ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 5, May 2018

Emotion Recognition from Hindi Speech Signal

Meenakshi Singh, Parul Khullar

M.Tech (CSE), Department of Computer Science and Engineering, Satya College of Engineering and Technology Palwal, Haryana, India

Assistant Professor, Department of Computer Science and Engineering, Satya College of Engineering and Technology Palwal, Haryana, India

ABSTRACT: This project is an effort to recognize emotion from Hindi speech signals. Emotions are our life's power and vitality, which are reflected through speech signals and these signals are those carrier wave upon which the projected thoughts are sent to create or modify the reality. Emotion explains much about a person like, its behavior, intentions at an instance, therefore, it can be used in many applications like: in call centers to detect customer emotions, in field of medical (psychiatry), in criminology, in entertainment electronics, in text to speech synthesizer etc. The database used was collected from various speakers belonging to different genders and age group. This work basically focused on eight emotions which comprises of fundamental emotions with some advance emotions and are listed as , Happy, Angry, Sad, Depressed, Bored, Anxious, Fear and Nervous . Classification of emotion is done by analyzing speech signals .These signals were preprocessed and analyzed using various techniques like: cepstral , linear prediction coefficient etc.. In feature extraction of speech signals , the system uses various parameter features are : fundamental frequency, pitch contour, formants, duration(pause length ratio) etc. to form a feature vector and then Knn Classifier were used to classify and recognize those emotions.

KEYWORDS: Emotion Recognition,Feature Extraction, standard deviation of fundamental frequency, standard deviation of energy and pitch contour, formant frequencies, duration, zero crossing result.

I. INTRODUCTION

The human speech or conversation convey various kind of information about a person through two different channels where explicit channel carrying *linguistic* content(‘tells what was said’)communication , and the implicit channel contains *paralinguistic* information as: gender , nationality, age, emotion, qualification, psychological, physical state(alcohol or drug consumption) etc. Among all these properties emotion play key role in defining various properties and therefore has many application includes call centers to detect customer’s emotions , criminology, in entertainment electronics to manipulate robotic voice and to gather emotional users feedback, in text-to-speech system to synthesize emotionally more natural speech etc.

In general , emotion is subjective term, describes feelings related to position, object and many other uncertainties , that’s why it is hard to define prototype for all real life emotions under one figure in psychological literature. In an approach Ekman identifies distinct fundamental emotion class define four basic emotions such as: anger, fear, disgust and sadness. And , in another of his own definition Ekman consider , as other advance emotions as nervousness, cold annoyance , depression are derived from these basic emotions only.

Although much effort has been posed on emotion recognition over various languages but still it is in its infancy when it comes to indian Indian languages like , hindi, gujrati, tamil, oriya, Marathi, Bengali,karnarda etc. This paper focuses on emotion recognition over Hindi speech signals for eight different emotions are: Happy, Sad, Anxiety, Depression, Nervousness, Boredom, Anger, Fear .The sample data has taken from thirty different person of different gender and age in silent environment in order to avoid any additional noise in signals, in which twelve samples were faulty , eight samples have shown erroneous results.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 5, May 2018

Emotion recognition is possible through analysis of various acoustic features of human speech . This involves feature extraction and classification for which various algorithms been proposed in previous researches.
In this project more features been covered as compared to previous other researches for better results, for feature extraction like duration, zcr, pitch detection , formant estimation ,techniques used are : cepstral , autocorrlelation, LPC(linear prediction coefficient) respectively for efficient and accurate estimation of results used in recognizing emotions. Further , emotion has been classified using Artificial neural network .

II. RELATED WORK

In 2011 , Shashidhar G. Koolagudi , Ramu Reddy, Jainath Yadav , K.Sreenivasa from IIT kharagpur , proposed *emotional hindi speech corpus (IITKGP-SEHSC)* ,where considered emotions were Anger, Disgust, Fear, happy, neutral, sad, sarcastic, and surprise . emotional classification were performed over prosodic and spectral features , where energy, pitch and duration were used to find prosodic content in speech and technique mel frequency ceptsral coefficient (MFCC) for determining spectral information associated with those speech signals.

