ISSN (Online) 2394-6849

International Journal of Engineering Research in Electronics and Communication Engineering (IJERECE)
Vol 4, Issue 9, September 2017

Novel Speaker Recognition System using GMM

[1] V.Srinivas,[2] Ch.Santhi Rani, [3] P.Hemakumar
[1][2] Swarnandhra Institute of Engineering and Technology, Narsapur
[3] D.M.S&S.V.H Engineering College, Machilipatnam

Abstract:-- **A text dependent speaker recognition system can be developed by using MFCC and Vector Quantization in a controlled environment. But MFCC with Vector Quantization cannot be useful for developing a text independent speaker recognition system and also does not provide accurate results. So, the main aim of this paper is to develop a text independent speaker recognition system using MFCC and GMM along with NLMS adaptive filter, such that the input utterance is given in real time using a microphone. NLMS adaptive filter is used to reduce the noise in the speech signal and then passed through the feature extraction phase. It is developed as Text- independent Speaker Recognition System with 50 speakers and also uses the locally recorded database for training. The performance of the proposed system tested in real time using Adaptive filter based on the log likelihood scores.**

Index Terms **— NLMS Adaptive Filter, Vector Quantization, Gaussian Mixture Model (GMM), FFT**

1. INTRODUCTION

One of the biometric applications is speaker authentication that can verify the unknown's voice. The motive of the Speaker Recognition System is to identify the person from MFCC features of his/her voice by comparing with trained data which can be useful in most security applications.

The problem in existing speaker recognition system is its text dependency that is speaker need to provide the same text or same word in both training and testing and also the number of speakers is the constraint. So, the non-parametric model may not give the accurate results. So, the GMM which is a parametric model for speaker identification is the empirical observation that a linear combination of Gaussian basis function is capable of representing a large class of sample distributions. Generally we recorded the speech signal in different environmental conditions, so the undesired features of a speaker results and this leads to degrade in the performance of the system drastically. That is, recording the unknown speaker voice through microphone dynamically and comparing the features with features of reference model in trained database leads to incorrect results especially in noisy environments.

So, the objective of this paper is developing a text independent system using MFCC and GMM such that it does not depend on the particular text to be spoken in training and testing phases and also to work with large data set with large number of speakers such that it uses the log likelihood detector algorithm for making decision of accepting or rejecting the speaker at the final stage. These experiments are performed in different environmental states for that purpose the noise cancellation can be done by using Normalised Least Mean Square (NLMS) Adaptive Filter which reflects to enhance the performance. Moreover, the GMM does not work properly under the mismatch conditions of training and testing. So, developing the speaker recognition system in real time through microphone is the major challenging task. The complete system involves MFCC and GMM model is developed through MATLAB.

The complete paper describes about the survey of literature on speaker recognition system using GMM in the next section. Also in the next subsequent sections, it deals about the pre-processing of the speech signal using adaptive filter such as NLMS adaptive filter. Then the description about GMM modelling with the proposed system and its implementation. Finally, described about the experiments conducted and the performance evaluation with results

2. REVIEW OF LITERATURE

S G Bagul , R.K.Shastri the paper was based on statistical model like Gaussian Mixture Model (GMM) and Features extracted from the speech signals (MFCC) and they concluded with FFT based algorithms are better compared to Mel scale based methods and Gaussian mixture models provides robust speaker recognition also computational less expensive in real time.

Rania Chakroun, Leila Beltaïfa Zouari1, Mondher Frikha, and Ahmed Ben Hamida were proposed reduced feature vector employing new information detected from the speaker's voice for performing text-independent speaker verification applications using GMM. They concluded that this will decrease the error rate and avoided the complicated calculations and gives the better results compared to baseline systems with GMM Models.

ISSN (Online) 2394-6849

International Journal of Engineering Research in Electronics and Communication Engineering (IJERECE)
Vol 4, Issue 9, September 2017

Sourjya Sarkar and K.Sreenivasa Rao proposed "Speaker Verification in noisy environment using GMM super vectors". They use combined approach of GMM- SVM. For that hybrid systems, the result for the noisy data was studied and they seen significant improvement in performance of the system was observed.

