hospital 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 [27], [28]. 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 1.5 sec duration was selected for the acoustic analysis. All the recordings were done in mono-channel mode and saved in WAVE format on the hard disk and acoustic analysis were done on these recordings.

2.2 Feature Extraction

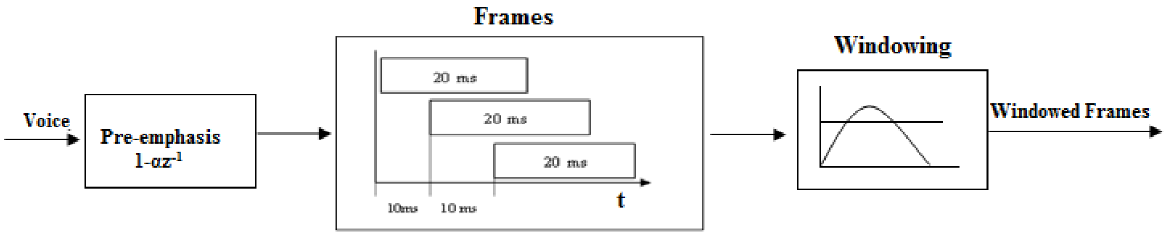

Fig.1 Framing of the voice signal.

The foremost step involved in feature extraction is to extract vectors of features which are uniformly spaced in the time domain based voice sample waveform. For this the framing [29] of the waveform is done as shown in fig. 1.

2.2.1 Pre- emphasis

The speech production system of human has the tendency of attenuating the high frequencies, hence to emphasise on the higher frequencies a 1st order high pass filter with filter function given by eq. (1) is used.

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

Where $x(t)$ is the input voice and $y(t)$ is the output.

2.2.2 Framing

The time-domain waveform is divided into overlapping fixed duration segments called frames. Here frames of 20 ms with 10 ms overlap is considered as shown in fig.1. [26], [30].

2.2.3 Windowing

The framing operation has a rectangular window effect which will generate undesirable spectral artefacts. Thereby each frame is multiplied by a window function to smooth the effect by tapering each frame at the beginning and end edges (Hamming window). This tapered window function creates a smoother and less distorted spectrum.

2.2.4 MFCC Features

The acoustic measurement of voice can be carried out in two dominant methods. First, the parametric modelling approach which is used to develop a model which matches closely the resonant structure of the human vocal tract that produces the corresponding voice/speech sound. This is derived from LPC analysis. The second is the nonparametric modelling method; the basis for this method is the human auditory perception system. The FFT based MFCCs are used to have knowledge on the human auditory perception system [30]. The term mel refers to a kind of measurement related to perceived frequency scale. The mapping between the real frequency scale (Hz) and the perceived frequency scales (mels) is approximately linear below 1 kHz and logarithmic at higher frequencies. The bandwidth of the critical band varies according to the perceived frequency.

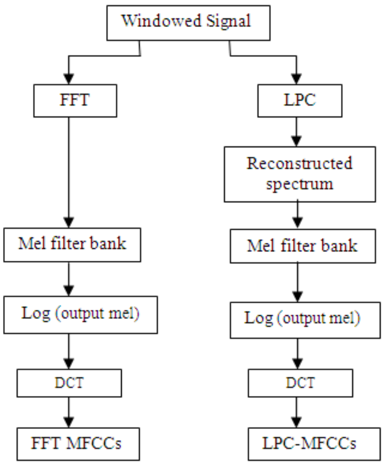

Fig.2: Extraction of MFCC features using FFT and LPC method

The MFCC parameters were calculated for both normal and neurological subjects for a dimension of 13. Figure 2 shows the extraction of MFCC parameters. The left hand flowchart in the figure shows the calculation of the conventional FFT based MFCCs. The right hand flowchart shows the LPC based MFCCs method. In our earlier work [31], we have discussed in detail the methodology of conventional FFT-MFCCs parameters calculation. In this, work the LPC-MFCCs were computed using the method as shown in fig.2. Here the LPC spectral estimate is of the spectral envelope of an AR filter resulting from a 10th order LPC analysis. In order to provide a representation of the speech signal, which is, as similar as possible for the conventional FFT based MFCC method, reconstruction of the spectral envelope from the LPC coefficients are done and the residual signal energy is used to scale them back to their original energy level. Thirteen MFCCs were derived from the log of the mel bank outputs using the discrete cosine transform (DCT). Cepstral mean subtraction (CMS) was applied as a channel normalization technique [26], [30]. The variation of MFCCs from frame to frame of normal and neurological disordered voices is shown in fig.3 In the case of normal voice, 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.

