Development in Emotion Recognition in Human Speech: A Review

Prof. Sandeep Khanna
Department of Computer Engineering
Dr.V.B.Kolte College of Engineering,Malkapur

Mr. Vipin J. Gawande
Department of Computer Engineering
Dr.V.B.Kolte College of Engineering,Malkapur

***Abstract:* - This paper presents guidelines to address the technical challenges in vocal emotion recognition in human machine interfaces which includes audio preprocessing, extraction of emotion relevant features and classification of it. Emotion recognition is the most challengeable and interestingtopic of research which is so far is dealt with offline evolution.Paper demonstrates the different issues related to online processing, database preparation, features dominancy according to emotion andpsychological changes during the emotion production.The overall objective of this paper is to help the reader to access the feasibility of human computer Interaction**

I. INTRODUCTION:

There exists considerable cultural variation in how emotions are conveyed, but there are universal properties in facial and vocal affective expressions.Emotions are traditionally classified as primary and secondary emotions, primary emotions are consistent across all human culture and among all social mammals. Secondary emotions are variations or combinations of primary one and unique to humans. There are considerable evidences that the emotion produces changes in respiration, phonation and articulation all of which determine the acoustic parameters of signal and listener also get infer affective state and speaker attitude from that signal. The speech stream is highly complex and variable signal that is most directly studied by Analyzing its acoustics properties or sound patterns from our everyday experience that talker provide about their emotional states through the acoustic properties of the speech In this paper we focus on technical issues and challenges that arises during equipped with human computer interaction with the ability to recognized the users vocal emotions. In this paper, we have studied discrete model of emotions as identified by psychological.

Remaining paper is organized as: the next section explains about the different emotional databases with some promising application fields. Three main parts of the automatic emotion recognition Speech preprocessing, feature extraction and classifications are discussed in each section respectively.At last we exemplify the major challenges and issues of vocal emotion recognition.

II. PHONATORY / ACOUSTIC MEASURES OF EMOTION IN SPEECH:

Physical measures of human speech and vocal sounds are based on three perceptual dimensions loudness, pitch, time information in emotion is decoded in all aspects of language. What we say and how we say that how is more important for effective communication. The vocal emotion carries throughprosodic and acoustic features of the speech and also depends on voice quality. The important research about acoustic parameter is done by murrey and arnott [1] they summarized several notable acoustic attributes for detecting primary emotions shown in the table 1, and classification of emotional states based on prosody pitch intensity energy duration and voice quality requires classifying the connections between acoustic features in speech and emotions It is important thing find suitable features and modeled for use in recognition, In particular pitch intensity energy seems to be correlated and activation [2] [3] . From the table the vocal emotion recognition seems to be straight forward and simple but unfortunately not the case [4] [5]

Table 1: Acoustic Parameters of Tone Sequences Significantly

Rating Scale	Tempo	Harmonic	Pitch variation	Pitch level	Envelope	Amplitude variation
Pleasantness	Fast	Few	Large	Low	Sharp	Small
Anger	Fast	Many	Small	High	Sharp	High
Boredom	Slow	Few	Small	Low	Round	Moderate
Disgust	Slow	Many	Small	Low	Round	Small
Fear	Fast	Many	Small	High	Round	High
Happiness	Fast	Few	Large	High	Sharp	Moderate
Sadness	Slow	Few	Large	High	Sharp	Moderate
Surprise	Fast	Many	Large	High	Sharp	High