M.S.Sinith, AnoopSalim, GowriSankar K, Sandeep Narayanan K V, Vishnu Soman proposed "A Novel Method for Text-Independent Speaker Identification Using MFCC and GMM". They focussed on text independent speaker recognition system using GMM. The experiments conducted for various speech time durations and achieved high recognition rate.

JyotiDhiman, ShadabAhmad, Kuldeep Gulia proposed a paper which illustrates the implementation aspects of LMS and NLMS adaptive filters and the performance of these filters with respect to computational complexity and Signal to Noise Ratio.

"Improved Recognition Rate of Language Identification System in Noisy Environment" was a paper presented by Randheer Bagi, JainathYadav, K. SreenivasaRao .In this paper, they focussed on language identification in noisy environment and for that they used GMMs to train the models. Also they used Spectral Subtraction and Minimum Mean Square Error (MMSE) methods to reduce the noise from the speech signal. They compared the recognition rate of the system for the clean signal and noise suppressed signal.

"Speaker identification and verification using Gaussian mixture speaker models" was a paper proposed by Douglas A. Reynolds. This paper was focussed to achieve the high performance speaker identification and verification; these were evaluated using different speech databases such as TIMIT, NTIMIT and YOHO. It uses different levels of degradations, speech quality with noisy signal, clean speech, and telephonic speech.

3. GAUSSIAN MIXTURE MODEL (GMM)

There are two types of methods, Deterministic methods and statistical methods. Here we used the statistical method that is parametric model called Gaussian Mixture Model (GMM) for the identification of speakers.
The weighted sum of M component densities gives the Gaussian mixture density.

The Gaussian mixture density has the parameters mean vectors, covariance matrices and mixture weights, such that these can be represented by

$$\lambda = \{Pi, \bar{\mu}, \sum i\} \qquad i=1,\ldots\ldots..M.$$

There are two main reasons for using the GMM as representation of speaker identity.
Individual densities of a multi modal density like GMM may model underlying set of acoustic classes also such that these set of acoustic classes are used to characterize the speech utterance of the particular speaker that represents some phonetics such as vowels, nasals and fricatives.
Another important reason involves, GMM is able to represent large class of samples or its distributions by means of linear representation of Gaussian functions.

GMM is a parametric model; the maximum likelihood estimation is one of the methods available for estimating the parameters of GMM. The algorithm called Expectation Maximization (EM) can be used for estimation of ML parameter iteratively and these parameters maximize the likelihood of the GMM.

For to guarantee the improvement in model's likelihood value, the following formulas can be used,

$$\overline{p_1} = \frac{1}{T}\sum_{t=1}^{T} p(i/\bar{x}_t, \lambda)$$

$$\bar{\mu} = \frac{\sum_{t=1}^{T} p(i/\bar{x}_t,\lambda).\bar{x}_t}{\sum_{t=1}^{T} p(i/\bar{x}_t,\lambda)}$$

$$\bar{\sigma}^2 = \frac{\sum_{t=1}^{T} p(i/\bar{x}_t,\lambda).\bar{x}_t{}^2}{\sum_{t=1}^{T} p(i/\bar{x}_t,\lambda)} - \bar{\mu}^2$$

The M component Gaussian mixture density forming a GMM can be represented by $p(\vec{x}/\lambda)$.
We have T independent training vectors for a given sequence, for these vectors calculate the log likelihood scores and search for the maximum likelihood. The log likelihood can be computed as

$$\text{Log } p(X/\lambda) = \frac{1}{T}\prod_{t=1}^{T} p(\vec{x}/\lambda)$$

The system uses log-likelihood scores to know whether the claimed speaker is true or false. Thus for an input vector and a claimed speaker model λ_1, the likelihood score can be given by $p(x/\lambda 1)$.

