International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 03 | March - 2023 Impact Factor: 7.185 ISSN: 2582-3930

MULTI MODAL EMOTIONAL RECOGNITION SYSTEM USING FACIAL RECOGNITION AND SKEW GMM

Dr.M.Chinnarao

Lingayya institute of management and Technology,Vijayawada

***Abstract*-Emotion recognition systems have a significant role in many areas including forensic application, teleconferencing, phone conversation and other areas. These emotions that are exhibited by the individuals can be sensitized through the facial expression and in remote conversations. These emotions can be figured out using the speech signals. These speech signals that are articulated from the voice can be generated and recorded. The recorded pattern using the features the emotions generated can be identified. Emotional recognition is a major research area in speech recognition. The features of the emotions will affect the recognition efficiency of the speech recognition systems. Several techniques are used in recognizing the emotions. This paper contributes towards a novel methodology based on Multi modal emotion recognition system using Skew Gaussian mixture model, the updated equations are derived from using EM Algorithm. The verbal communication signals are recorded into WAV format, which are used for research, collectively with the facial gestures. The emotions of both male and female are obtained using LPC, MFCC, and SDC.**

***Keywords*- Emotional recognition, Skew Gaussian mixure model, EM algorithm,LPC, MFCC and SDC.**

1.1 INTRODUCTION

Emotion recognition systems have a significant role in many areas including forensic application, Teleconferencing, phone conversation and other areas. These emotions exhibited by the individuals can be sensitized through the facial expression and in remote conversations[1]. These emotions can be figured out using the speech signals, these speech signals articulated from the voice can be generated and recorded, and from the recorded pattern using the features the emotions generated can be identified[2]. Every individual will have a range of frequency to identify the emotions of a person, these emotions which are recorded generally through the acting sequences. Most of the works in the literature are coined with the acting sequences generated from both genders[3]. In this paper experimentation is conducted to identify the physical expression of the faces and also identify the emotions through the speech signal and fusing this one can interpret the actual form of a person[4]. These helps in many realistic applications where in if either a facial expression is identified are the speech template is identified, the actual mind of a person can be accurately identified. Every lifetime is driven in this direction, this chapter of the thesis attempts to bring out using model using Skew GMM[5].

The voices of each of the speaker are recorded in a realistic way along with the expression of face at the time of the speech in order to experiment the model. We have collected a real time data generated from recorded sequence from students of KIET with both gender. The system is also tested on real-time facial expression data obtained from the data COHN – KANADE. The emotions generated are recorded in tune with the facial expression[6-9]. This model helps to understand the emotions of the individual in the data base either in the presence of facial expression without voice or in the presence of only availability of voice. The data base considered consist of thousand persons facial data along with the emotion speech of which we have considered 50% data for the testing purpose.

1.2 SKEW GAUSSIAN MIXTURE MODEL:

It is asymmetric distribution well suited for analyzing the asymmetric nature of the speech Signal. The probability density function and updated equation are as follows

The PDF of Skew GMM is given by

$$f(z) = 2.\phi(z).\Phi(\alpha z); \qquad -\infty < z < \infty \qquad \ldots 1.2.1$$

Where, $$\Phi(\alpha z) = \int_{-\infty}^{\alpha z} \phi(t)dt \qquad \ldots 1.2.2$$

And, $$\phi(z) = \frac{e^{-\frac{1}{2}z^2}}{\sqrt{2\pi}} \qquad \ldots 1.2.3$$

Then, $y = \mu + \sigma z$

$$z = \frac{y-\mu}{\sigma} \qquad \ldots 1.2.4$$

Replacing equations (1.2.2),(1.2.3), and (1.2.4) in equation (1.2.1)

$$f(z) = 0.49.e^{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2}\left[\int_{-\infty}^{\alpha\left(\frac{y-\mu}{\sigma}\right)} e^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2} * 0.2dt\right]$$