Fig. 4 shows the spectrum of a voice signal. The most prominent difference between LPC and FFT spectral estimators is related to the way in which they describe spectral peaks and valleys. The LPC estimation of a spectrum which is a parametric representation yields a spectral envelope with good description of the peaks and valleys in the spectrum which describes the energy level of the signal. This spectral envelope marks the peaks of the formants of the voice frame as shown in fig. 4. Hence, it can be said that the LPC will enhance the formant energy which may be a predominant descriptive feature for identification of neurological disordered voice [30]. Hence, in the present work both FFT based MFCCs and LPC based MFCCS are considered as the features to GMM classifier for a comparative study. Figure 5 shows an important observation in the power spectrum of the collected data sample, that is, the formants energy of neurological disordered voice is more compared to normal voice. This may be due to the existence of vocal pathology; the subject has to use more amount of energy to produce the desired level of voice signal.

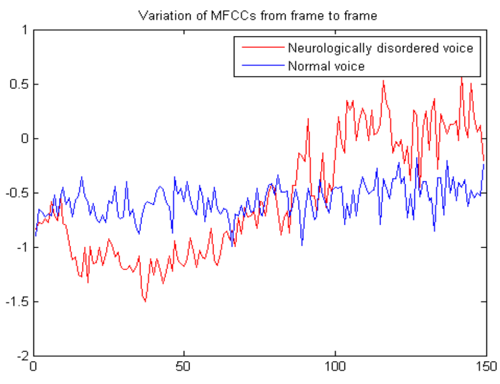

Fig.3 Variation of MFCCs from frame to frame

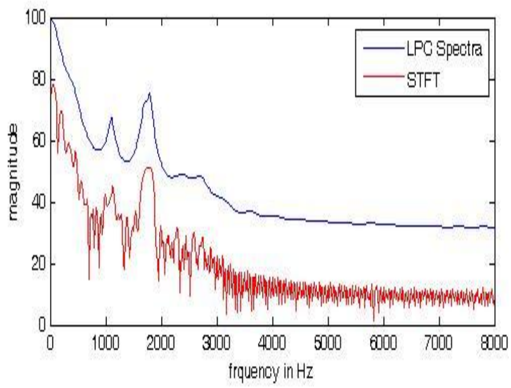

Fig. 4 Spectrum of the voice signal (ah)

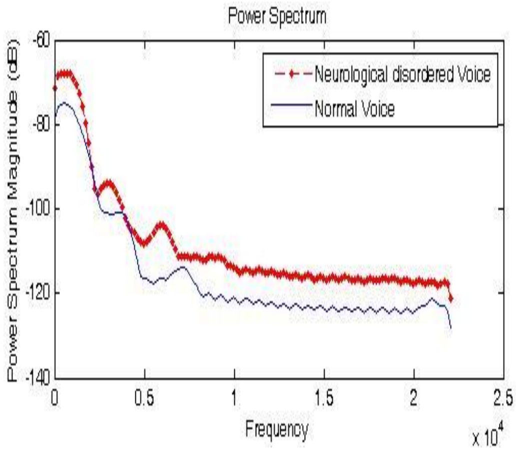

Fig. 5 Power spectrum of Normal and Disordered voice

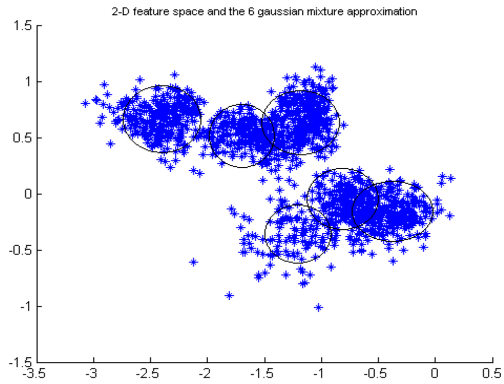

Fig.6. Scatter plot of 2-D cepstral vector by means of Gaussian mixture.

