Inferring the Human Emotional State of Mind using Assymetric Distrubution

N. Murali Krishna[1]

Asst Professor
Dept of CSE, GITAM University
Vishakhapatnam, INDIA

P.V. Lakshmi [2]

Professor
Dept of IT, GITAM University
Vishakhapatnam, INDIA

Y. Srinivas [3]

Professor
Dept of IT, GITAM University
Vishakhapatnam, INDIA

Abstract— **This present paper highlights a methodology for Emotion Recognition based on Skew Symmetric Gaussian Mixture Model classifier and MFCC-SDC ceptral coefficients as the features for the recognition of various emotions from the generated data-set of emotional voices belonging to students of both genders in GITAM University. For training and testing of the developed methodology, the data collection is carried out from the students of GITAM University of Visakhapatnam campus using acting sequence consisting of five different emotions namely Happy, Sad, Angry, Neutral, Boredom; each uttering one short emotional base sentence. For training the data we have considered fifty speakers from different regions (30 male & 20 female) and one long sentence containing an emotional speech from each speaker. The experimentation is conducted on text dependent speech emotion recognition and results obtained are tabulated by constructing a Confusion Matrix and comparing with existing methodology like Gaussian mixture model.**

Keywords— *Skew Gaussian Mixture Model; MFCC; SDC; Emotion Recognition; Confusion Matrix*

I. INTRODUCTION

For human communication system, speech acts as the medium from his/her voices. The main advantage of speech is that we can identify the interacting person voice. Emotion is a integral part of the speech which helps to identify the internal feelings of the speaker and in other words, emotion helps to understand the listeners state of mind. In many of the practical situations like BPO, telephonic communication etc it is desirable to understand the emotions of the speaker [1]. The Emotion Recognition is carried out by an acting sequence generated by each speaker varying different emotion back grounds, these speeches are therefore termed as acting sequences.

However, while generating different emotions the same step is to be applied. The emotions of speech narrate the prosody in a speech. The prosody of speech depends upon many characters, which include rules of the language, condition of living, place of living, culture of community etc [2]. Many models have been evolved for the recognition of the emotions based on GMM [3, 4, 5, 6], SVM [7, 8], HMM [9, 10], Truncated GMM [11, 12]. In order to have effective speaker emotion recognition system, the features that help in recognition the emotions are to be identified effectively. Many of the models in Literature use MFCC coefficients for the recognition of speaker emotions.

However, these MFCC coefficients will be useful, if speech is of short duration, for long-term speech Signals Shifted Delta Coefficients (SDC) features are more appropriated since they identify the dynamic behavior of the speaker along with the prosodic features of the emotions. More over the models discussed mainly in the Literature Review are aimed towards the identification of the speaker's emotional speech, assuming that the emotion speech signals are symmetric. However in reality, the speech signal is asymmetric in nature.

Hence in order to interpret the speech signal effectively, it is advantageous to use asymmetric distributions like Skew Gaussian Distribution, Log Distribution, Gamma Distribution, etc. In this paper we have utilized Skew Gaussian Distribution, since it contains Gaussian distribution as a particular case. The rest of the paper is organized as follows, in Section-2, the feature extraction methodology is presented, section -3, deals with the skew Gaussian mixture model, methodology is presented along with performance evaluation in section -4, the section-5 of the paper deals with results arrived and comparisons.

II. THE FEATURE EXTRACTION

For effective recognition of emotional speech it is important to extract the features effectively. In this paper we have considered the MFCC-SDC features for the extraction of features from the speech database. MFCC are preferred because of its ability to interpret the signal in short duration and SDC is used to extract the features from the long duration speech samples or dynamically changing samples. The processes carried out for the extraction of features are

1. The speech voices are fragmented in to small frames and these frames are given to MFCC.

2. Long term speeches are also considered and segmented the speech samples of windows sizes of 30ms, 60 ms etc and these sequences are given to SDC.

3. Using the combo effect of MFCC-SDC, the features are extracted for different emotions.

4. Classification of the emotion is carried out using Skew Gaussian mixture model.

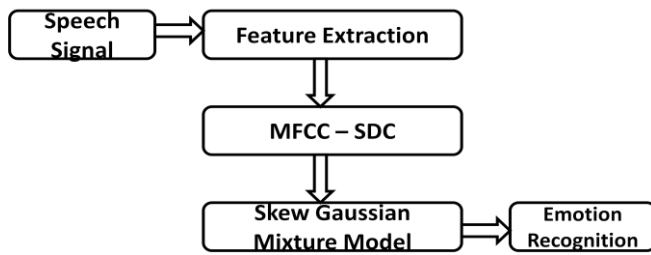5. The developed model is compared to that of GMM.

Fig-1 Process of Emotion Recognition Model

The speech signal in the database containing different emotions such as Happy, Sad, Angry, Neutral, Boredom are extracted and are trained using Skew Gaussian mixture model. The feature extraction process involves the generation of emotions samples in .WAV form extracting the amplitude values for each of these emotions signals and these values are given as input to be Skew Gaussian distribution and the probability density function values of the Skew Gaussian distribution are generated. For the testing purpose, test signal is considered, amplitude are values generated and these feature values are to be given as input to the speech signal for classification of the emotion.