After then, In same year another research related to emotion recognition were made by same former center by Rahul chauhan ,Jainath yadav , S.G.Koolagudi and K.Sreenivasa Rao , who worked upon *Text independent emotion recognition using spectral features* over hindi speech database achieved through IITKGP-SEHSC , where the techniques used were mel frequency cepstral coefficient (MFCC) and Gaussian mixture model (GMM) for emotion recognition and classification of text independent and text dependent cases for considered emotions were: anger, happy,disgust, fear,neutral, sarcastic and surprise ,respectively.

After which many other researches were made like , In 2011 , Sujata B.Warkhade, Pritish Tijare , yashpalsingh chauhan presented paper over speech emotion recognition system : using SVM and LIBSVM , where emotion database in hindi , berlin were used for analyzing the efficiency of techniques are SVM and LIBSVM by examining their recognition rate. Here , prosodic features were extracted using MFCC and MEDC technique for emotional classification.

Then , In 2012 , Peerzada Hamid Ahmad proposed *transformation of emotions using pitch as parameter for hindi speech*, this paper focused on analysis of different emotion from neutral emotion based on pitch contour and proposed an algorithm for emotion conversion based on pitch factor involved simple rule for converting pitch points.

After that an study were made over *transformation of emotion based on intonation patterns for hindi speech* given by S.S.Agrawal, Nupur Prakash , Anurag Iain from kalinga institute of industrial technology and guru govind singh indraprastha university , in which they worked to transform neutral sentences into emotion rich sentences with an idea that changing an intonation pattern or structure of sentence can change the emotion associated with the sentence or phrase. And , for this fundamental frequency(f_0), energy contour were used as parameters to convert intonation emotion . The emotions under consideration during the experiment were: surprise, sad, happiness, anger, sadness.
Then , In 2013 , Sushma Bahugama and Y.P Raiwani from BCIIT and HNB Garhwal university proposed a work over emo-voice model for speaker's emotion identification using MFCC and vector quantization techniques in hindi database for four basic emotions were: Happy, sad , anger, neutral.

III. PROPOSED METHOD

As in previous researches related to emotion recognition in hindi speech didn't include all acoustical features such as formants , zcr etc. which are also involved here in this paper. The flow of processes associated with emotion recognition is designed below:

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 5, May 2018

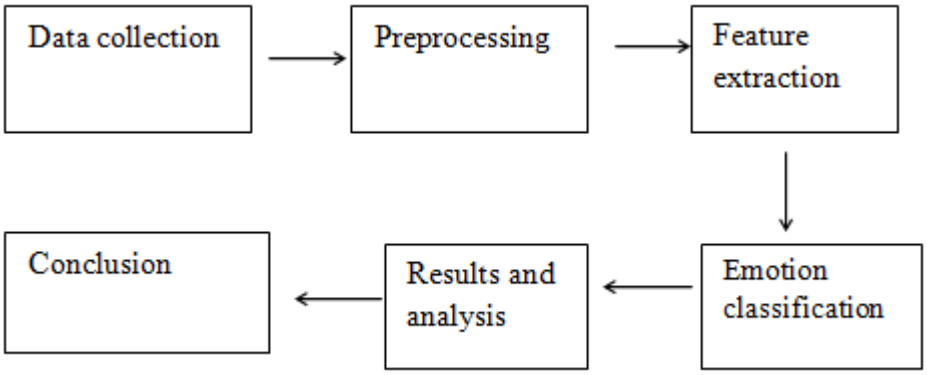

Figure 1. The flow of processes associated with emotion recognition

1. **Data collection**: source data is collected from 20 non-professional speakers (include both genders) ,each person uttered the same sentence eight times in eight different emotions, which made total of 160 samples.
2. **Preprocessing:** for analyzing achieved speech signals ,preprocessing of signal is required to extract errorless information regarding:

-duration of speech signals, zero crossing rate etc.

-fundamental frequency variation, pitch variation using two different techniques (cepstral) in order to analyze difference.

