International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249-8958 (Online), Volume-4 Issue-2, December 2014

Emotion Recognition from Chhattisgarhi Speech using Neural Network

Yachana Gaikwad, Yogesh Rathore

Abstract*—Speech Emotion Recognition (SER) is emerging as a crucial research area. Many works have been done in field of SER for example speaker dependent / independent SER system, language dependent/independent SER system, extracting different emotions like happiness, anger, sadness, disgust, boredom, neutral. All these works have been achieved by working on speech sample, for which we need speech emotion database. In this work we have developed speech emotion database in Chhattisgarhi language. Speech emotions can be recognized by using different features of speech, which may be prosodic feature (pitch, energy) or phonetic features (MFCC, Format Frequency) etc. Thus for selecting feature of speech for emotion identification, a review of works on speech emotion recognition is given in this paper. The aim of this paper is to present the works which are important to design and develop SER system for Chhattisgarhi language using neural network and analyze it.

***Index Terms*— Energy, Format Frequency, MFCC, Pitch, Neural Network.**

I. INTRODUCTION

Software Engineering and Technology has turned into a fundamental part of our life. An idea of Biometric is utilized as a manifestation of identification and access control. This combination is valuable to recognize people in gatherings that are under observation. Over the evolution of human machine interaction technology, a user friendly interface is going forth as more and more paramount for speech oriented application. Therefore it will be seen that, ubiquitous computing environment will require to deliver human – centered design in lieu of machine – centered one. In human interaction, emotions play consequential roles for pragmatic & semantic role. The same thing happens with human – machine interaction or developing any artificial intelligence in field of speech. Therefore research on verbalization emotion apperception has got much attention by speech researchers [1]. With time many researchers have work in application of human – machine interaction in various field like information retrieval, medical analysis etc[2]. In this paper we will work on one of the applications of biometric science, which is SER system in Chhattisgarhi language. Speech is vocalized form of human communication. Research on speech can be in terms of either speech production or speech perception or both. Here for developing speech emotion recognition system we will consider both. The study of perception is nearly joined to the field of phonetic and linguistics.

Speech perception can be defined as the process by which humans become able to intercept and understand the sound used in languages. Research studies how human listener's recognition speech sounds and uses this information to understand spoken language. Building a computer system that recognizes speech emotion is hot topic in speech research. The first book on expression of emotions in animals and human was written by Charles Darwin in the nineteenth century [3]. After this milestone work psychologist have gradually accumulated knowledge in this field. Dealing with speaker's emotion is one of the latest challenges in speech technologies. Speech is most basic and main communication tool in human to human interaction. Emotion can make its meaning more complex and the listeners can react differently according to what kind of emotion the speaker transmit, e.g. consoling a sad one with soft words. Speech signal contain different type of information including not only the information about message but also speaker's identification, emotions' identification and identification of language and so on [4][5].The word emotion is inherently uncertain and subjective, as it is an individual mental state that arises spontaneously rather than through conscious effort. Therefore there is no common objective definition and agreement on the term emotion. SER is implemented for different fields like in call center, an agent is used as a part of a decision support system for prioritizing voice messages and assigning a proper human agent to response the message at call center environment. A large number of applications exist reaching from the discipline of information retrieval to medical analysis, our daily life activities, like mobile applications, healthcare, in human-computer interaction systems etc.

A. Speech

First form of speech can be viewed as sound. But when we observe it as speech it starts, when a speaker formulates a message in its mind to transmit to the listener via movement of vocal tract. Sound can be produced by any vibrated medium. Speech is a form of sound and its production is measure by unit of phoneme.

B. Speech Production

Human verbalization is engendered by vocal organs. Any vibration medium can engender sound waves. It can be explained by one example of violin, when strings of violin's are pressed, the air above it peregrinate from their position and coerced to other air molecules. By this way they amassed together and waves are engendered. Following particles of human body are known as vocal organs, which are engenderer of sound waves.

Manuscript published on 30 December 2014.
* Correspondence Author (s)
Yachana Gikwad*, Department of Computer Science & Technology, C.S.V.T.U. RIT College, Raipur, India.
Yogesh Rathore, Department of Computer Science, C.S.V.T.U. RIT College, Raipur, India.