…1.2.5

For updating the μ:

$$\mu^{(l+1)} = y + \sigma^{2^{(l)}} + \frac{1}{\int_{-\infty}^{\infty^{(l)}(\frac{y-\mu^{(l)}}{\sigma^{(l)}})} e^{-\frac{1}{2}(\frac{t-\mu^{(l)}}{\sigma^{(l)}})^2} dt} + \int_{-\infty}^{\infty^{(l)}(\frac{y-\mu^{(l)}}{\sigma^{(l)}})} (t-\mu^{(l)}) e^{-\frac{1}{2}(\frac{t-\mu^{(l)}}{\sigma^{(l)}})^2} dt - \sigma^{(l)}\alpha^{(l)} e^{\frac{\left[(\alpha^{(l)}+\sigma^{(l)})\mu^{(l)}-\alpha^{(l)}y\right]^2}{2\sigma^{4^{(l)}}}}$$

….1.2.6

Updating Σ^2:

$$\sigma^{(l+1)} = 1 \Bigg/ \frac{(y-\mu^{(l)})^2}{\sigma^{3^{(l)}}} + \frac{1}{\int_{-\infty}^{\infty^{(l)}(\frac{y-\mu^{(l)}}{\sigma^{(l)}})} e^{-\frac{1}{2}(\frac{t-\mu^{(l)}}{\sigma^{(l)}})^2 dt}} + \int_{-\infty}^{\infty^{(l)}(\frac{y-\mu^{(l)}}{\sigma^{(l)}})} \frac{(t-\mu^{(l)})^2}{\sigma^{3^{(l)}}} e^{-\frac{1}{2}(\frac{t-\mu^{(l)}}{\sigma^{(l)}})^2} dt + \alpha^{(l)}(\frac{\mu^{(l)}-y}{\sigma^{2^{(l)}}}) e^{\frac{\left[(\alpha^{(l)}+\sigma^{(l)})\mu^{(l)}-\alpha^{(l)}y\right]^2}{2\sigma^{4^{(l)}}}}$$

….1.2.7

Updating a:

$$\alpha^{(l+1)} = \frac{\sqrt{2}\sigma^{2^{(l)}}}{\mu^{(l)}-y}\left[\log\left(\int_{-\infty}^{\alpha^{(l)}(\frac{y-\mu^{(l)}}{\sigma^{(l)}})} e^{-\frac{1}{2}(\frac{t-\mu^{(l)}}{\sigma^{(l)}})^2} dt\right) - \log\left(\frac{y-\mu^{(l)}}{\sigma^{(l)}}\right)\right]^{\frac{1}{2}} - \frac{\sigma^{(l)}\mu^{(l)}}{\mu^{(l)}-y}$$

…1.2.8

1.3 FEATURE EXTRACTION

Feature extraction plays a vital role for recognizing the faces and emotions accurately. In this chapter of thesis we have considered MFCC feature for recognizing the features from given data set of the emotions and PCA is considered for identifying the Eigen values and dimension of the reduction of the faces[10-12]. These features are focused using alpha scope fusion.

1.3.1 Algorithm for identification of emotion from speech signal

Step 1: The speech signals are recorded into wav format.
Step 2: To eliminate the noise through preprocessing the voices generated by the speaker.
Step 3: Obtain the amplitude sequences from these verbal communication signals using MFCC coefficients.
Step 4: Recognize the collection of each emotion based on the amplitude information.
Step 5: Classify each stored emotions and store each emotion in opposition to the speaker.
Step 6: Replicate the process for both male and female.

1.4 PCA FOR THE FACE RECOGNITION

PCA is one of the process dimensionality reductions of method and feature extraction technique. It is considered mostly in the literature for the facial verifications. The PCA algorithm is given as the following.

1.4.1 Algorithm for recognizing the Eigen faces:

Step 1: Regularize the faces in the stored data bases
Step 2: Eradicate the background information.
Step 3: Eradicate the noise by deducting the mean value of every image.
Step 4: Evaluate the co variant matrix
Step5: Develop the distinctive equation which gives Eigen vectors.
This Eigen vectors are stored against each focal point in the data bases.

1.5 MULTI MODEL APPROACH FOR EFFECTIVE EMOTION SYSTEM BASED ON FACIAL EXPRESSION TOGETHER WITH EMOTIONAL SPEAKER

In order to have in effective emotional recognized system, in this chapter of paper we have highlighted contribution based on fusion. In this chapter we have utilize ∞ level score fusion. The facial data base is obtaining (www.cohnkanade.com) containing different emotions from both genders. This data base is considered for the experimentation. The data base is consisting of 500 facial expressions pertaining to different emotions viz. angry, happy, boring, neutral and sad. Each of these emotions is prerecorded against a subject[13]. The unique features from the facial samples are extracted using principal component analysis presented in section 1.4.