-formant frequency for detecting syllables which reflect changes in voice feature due to change in vocal tract shape on utterance of phrase with different emotions.

-energy or amplitude variation with in same phrase but with different emotions.

Here the signal is segmented in frames of 30 ms and overlapping of 10ms between frames was made for interpolation to reduce the loss of information while processing of data. Then , signal were filtered using low pass filter and after that framed signal was further divided into voiced and unvoiced signal as voiced frames differentiate themselves from unvoiced in signals by means of energy, zero crossing rate and classifying it favors accurate measurements.

3.Feature extraction

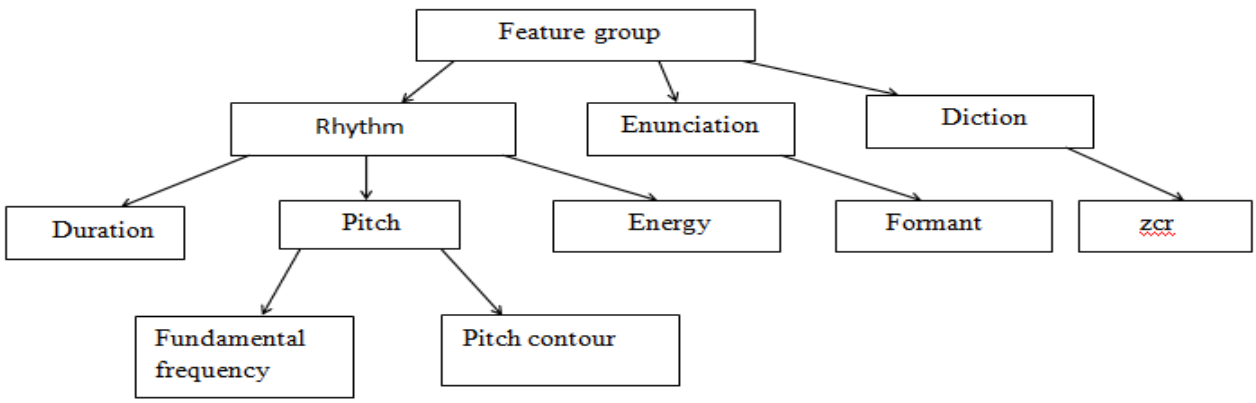

Figure 2.Feature Extraction Of Speech

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 5, May 2018

Duration : Duration of signal varies with number of pauses in the signal which depend upon emotions associated with uttered phrase, which can be calculated as:

$$T=(N-\sum_{p=0}^{p=n}(P))/dt$$

Where, T= duration of sample
N= length of sample
P=length of pause
dt= time rate

ZCR: zcr i.e zero crossing rate, which defines number of zero crossing per time unit, and it can be measured using:

$$Z= n\_c.f/n$$

Here,n_c=number of zero crossing per frame
Z=zero crossing rate per sample
f=sampling frequency., 44100
n=length of frame(here 30ms)

Energy: energy defines how much force been posed over a phrase while reciting it, it basically used to identify the intensity of signal, for maintaining accuracy , energy is calculated for short framed signal in order to avoid unvoiced part of signal. Energy can be calculated using:

$$E(x)=\sqrt[2]{\sum s_f^2}/n$$

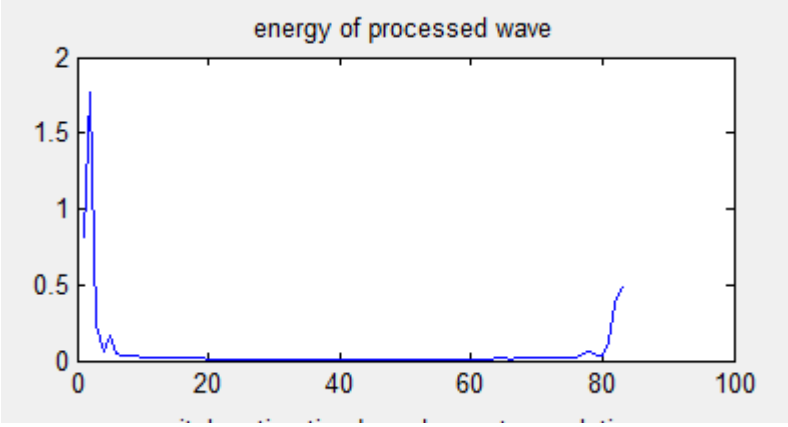

Figure :3 energy of filtered(butter filter) speech signal of phrase *Yahan aao*

E = energy of sample
s= sample value of f^{th} frame
n= frame length, here 30ms.

