K SuriBabu, SrinivasYarramalle, Suresh VarmaPenumatsa/ International Journal of Engineering Research and Applications (IJERA) ISSN: 2248-9622 www.ijera.com
Vol. 2, Issue 3, May-Jun 2012, pp.1522-1526

EMOTION CLASSIFICATION SYSTEM BASED ON GENERALIZED GAMMA DISTRIBUTION

***K SuriBabu, **SrinivasYarramalle, ***Suresh VarmaPenumatsa**

*Scientist, NSTL (DRDO),Govt. of India,Visakhapatnam, India
**Department of IT,GITAM University,Visakhapatnam, India
***Dept. of Computer Science,AdikaviNannaya University,
Rajahmundry, India

ABSTRACT

This paper deals with an automatic emotion recognition system based on Generalized GAMMA Distribution. Automatic emotion recognition is a research potential area in speech processing having major applications in the human interaction systems. This system comprises of three basic operations namely, Preprocessing (signal cleaning), Feature Extraction and the emotion classification. For training and testing in our proposed method, we have generated a speech database obtained from the acting sequence of one short emotionally based speech sentence comprising of 5 different emotions, from 200 speakers of both gender, for training and another 50 speakers emotions for testing. The classification is done by extracting features like MFCC, LPC from the speech signal.

Key Words:*Emotion Recognition, Feature extraction, Generalized Gamma Distribution, preprocessing, classification, MFCC, LPC.*

1.INTRODUCTION:

Emotion plays a dominant role in the area of psychology, to understand the human behavior. It also have a significant role in the field of human computer interaction, robotics and in BPO sector. With the advancements in the field of communication, communication can be established within few seconds across the globe. Since most of the communication channels are public, the data transmission may not be authenticated. Also in such situations, before interacting, it is essential to recognize the speaker and the speech. A speaker can modulate his/her voice and hence it is very essential to understand the behavior pattern of the speaker along the speech for effective recognition. In cases such as telemedicine, call centers the key step is to identify the speaker and his emotions and act basing on the emotions. Emotion recognition is very challenging task having its routes in security. The main goal of the speaker recognition system is to effectively recognize the speaker's ad mist different emotions. In particular situations, if any speaker tries to mimic the other speaker's voice or in the situations where the original recorded sequence differ from the test sequence where emotion is a integral part, then in such situations, the speaker identification should identify the speaker irrespective of different behavioral patterns. The main emotions that are considered for the analysis are happy, sad, angry, neutral, boredom. In order to identify the speaker with emotional speeches and articulated speeches, it is necessary to extract the features. The main speech features that can be of vital important include pitch, energy, and amplitude and speech rate. To recognize the speaker's with emotions, one need to consider both temporal and spectral features of speech signal from which amplitude, pitch, and formants of the speech are extracted and these features can be effectively used for classification of different speakers irrespective of the behavioral pattern of the speech.

Lot of research has been projected in the area of emotion speech recognition with objective to effectively recognize a speaker, irrespective of the emotions.

In most of the cases parametric models are preferred since the speeches extracted from the speech signals generated from the wave file are converted in to parametric form. With this assumption Gaussian mixture model mostly used in the literature [5] [6] [7].the main advantage of utilizing Gaussian mixture model is of the fact that any signal by default will be in Gaussian in shape and forms a bell shaped curve also due to its infinite range, it is also assumed that Gaussian distributions are well effective for recognizing speech in presence of noise and speech spectra, because of its symmetry. However the voices generated from the speech signals are not symmetric and either the shape is not a bell shape.

In the speaker recognition system the speech signal considered for the recognition is finite in size hence and using Gaussian distribution will yield fruitful results, also in speech signals to recognize a minimum of 0.2 sec speech duration is necessary and Gaussian distribution is more appropriate if the speech sample size is more , for effective recognition of the speech samples inn shorter duration sizes will be much needed for the identification of the speaker and together with elimination of noise and silences.

In order to classify a speaker different other classification techniques such as SVM, HMM are preferred however these techniques have their own disadvantages. The main disadvantage of SVM is that non-parametric techniques such as SVM are the lack of transparency of results. SVMs cannot represent the score of all companies as a simple parametric function of the financial ratios, since its dimension may be very high. It is neither a linear combination of single financial ratios nor has it another simple functional form. HMMs only use positive data to train. In other words, HMM training involves maximizing the observed probabilities for examples belonging to a class. But it does not minimize the probability of observation of instances from other classes. Hence for the effective recognition of the speech signals associated with the behavioral pattern, generalized gamma distribution is utilized. The main advantage of using the generalized gamma distribution includes the modeling of both clean and noise speeches can be done along with the identification of the speaker from the stationary speech signal having small frame duration the feature vector used for the identification of the speaker from the speech signals include MFCC-SDC-LPC,MFCC-LPC, The results obtained are compared with that of existing model ,Gaussian distribution.