So, based upon the log likelihood scores of the unknown speaker model and the trained speaker models, the decision

ISSN (Online) 2394-6849

International Journal of Engineering Research in Electronics and Communication Engineering (IJERECE)
Vol 4, Issue 9, September 2017

of accept or reject of the speaker can be made by considering a given threshold.

Selecting the order M of the mixture and initializing the parameters before EM algorithm plays a dominant role in the speaker modelling.

By using likelihood algorithm, likelihood scores for all speakers are calculated for the corresponding GMM model which forms the trained database. Finally, the likelihood score of unknown speaker compared with the trained database.

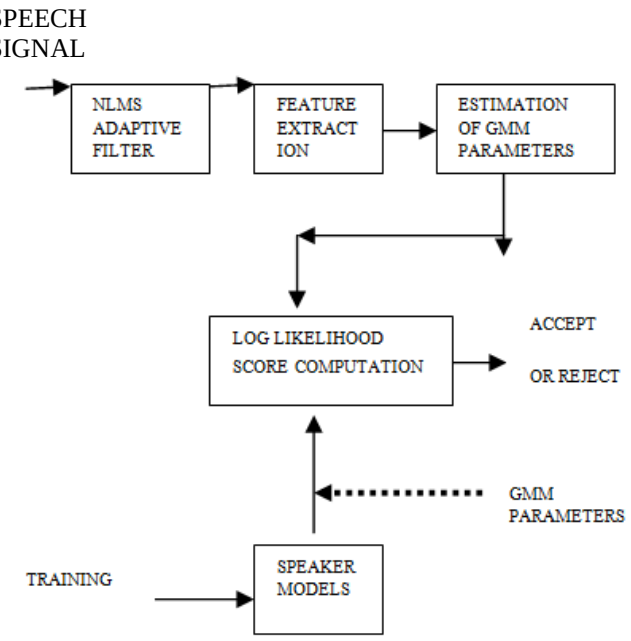

Fig : Block diagram of the proposed system

Normalized Least Mean Square (NLMS) Adaptive Filter:
The method of passing a signal corrupted with some additive noise is to allow passing through a adaptive filter which reduces or suppresses noise without effecting the actual speech signal. Moreover, the Adaptive filter is self adjusting the parameters automatically and does not need any knowledge of past data or characteristics.

Least Mean Square (LMS) filter is a kind of adaptive filter used to make a desired filter by finding filter coefficients that produces the least mean squares of the error signal. Finding the gradient of mean square error can be useful in modifying the weights. Based upon this value, the error would increase positively or negatively. so increasing or decreasing the weights depends upon polarity of the gradient. So, the Adaptive filter adjusts the parameters such that output approximates the unknown which in turn reduces or minimizes the error signal. Also the problem is under varying operating conditions, it leads to misleading of convergence in mean, this can be somehow overcome by NLMS Adaptive filter. Ofcourse, Adaptive filter can be used in many signal processing areas such as echo cancellation, signal prediction etc., but here we used Adaptive filter in this paper for Echo Cancellation.

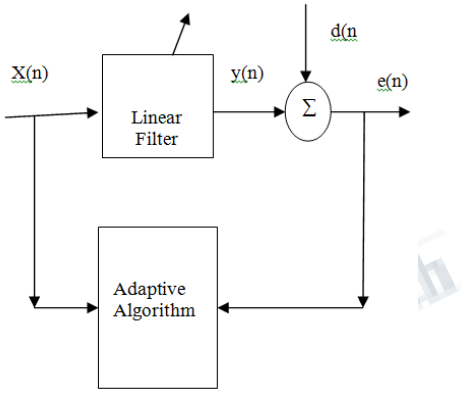

Fig 2. Typical Adaptive Filter

There are number of parameters in Adaptive filters that plays an important role in reducing or eliminating the noise but noise cancellation can be obtained by variation in step size and thus adaptive filter deals with noise cancellation for signal corrupted with white Gaussian noise [2]. For the small amount of changes in weight, it changes about the optimal weights and thus if variance changes large extent due to weights change the convergence in mean would be misleading. So, the proper selection of step size is very much needed in order to avoid the problem [2].