Pitch detection: One of the essential acoustic feature for emotion recognition is pitch , defines the rate of vibration of human vocal chord while uttering a phrase. It has its various subfeatures as fundamental frequency, harmony, pitch contour. In this paper, fundamental frequency contour,f_0,pitch contour are taken into consideration .pitch contour gives variation of pitch along the signal on utterance of a phrase with particular emotion .The fundamental frequency(f_0),that is first harmonic define as greatest common divisor of all harmonics ,plays vital role for determination of pitch as it gives highest peak in a period of signal, therefore it often called pitch .Various algorithms have been proposed for reliable estimation of pitch but none of them are able to give accurate estimation of pitch.

Commonly autocorrelation pitch detection algorithm is used for pitch estimation , but in this paper pitch been estimated through cepstral pitch determination algorithm , as this algorithm has advantages over autocorrelation based PDA ,because in autocorrelation method effect of vocal tract and vocal source are convolved with eachother but in cepstral approach they are independent and easily identifiable.

For better results, voice signals are first filtered with low pass filteration , and then separated in to voiced and unvoiced signals . further,signals were segmented into short frames of 30ms with an overlapping of 10ms. It can be estimated as:

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 5, May 2018

$$C(t)=F^{-1}(|\log(F(x(t))|^2)$$

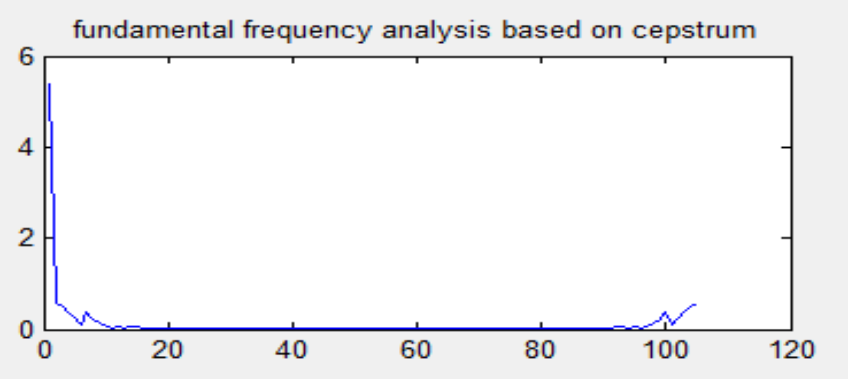

Figure 4: cepstral fundamental frequency of speech signal (containing phrase yahan aao) from frame size 30ms.

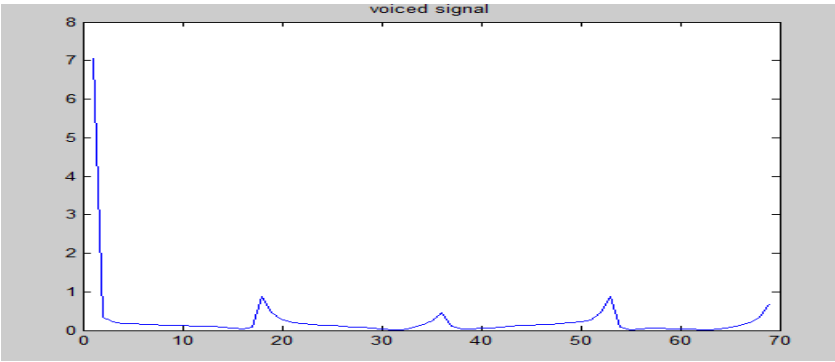

Figure 5: Cepstral fundamental frequency of speech signal using frame size 441.