In the similar manner the voice template of the speakers whose facial expression are considered or recorded the feature from these speech samples have been obtained 1.3.1

Each of the speech sample along with the facial feature vector or considered the speech template are given as input to Skew symmetric Gaussian model. The updated equations are generated using the EM algorithm[14-15].

1.6 EXPERIMENTATION

The proposed model data base of the facial expressions are obtained from COHN KANADE database is considered and the emotion speech samples are collected from the KIET engineering students' database is considered. In this database consists of 500 voice samples collected from acting sequences from both male and female with 5 different emotions namely anger, sadness, neutrality, happiness and boring from each of the faces are to be normalized into a square. Only the frontal portion of the faces is measured and the background is subtracted. The co features are identified after the dimensionality reduction using the PCA is presented in section 1.4 of this paper. The Eigen vectors equivalent to individual emotion is obtained and that are stored into database against a face. The corresponding signals of the same individual emotion are extracted and stored in the database. These emotions that are subjected to feature extraction presented in section

1.3 of the paper MFCC features, delta features, delta–delta features are obtained against individual speech pattern concern to a particular emotion. This process is repetitive for the speech pattern of both genders. These features are given as input to the Skew Gaussian distribution presented in section 1.2 of the paper.

The PDF is extracted against each of speech pattern are stored. The concept of fusion is acquired to regularize the emotional condition of mind by fusing the PDF values extracted against the MFCC features. The main benefit of this method is to find individual emotion either to comprehend by the innermost physical appearance or outward intrinsic feelings, however in the most of practical cases the inward appearance may not coincided with the inherent emotions and there is no guarantee that each individual should respond with the same expression at a particular instance. Hence in this chapter of the thesis a multimodal system is highlighted where in the real physical expression and the inherent expression can be correlated before taking a decision. This developed model can be used as substitute for NACTO analysis where the crime forensics can be well understand.

Missed Detection Rate (MDR) = (Total no of missed recognition/Total Template)*100;

False Acceptance Rate (FAR) = ((Total considered-Total Accepted)/Total Template)*100;

Acceptance Rate (AR) = (Total Number of Accepted Traits /Total Numbers of Traits)*100;

To authenticate the derived results using the metrics are True positive (+ve), False negative (-ve) and Precision and Recall.

i) The true positive considers all the emotions and is classified as particular emotions among all the emotions which truly have these particular emotions.

True positive Rate = True positive / (True positive (Tp) + False negative (Fp))

ii) False Negative, are those emotions which are wrongly classified and it is denoted as

Functional Point (Fp) = (Addition of the total emotions / the particular emotion) * overall No of emotions

Precision: It is represented as

(The rate of a specific emotion) / (summation of remaining emotions + the specific emotion)

iv) The Recall is probability that the particular speech signal belongs to a particular emotions as being in a class Z, if it is actually belongs to that class

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 03 | March - 2023 Impact Factor: 7.185 ISSN: 2582-3930

Table 1 showing the formula for calculating precession and Recall

		Actual class	
		Z	Not Z
Predicated class	Z	E	F
	Not Z	G	H

Recall =E / E+G

Precision =E / E+F

- Z-Score Normalization process is used
- Using the formula:- Z-Score:

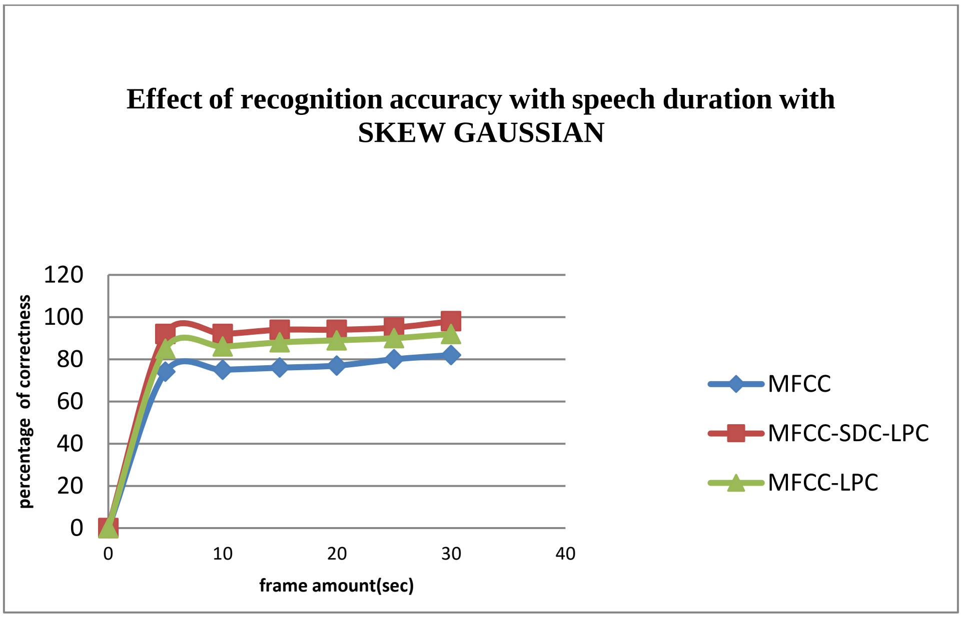

figure 1 representing the male effect of Emotion recognition using Skew-GMM

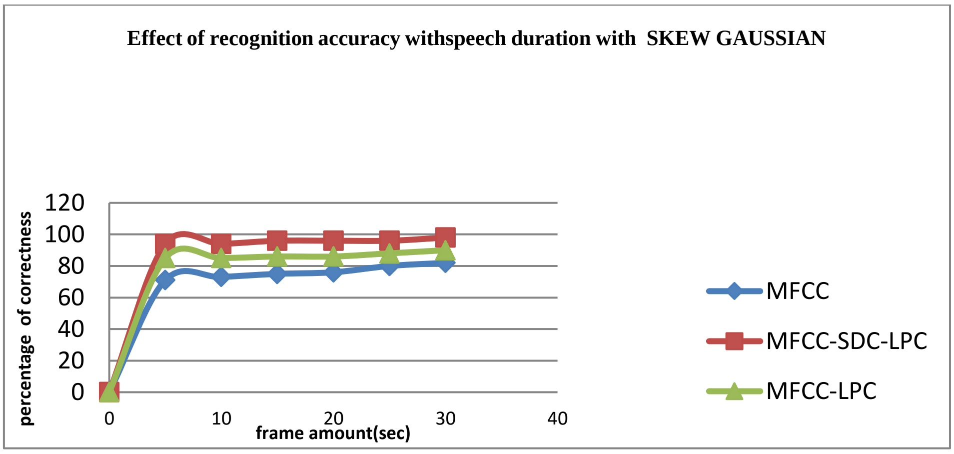

figure 2 representing the female effect of Emotion recognition using Skew-GMM.

The proposed model is tested for recognition accurateness on both the Berlin databases and KIET Students database.

Table 2 Evaluate of Confusion Matrix to recognize dissimilar emotions of Male-using KIET Students database

Stimulation	Identification of Emotion (%)/Developed model				
	Anger	Bore	Happiness	Sadness	Neutrality
Anger	93	0	0	7	0
Bore	5	87	0	10	0
Happiness	0	0	91	0	9
Sadness	0	11	0	87	5
Neutrality	0	11	0	8	83

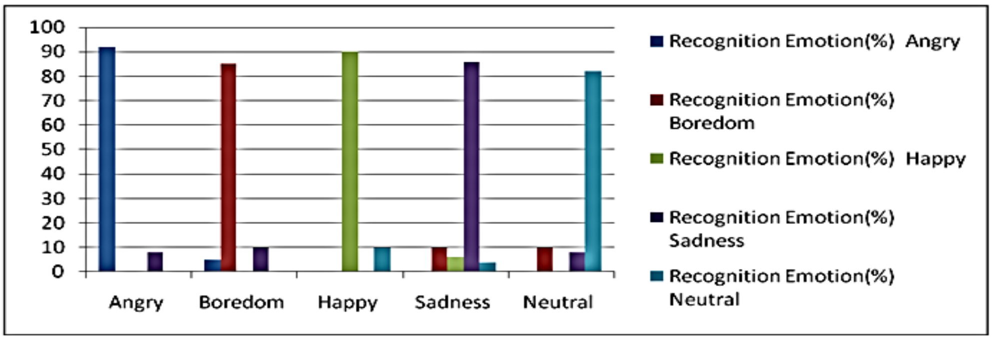

figure 3 Bargraph representation the identification rates from Male database based on Skew-GMM using KIET Students database.