The main drawback of LMS algorithm is it is very sensitive to scaling of its input. The normalised Least mean square adaptive filter is another class of LMS algorithm that clearly overcome the drawback of LMS algorithm by normalising the power of the input. Also the range of step size and mean square deviation depends on the power of input signal; this makes clearly the NLMS adaptive filter can be effectively used than that of LMS Adaptive filter. The convergence speed of NLMS adaptive filter depends upon the statistics of input signal.
Also the experiments in noisy environments with using NLMS adaptive filter also conducted and also observed the

ISSN (Online) 2394-6849

International Journal of Engineering Research in Electronics and Communication Engineering (IJERECE)
Vol 4, Issue 9, September 2017

various performance metrics that shows the improvement in present system when compared to those previous experiments.

4. EXPERIMENTS

For this Text independent speaker recognition system using Gaussian mixture models, the experiments conducted using Matlab 7 language environment. Here about 50 numbers of speakers were trained and uses the locally recorded database. Here, experiments are conducted such that training and testing in the same environment.
Here, Mel Frequency Cepstral Coefficients (MFCC) is used to extract features from the speech utterance. These features are widely used in speaker recognition to obtain speaker specific information which gives frequency distribution of sounds along with vocal tract shape and length. For each frame and Every 10 ms, 12 MFCC together with log energy were calculated.

Experiments conducted in different environments with using NLMS adaptive filter. The effect of using the adaptive filter on the input speech utterance can be observed below. The plots of the speech signal with reduced noise and the speech signal with noise can be obtained using Mat lab.
The filtered signals and the speech signal without filtered can be observed as shown below.

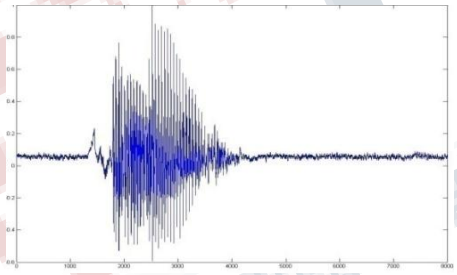

Fig 3: Speech signal without filter

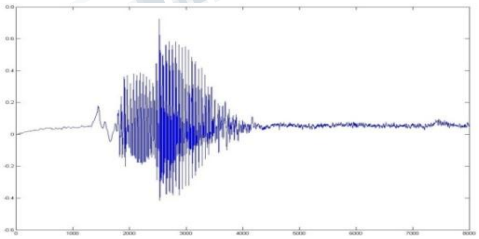

Fig 4: Speech signal with filter

5. RESULTS

In this system, the locally recorded database can be used for speaker recognition task and the speech signal was sampled at 8 KHz sampling frequency. In these experiments, the use of adaptive filter may slightly increase the system performance for different speakers. The speaker verification involves accepts or rejects the identity, the task can be achieved by using log likelihood scores of each speaker and also here the accuracy or recognition rate can be used to measure the performance of the system. The recognition rate for the speaker recognition system with NLMS adaptive filter is better compared to the same system without using NLMS adaptive filter.

The recognition rate of 96.96% was achieved for the system with adaptive filter; also this can be computed based on log likelihood scores. Moreover, the unknown speaker can be accepted as a true speaker when the log likelihood ratio is below the predetermined threshold, otherwise the speaker was rejected and does imply as a imposter.
The below table shows the log likelihood ratio and recognition rate for the speaker recognition system in real time with adaptive filter of the unknown speaker with the duration of the speech utterance is 5sec. The same utterance for 3 times by the same speaker.

Table 1.Output for speaker recognition system based on recognition rate.