2.3 GMM classifier

A GMM is a parametric probability density function represented as a weighted sum of Gaussian component densities. GMMs are commonly used as a parametric model of the probability distribution of continuous measurements or features in a biometric system, such as vocal-tract related spectral features in a speaker recognition system [32].

A Gaussian mixture model is a weighted sum of M component Gaussian densities as given by the equation,

$$P(x|\ \lambda) = \sum w_i g(x|\ \mu_i \Sigma_i) \qquad (2)$$

where x is a D-dimensional continuous-valued data vector (i.e. measurement or features), w_i , i=1,…..M , are the mixture weights, and $g(x \mid \mu_i \Sigma_i)$, i=1, …..,M are the component Gaussian densities. Each component density is a D-variant Gaussian function of the form,

$$g(x|\ \mu_i \Sigma_i) = \frac{1}{2\pi^{D/2}|\Sigma_i|} \exp\left\{-\frac{1}{2}\ (x-\mu_i)'\right\} \Sigma_i^{-1}(x-\mu_i) \qquad (3)$$

With mean vector μ_i and covariance matrix Σ_i. the mixture weights satisfy the constraint that $\sum_{i=1}^{M} w_i$=1.

The mean vectors, covariance matrices, and mixture weights from all component densities parameterize the complete GMM. The notation collectively represents these parameters [30].

$$\lambda = \{w_i, \mu_i, \Sigma_i\} \qquad i = 1, \ldots\ldots\ldots,M \qquad (4)$$

GMMs are often used in biometric systems, most notably in speaker recognition systems, due to their capability of representing a large class of sample distributions. One of the powerful attributes of the GMM is its ability to form smooth approximations to arbitrarily shaped densities. The use of a GMM for representing feature distributions in a biometric system may also be motivated by the intuitive notion that the individual component densities may model some underlying set of hidden classes. For example, in speaker recognition, it is reasonable to assume the acoustic space of spectral related features corresponding to a speaker's broad phonetic events, such as vowels, nasals, or fricatives. These acoustic classes reflect some general speaker dependent vocal tract configurations that are useful for characterizing speaker identity. The spectral shape of the i^{th} acoustic class can in turn be represented by the mean μ_i of the i^{th} component density, and variations of the average spectral shape can be represented by the covariance matrix Σ_i. because all the features used to train the GMM are unlabeled, the acoustic classes are hidden in that the class of an observation is unknown. Assuming independent feature vectors, the observation density of feature vectors drawn from these hidden acoustic classes is a Gaussian mixture [33]. MFCCs follows GMM distribution is as shown in fig.6. Hence, D-dimensional MFCCs can be modelled by GMM model of M-mixtures.

2.3.1 Maximum Likelihood Parameter Estimation

Given training vectors and a GMM configuration, it is necessary to estimate the parameters of the GMM, λ, which in some sense best matches the distribution of the training feature vectors. There are several techniques available for estimating the parameters of a GMM [4]. By far the most popular and well-established method is maximum likelihood (ML) estimation. The aim of ML estimation is to find the model parameters, which maximize the likelihood of the GMM given the training data. For a sequence of T training vectors $X = \{x_{1} \ldots x_{t}\}$, the GMM likelihood, assuming independence between the vectors, can be written as,

$$p(X|\lambda) = \prod_{t=1}^{T} p(x_t|\lambda) \quad (5)$$

Unfortunately, this expression is a non-linear function of the parameters λ and direct maximization is not possible. However, ML parameter estimates can be obtained iteratively using a special case of the expectation-maximization (EM) algorithm [34].

The basic idea of the EM algorithm is, beginning with an initial model λ, to estimate a new model λ', such that $(X|\lambda') \geq p(X|\lambda)$. The new model then becomes the initial model for the next iteration and the process is repeated until some convergence threshold is reached. The initial model is typically derived by using some form of binary vector quantization estimation.