The human vocal organs are (1) Nasal cavity, (2) Hard palate, (3) Alveoral ridge, (4) Soft palate (Velum), (5) Tip of the tongue (Apex), (6) Dorsum, (7) Uvula, (8) Radix, (9) Pharynx, (10) Epiglottis, (11) False vocal cords, (12) Vocal cords, (13) Larynx, (14) Esophagus, and (15) Trach. This organs are shown in figure 1.

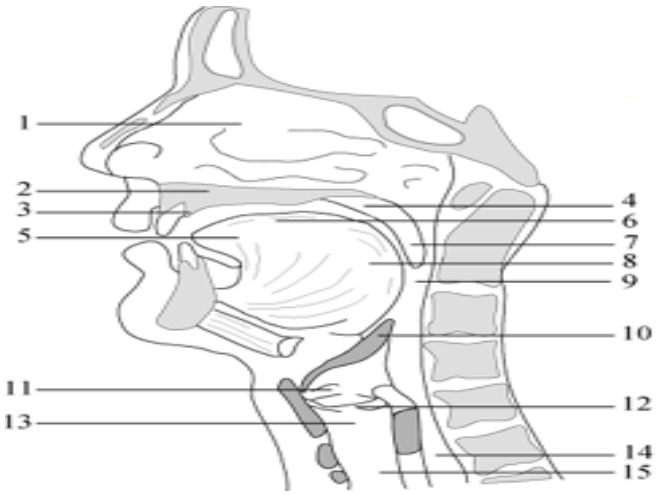

Figure 1: Human Vocal Organs

Speech production starts from lungs with the diaphragm. When verbalizing is done, the air flow is coerced through the glottis between the vocal cords and the larynx to the three main cavities of the vocal tract, the pharynx and the oral and nasal cavities. From the oral and nasal cavities the air flow exits through the nasal discerner and mouth, respectively [6]. Speech contain different types of emotions like speaker identification speech content, speaker's emotions etc [7].

C. Emotion Expression

When we talk about communication, it is striking that "what" we are talking, but it is more consequential that "how" we are expressing. Subjective & objective factors interaction form a complex set, medicated by neural / hormonal system of human body, can give rise experience such as feeling of happiness, arousal, and pleasure/ disappear. Although emotion word used by most of the psychologist but for super ordinate label "affect" or "affective process" may be used[8]. Emotions can be classified into two categories primary which are basic emotions including fear, anger, sadness, disgust and second one is secondary or derived emotions such as pride, gratitude, sorrow, irony and surprise. Primary emotions are normally experienced by all social mammals whereas secondary emotions are combination of primary ones [9] [10].There may be different types of sign that indicate emotion. In human – human communication, emotion expressed in terms of verbal or facial or by somatic cues. In verbal expression two types information are found. First is linguistic information, which is related to language of any religion and acoustic information included prosody or voice quality. Facial expression is done by movement of facial features like movement of lips in happy emotion. Somatic cues are expressed by heart rate, skin respectively, swear etc. In developing process SER system deals with verbal expression. Throughout this paper we will discuss great detail of development of SER system for Chhattisgarhi speech. Section I is introduction part. Previous works done on SER is discussed in Section II. Section III will explore problem identification. Section IV talks about methodology. In Section V we discuss result. Section VI is giving the conclusion and future scope.

