An Emotion Recognition System based on Right Truncated Gaussian Mixture Model

N. Murali Krishna[1]
Asst Professor
Dept of CSE, GITAM University
Vishakhapatnam, INDIA

Y. Srinivas[2]
Professor
Dept of IT, GITAM University
Vishakhapatnam, INDIA

P.V. Lakshmi [3]
Professor
Dept of IT, GITAM University
Vishakhapatnam, INDIA

ABSTRACT
This article address a novel emotion recognition system based on the Truncated Gaussian mixture model .The proposed system has been experimented over an gender independent emotion recognition database In the recent past, many models have been listed in the literature based on the emotion recognition, but these papers are more focused towards the speech, ignoring the emotion of the speaker at the time of speech which may be of significant importance at some particular instances such as BPO. To overcome this, we present a model using Right Truncated Gaussian mixture model and K-means algorithm to classify the emotion speeches. MFCC features of the emotions are extracted. The proposed system has been experimented over an gender independent emotion recognition database. The results obtained are evaluated using a confusion Matrix and compared with that of the Gaussian mixture model .Our model achieved a recognized rate of above 90 %.

Keywords: TGMM, K-means, MFCC, Feature Extraction, Confusion Matrix.

1. INTRODUCTION
Speech is the medium of communication in natural form. It conveys the message regarding the identity of the speaker. In many practical situations such as telephone Conversation, Business Process Outsourcing (BPO), and other areas the speakers message along with the emotion is of crucial important. This emotion helps in understanding the inherent feeling of the of an individual and also helps in many situations where, for instance, in the police stations or railway stations about an unidentified call regarding a mishap that is going to take place, can be well interpreted with the identification of the emotions. In order to identify an individual speaker emotions we can use either prosodic feature or acoustic features. Every speaker has his own articulation rate by which the identity of a speaker can be established. Ceptral coefficients such as MFCC help to identify a speaker more exactly. The review work is this area is mostly projected using GMM, [1][2][3][4] but the main disadvantage with the GMM is the infinite range [5]. Most of the uttered speech emotions, in reality are of finite range hence, considering the infinite range and modeling the emotion recognition system using GMM is a crude approximation [6]. In general, the Emotion signal will always be of finite range and therefore, it needs to truncate the infinite range. Hence it is always advantage to consider the truncations of GMM into finite range, also it is clearly observed that the pitch signals along the right side are more appropriate. Hence in this paper we have considered Right Truncated Mixture Model. In order to experiment our model we have created a data base with 200 speakers of both the genders with acted sequences of 5 different emotions, namely happy, sad, angry, boredom, neutral. In order to test the data 50 samples are considered and a database of audio voice is generated in .wav format. The performance of the developed method is compared to that of the GMM. The next section of paper deals with the Feature extraction, section-3 deals with Right Truncated Gaussian mixture model, section-4 deals with K-means algorithm ,section-5 deals with the performance evaluation and finally section-6 concludes the paper.

2. FEATURE EXTRACTION:
Feature extraction and selection are important in achieving high recognition rate in this paper we investigate a representation based on MFCC features, on Formants and on hybrid representation MFCC/Formants. MFCC features are considered due to the fact that it can identify the speech from small sample rates also effectively.

2.1 Mel frequency cepstral coefficients (MFCC)
Generally the emotion speech sound frequency are measured in Mel-Scale rather than linear scale ,which is a linear frequency spacing below 1000Hz and logarithmic spacing above 1000Hz and is computed using

$$\text{Mel}(f)=2595*^{\log_{10}(1+\frac{f}{700})} \quad (1)$$

The next step is to calculate the Mel frequency cepstral coefficients, where the log Mel spectrum coefficients are converted to time domain using the discrete cosign transform (DCT);and the MFCC [6] is the process of reconverting the log Mel spectrum into time frequency .The process[6] is indicated in figure –1