In order to evaluate our developed model, a speech data base of 200 speakers of both the genders, with different emotions: happy, angry, sad, neutral, boredom are considered. For the test sequence 40 test speeches are considered. The results obtained are computed using a confusion matrix

2.Feature extraction:

In this study from the speaker data base generated with speech associated with emotions we have considered prosodic, MFCC and formant frequency features as features for classification. Prosody is concern with rhythmic, melody part of speech, which is considered as the basis of emotion state. The extracted prosody features include mean, median, standard deviation, minimum and maximum range pitch. This pitch is estimated basing on cepstral coefficients. MFCC-LPC are utilized in this work, because of the fact that these features help to identify the speech signals in presence of noise and short duration frame length.

3. GENERALIZED GAMMA DISTRUBUTION

Generalized Gamma distribution is used for classification of the emotion signals in this paper. Lot of research has been projected in the area of emotion recognition using GMM, but the main disadvantage with respective GMM is that the inference derived relies exclusively on the asymptotic approximation and slow in convergence .In order to overcome the disadvantage Generalized Gamma distribution[8] is utilized .In order to modeled the emotion signal effectively. Generalized Gamma distribution represents the sum of n exponentially distributed random variables both the shape and scale parameters can have non integers values. Generalized Gamma distribution is defied in terms of scale factor and shape factor .The generalized Gamma distribution is of the form

$$f(x,k,c,a,b) = \frac{c(x-a)^{ck-1}e^{-\left(\frac{x-a}{b}\right)^{c}}}{b^{ck}\Gamma(k)}$$

4.EXPERMENTAL EVALUTION

To demonstrate our method we have generated a database with 200 different speakers for testing with five different emotions namely Happy, Sad, Boredom, Neutral and Angry, for training purpose, we utilize 50 speakers. The algorithm for our model is given under

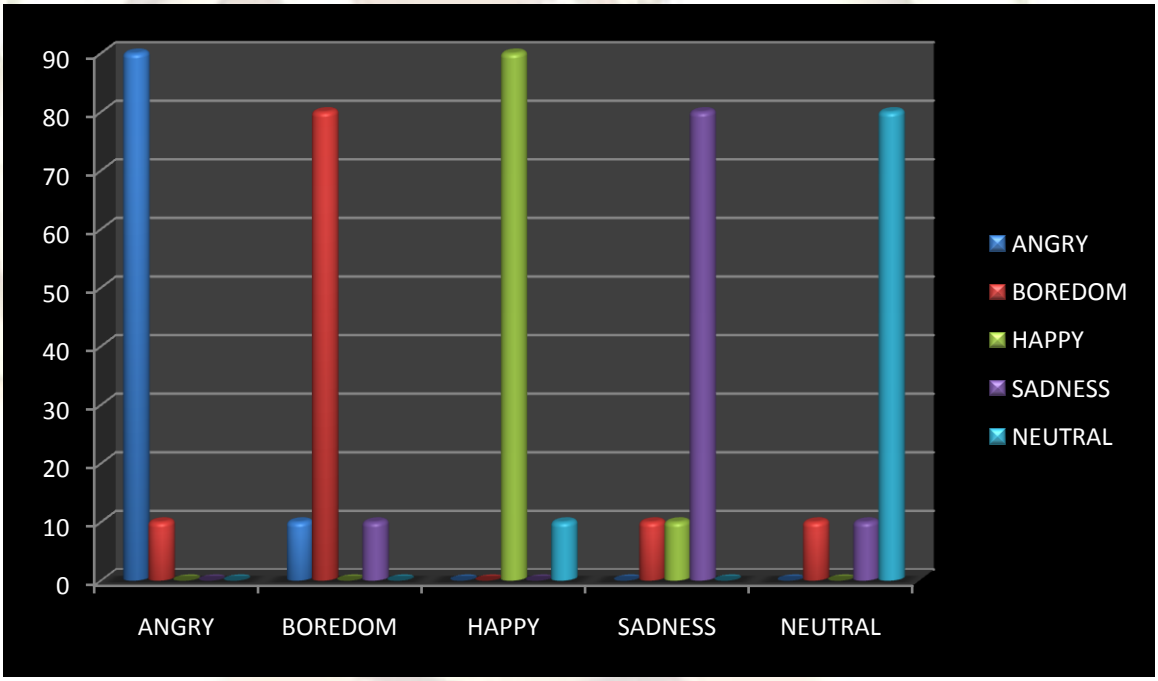
Phase -1: Extract the LPC coefficients from MFCC coefficients
Phase -2: process the LPC coefficients using the LPC Filter
Phase -3: Calculate the Generalized Gamma pdf
Consider the test emotion and follow steps 2 to 4
The speeches are recorded each containing of 30 sec for training and minimum of one sec of data for testing.