Contributing to the Variance of Attributions of Emotional States

A variety of acoustic features have also been explored M Lee et al[6] worked for classification of negative and non-negative emotion using linear discriminate classification with Gaussian class conditional probability distribution and k nearest neighborhood methods ,the features used by the classifier are utterance level statistics of the fundamentals frequency and energy of the speech signal, for improvement of performance of classifier they also used promising first selection method and forward selection, PCA is s used to reduce the dimensionality of the features. Overall classification error rates in female speechdata were 20 % for PFS and 22.5 % for FS and 26 % for PCA with KNN and results for male speech were 24.19% for PFS, 24.19 for FS and 33.87% PCA. Valery petrushin et al [7] used some statistics of pitch, the first and second formants ,energy and speaking rate as featured and classified it using ensembles of neural network recognition the demonstrate the accuracy for neutral is 55 -75 % ,happiness is 60 70% anger is 70 80% sadness is 75 80 % fear is 35 55 % total average accuracy is about 70 % they also analyses telephone quality speech to distinguish calm and agitation with the accuracy if 77% Tsang long et al [8] used LPC and MFCC for feature extraction and mini distance and nearest class mean for classification of mandarin data and they classified anger boredom, happiness neutral sadness and they got average accuracy of 79.1% .Y Lin et al [9] used MFCC within the method based on S.V M a new vector measuring the difference between Mel frequency scale sub bands energies is proposed .The performance of KNN classifier using proposed vector was also investigated both gender dependent and gender independent experiments were conducted on Danish emotional speech the recognition rate by HMM were 98.5% for female and 100 for male and 99.5% for gender independent class Liqin at el [10] worked for speaker independent emotion recognition relative features obtained by calculating the features change of emotion speech relative to natural speech adopted to weaken the influences from the individual differences improved ranked voting fusion system is proposed to combined the decisions from four HMM classifiers. which are different feature vector respectively the recognition of results of the provided algorithm have been compared with isolated HMM with absolutely features by Berlin database of emotional speech and average recognition rate has reached 78.4% is speaker dependent case. Christos Nikolas et al[11] found 133 sound features extracted from pitch MFCC energy and formant were evaluated in order to create feature set sufficient to discriminate between seven emotions in acted speech multilayer perceptron were trained for emotion recognition on the basis of 23 input vector which provide prosody of the speaker over the Entire sentence, speaker are not to classifier the proposed featured vector achieved promising results 51% for speaker independent recognition in seven emotion emotions classes for low and high arousal emotion it reaches 86.6% successful recognition. Ling he at el [12] conducted experiments on SUSAS with three classes with high stress moderate stress low stress and ORI with five classes angry ,happy, anxious, dysphonic and neutral and features extracted using teager energy operator perceptual wavelet packet MFCC model using two classifier the GMM probabilical neural network ,for SUSAS they got 95%-61% results and for ORI it is 37 to 57 % Mingli song et al[13] used tripled HMM for audio visual based emotion recognition this experimental results shows that this approach out performs only using visual and audio separately Schuller et al[14] introduced speech emotion recognition by use of continuous HMM and global statistics frame work of an utterance is classified by GMM using derived features of the raw pitch energy contour of the speech signal 86% .Kyung Hak Hyun at el[15] gives idea about the phoneme dominant features and emotion reflective features were extracted based on the same phoneme information which was classified by phoneme dominant features which are more sensitive to emotion but less sensitive to phoneme Manish gaurav at el [16] analyze the performance of the spectral and prosodic features and the fusion of GMM and SVM and combined the scores from the different model Carlos Busso et al[17] used two databases, one recorded with a microphone and another recorded from the telephone application they got accuracy level up to 78 % and 65 % respectively and raw melfilterbank outperformed than conventional MFCC with both in broad band and telephone band speech. Shashidharkoolagudi et al [18] used LPCC MFCC LFPC are explored for classification of emotion for capturing the emotion specific knowledge from the above short term speech features with VQ model for simulated emotion speech corpus is used

II. MODELING OF DISCRETE EMOTIONS:

As all the people model their emotions differently it is not easy to judge or to model human emotions. Researchers used two different methods as Discrete model like joy ,sad ,happy ,surprise ,anger ,love ,fear ,etc. but it contains blended emotions that cant be adequately expressed in words. The choice of the word is too restrictive and culturally dependent. Another way is to have multiple dimensions or scale to categorize emotions like pleasant, unpleasant, attention, rejection, simple, complicated, positive, negative, etc.

Two common scales are valence and arousal the valence represents the pleasantness of stimuli at positive at one end and negative at another end. for example, happiness has positive arousal and sadness has a low arousal where as surprise has a high arousal. The different emotion label can be plotted at various positions on two dimension plane spanned by the two axes to construct 2D emotion model in figure 1 [19] [5]. As an emotion is function of time content space, culture and person, physiological patterns may widely differ from user to user and from situation to situation.