II. REVIEW OF PREVIOUS WORK

Northbrook describes an experimental study on emotion recognition by developing 140 utterances per emotional state. Each utterance was recorded using close talk microphone[11]. Vocal energy, frequency, formats were used for feature extraction using neural network. He presented result with accuracy of 61.4 % for happiness, 72.2% for anger, 68.3 % for normal. Nogueries et al. presented work on Spanish corpus of interface with low level features and HMM structure [12]. The results obtained were similar to those given by human judges. Pierre-Yves Oudeyer suggested an algorithm to allow a robot to express emotion [13]. Comparison between different classification methods is given. It focused on speaker dependent system. Schuller et al. suggested a novel approach to the combination of acoustic features and language information for seven emotional states [14]. A dynamic AKG – 1000 MK- II microphone was used in an acoustically isolated room to record the emotional utterances. Result is gain reducing error rate up to 8.0 % with above mentioned combination. Drages Dtatus used DES (Danish Emotional Speech) Berline and German Emotional Database [15]. Gentle Boot classification was used for neutral, surprise, happiness, sadness emotions. This research's aim was to analysis of segmentation methods and of the performance of the Gentle Boost classifier on emotion recognition from speech. Nobuo Sato and Yasunari Obuchi proposed a new approach using phonetic features where MFCC was used by describing new algorithm for its calculation [16] .Comparison is given between proposed and previous approach. Singh et al. describe database development for Hindi Hybrid word. Main focus of this paper was to analysis database using end point detection [17]. Kuldip kumar, R.A. Agrwal presented work for Hindi speech with 30 Hindi words. HTK toolkit was used and HMM was used for classification[18]. Alexandros Georgogiannis, Vassilis Digalakis introduced one of speech's features Teager MFCC, which can also work in noisy environment [19]. Bhoomika et al. Mel Frequency cepstrum coefficients (MFCC) was extracted from speech utterance in this paper [20]. The Support Vector Machine (SVM) was used as classifier to classify different emotional states such as anger, happiness, sadness, neutral, fear, from a database of emotional speech collected from various emotional drama sound tracks. The SVM was used for classification of emotions. 93.75% classification accuracy for Gender independent case was found and 94.73% for male and 100% for female. Bageshree et al. Presented analysis of speech signal to identify the 3 emotional states Neutral, Angry and Happy from the speech using Pitch and Formant Frequencies as the basic features from the results obtained it is shown that Pitch is the best feature to identify the two emotions Neutral state and Angry [21]. K SuriBabu et al. describes development of SER system for 5 emotions anger, happy, neutral, boredom, and sadness. For classification gamma distribution was used. Result is given for all emotion individually [22].

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

Mina et al. reported an effort towards automatic recognition of emotional states from continuous Persian speech by building database of emotional speech in Persian [23]. This database consists of 2400 wave clips modulated with anger, disgust, fear, sadness, happiness and normal emotions. Prosodic features, including features related to the pitch, intensity and global characteristics of the speech signal were used for feature extraction. Neural networks were used for automatic recognition of emotion. The resulting average accuracy was about 78%.

III. PROBLEM IDENTIFICATION

SER system can be defined as pattern recognition task. Pattern recognition is the assignment of a label to a given input values. For example determine whether a given email is "spam" or "non – spam". Pattern recognition algorithms generally aim to provide a reasonable answer for all possible "most likely", matching of the input, taking into account their statistical variation. Speech emotion recognition follows following stages: feature extraction, feature selection, classifier choice and testing. This sequence is called pattern recognition cycle, and is depicted in figure2.