Speaker	Log likelihood ratio	Recognition rate
1st time	5.59	96.73%
2nd time	7.68	95.56%
3rd time	5.18	96.96%

The below figure 5 shows the maximum likelihood score for the particular speaker. Similarly, we can plot the maximum likelihood scores for remaining speakers also such that the speaker can provide any text.

ISSN (Online) 2394-6849

International Journal of Engineering Research in Electronics and Communication Engineering (IJERECE)
Vol 4, Issue 9, September 2017

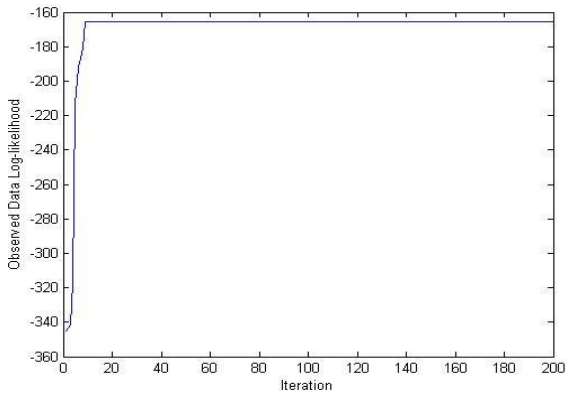

Fig 5: Plot of iteration vs log likelihood for the unknown speaker

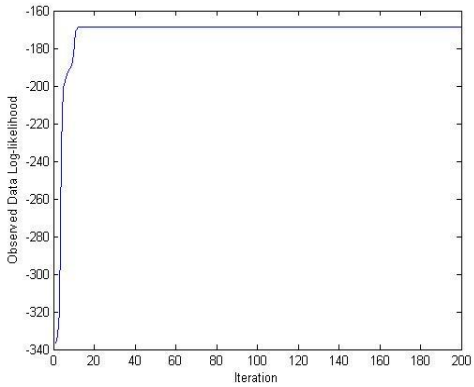

Fig 6: Plot of Iteration vs log likelihood for one of the speaker in the trained database.

Also the results obtained from different systems with mismatched conditions for training and testing and without adaptive filter were also observed. The recognition rate was decreases and also log likelihood ratio was also not within the threshold was observed, so there was significant degradation of the performance of the system in real time was clearly observed. But , if the speaker recognition system uses the same database for training and testing rather than real time system then the recognition rate was improved.
The below table shows the performance of the system in real time without any adaptive filter in terms of recognition rate

Table 3. Output for the system without adaptive Filter

Speaker	Log likelihood ratio	Recognition rate
1st time	>8	57 % with the result "Speaker does not found"
2nd time	>8	65% with the result "Speaker does not found"
3rd time	>8	55 % with the result "Speaker does not found"

Here, the recognition rate for the system with enhanced speech signal and system with noisy speech signal are compared as shown in table 1 and table 2. Also, the speech utterance of duration of 5sec is used for each speaker (3 times) and computed the log likelihood scores. So the task of recognizing the person in real time was somehow improved, that is shown in above tables.

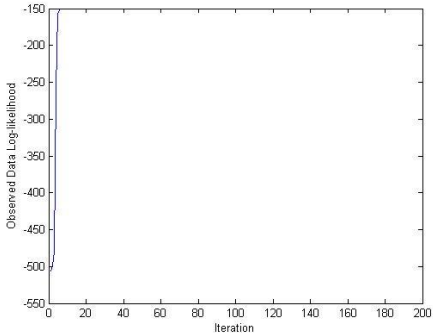

Fig 7. Plot of iteration vs log likelihood for an imposter

The log likelihood score for each iteration was computed. So for all 200 iterations , computed log likelihood scores and GMM parameters called mean, variance also computed. The plot of log likelihood scores for the 200 iterations is shown in the below figure.

6. CONCLUSION

This paper clearly shows the text independent speaker recognition system using GMM along with NLMS adaptive filter improves the performance, the recognition rate in adverse conditions. The recognition rate of 96.96% was achieved using this proposed method, when using the adaptive filter. However, the accuracy drastically reduces under mismatched conditions. Further the system can be improved by using voice activity detector (VAD) in the pre-processing stage and also using shifted MFCC in place of MFCC for feature extraction may increase the performance of the speaker recognition system.