Table 3 Evaluate of Confusion Matrix to recognize dissimilar emotions of Female-using KIET Students database

Stimulation	Identification of Emotion (%)/Developed model				
	Anger	Bore	Happiness	Sadness	Neutrality
Anger	88	04	00	08	00
Bore	00	85	00	05	10
Happiness	00	00	90	00	10
Sadness	05	07	00	90	00
Neutrality	00	05	10	00	85

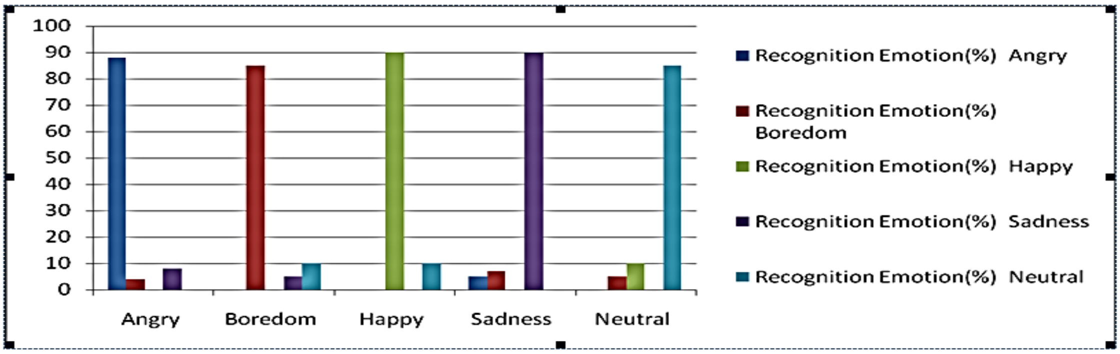

figure 4 Bargraph representation the identification rates from Male database based on Skew-GMM using KIET Students database.

Table 4 Evaluate of Confusion Matrix to recognize dissimilar emotions of Female-using KIET Students database

Stimulation	Identification of Emotion (%)/Developed model				
	Anger	Bore	Happiness	Sadness	Neutrality
Anger	88	04	00	08	00
Bore	00	85	00	05	10
Happiness	00	00	90	00	10
Sadness	05	07	00	90	00
Neutrality	00	05	10	00	85

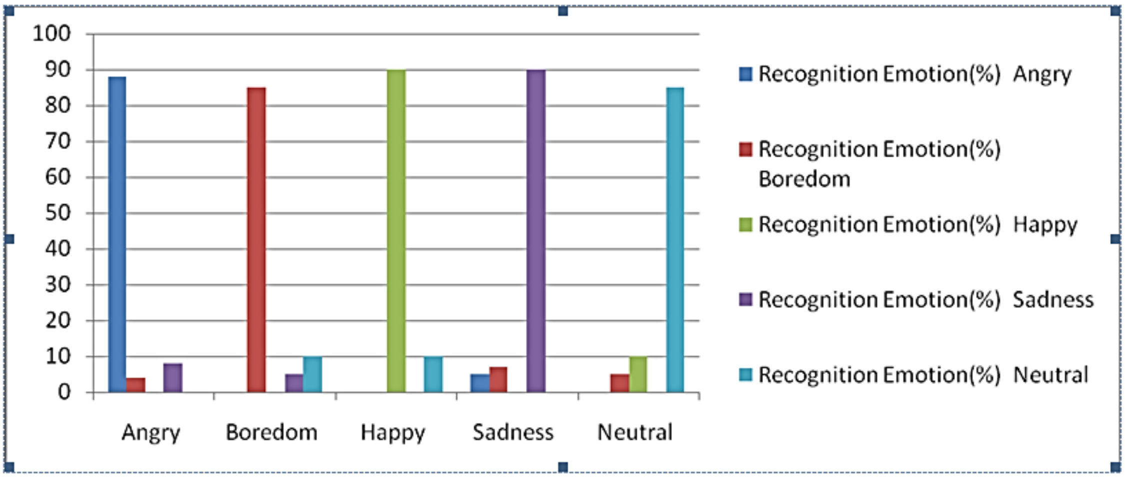

figure 5 Bargraph representation the identification rates from Female database based on Skew-GMM using KIET Students database