On each EM iteration, the following re-estimation formulas are used which guarantee a monotonic increase in the model's likelihood value,

Mixture weights $$w'_i = \frac{1}{T}\sum_{t=1}^{T}(i|x_t,\lambda) \quad (6)$$

Mean $$\mu'_i = \frac{\sum_{t=1}^{T}\Pr(i|x_t,\lambda)\lambda_t}{\sum_{t=1}^{T}\Pr(i|x_t,\lambda)} \quad (7)$$

Variance (diagonal covariance) $$\sigma'^2_i = \frac{\sum_{t=1}^{T}\Pr(i|x_t,\lambda)x_t^2}{\sum_{t=1}^{T}\Pr(i|x_t,\lambda)} - \mu'^2_i \quad (8)$$

2.3.2 MFCC-based GMM method

The voice samples were analyzed with 20 ms interval with a overlapping of 10 ms with the previous frame and multiplied by a hamming window. A 13 dimension FFT based MFCCs were extracted and fed into a GMM-based detector enabling a final decision about the absence or presence of pathology. The GMM model for normal and neurologically disordered subjects voices (i.e. λ_N and λ_D) having 4, 6, and 8 mixtures were trained separately with the expectation-maximization (EM) algorithm to determine the model parameters such as mean vectors, covariance matrices, and mixture weights [35]. During the testing phase the log-likely hood of the test feature vector X={x_1, x_2, ... x_T) is computed for both normal and neurologically disordered subjects voices models. Log likely hood ratio (LLR) for feature vector X is given by the equation [35] and this was repeated for LPC based MFCCs also

$$\Lambda(X) = \log\left[p\left(\frac{X}{\lambda_N}\right)\right] - \log\left[p\left(\frac{X}{\lambda_D}\right)\right] \quad (9)$$

The flow of the training and testing of the GMM model is as shown in fig.7. The histogram of the LLR estimated from normal and pathological voices after training process is shown in fig.8. The decision threshold Λ_{TH} is then set to adjust the tradeoff between rejecting pathological voices (false rejection) and accepting normal voices (false acceptance). The log- likely hood ratio is compared with Λ_{TH} and the voice is said to be normal if $\Lambda(X) > \Lambda_{TH}$ and disordered if $\Lambda(X) < \Lambda_{TH}$. Fig. 9 shows false acceptance and false rejection plots versus threshold Λ_{TH} . Both lines cross over the equal error rate point (EER).

III. Results And Discussions

The performance of the system was assessed by averaging the results obtained from fivefold cross-validation scheme [29]. Table 1 shows FFT-MFCC based GMM classifier confusion matrix, accuracy, sensitivity and specificity for different number of the Gaussian mixtures. Specificity and sensitivity means the test's ability to identify negative and positive results, respectively. The accuracy is the proportion of true results (both true positives and true negatives) in the dataset. The GMMs were trained using 4, 8, and 16 mixtures. The classification performance for the different mixtures is tabulated in table 1. It can be observed that the classifier performance for 8-mixture configuration is found to be 88.5%, and it is observed that as the mixtures are increased, the classifier goes into saturation. The equal error rate (EER) of the MFCC-based GMM method is shown in Fig.9. Figure 10 and 11 shows the ROC curve and the DET curve respectively when MFCC-based GMM method shows the best accuracy. The experimentation was repeated for LPC-MFCCs features as input to the GMMs for different mixtures and is tabulated in table 2. It can be observed that the classification accuracy for 8 mixtures is enhanced to 89.55% compared to the MFCCs based GMM, which is 88.5%.