7. REFERENCES

[1] M.S.Sinith, AnoopSalim, GowriSankar K, Sandeep Narayanan K V, Vishnu Soman "A Novel Method for Text-Independent Speaker Identification Using MFCC and GMM" ICALIP2010 ©2010 IEEE.

[2] Roshny Jose George "Design Of an Adaptive Filtering Algorithm for Noise Cancellation "International Research Journal of Engineering and Technology (IRJET), Volume: 02 Issue: 04 | July-2015.

[3] John Creighton and R.Doraiswami"Real Time Implementation Of anAdaptive Filter For Speech Enhancement" 2004 IEEE.

[4] Wenyong Lin "An improved GMM based clustering algorithm for efficient speaker identification" 2015 4th International Conference on Computer Science and Network Technology (ICCSNT 2015) ©2015 IEEE.

[5] Michael Lutter "Mel Frequency Cepstral Coefficients (feature extraction/ Mfcc)" The Speech Recognition Wiki25 November 2014.

[6] J.P.Campbell, "Speaker Recognition: A Tutorial", Proc. Of the IEEE, Vol 85,No. 9, September 1997,pp. 1437-1462.

[7] VibhaTiwari "MFCC and its applications in speaker recognition "International Journal on Emerging Technologies 1(1): 19-22(2010) ISSN : 0975-8364.

[8] Rania Chakroun, Leila BeltaïfaZouari, MondherFrikha, and Ahmed Ben Hamida " Improving Text-independent speaker recognition with GMM" 2nd International Conference on Advanced Technologies for Signal and Image Processing - ATSIP'2016 March 21-24, 2016, Monastir, Tunisia ©2016 IEEE.

[9] Prof.Vaishali M. Karne, Prof.Akhilesh Singh Thakur , Dr.VibhaTiwari " Least Mean Square (LMS) Adaptive Filter For Noise Cancellation "International Journal of Application or Innovation in Engineering & Management (IJAIEM) , ISSN 2319 – 4847.

[10] SourjyaSarkar, K.SreenivasaRao "Speaker Verification in Noisy Environment Using GMM Super vectors" © 2013 IEEE.

[11] Sheng Zhang, student Member, IEEE, Jiashu Zhang, and HongyuHan "Robust Variable Step-Size Decorrelation Normalized Least-Mean Square Algorithm and its Application to Acoustic Echo Cancellation" IEEE/ACM Transactions on Audio, Speech, and Language Processing.

[12] Xin-xing ling, Ling Zhan, Hong Zhao, Ping Zhou "Speaker Recognition System Using the Improved GMM-based Clustering Algorithm" ©2010 IEEE.

[13] Yuan Liu, Tianfan Fu, Yuchen Fan, YanminQian, Kai Yu "Speaker Verification with Deep Features"2014 International Joint Conference on Neural Networks (IJCNN) , July 6-11, 2014, Beijing, China, ©2014 IEEE.

[14] SourjyaSarkar, K. SreenivasaRao "Significance Of Utterance Partitioning In GMM-SVM Based Speaker Verification In Varying Background Environment "

[15] RandheerBagi, JainathYadav, K. SreenivasaRao "Improved Recognition Rate of Language Identification System in Noisy Environment" ©2015 IEEE.

[16] JYOTI DHIMAN, SHADAB AHMAD, KULDEEP GULIA " Comparison between Adaptive filter Algorithms (LMS, NLMS and RLS)"International Journal of Science, Engineering and Technology Research (IJSETR) Volume 2, Issue 5, May 2013, ISSN: 2278 – 7798.

[17] Douglas A. Reynolds "Speaker Identification and verification using Gaussian Mixture speaker models" speech communication 17 (1995), ELSEVIER