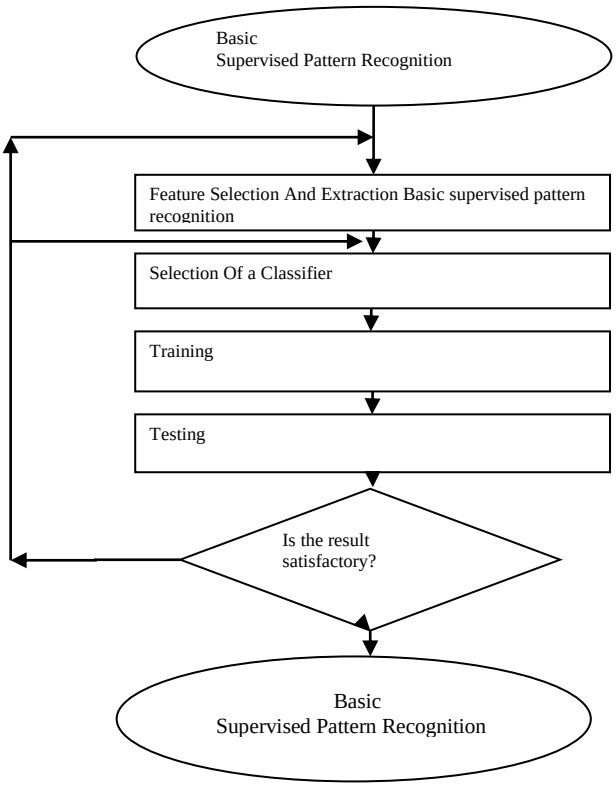

Figure 2: Pattern recognition cycle

IV. METHODOLOGY

If we sub categories the time line of SER system, it can be easily noticed that, they all are following one set of sequence for detecting emotions from speech signal, although they can use different features and different classification models. If we study the papers during 1990's (less than 10 papers) and until 2004 (30 papers per year) until today (greater than 100 paper per year) [24], we can find following methodology for developing SER system

A. Preprocessing

This is the first step in which acoustic sound pressure wave is converted into a digital signal. This paper is presenting development process for SER system foe Chhattisgarhi speech. As discussed in above paragraph it can be noticed that it is very important factor to select the dataset for which we have to work. Here the presented work is for Chhattisgarhi speech. Following are the step by step process of this work. Chhattisgarhi Emotion Speech Corpus is used in this paper. The speech corpus is recorded using 6 (3 male and 3 female) professional artists , India. The artists had sufficient experience in expressing the desired emotions from the neutral sentences. All the artists were in the age group of 25-40 years, and had the professional experience of 8-12 years. Four emotions recorded for this database are anger, happy, neutral, and sad. For expressing the emotions, 15 Chhattisgarhi sentences are considered. Each of the artists had to speak the sentences for different emotions. The number of sessions considered for preparing the database is 4. The total number of speech sample in the database is around 180 . The total duration of the database is around 3 min. The speech is recorded using digital audio system and sampled at 48 kHz. Each sample is represented as 16 bit number.

B. Feature Extraction Process

Some specific parameters are termed as features, which has goal to find a set of properties of an utterance that have acoustic correlation to the speech signal, which can somehow to computed or estimated through processing of the signal waveform. Feature extraction can also be defined as the process of measuring some important characteristics of the signal as energy or frequency response. SER methodology deals with two operations (process) signal modeling and pattern matching where signal modeling is a process of converting speech signal into a set of parameters and in pattern matching, we find parameters sets from memory which closely matches the parameter set obtained from the input speech signal. Before going to discuss about feature extraction we will see one important process, which is feature selection. There are number of features of speech pitch, formats, energy, intensity, LP (Linear Perception). Every feature of speech signal contains information related to speech signal. It is very important to choose feature of signal which carry information about transmitted emotion and is also required to fit in used classified model [25]. Feature of speech can be broadly classified into 2 categories acoustic and linguistic feature. Intonation (F0 or pitch modeling), Intensity (energy, teager), Mel Frequency Cepstrum Coefficient (MFCC), Linear prediction (LPCC, PLP), Formats (amplitude) and Tf-transmition (wavelength) are acoustic feature of speech signal. Linguistic (phenomena, word), Para- Linguistics (laughter, sigh) and Disfluencies (pause). In this paper we will study MFCC in detail. In this section we will deal with MFCC because feature extraction is the first step in developing SER system. A filtered process is adopted by shape of vocal tract which include tongue, teeth etc. This is the very important point about the speech. This shape is used when sound comes out.

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

It forms an envelope of short time power spectrum, and MFCC's job is to present it accurately to this envelope. This feature is generally used in automatic speech recognition and speaker recognition. Daris & Mermelsten introduced this feature in the 1980's, even avant – grade. MFCC can be calculated by following calculation.

C. *Steps for Calculating MFCC*

(i) Divide the signal into number of frames.

(ii)Period gram estimate of the power spectrum is calculated by each frame.For length N input vector x, the DFT is a length N vector F, with elements

$$F(k) = \sum_{n=1}^{N} x(n) * \exp(-j * 2 * pi * (k-1) * (n-1)/N) \quad (1)$$

Where, $1 <= k <= N$

iii) Mel filter bank is applied to the power spectrum, then sum the energy in each filter.Relationship between frequency f and Mel m is given by

$$m = \ln(1 + f/700) * 1000 / \ln(1 + 1000/700) \quad (2)$$

iv) Log of filter bank engineering is find.

v) Then calculate DCT of the log of filter bank engineering.

vi) DCT coefficients 2 – 23 are kept, rest are discarded.