5. RESULTS

After extracting the emotion features and training the features using Generalized Gamma distribution, the results obtained are stored in the database. 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 LPC coefficients . The features of the test emotion are classified using Generalized Gamma 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 Barcharts 1 &2.speakers

Table-1 Confusion matrix for identify different emotion of Male

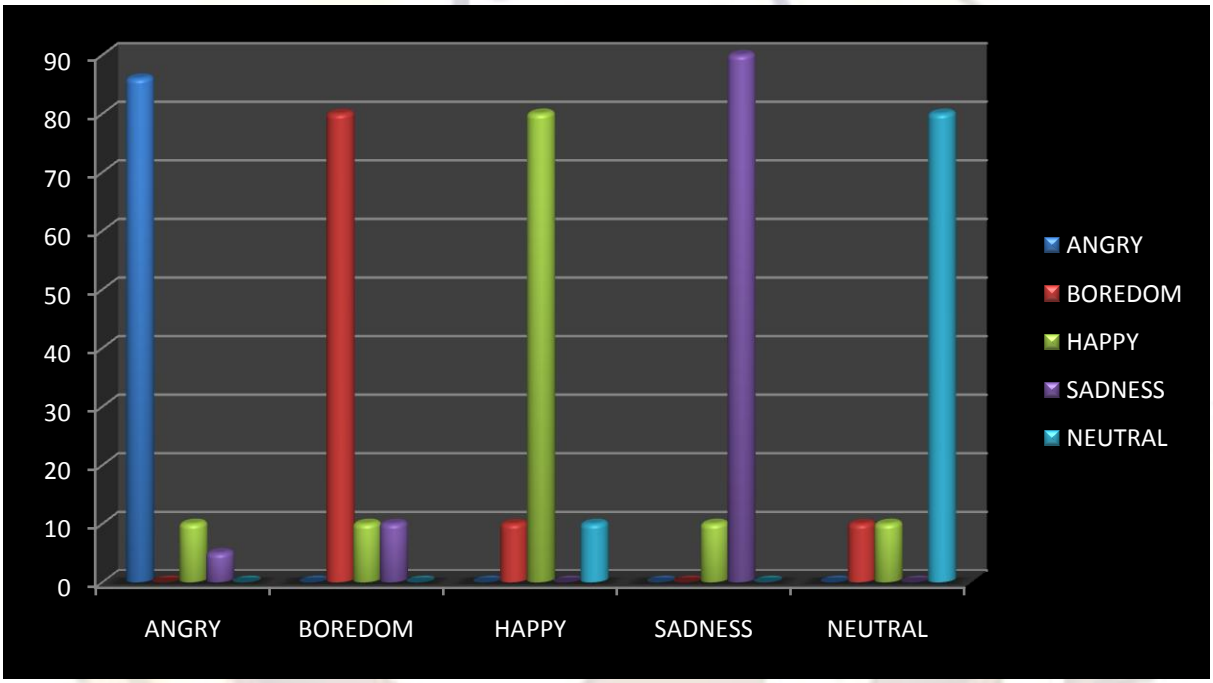
stimulation	***Recognition Emotion(%)***				
	Angry	Boredom	Happy	Sadness	Neutral
Angry	90	10	0	0	0
Boredom	10	80	0	10	0
Happy	0	0	90	0	10
Sadness	0	10	10	80	0
Neutral	0	10	0	10	80



Barchart-1, representing the recognition rates from Male database

Table-2 Confusion matrix for identify different emotion of female

stimulation	***Recognition Emotion(%)***				
	Angry	Boredom	Happy	Sadness	Neutral
Angry	85	0	10	5	0
Boredom	0	80	10	10	0
Happy	0	10	80	0	10
Sadness	0	0	10	90	0
Neutral	0	10	10	0	80



Barchart-2, representing the recognition rates from Female database

6. CONCLUSION

In this paper a novel methodology for emotion recognition is using generalized gamma distribution was developed. The emotions were considered from the database generated with 200 speakers from both the genders and from 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 Generalized Gamma 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 overall emotion rate if 80% .This shows that, the developed model performs well in identifying the emotion.

7.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

[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 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

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

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

[7]. George Almpanidis and Constantine Kotropoulos "Voice Activity Detecting using Generalized Gamma distribution"

[8]. Prasad Reddy et al, Gender based emotional recognition system for telugu rural dialects using Hidden markov models, Journal of computing, vol-2, issue-6, june-2010, pp 94-98

[9]. Ian Mcloughlin" Applied speech and audio processing "Cambridge press, 2005.

[10]. Sunil Agarwal et al (2010), Prosodic feature based text dependent speaker recognition using machine learning algorithms, International journal of Engineering Science and Technology, Vol 2(10),2010, pp 5150-5157