III. SKEW GAUSSIAN Distribution

For the recognition of speech samples it is essential to understand the behavioral pattern of the speech signal. In many of the cases, the speech signal assumed to be symmetric but speech signal mainly depends upon the prosodic and also depends upon the pitch, energy and other factors associated with each speaker. Since these features are not symmetric always, asymmetric features are to be considered. Hence in this paper we have considered speech signal which caters the speeches which are both symmetric and asymmetric.

The probability density function of Skew Gaussian distribution is given by

$$f(z) = 2.\emptyset(z).\emptyset(\alpha Z); \quad -\alpha < Z < \alpha \qquad (1)$$

Where, $\emptyset(\alpha z) = \int_{-\alpha}^{\alpha z} \emptyset(t)dt$ \qquad (2)

And, $\emptyset(z) = \dfrac{e^{\frac{1}{2}z^2}}{\sqrt{2\pi}}$ \qquad (3)

Let $y = \mu + \sigma z$

$Z = \dfrac{y-\mu}{\sigma}$ \qquad (4)

Here we are submitting the equations (2),(3), and (4) in equation (1), we have

$$f(z) = \sqrt{\frac{2}{\pi}} \; e^{-\frac{1}{2}(\frac{y-\mu}{\sigma})^2} \left| \int_{-\alpha}^{\alpha(\frac{y-\mu}{\sigma})} \frac{e^{-\frac{1}{2}(\frac{t-\mu}{\sigma})^2}}{\sqrt{2\pi}} dt \right| \qquad (5)$$

IV. EXPERIMENTATION Evaluation

To demonstrate the model a database is generated with 50 students (30 male and 20 female) of GITAM University containing different dialects of Andhra Pradesh. The voices are recorded in acting sequences with different emotions Happy, Sad, Angry, Neutral and Boredom. The process of the emotion classification is as follows

A. *Phase-1Extract the MFCC-SDC coefficients*

B. *Phase-2: Train The Data By The Probability Density Function Of Skew Gaussian Distribution Model*

C. *Phase-3: Consider A Test Signal From Voice Database, Apply The Step-1 & Step-2 And Classify The Emotion*

D. *Phase-4: The Output Generated Is Depicted In The Form Of Confusion Matrix Shown In Below Table-1 And Table - 2*

TABLE 1. COMPARISON OF CONFUSION MATRIX FOR IDENTIFY DIFFERENT EMOTION OF MALE

stimulation	Recognition Emotion (%)/*proposed model*					Recognition Emotion (%)/*GMM*				
	Angry	Boredom	Happy	Sadness	Neutral	Angry	Boredom	Happy	Sadness	Neutral
Angry	92	0	0	8	0	80	0	10	10	0
Boredom	5	85	0	10	0	10	70	0	20	0
Happy	0	0	90	0	10	10	10	70	0	10
Sadness	0	10	0	86	4	10	0	10	60	20
Neutral	0	10	0	8	82	0	10	0	20	70

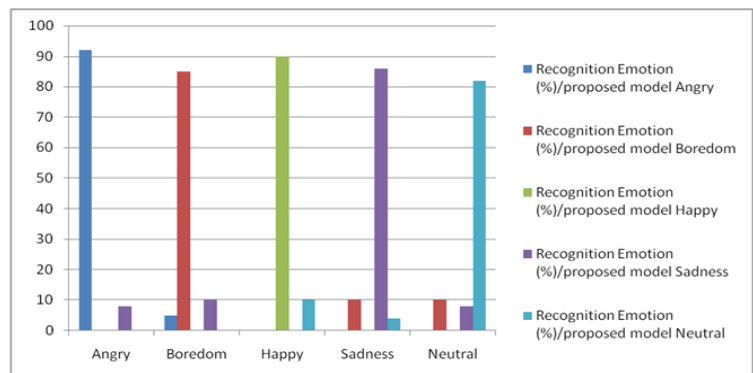Figure 1. BARCHART-1, REPRESENTING THE RECOGNITION RATES FROM MALE DATABASE

TABLE 2. CONFUSION MATRIX FOR IDENTIFY DIFFERENT EMOTION OF FEMALE

stimulation	Recognition Emotion (%)/*proposed Model*					Recognition Emotion (%)/*GMM*				
	Angry	Boredom	Happy	Sadness	Neutral	Angry	Boredom	Happy	Sadness	Neutral
Angry	85	0	0	10	5	92	0	8	0	0
Boredom	0	85	0	5	10	10	70	20	0	0
Happy	0	5	85	0	10	10	0	82	0	08
Sadness	0	0	10	90	0	10	10	0	70	10
Neutral	0	5	10	0	85	10	0	10	0	80