D. *Classification*

There are different types of classification's methods are explored. Some past references can be use for choosing best classification method. They are basically two types of classification method [2]:

- Linear classification
- Non Linear classification

Linear classification uses linear combination of the object characteristic and to develop nonlinear classifier characteristic nonlinear weighted combination of object is used. Artificial neural network, Gaussian mixture models, Hidden Markov models, Decision trees, nearest neighbor algorithm are comes under nonlinear classifier. Support vector machine (SVM), Perceptron classifier, Navie Bays classifier is comes under linear classifier. One layer feed forward back propagation neural network is used. This type of neural network is selected because they have been applied successfully to a wide range of information processing tasks in such diverse fields as speech recognition, image compression. Layer consist artificial neurons or node. Network pattern recognition (NPR) tool is used for result estimation. Traing data is provided with different Chhattisgarhi speech sentences with 5 person's voice and 4 emotion happy, neutral, sad anger. Target data is used in pattern of 7 bit of 0 and 1. Testing is done by different person with different sentence and same person used in training with difference sentence.

V. RESULT ANALYSIS

Result is given with compression of human perception (HP) and machine perception (MP). Human perception is done by different person's intelligency of speech emotion recognition. A testing dataset consisting 40 sentences of Chhattisgarhi speech,10 speech per emotion. Human intelligence is used for predicting emotion identification. Here the result's average value, predicted by human mind is termed as human perception. Machine perception is done by one of the tool of neural network, NPR tool. It takes input and target value and one hidden layer consisting 10 neurons. Table 1 is showing the accuracy comparison between HP and MP. Figure 3 is screen shot of result for anger emotion with accuracy 90 % and performance value is .0087. Human intelligence is used for predicting emotion identification. Table 2 is showing comparison of accuracy result of performance value .094 with different count value. Here count value is indicating number of times for which training is performed for getting performance value less than .01. Figure 4 and 5 shows different result for same performance value with different number of count. Table 3 is showing fault acceptance ratio of this Chhattisgarhi SER system by giving the wrong identification percentage of emotion.

Table 1: Comparison table for Accuracy between HP and MP (Performance value .0087)

Emotion	Anger		Happiness		Neutral		Sad	
Prediction Type	MP	HP	MP	HP	MP	HP	MP	HP
Result	90%	. 96%	90%	96%	90%	92%	80%	96%

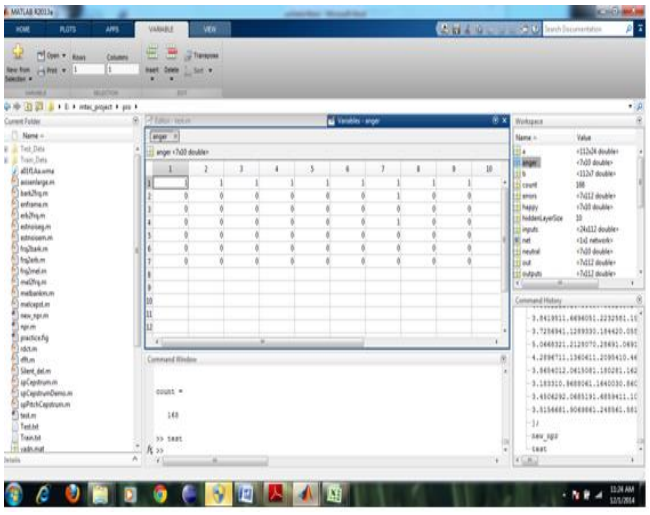

Figure 3: Figure 5.6 Screen shot of result for anger with 90 % accuracy with performance value .0087

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

Table 2 Result for Performance value .094 for different count

Emotion	Anger		Happiness		Neutral		Sad	
Count for Performance value .094	168	3	168	3	168	3	168	3
RESULT	100%	90%	90%	90%	80%	50%	90%	100 %

