IJIRSET
International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY

e-ISSN: 2319-8753 | p-ISSN: 2320-6710

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 10, Issue 4, April 2021

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 7.512

9940 572 462 | 6381 907 438 | ijirset@gmail.com | www.ijirset.com

International Journal of Innovative Research in Science, Engineering and Technology (IJIRSET)

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 7.512|

|| Volume 10, Issue 4, April 2021 ||

DOI:10.15680/IJIRSET.2021.1004074

Automated Voice Recognition System for Speaker Emotion Classification

Mr.C.Senthil Kumar, T. Aravinth, K. Dayanandhan, J.Neejay Prasanth, M. Arun Prasath

Assistant Professor, Department of Information Technology, SNS College of Technology, Coimbatore, India

B.Tech, Department of Information Technology, SNS College of Technology, Coimbatore, India

B.Tech, Department of Information Technology, SNS College of Technology, Coimbatore, India

B.Tech, Department of Information Technology, SNS College of Technology, Coimbatore, India

B.Tech, Department of Information Technology, SNS College of Technology, Coimbatore, India

ABSTRACT: Over this decade the speech recognition plays an important role for speechmaker identification and identification of the various characteristics of a person involved in a particular section of the voice. The information obtained from those systems are used for interaction between the user and the machine. The emotion detection through face recognition needs to person's face captured in the sensor for the detection of the emotion who are involved in the session. But voice recognition can be done without the knowledge of the that person. The application is used in the client service, education and health care. The high accuracy system of this type can be used in the intelligence services to know the truthfulness in investigation. For this task various parameters of the audio is obtained as feature called MFCC. The GMM modelling is successfully modelled by the feature MFCC. In addition to that UBM it gives better result in emotion of the speechmaker.

KEYWORDS: Emotion recognition, GMM, GMM-UBM, MFCC.

I. INTRODUCTION

The needs of the user can be satisfied easily by understanding the category of the people the belong. The user category can be classified based on the age, gender and feeling. Machine-driven feeling characteristic classification is the tactic of characteristic human feeling by semi-automated approach. The foremost application of this technology is to help people with emotion recognition may be a relatively risen analysis house. Basically, the technology works best if it has multiple modalities in context. The foremostwork has been conducted on identifying recognition of facial expressions from image ad written expressions from the text, but the emotion recognition is not done so much. This application aims to provide applications in the customer service, education and health care. Data regarding the speaker's feeling is also a vital ingredient that ends up in the affluent activity information processing.

This system is based on the MFCC and GMM modelling. Over this decade the MFCC is widely used for voice based classification of the speaker. The dominance of the MFCC is due to its high performance and accuracy over various classifiers. The GMM is well good for the classification and gives comparatively higher accuracy. Along with UBM it shows better results. This gives the strength to build the system for the emotion recognition.

II. PROPOSED METHODOLOGY

The project involves two broad phases. The first phase is of the feature extraction and the second phase will be of modelling over the feature extracted. Here the MFCC serves us the reduction of noise and the interference so that system would of more effective. The UBM serves us the cancellation of over-fitting problem.

A. MFCC

The MFCC involves various steps that need to be performed before modelling. The initial step is pre-emphasis which is to separate the voice and blank spaces in the audio. The nextstep is framing of the signals. The signal is framed so that signal does not vary in the shorter scale. When the frame is large the signal varies abruptly but when the frame is shorter sample cannot be collected so that it cannot be used for modelling. The next step in the MFCC is to calculate the power spectrum of the frame, which is done in the human ears also to find which indicates frequencies in those frames. The human ear vibrates at different spots for different frequencies, depending on those vibrations on the

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 7.512|

|| Volume 10, Issue 4, April 2021 ||

DOI:10.15680/IJIRSET.2021.1004074

different location in cochlea we can denote what frequency is present. The next is to find the periodogram spectral estimate and sum of those periodogram bins to find how much energy is present in those frames. To make loudness in the audio logarithm is suggested. The final step is to compute the DCT (Direct Cosine Transform), it is motivated to decorrelates the energies which are overlapping.

B. GMM

The GMM (Gaussian Mixture Model) is a probabilistic model for normal distribution sub model of overall dataset. It allows model to learn from sub model automatically. Since sub model assignment is not known, it constitutes unsupervised learning.

The GMM is modelled by the use of the Maximum Likelihood and the EM (Estimation Maximization). Maximum likelihood is the process of maximizing the likelihood described by the model which is observed.

The EM is the optimization of the Maximum likelihood method. In this method random values for the missing data points are assigned and estimates the new set of data. Those values are used recursively to estimation.