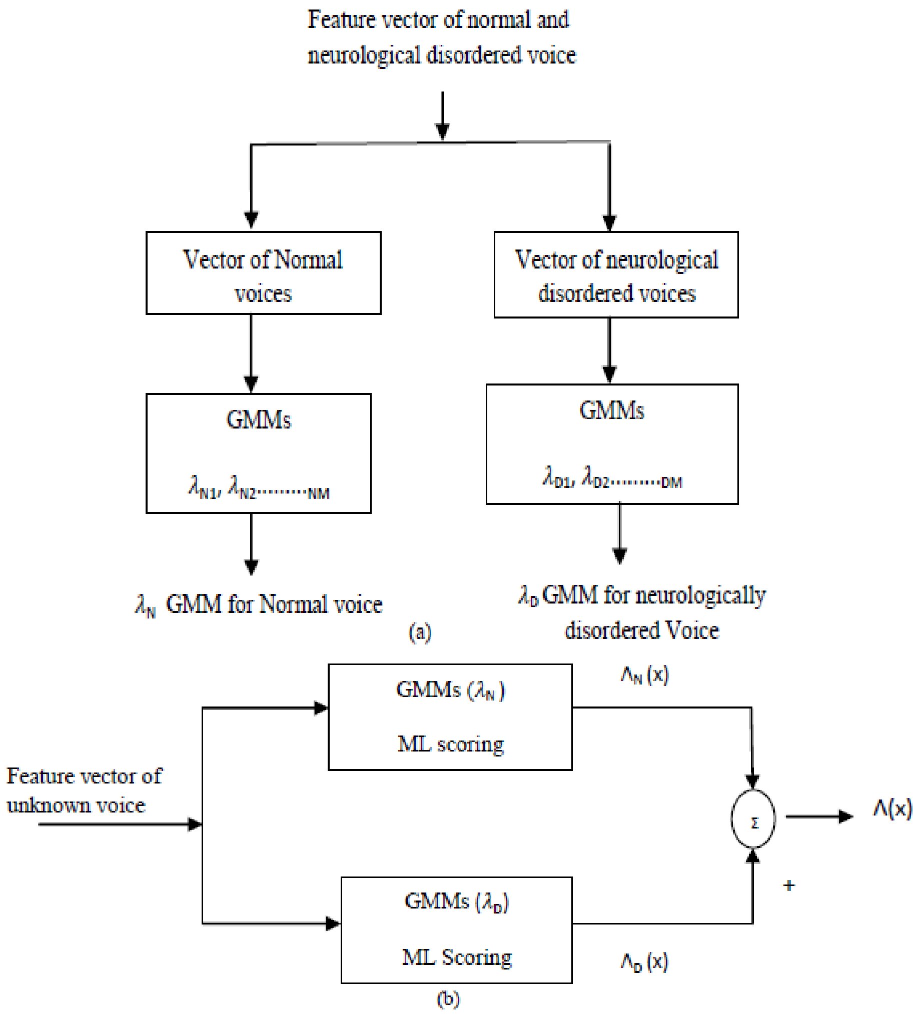

Fig. 7. (a) GMM training framework. **(b)** GMM testing framework

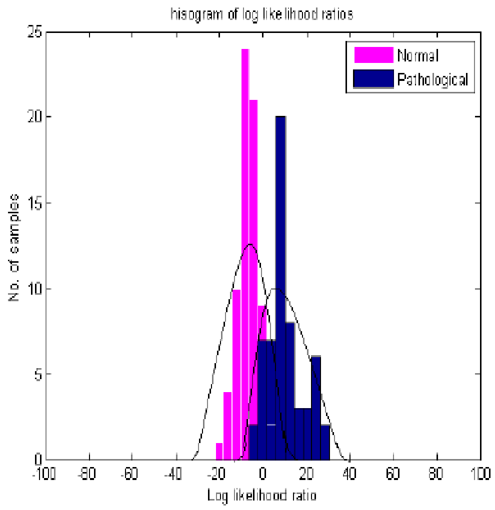

Fig. 8. LLR histogram estimated from normal and pathological (left). Voice in the training procedure.

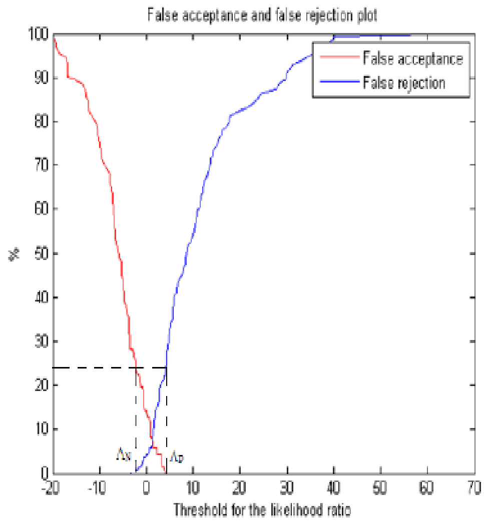

Fig.9. Cumulative false rejection (right) and false acceptance Λ_N and Λ_P indicate the thresholds of LLR estimate each GMM for normal and pathological voices in the t