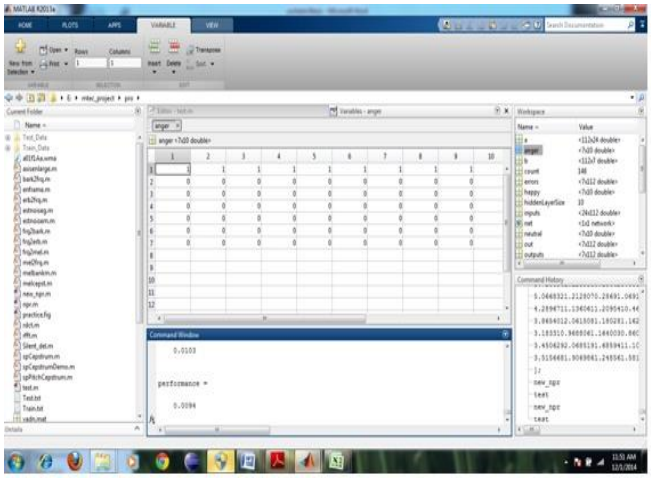

Figure 4: Result for anger with performance value= .0094 count = 148

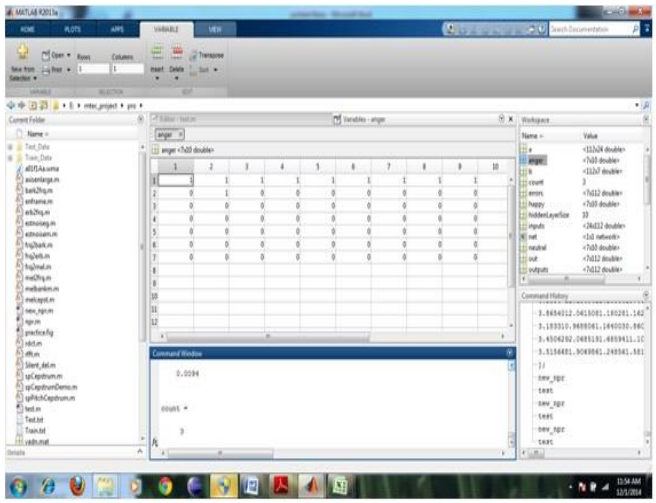

Figure 5: Result for anger with performance value .0094 counts = 3

Table 3 Fault acceptance ratio table: (PERFORMACE VALUE .0087)

	Anger	Happy	Neutral	Sad
Anger		10%	0%	0%
Happy	10%		0%	0%
Neutral	0%	0%		10%
Sad	0%	0%	20%	

VI. CONCLUSION AND FUTURE SCOPE

Result is giving 87% accuracy for four emotion namely anger, happy, sad neutral and sad emotion for neural network performance value .0087. It is found that performance value less than .01 give good accuracy result NPR tool. Result is much similar to human perception. Code used for emotion detection is not dependent on gender or sentence parameter. It classify sentence according to different emotion's feature. To create one perfect training dataset it's very important to verify it by human mind. It is concluded by this experiment that confusion rate between anger and happy emotion is high and also for neutral and sad emotion confusion rate is high.

VII. APPENDIX

SER- Speech Emotion Recognition

REFERENCES

[1] R. Cowie, E. Douglas-Cowie, N. Tsapatsoulis, G. Votsis, S. Kollias, W.Fellenz, and J. Taylor, "Emotion recognition in human-computer interaction," Signal Processing Magazine, IEEE, vol. 18, no. 1, Jan 2001.

[2] D. Ververidis and C. Kotropoulos, "Emotional speech recognition: Resources, features, and methods," Speech Communication, vol.48, no.9, pp.1162–1181, 2006. [Online].

3] Darwin, Ch "The expression of the emotions in man and animals." University of Chicago Press, 1965

[4] Shashidhar G. Koolagudi · K. Sreenivasa Rao "Emotion recognition from speech: a review " Int J Speech Technol (2012) 15 :99 – 117 DOI 10.1007/s 10772-011-9125-1.