From the above graph it can be easily seen that developed model exhibits better recognition rate and recognizing the emotions of female speakers. The identification rate of about 90% is recorded for Happiness and Sadness and about 80% for further emotions. The recognition rates are also tested using standard dataset of BERLIN and the developed method exhibits good recognition accuracy.The performance of the derivative results are tested for accurateness using metrics like precision and recall. The formulas are given in section 5.6 and the outputs are given in Table 4 to 5

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 03 | March - 2023 Impact Factor: 7.185 ISSN: 2582-3930

Table 5 Precision & Recall values of Male Emotion using KIET Students database

	Precision	Recall
Anger	0.93	0.95
Bore	0.87	0.85
Happiness	0.90	0.99
Sadness	0.87	0.78
Neutrality	0.83	0.86

Table 6 Precision & Recall values of Female Emotion using KIET Students database

	Precision	Recall
Anger	0.89	0.95
Bore	0.86	0.85
Happiness	0.92	0.91
Sadness	0.89	0.88
Neutrality	0.86	0.81

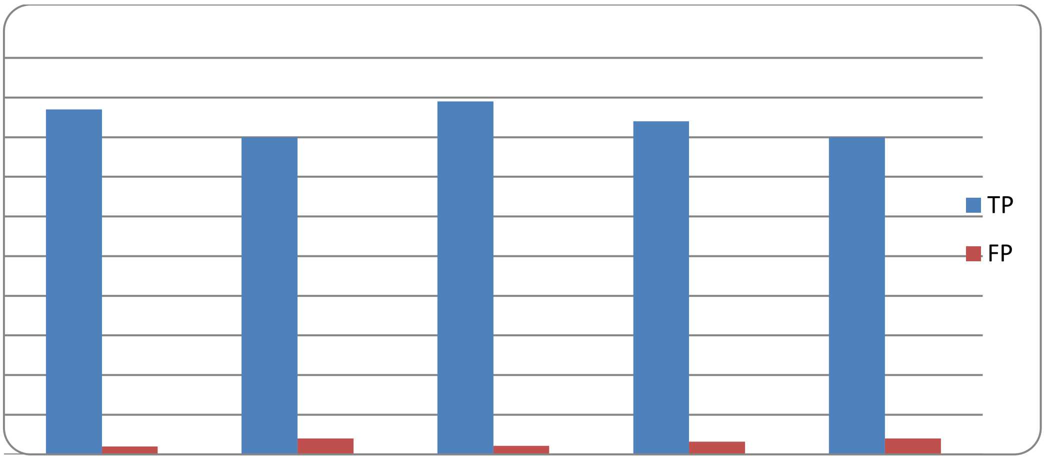

figure 6 Bargraph representing the values of Tp & Fp for Male Emotion

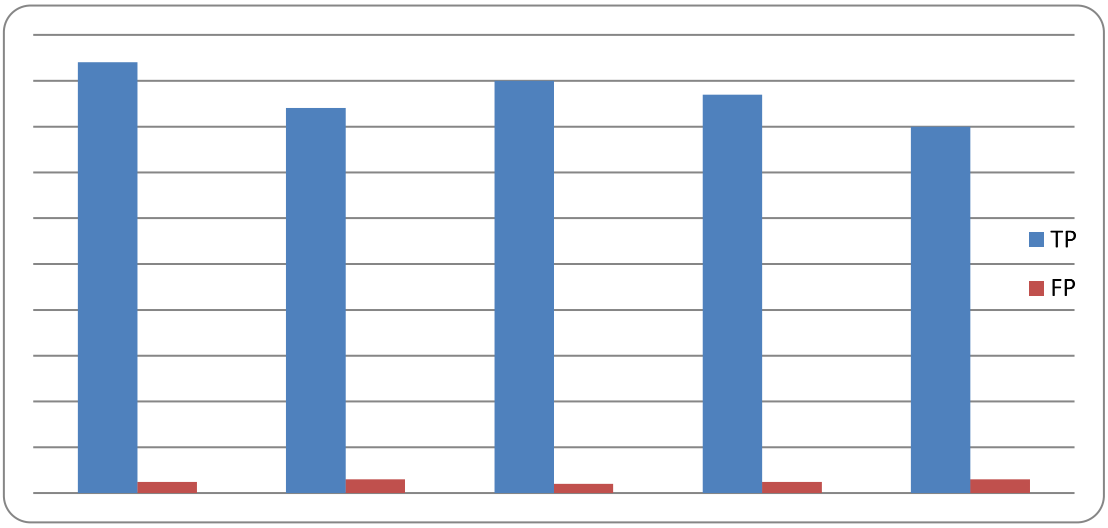

figure 7 Bargraph representing the values of Tp & Fp for Female Emotion

1.7 CONCLUSION

This paper contributes towards a novel methodology based on Multimodal emotion recognition system using Skew GMM, the updated equations are derived using EM Algorithm. The verbal communication signals recorded into .WAV format, which are used for research, collectively with the facial gestures. The emotions of both Male and Female are obtained using LPC, MFCC, and SDC. The testing is conceded with the database produced from the KIET Students database, Kakinada. The obtained results are evaluated using metrics like FAR, MDR, and FRR. The obtained results demonstrate that the proposed model performs resourcefully in identifying the emotions.

1.8 REFERENCES

[1] M.ChinnaRao, A.V.S.N.Murthy, Ch.Satyanarayana,"Emotion Recognition System Based On Skew Gaussian Mixture Model and MFCC Coefficients", IJIEEB, vol.7, no.4, pp.51-57, 2015. DOI: 10.5815/ijieeb.2015.04.07

[2] Arvid C. Johnson, "Characteristics and Tables of The Left-Truncated Normal Distribution" International Journal of Advanced Computer Science and Applications (IJACSA), pp133-139, May 2001.