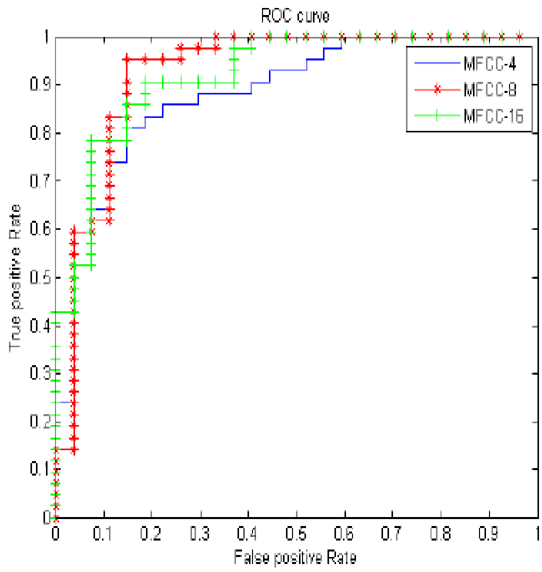

Fig.10 ROC curves for GMM with mixtures 4,8,16.

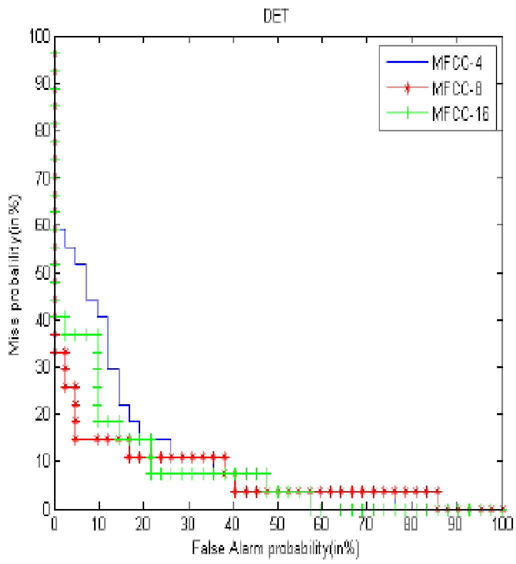

Fig.11. DET curves for GMM with mixtures4,8,16.

Table 1: Performance Of FFT-MFCC Based Gmm Classifier

Method	Confusion Matrix	Sensitivity (%)	Specificity (%)	Accuracy (%)
GMM-4 mixtures	90.5 9.5 25.9 74.1	90.5	74.1	82.3
GMM-8 mixtures	**88.1 11.9** **11.1 88.9**	**88.1**	**88.9**	**88.5**
GMM-10 mixtures	97.6 2.38 22.2 77.7	97.6	77.7	87.1
GMM-12 mixtures	92.9 7.14 18.5 81.5	92.9	81.5	87.2
GMM-16 mixtures	100 0 25.9 74	100	74	87

The best performance is highlighted in bold.

Table 2: Performance Of LPC-MFCC Based Gmm Classifier

Method	Confusion Matrix	Sensitivity (%)	Specificity (%)	Accuracy (%)
GMM-4 mixtures	90.47 9.5 18.5 81.5	90.5	81.5	85.97
GMM-8 mixtures	**97.61 2.38** **18.5 81.5**	**97.6**	**81.5**	**89.55**
GMM-10 mixtures	95.23 42.76 18.51 81.48	95.2	81.48	88.36
GMM-12 mixtures	97.6 2.38 22.2 77.7	97.6	77.7	87.7
GMM-16 mixtures	100 0 33.3 66.6	100	66.6	83.3

The best performance is highlighted in bold.

IV. Conclusions

The MFCC based GMM classifier has shown to be a good classifier for detection of voice disorder showing a classification accuracy of 88.5 % , but it can be observed that the classification accuracy has been enhanced to 89.55% for 8 mixture LPC-MFCC based GMM showing an improvement of 1%. Hence, it can be concluded that LPC-MFCC GMM is a better classifier for classifying neurological disorder voice from normal voices. There is a scope for the future work for implementation of multi class classifier to further classify the different neurological disorders.