[5] Anand singh, Dr Dinesh Kumar Rajoriya , Vikash Singh "Database Development and Analysis of Spoken Hindi Hybrid Words Using Endpoint Detection"International Journal of Electronics and Computer Science Engineering ISSN- 2277-1956

[6]http://www.acoustics.hut.fi/publications/files/theses/lemmetty_mst/chap3.html

[7] A. Reynolds and R. C. Rose, "Robust Text-Independent Speaker Identification Using Gaussian Mixture Speaker Models", IEEE transactions on speech and audio processing, vol. 3, pp. 72-83, January 1995

[8] Kleinginna, P. R., & Kleinginna, A. M. (1981). A categorized list of emotion definitions, with suggestions for a consensual definition. Motivation and emotion, 5(4), 345–379. Cited by 445.

[9] Murray, I. and Arnott, J. L., Towards the Simulation of Emotion in Synthetic Speech: A Review of the Literature on Human Vocal Emotion, in Journal of the Acoustic Society of America, pp.1097-1108 (1993).

[10] Stibbard, R. M., Vocal Expression of Emotions in Non-laboratory Speech: An Investigation of the Reading/Leeds Emotion in Speech Project Annotation Data, Unpublished PhD Thesis. University of Reading, UK. (2001).

[11] Valery A. Petrushin **"**Emotion Recognition In Speech Signal: Experimental Study, Development, And Application" 3773 Willow Rd., Northbrook, IL 60062, USA.

[12] Albino Nogueiras, Asunción Moreno, Antonio Bonafonte, and José B. Mariño **"**Speech Emotion Recognition Using Hidden Markov Models" Eurospeech 2001 – Scandinavia.

[13] Pierre-Yves Oudeyer "The production and recognition of emotions in speech: features and algorithms" Sony CSL Paris, 6, rue Amyot, 75005 Paris, France.

[14] Björn Schuller, Gerhard Rigoll, and Manfred Lang "Speech Emotion Recognition Combining Acoustic Features And Linguistic Information In A Hybrid Support Vector Machine - Belief Network Architecture"

[15] Drago_ Datcu, Leon J.M. Rothkrantz "The recognition of emotions from speech using Gentle Boost classifier. A comparison approach" International Conference on Computer Systems and Technologies - CompSysTech'06

[16] Kuldip ,Nobuo Sato and Yasunari Obuchi "Emotion recognition using mel frequency cepstram**"** International Journal of Computing and Business Research ISSN : 2229-6166

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)

[17] Alexandros Georgogiannis, Vassilis Digalakis Speech Emotion ecognition Using Non-Linear Teager Energy Based Features in Noisy Environments 2012.
[18] Kuldeep Kumar R. K. Aggarwal "Hindi Speech Recognition System Using Htk" ISSN (Online) : 2229-6166
[19] Anand singh * , Dr Dinesh Kumar Rajoriya , Vikash Singh Database Development and Analysis of Spoken Hindi Hybrid Words Using Endpoint Detection ISSN- 2277-1956.
[20] Bhoomika Panda* , Debananda Padhi2, Kshamamayee Dash3, Prof. Sanghamitra Mohanty4 " Use of SVM Classifier & MFCC in Speech Emotion Recognition System" Volume 2, Issue 3, March 2012 ISSN: 2277 128X.
[21] Bageshree V. Sathe-Pathak, Ashish R. Panat "Extraction of Pitch and Formants and its Analysis to identify 3 different emotional states of a person" IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 4, No 1, July 2012.
[22] K SuriBabu, SrinivasYarramalle, Suresh VarmaPenumatsa Emotion Classification System Based On Generalized Gamma Distribution Vol. 2, Issue 3, May-Jun 2012, Pp.1522-1526.
[23] Mina Hamidi1and Muharram Mansoorizade2 "Emotion Recognition From Persian Speech With Neural Network" Vol.3, No.5, September 2012.
[24] review paper
[25] Bjorn Schuller and Gerhard Rigoll "Timing Levels in Segment – Based Speech Emotion Recognition " Institute for Human-Macjine Communication Technische University INTERSRSPEECH 2006-ICSLP.
[26]http://in.mathworks.com/matlabcentral/fileexchange/27059-speaker-recognition-system /content / sharks _1.0/melcepst.m

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)