C. GMM-UBM

The UBM (Universal Background Model) is associate improvement to the probabilistic model. It implements the MAP (Maximum A Posteriori) to reduce the overfitting caused in the Maximum Likelihood of the GMM modelling. MAP is used to obtain the point estimate of unobserved quantity to optimize the model.

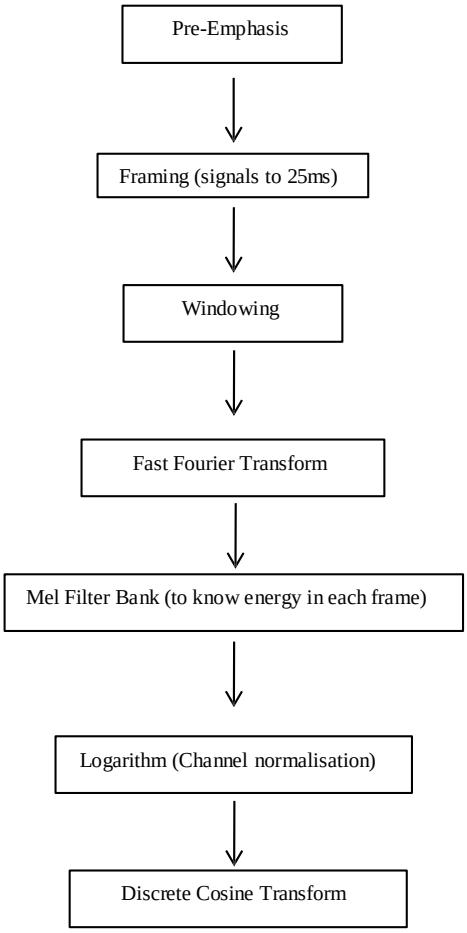

Fig. 1. Stages involved in MFCC feature extraction

III. WORK CARRIED OUT

The first step in the project is to extract the feature from the audio file that is the extraction of the MFCC. The next step in the project is Modelling. For Modelling we need to perform two steps the first step in the modelling is

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 7.512|

|| Volume 10, Issue 4, April 2021 ||

DOI:10.15680/IJIRSET.2021.1004074

training the GMM-UBM model and then to perform the MAP (Maximum A Posteriori) method to remove outfitting. Then the target can be obtained. For Testing we need extract the feature and perform Maximum Likelihood for that feature. The Testing also involves the extraction of the MFCC feature from that test file as extracted for the training of the model so that there would not be any mismatch in the frames or in the model feature array size so that the model cam be compared with the feature extracted from the test file. Based on the predicted output the choice will be given to the person so that person can select the option and continue the program to the next detection of the emotion state.

IV. SYSTEM ARCHITECTURE

Fig.2 shows the design for our system. The system consists of 2 phases coaching and testing. The coaching cons for Modelling we want to perform 2 steps the primary step within the modelling is coaching the GMM-UBM model then to perform the MAP (Maximum A Posteriori) methodology to get rid of armament. Then the target may be obtained. For Testing we want extract the feature and perform mostprobability for that feature. The Testing additionally involves the extraction of the MFCC feature from that check file as extracted for the coaching of the model so there wouldn't be any couple within the frames or within the model feature array size so the model can be compared with the feature extracted from the check file. The most likelihood is that calculated from the obtained check feature characteristics and also the trained model.