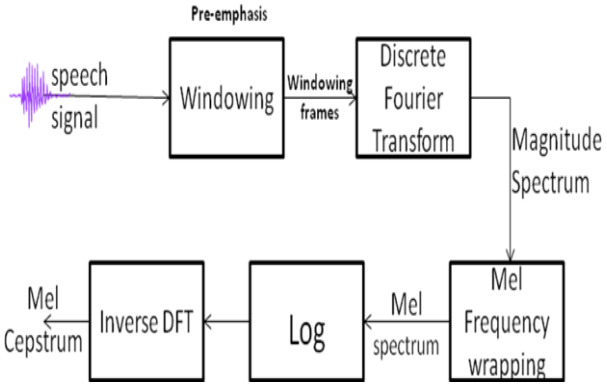


Figure –1 process of Mel Cestrum values of speech signal

The important modules considered in this paper are
1.Segmenting the speech into frames of Voice
2.Extraction of features from different emotion
3. Classification of emotion using Right Truncated GMM
4. Determining the accuracy of emotion
The emotion features like happy, sad, angry, neutral, boredom are extracted from the speech samples and are trained using Right Truncated Gaussian Mixture model. The feature extraction steps include, generating the emotion samples in .wav format and converting these into amplitude values, after transforming these signals into amplitude sequence, we get the values for the emotions like, happy, sad, angry, neutral, boredom . Using these amplitude values, the Probability Density Function (PDF) values of the Right Truncated Gaussian mixture are generated, the test signal is considered and the PDF values of the test signals are classified to ascertain the emotion.

3. RIGHT TRUNCATED NORMAL DISTRIBUTION

The probability density function of a Gaussian distribution is given by

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, -\infty \le x \le \infty \quad (2)$$

If the values of x below some interval x_R is truncated, the resulting distribution is a Right-Truncated normal distribution with probability density function $f_{RTN}(x)$

$$f_{RTN}(x) = \begin{cases} \frac{f(x)}{\int_{-\infty}^{K_R} f(x)dx}, -\infty \le x \le x_R, x_R \le x \le \infty \end{cases} \quad (3)$$

Where f(x) is as given in Eq 5.
The probability density function (PDF) of Gaussian distribution in terms of standard Normal distribution is given below

Where $Z=\frac{x-\mu}{2a\sigma}$ (4)

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}, -\infty \le z \le \infty \quad (5)$$

The Right Truncated then the normal distribution using the standard normal variate is represented as

$f_{SRTN}(t)$ and is given by

$$f_{SRTN}(t) = \begin{cases} \frac{f(t+K_R)}{0, \int_{-\infty}^{K_R} f(Z)dz} & t \le 0 \\ & t \ge 0 \end{cases} \quad (6)$$

Where t =Z-K_R (7)

Where $K_R = \frac{x_R-\mu}{\sigma}$ (8)

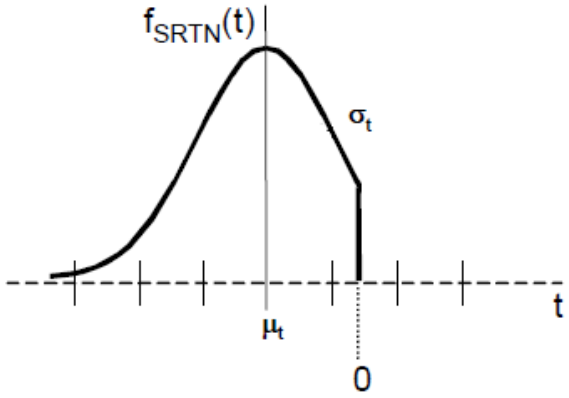


Figure-2 Standardized, Right –Truncated Normal Distribution

4. GENDER IDENTIFICATION USING K-MEANS CLUSTERING

In order to extract the feature vectors from the database, K-Means clustering algorithm is used. The most important aspect considered for any speaker identification is Gender identification . Unsupervised machine learning algorithm, such as K-Means algorithm is preferred, due to the fact that no gender information is available in the database. The dataset is clustered basing on the speech sample size for both male &female speech samples . Two different cancroids are considered, for male and female. Based on the distance between the pitch values and each of the cancroids (μ) the male or the female data is classified. The new mean μ_c of each cluster C_c is calculated by using equation (9).

$$\mu_c = \frac{\sum x_i C_c x}{[C_c]} \quad (9)$$

Where X_i is the pitch value of the ith sample[6], The process is applied iteratively until cluster convergence is attained. Once the training process of k-means clustering is completed, the classification and identification is carried out by using feature vector.