Figure 2. BARCHART-2, REPRESENTING THE RECOGNITION RATES FROM FEMALE DATABAS
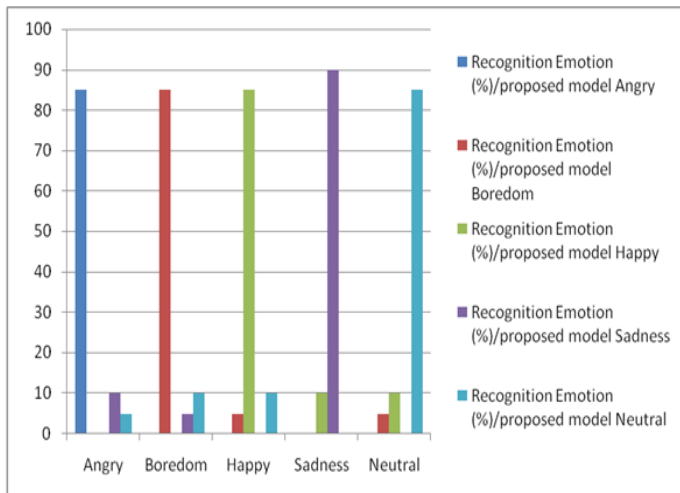
The results obtained are compared with that of existing model like Gaussian mixture model and results obtained in the presented in the above Table-1 & Table -2

V. CONCLUSION

In this paper a novel methodology for Emotion Recognition is done by using Skew Gaussian Mixture Model is developed. These emotions are recorded at 30 ms with five different emotions. The speech database is generated from the acting sequence of one short emotionally based speech sentence comprising of 5 different emotions from50 students (speakers) from different dialects of Andhra Pradesh. The features are extracted and for recognizing, the test speaker's emotion is considered and classified using Skew Gaussian Mixture Model. The results obtained are presented in the confusion matrix for both genders in Table-1 and in Table -2, and Bargraphs-1 &2. From the above Tables and Bar-Graphs, it can be see that the recognition rate is 90%.in case of certain emotion and for the other emotion. The output is compared with that of the existing model based on GMM and from the

Table-1 & Table -2, it can be clearly seen that our method outperforms the existing model. The overall emotion rate is above 85%.

REFERENCES

[1] N. Murali Krishna et al "Emotion Recognition using Dynamic Time Warping Technique for Isolated Words",IJCSI-vol-8,Issue-5,Sept-2011.

[2] björn schuller, et al" speech emotion recognition combining acoustic features and linguistic information in a hybrid support vector machine - belief network architecture" IEEE, icassp 2004,pp 577-580.

[3] Y.G. Kang et al "A hybrid GMM and codebook mapping method for spectral conversion" in Proc. The 1st International Conference on Affective Computing and Intelligent Interaction, 2005, pp.303-310.

[4] RajeswaraRao.R.,Nagesh et al "Source Feature Based Gender Identification System Using GMM" International Journal on computer science and Engineering,Vol:3(2),2011,pp-586-593.

[5] Xianglin Cheng et al "Speech Emotion Recognition Using Gaussian Mixture Model" International Conference on Computer Application and System Modeling-(2012)

[6] Bihar Kandali Emotion recognition from Assamese speeches using MFCC features and GMM classifier "TENCON, 2008 , Page(s): 1 - 5

[7] Meng-Ju Han"A new information fusion method for SVM-based robotic audio-visual emotion recognition "IEEE International Conference, 2007 , Page(s): 2656 - 2661

[8] Peipei Shen et al "Automatic Speech Emotion Recognition using Support Vector Machine "EMEIT, 2011 , Page(s): 621 – 625

[9] Xia Mao et al "Multi-level Speech Emotion Recognition Based on HMM and ANN", 2009 WRI World Congress,Computer Science and Information Engineering, pp.225-229, March2009.

[10] K. Choi and J.N. Hwang, "Baum-Welch HMM inversion for reliable audio-to-visual conversion", in Proc. IEEE International Workshop Multimedia Signal Processing, 1999, pp.175-180

[11] N. Murali Krishna et al "An Emotion Recognition System based on Right Truncated Gaussian Mixture Model" IJCA Journal, Volume 42 – issue Number 10,March- 2012

[12] V Sailaja et al "Text Independent Speaker Identification Using Finite Doubly Truncated Gaussian Mixture Model" International Journal of Information Technology and Knowledge Management, July-December 2010, Volume 2, No. 2, pp. 475-480

[13] Suri K Babu et al "Text-Independent Speaker Recognition using Emotional Features and Generalized Gamma Distribution" Volume 46 - Number 2, 2012

[14] Nagesh Vadaparthi et al "Unsupervised medical image segmentation on brain MRI images using Skew Gaussian distribution" International Conference on Recent Trends in Information Technology - ICRTIT , 2011

[15] Jianhua Tao, at al" Utterance Independent Bimodal Emotion Recognition in Spontaneous Communication" EURASIP Journal on Advances in Signal Processing,Volume-4, May 2011