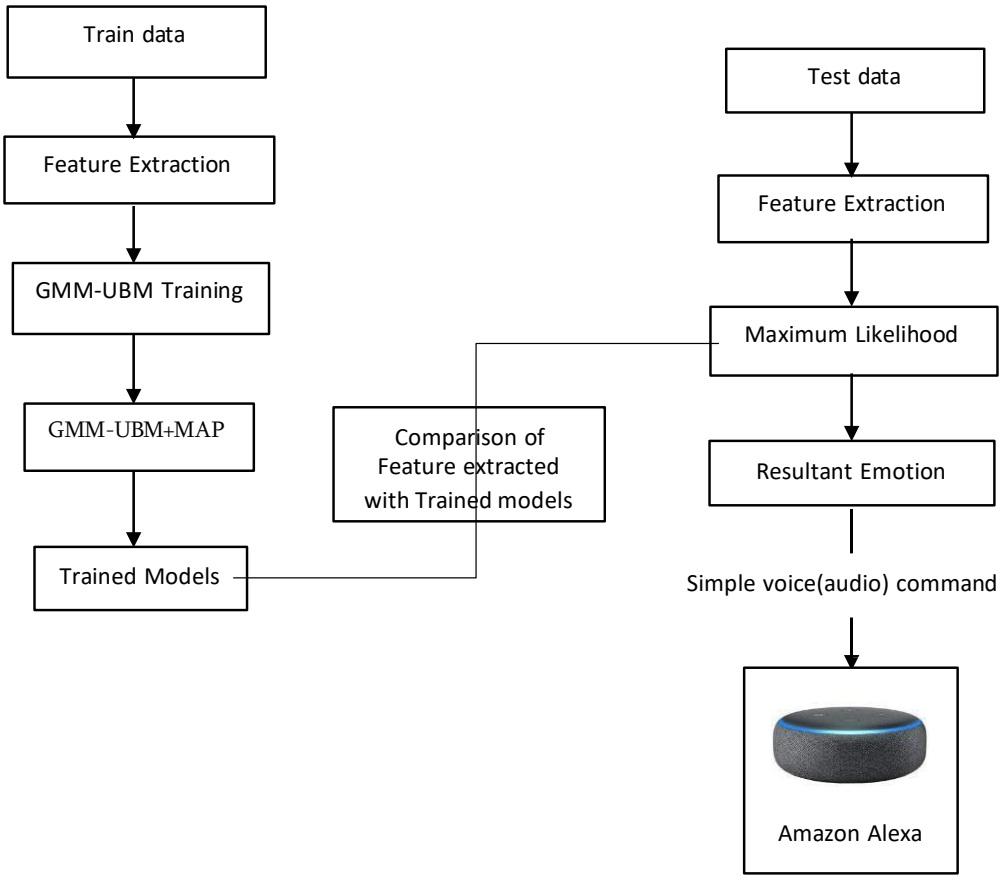

Fig. 2. Architecture for Training and testing of the Emotion Recognition System

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 7.512|

|| Volume 10, Issue 4, April 2021 ||

DOI:10.15680/IJIRSET.2021.1004074

TABLE I. PRECISION

	Angry	Disgust	Fear	Happy	Neutral	Pleasant Surprise	Sad
GMM	98%	70%	98%	98%	63%	90%	76.47%
GMM-UBM	98%	92.8%	92.8%	97%	87.5%	87.5%	91.67%

TABLE II. RECALL

	Angry	Disgust	Fear	Happy	Neutral	Pleasant Surprise	Sad
GMM	50%	100%	85.71%	71.48%	100%	64%	92.8%
GMM -UBM	100%	92.8%	92.8%	85.71%	100%	100%	78.5%

TABLE III. ACCURACY

	Accuracy	Precision	Recall
GMM	80.6%	84.01%	79%
GMM-UBM	92.8%	91.3%	91%

V. PERFORMANCE EVALUATION

We had study on the both GMM and GMM-UBM Modelling in those we had performed an accuracy test which shows that the GMM-UBM perform well for our dataset and gives higher accuracy. It is shown in the Table-I, Table-II and Table-III.

A. Confusion Matrix

A confusion matrix is associate analysis outline of prediction results on a modelling classification drawback. The quantity of correct and incorrect predictions square measure summarized with count values and countermined by every feeling category. This can be the key to the confusion matrix. The confusion matrix size depends on variety of categories we tend to have. In out project we've seven totally different categories, therefore the size of the confusion matrix is of 7*7.

For GMM the Confusion matrix is:

$$\begin{bmatrix} 7 & 2 & 0 & 0 & 3 & 0 & 2 \\ 0 & 14 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 12 & 0 & 2 & 0 & 0 \\ 0 & 1 & 0 & 10 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 & 14 & 0 & 0 \\ 0 & 3 & 0 & 0 & 2 & 9 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 13 \end{bmatrix}$$

For GMM-UBM the Confusion matrix is:

$$\begin{bmatrix} 14 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 13 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 13 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 12 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 14 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 14 & 0 \\ 0 & 1 & 0 & 0 & 2 & 0 & 11 \end{bmatrix}$$

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 7.512|

|| Volume 10, Issue 4, April 2021 ||

DOI:10.15680/IJIRSET.2021.1004074

B. Precision

Precision is additionally referred to as positive prognosticative price. Exactitude is that the fraction of relevant instances among the retrieved instances. exactitude measures however sensible the feeling model is at distribution positive events to the positive category. That is, however correct the spam prediction is. The Formula for locating the exactitude of sophistication (Pclass) is:

$$\text{Pclass} = \frac{\text{Correctly Predicted}}{\text{Total Actual of that class}}$$

The formula for finding the average weighted precision (P) is

P= Sum of Product of Total actual of classes and Precision of classes

C. Recall

Recall is additionally called sensitivity and it's the fraction of the entire quantity of relevant instances that were really retrieved. Recall measures however sensible the model is in detective work positive events. The Formula for locating the Recall of sophistication (Rclass) is.

$$\text{Rclass} = \frac{\text{Correctly Predicted}}{\text{Total Prediction of that class}}$$