5. EXPERMENTAL EVALUTION

To demonstrate our method we have generated a database with 200 different speakers of Gitam University, INDIA, with different dialects , having five different emotions namely Happy, Sad, Boredom, Neutral and Angry.
The algorithm for our model is given under
Phase -1: Extract the MFCC coefficients
Phase -2: cluster the data with different sample sizes
Phase -3: train the data by PDF of Right Truncated GMM
Consider the test emotion and follow steps 2 and 3

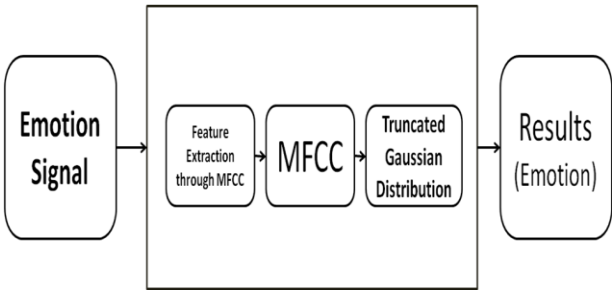


Figure-3 The Emotion Recognition process Model

The speeches are recorded each containing of 30 sec for training and minimum of one sec of data for testing.

5. RESULTS:

After extracting the emotion features and training the features using Right Truncated Gaussian distribution, the results obtained are stored in the database in Excel format. The emotion speech signal to be tested is trained as specified in section-4 of the paper, and the obtained features are compared with the existing emotions, based on MFCC coefficients . The features of the test emotion are classified using Right Truncated Gaussian distribution using the emotions in the database and the results obtained are tabulated using a confusion matrix and are presented in Table1 and Table-2 and Barchart 1 &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	90	10	0	0	0	80	0	10	10	0
Boredom	8	82	0	10	0	10	70	0	20	0
Happy	0	0	90	0	10	10	10	70	0	10
Sadness	0	10	10	80	0	10	0	10	60	20
Neutral	0	10	0	10	80	0	10	0	20	70

Barchart-1, representing the recognition rates from Male database

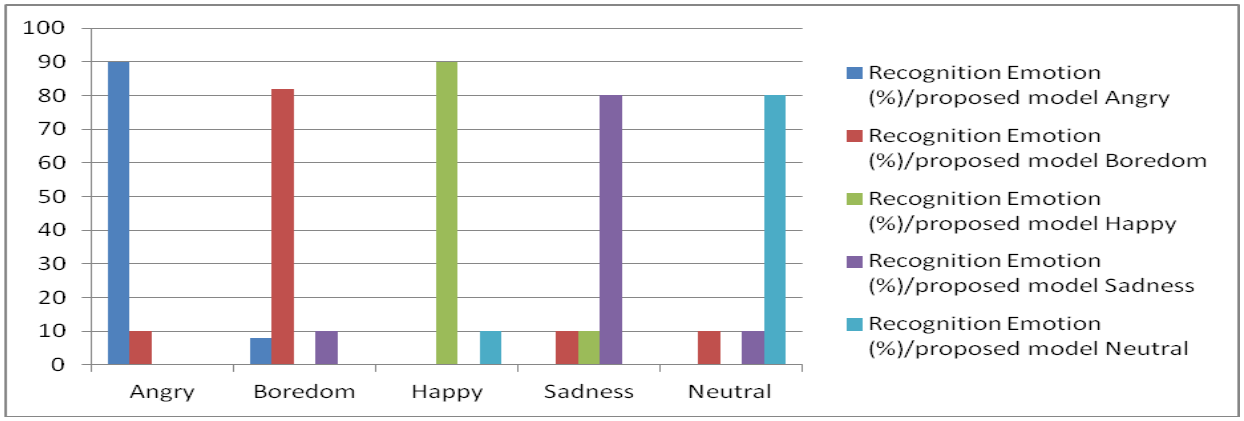
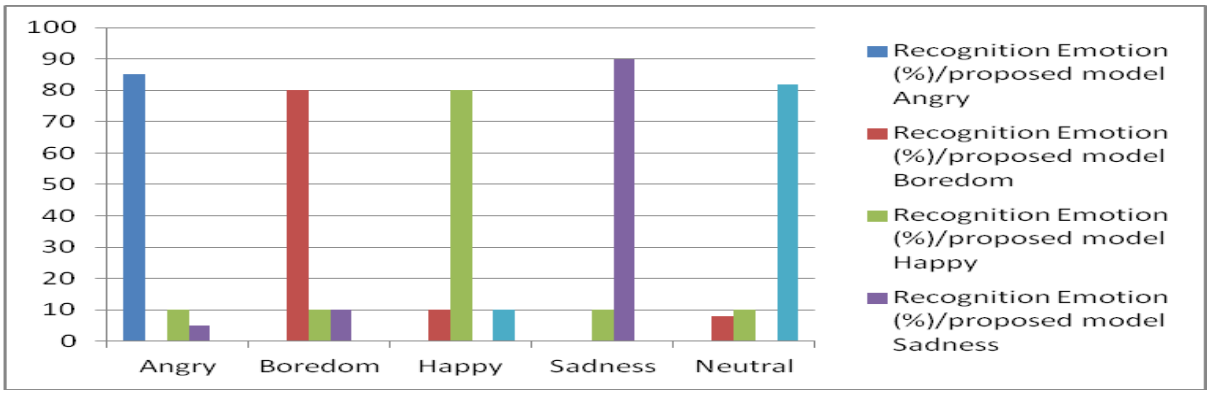