Acknowledgment

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

References

[1] R.J.Baken, R.F.Orlikoff, Clinical measurement of speech and voice (2^{nd} edition, singular Thomson learning, 2000).
[2] P. Henriquez, J. B. Alonso, M. A. Ferrer, C. M. Travieso, J. I. Godino-Llorente, and F. Diaz-de-Maria, Characterization of healthy and pathological voice through measures based on nonlinear dynamics, IEEE Trans. Audio, Speech, Lang. Process, vol., 17,no.6, 2009,1186–1195.
[3] Nicolas, Saenz-Lechon, Juan I. Godino-Llorente,Vıctor Osma-Ruiz , Pedro Gomez-Vilda Methodological issues in the development of automatic systems for voice pathology detection , J.of Biomed. Sig.Proc. and Cont., vol. 1, no.2, 2006, 120–128.
[4] Stefan Hadjitodorov, Petar Mitev, A Computer System For Acoustic Analysis Of Pathological Voices And Laryngeal Diseases Screening, J.of Med. Eng.Phy., vol. 24, no.6, 2002, 419-429.
[5] Paulo R. Scalassara, Carlos D. Maciel, And Jos_E C. Pereira Predictability Analysis Of Voice Signals, Analysing Healthy And Pathological Samples, Eng.in Med. and Bio. Mag., vol.28, no.5, 2009,30-34.
[6] Juli A N David Arias-Londo N O , Juan I. Godino-Llorente , Maria Markaki & Yannis Stylianou, On combining information from modulation spectra and mel-frequency cepstral coefficients for automatic detection of pathological voices, J. of Logopedics Phoniatrics Vocology, vol.36, no.2,2011,60-69.
[7] Lindasalwa Muda, Mumtaj Begam & I. Elamvazuthi, Voice Recognition Algorithms using Mel Frequency Cepstral Coefficient (MFCC) and Dynamic Time warping (DTW) Techniques, J.of Comp., vol. 2, no.3, 2010, 138-143.
[8] M. A. Little, P. E. Mc Sharry, S. J. Roberts, D. Costello, and I. M. Moroz, Exploiting nonlinear recurrence and fractal scaling properties for voice disorder detection, Biomed. Eng. [Online]. vol. 6, no. 23, 2007.
[9] Max A. Little., Patrick E. McSharry, Eric J. Hunter, Jennifer Spielman, and Lorraine O. Ramig, Suitability of Dysphonia Measurements for Telemonitoring of Parkinson's Disease, IEEE Trans. Biomed. Eng., vol. 56, no. 4, 2009, 1015- 1022,.
[10] Goetz, C.G., Stebbins, G.T., Wolff, D., DeLeeuw, W., Bronte-Stewart, H., Elble, R., Hallett, M., Nutt, J., Ramig, L., Sanger, T., Wu, A.D., Kraus, P.H., Blasucci, L.M., Shamim, E.A., Sethi, K.D., Spielman, J., Kubota, K., Grove, A.S., Dishman, E., Taylor, C.B., Testing objective of motor impairment in early Parkinson's disease: Feasibility study of an at-home testing device, J. Mov. Disorders, vol.24, 2009, 551-556,.
[11] M.Hariharan, M. P. Paulraj, Sazali Yaacob ,Time-Domain Features And Probabilistic Neural Network for the Detection Of Vocal Fold Pathology, Malaysian J. of Comp. Sci., vol. 23, no. 1, 2010, 60-67.
[12] Uma Rani.K and Mallikarjun.S.Holi, Analysis of Speech Characteristics of Neurological Diseases and their Classification , Proc.of IEEE Inter. Conf. on Comp. Comm. & Networking Technologies (ICCCNT), 2012 Coimbatore, India, 1-6.
[13] Antoine Giovanni, Ping Yu, Joana Revis, Marie-Dominique Guarella, Bern a rd Teston, Maurice Ouaknine, "Objective dysphonia evaluation using the EVA® workstation. A review", Fr ORL, vol. 90, , 2006, 183-19.
[14] Mehmet Fatih CAGLAR, Bayram CETISLI, Inayet Burcu TOPRAK, Automatic Recognition of Parkinson's Disease from Sustained Phonation Tests Using ANN and Adaptive Neuro-Fuzzy Classifier, J. Eng. Sci. and Design, vol.1, no.2, 2010, 59-64.
[15] 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, 2004,380–384.