The formula for finding the average weighted Recall (R) is

R= Sum of Product of Total actual of classes and Recall of classes

D. Accuracy

The Accuracy(A) of the classifiers can be obtained by the formula

$$A = \frac{Sum\ of\ Correctly\ Predicted}{Total\ number\ of\ Prediction}$$

VI. CONCLUSION AND FUTURE WORK

This paper forwarded efficient system for effective and accurate classification and recognition of the emotion of a person. After implementation of the proposed system using the GMM-UBM modelling, it has shown that the proposed system provides very high accuracy as compared to the available techniques. This proposed system is used as an application to find the emotion of an individual which can be used by amazon alexa to determine the emotion of the individual. The dataset that we have used in the system has an accent which when compared with the real timer data gives a varied output because accent of real time data is different. Hence the dataset should be improved in the future work to give better result.

REFERENCES

[1] Ravi Kumar V., Hari Krishna Vydana and Anil Kumar Vuppala “Significance of GMM-UBM based Modelling for Indian Language Identification” Eleventh International Multi-Conference on Information Processing-2015 (IMCIP-2015).

[2] Senjaliya, N., & Tejani, A. (2020). Artificial intelligence-powered autonomous energy management system for hybrid heat pump and solar thermal integration in residential buildings. International Journal of Advanced Research in Engineering and Technology (IJARET), 11(7), 1025-1037.

[3] Ashwini Rajasekhar and Malaya Kumar Hota, “A Study of Speech, Speaker and Emotion Recognition using Mel Frequency Cepstrum Coefficients and Support Vector Machines”, International Conference on Communication and Signal Processing, April 3-5, 2018, India

[4] Zhao, Xiaojia, and DeLiang Wang. “Analyzing noise robustness of MFCC and GFCC features in speaker identification”, IEEE International Conference on Acoustics, Speech and Signal Processing,2013

[5] R.Natarajan, R.Sugumar, "Highly Secure Trust Based Multicast Tree Routing Protocol for MANET", Journal of Advanced Research in Dynamical & Control Systems, Vol.9, Issue 3, March 2017

[6] Davis, Steven, and Paul Mermelstein, "Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences," Acoustics, Speech and Signal Processing, IEEE Trans, pp. 357 -

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 7.512|

|| Volume 10, Issue 4, April 2021 ||

DOI:10.15680/IJIRSET.2021.1004074

366, 1980.

[7] K. Anbazhagan, R.Sugumar, A Proficient Two Level Security Contrivances for Storing Data in Cloud, Indian Journal of Science and Technology, Vol.9, Issue 48, October 2016.

[8] HrishiRakshit and Muhammad Ahsan Ullah, "A Comparative Study on Window Functions for Designing Efficient FIR Filter", The 9th International Forum on Strategic Technology (IFOST), October 21-23, 2014, Cox's Bazar, Bangladesh.

[9] Kritagya Bhattarai, P.W.C. Prasad, AbeerAlsadoon, L. Pham, A. Elchouemi, "Experiments on the MFCC application in speaker recognition using Matlab", Seventh International Conference on Information Science and Technology, April 16-19, 2017, Da Nang, Vietnam

[10] S. Maity, A. K. Vuppala, K. S. Rao and D. Nandi, IITKGP-MLILSC Speech Database for Language Identification, In National Conference on Communications (NCC), 2012, IEEE, pp. 1–5, (2012).

Gunjan Jhawar, Prajacta Nagraj, and P. Mahalakshmi, "Speech Disorder Recognition using MFCC", International Conference on Communication and Signal Processing, India, April 6-8, 2016

[11] H. K. Palo, M. N. Mohanty, and M. Chandra. Computational Vision and Robotics. Advances in Intelligent Systems and Computing, 332:63–70,2015.

[12] C. H. Wu and W. B. Liang. Emotion recognition of affective speech based on multiple classifiers using acoustic-prosodic information and semantic labels. IEEE Transactions on Affective Computing, 2(1):10– 21, 2011.

[13] A. K. Samantaray and K. Mahapatra. A novel approach of speech emotion recognition with prosody, quality and derived features using SVM classifier for a class of North-Eastern Languages. Recent Trends in Information Systems (ReTIS), 2015 IEEE 2nd International Conference on, pages 372–377, 2015.

[14] I. A. Lima, M. S. Alencar, W. T. A. Lopes, and F. Madeiro. Evaluation of optimal and sub-optimal speech noise reduction wiener filters. In IWT 2015 - 2015 International Workshop on Telecommunications, pages 1– 5. IEEE, June 2015.

INNO SPACE
SJIF Scientific Journal Impact Factor

Impact Factor:
7.512

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर
NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

9940 572 462 6381 907 438 ijirset@gmail.com

www.ijirset.com

Scan to save the contact details