Table-2

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	10	5	0	92	0	8	0	0
Boredom	0	80	10	10	0	10	70	20	0	0
Happy	0	10	80	0	10	10	0	82	0	08
Sadness	0	0	10	90	0	10	10	0	70	10
Neutral	0	8	10	0	82	10	0	10	0	80

Barchart-2, representing the recognition rates from Female databas



6. CONCLUSION:

In this paper a novel methodology for emotion recognition is using Right Truncated Gaussian Distribution is developed. The emotions were considered from the students of Gitam University with different dialects. These emotion are recorded at 30 ms with five different emotion. The speech database is generated from the acting sequence of one short emotionally based speech sentence comprising of 5 different emotions from 200 students (speakers) from different parts of India. The features are extracted and for recognizing, the test speaker's emotion is considered and classified using Right Truncated Distribution. 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 graphs, it can be see that the recognition rate is 90%.in case of certain emotion and for the other emotion. The developed method is also tested by varying the sample sizes. 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 80% .This shows that, the developed model performs well in identifying the emotion.

REFERENCES

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

[2] George Almpanidis and Constantine Kotropoulos" Phonemic Segmentation Using the Generalised Gamma Distribution and Small Sample Bayesian Information Criterion.Vol 50,Issue-1,pp 38-55,January-2008.

[3] 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,pp4898-4901,18-aug-2005.

[4] Prasad Reddy P.V.G.D et al "Gender Based Emotion Recognition System for Telugu Rural Dialects Using Hidden Markov Models" journal of advanced research in computer engineering: journal of computiong Vol-2,c`issue 6,pp 94-98 ,june 2010.

[5] Forsyth M. and Jack M.," Discriminating Semi-continuous HMM for Speaker Verification "IEEE Int.conf.Acoust .,speech and signal processing ,Vol.1,pp313-316,April-1994.

[6] Forsyth M.," Discrimination observation probability hmm for speaker verification ",speech communication

, Vol.17,pp.117-129,1995.

[7] vibha Tivari"MFCC and its application in speaker recognition "international journal of emerging technology ISSN:0975-8364pp,19-22,2010.

[8] 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.

[9] N. Murali Krishna, P.V. Lakshmi, Y. Srinivas and J. Sirisha Devi " emotion recognition using dynamic time warping technique for isolated words" International Journal of Computer Science(IJCSI)vol-8,Issue-5,Sept-2011