[3] Forsyth M. and Jack M., "Discriminating Semi-continuous HMM for Speaker Verification"IEEE Int.conf.Acoust.,speech and signal processing,Vol.1,p313-316,1994.

[4] Forsyth M., "Discrimination observation probability hmm for speaker verification, speech communication", Vol.17, pp.117-129, 1995.

[5] George A and Constantine K "Phonemic Segmentation Using the Generalized Gamma Distribution and Small Sample Bayesian Information Criterion, speech communication" DOI: 10.1016/j.specom.2007.06.005, June-2007.

[6] Gregor D et al "Emotion Recognition in Borderline Personality Disorder- A review of the literature" journal of personality disorders, 23(1), pp6-9, 2009.

[7] Lin Y.L and Wei G "Speech Emotion Recognition based on HMM and SVM" 4th international conference on machine learning and cybernetics, Guangzhou, Vol.8, pp4898-401, 18-Aug-2005.

[8] Meena K, Subramanian U, and Muthusamy G "Gender Classification in Speech Recognition using Fuzzy Logic and Neural Network" The International Arab Journal of Information Technology, Vol. 10, No. 5, September 2013, PP477-485.

[9] Prasad A., Prasad Reddy P.V.G.D., Srinivas Y. and Suvarna Kumar G "An Emotion

Recognition System based on LIBSVM from telugu rural Dialects of Andhra Pradesh" journal of advanced research in computer engineering: An International journal ,volume 3, Number 2, july-December 2009.
[10] Vibha T "MFCC and its application in speaker recognition" international journal of emerging technology ISSN: 0975-8364 pp19-22.
[11] Kasiprasad Mannepalli, Panyam Narahari Sastryand V. Rajesh "Modelling And Analysis Of Accent Based Recognition And Speaker Identification System" ISSN 1819-6608, Dec 2014 pages 2807-2815.
[12] GSuvarna Kumar et. Al "SPEAKER RECOGNITION USING GMM." International Journal of Engineering Science and Technology, Vol. 2(6), 2010, 2428-2436.
[13] Stavros Ntalampiras and Nikos Fakotakis" Modeling the Temporal Evolution of Acoustic Parameters for Speech Emotion Recognition" IEEE Transactions on affective computing, vol. 3, no. 1, january-march 2012.
[14] K.Sreenivasa Rao Hindi Dialects and Emotions using Spectral and Prosodic features of Speech" Systems, Cybernetics and Informatics Volume 9 - Number 4 .ISSN: 1690-4524.
[15] Chung-Hsien Wu and Wei-Bin Liang "Emotion Recognition of Affective Speech Based on Multiple Classifiers Using Acoustic-Prosodic Information and Semantic Labels" IEEE Transactions on ffective computing, vol. 2, no. 1, january-march 2011.