4.Formant frequency: Formant frequencies has its significance in determining the phonetic content of speech signals as phonetic content cause peaks of vocal tract resonance in spectral envelop. Hindi is more phonetic content as compared to English . hindi phenomes bears about ten pure vowels (i, I, e, E, ɑ, ə, o, ʊ, u),all of these vowels also has nasalized form .Creaky and whispered vowels are rarely used.Several algorithms are in existence for finding formant frequencies such as ,analysis by analysis with fourier spectra, peak picking on cepstrally smoothed spectra, linear prediction. However , linear prediction is best approach for determining formant frequencies. The first three formants(f1),f2,f3 are usually considered to favor clear and effective analysis.

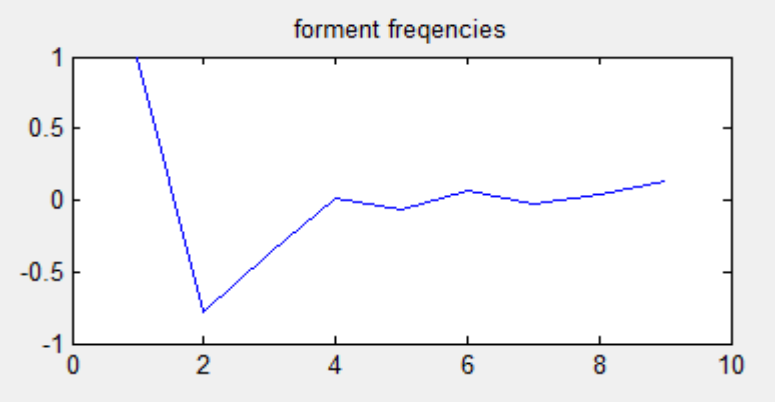

Figure :6 Formant frequency of speech signal.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 5, May 2018

5.Emotion classification: For emotion recognition , classification of emotion is obvious and inevitable requirement to train and test the data. In order to classify each parameter feature such as standard deviation of fundamental frequency, standard deviation of energy and pitch contour, formant frequencies, duration, zero crossing result , Artificial Neural Network was used , where all features parameter were used for input database and emotions as target and it has been shown through confusion matrix of 170 feature vector of each input and output sample.

IV. SIMULATION RESULT

The embedded emotions in utterances can be analyzed using graphical visualization.

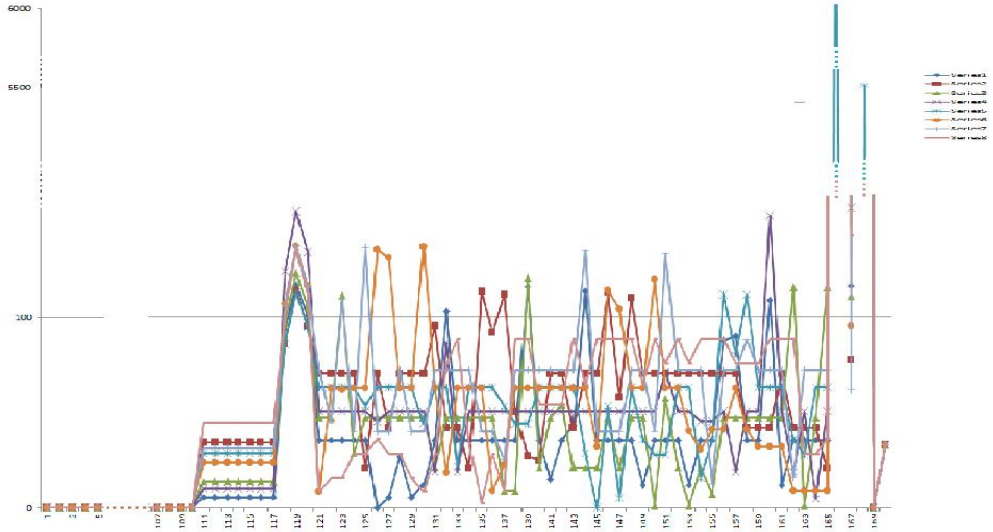

Figure 1. Graph showing variation of 170 features of total feature along different emotions.