[16] B. Boyanov and S. Hadjitodorov, Acoustic analysis of pathological voices, IEEE Eng. Med. Biol. Mag., 1997, 74–82.
[17] D. Michaelis, H. Gramss, and W. Strube, Glottal-to noise ratio- a new measure for describing pathological voices,"Acustica/Acta Acustica, vol. 83,1997,700–706.
[18] H. Kasuya, S. Ogawa, K. Mashima, and S. Ebihara, Normalized noise energy as an acoustic measure to evaluate pathologic voice, J. Acoust. Soc. Amer., vol. 80, no. 5, 1986,1329–1334.
[19] G. de Krom, A cepstrum-based technique for determining a harmonic- to-noise ratio in speech signals, J. Speech Hearing Res., vol. 36, 1993, 254–266.
[20] C. Manfredi, Adaptive noise energy in pathological signals, IEEE Trans. Biomed. Eng., vol. 47, 2001,1538–1543.
[21] W. Winholtz, Vocal tremor analysis with the vocal demodulator, J. Speech Hearing Res., vol. 35, 1992, 562–573.
[22] E. Yumoto, Y. Sasaki, and H. Okamura, Harmonics-to-noise ratio and psychophysical measurement of the degree of hoarseness, J. Speech Hearing Res., vol. 27, 1984,2–6.
[23] Petra Zwirner, Thomas Murry, Gayle E. Woodson, Phonatory function of neurologically impaired patients, J. of Comm. Disorders vol. 24, no.4, 1991, 287–300.
[24] Juli A N David Arias-Londo N O , Juan I. Godino-Llorente ,Maria Markaki & Yannis Stylianou, On combining informationfrom modulation spectra and mel-frequency cepstral coefficients for automatic detection of pathological voices, J. of Logopedics Phoniatrics Vocology, vol.36, no.2, 2011,60-69.
[25] Pravena.D, Dhivya.S, DurgaDevi.A , Pathological voice recognition for vocal fold Disease, Internal. J.of Comp. App., vol.47, no.13, 2012, 31-37.
[26] Febe de Wet , Bert Cranen , Johan De Veth , Loe Boves, A comparison of LPC and FFT-based acoustic features for noise robust ASR, in Proc. of Eurospeech 2001,868-868.
[27] Luis M. T. Jesus, Anna Barney, Ricardo Santos, Janine Caetano, Juliana Jorge, Pedro Sa Couto, Universidade de Aveiro's Voice Evaluation Protocol, in Proc. of Interspeech 2009, Brighton, UK, 2009, 971-974.
[28] Andreas Maier , Stefan Wenhardt , Tino Haderlein , Maria Schuster , Elmar Noth, A Microphone-Independent Visualization Technique for Speech Disorders, in Proc. of Interspeech 2009, Brighton, UK, 2009, 951-954
[29] Roberto Tongeri, Deniel Pullella, An Overview of Speaker Identification: Accurancy and Robestness Issues IEEE Circuits and Sys.Mag., 2011, 23-61.
[30] Tomi Kinnunen, Haizhou Li,"An overview of text-independent speaker recognition: From features to supervectors" J.of Speech Comm. vol.52,no.1,2010,2010.
[31] Uma Rani K and Mallikarjun S. Holi, Automatic Detection of Neurological Disordered Voices Using Mel Cepstral Coefficients and Neural Networks, in Proc. of IEEE-EMBS Special Topic Conference on Point-Of-Care (POC) Healthcare Technologies, Bangalore, India, 2013,76-79.
[32] Reynolds, D, Speaker identification and verification using Gaussian mixture speaker models. J. of Speech Comm. vol. 17, 1995, 91–108.
[33] Reynolds, D.A., Rose, R.C., Robust Text-Independent Speaker Identification using Gaussian Mixture Speaker Models, IEEE Trans. on Acoustics, Speech, and Signal Processing vol.3, no.1, 72-83.
[34] Reynolds, D.A., Quatieri, T.F., Dunn, R.B., Speaker Verification Using Adapted Gaussian Mixture Models J. of Digital Sigal Processing., vol. 10 no.1, 2000,19-41.
[35] M de Oliveira Rosa, JC Pereira, M Grellet, Adaptive estimation of residual signal for voice pathology diagnosis, IEEE Trans. Biomed. Eng. vol.47,no.1, 2000, 96-104.