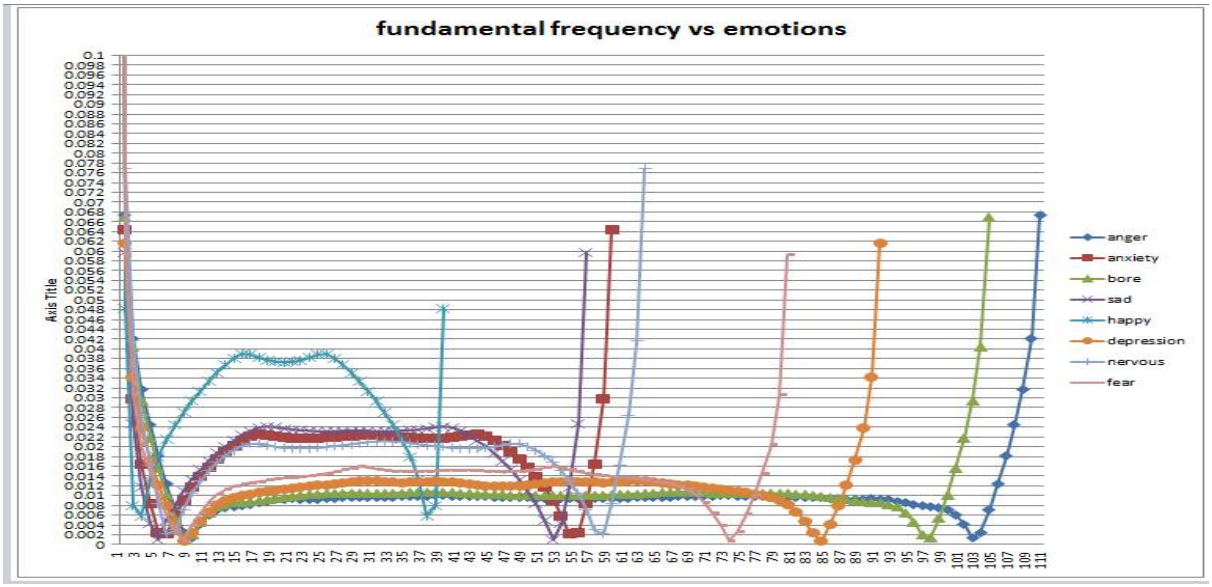

Figure 2. Variation of fundamental frequency along different emotions:

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 5, May 2018

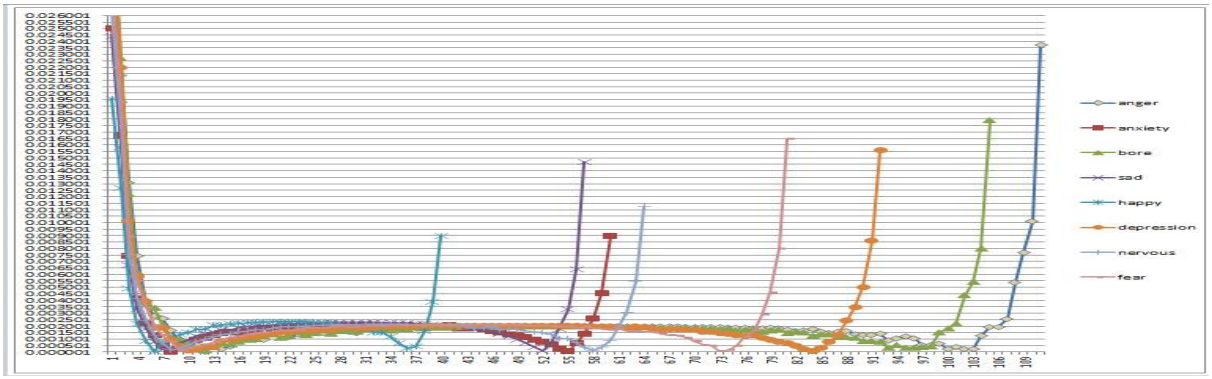

Figure 3: Showing energy variation with different taken emotions

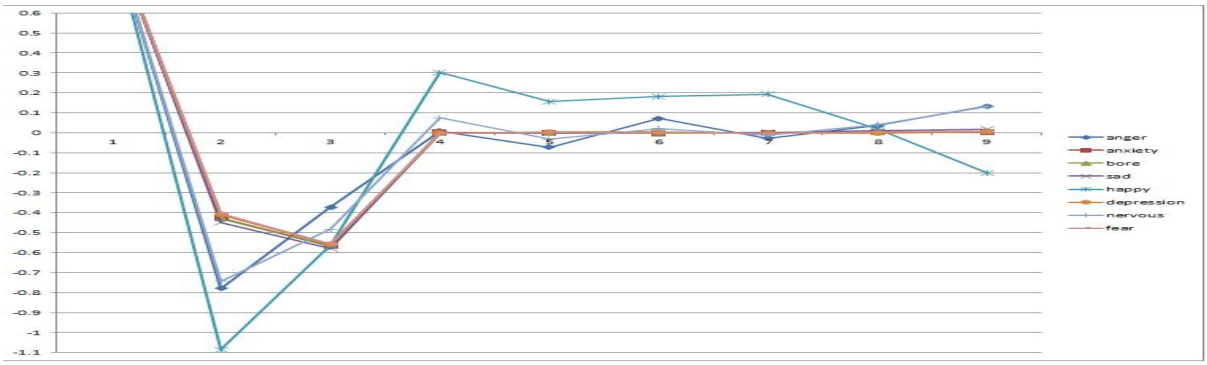

Figure 4 :Shows formant frequency variation along all eight different emotions.

Result From Above Graphs

	Anger	Anxiety	Bore	Sad	Happy	Depressio	Nervous	Fear
Anger	75.00	25.00	0	0	0	0	0	0
Anxiety	25.00	50.00	0	0	25.00	0	0	0
Bore	0	0	25.00	25.00	0	0	25.00	25.00
Sad	0	0	0	25.00	0	50.00	25.00	0
Happy	0	0	0	25.00	50.00	25.00	0	0
Depressio	0	0	25.00	0	0	50.00	25.00	0
Nervous	0	0	0	25.00	0	0	75.00	0
Fear	25.00	0	25.00	0	0	0	0	50.00

To recognize emotion , classification is required. Here, Knn classifier is used for classification purpose.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 5, May 2018

V. CONCLUSION AND FUTURE SCOPE

This will show great use in field of psychiatry, electronics and communication etc.In this paper, various acoustic features comprises of features like rhythm(energy, pitch, duration), enunciation(formant) , and diction(ZCR) are taken in to consideration for emotional classification used for recognizing emotions. Here the database for emotion recognition is over hindi speech signals in eight different emotions . The limitation it has that it can be use for .wav file only. The result shown that feature vector for each sample formed from large set features of different voice has an accuracy at an average of about 50%, and it can further be improved by using improved algorithm for pitch estimation, formant estimation etc.

Future Scope

Here it can not work over dynamic time with that its accuracy can be improve by applying the more modified and efficient algorithms.

But,

It has application in multidisciplinary areas as:

- In electronic and communication.
- In security.
- In human machine natural interaction (robots).
- In call centers.

REFERENCES

1.Moushumi Sharmin, Shameem Ahmed,Sheikh I. Ahamed, and Munirul M. Haque Wisconsin, USA
2 Sheikh I. Ahamed, Munirul M.Haque, Karl Stamm, Ahmed J Khan Wiscosin ,USA
Published by:
Originalarbeiten
In April 2010
3. Kotranza, A.Florida university,Gainesville Lok,Benjamin
Published in:
Virtual Reality Conference, 2008. VR '08. IEEE
4.http://www.sciencedirect.com/science/article/pii
5. http://scholar.google.co.in/scholar
6. http://ieeexplore.ieee.org/xpls/abs_all.jsp