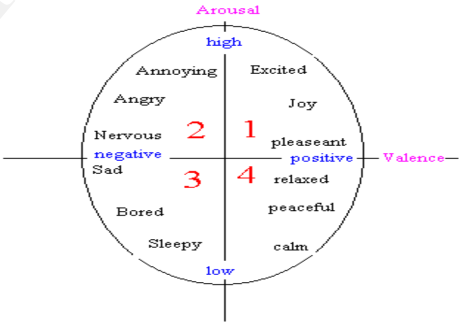

Fig1: A two dimensional vocal emotional space with valance and arousal

III. AUTOMATIC VOCAL EMOTION RECOGNIZER:

A vocal emotion recognizer consists of three parts speech processing, feature calculation and classification is shown in figure 2. The preprocessing involve the digitization and acoustic preprocessing like de-emphasis as well an framing and segmentation of the input feature calculation is concerned with identifying the relevant feature of acoustic signal with repeat to the emotion and the final stage is classification pointing out the differences between acted and spontaneous speech which is highly relevant for human machine interface[20] .

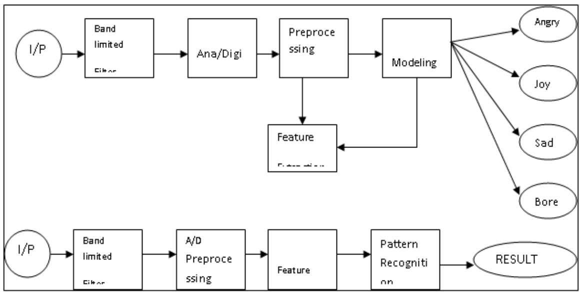

Figure 2: Vocal emotion recognition system

i) Preprocessing:

Most audio sample operates on analysis widows or frames or input audio the choice of window size as well as analysis method depend strongly upon the identity of the signal being analyzed. Speech noise and music all have different characteristics, at first Rabiner-sambur algorithm used to detect the end point and silence period are removed if the energy of 10-15ms frame is less than of average energy of 10-15ms. Such frames are taken consecutively all over each speech sample the frame is considered as silence frame and removed then speech samples are passed through high pass filter which gives a spectral tilt to the speech sample then speech samples is segmented in to 20-24ms frames with end frames having 30-50 us overlapping with the adjacent frame each frame is then multiplied by window of the same length [21].

ii) Relevant feature extraction

The next step for an emotion recognition system is the extraction of relevant features emotion specific information is present at all level of the speech signal and need to extract the features at different levels these features are used to build the models for capturing emotion specific knowledge. [Most approaches so far have deal with utterances of acted emotions where the choice unit is obviously just one utterance a well-defined linguistic unit with no changes of emotion within in this case[22]. However in speech this kind of observation unit does not exist] those feature are the characteristics for emotion and represent them in the form of n-dimensional feature vector and all the feature are not useful and good feature seems to be highly data dependent ,however high no. of feature is often not beneficial because most classifier are negatively influenced by redundant correlated or irrelevant features as a consequence most approaches compute high no of feature and apply them in order to reduce the dimensionality of the input data a feature selection algorithm that choose the most significant feature of training data for the given task alternatively a feature reduction algorithm like. can be used to each code the main information of the feature space more compactly .the start set of the features consisted originally mainly of pitch and energy related feature, formant and MFCC are also frequently found duration and considered in same paper [23].Voice quality, spectral measure and parametric represents other than MFCC 00 included wavelets, wavelet packet, teaser energy operator base feature, LPCC and LFPC [24].

Some super segmental acoustic feature also considered as global emotion features Bat liner el al and delivers et al used among them hyper clam speech pauses inside word syllable lengthening off talk respectively , dissiliency use inspiration,expiration mouth nose ,laughing ,crying , unintelligible voice, those there have mainly unnoted by hand ,automatic extraction would also be possible. The raw pitch energy continuous can be used as is and are these called short term feature or more often the actual features are derived from these acoustic variable by applying functions over the segment of values within an emotion segment called global statistic features The dynamic properties of emotion should be captured by the feature while in the letter case they are dealt with by the classifier [21][25].

iii) Emotional Classification

The emotional characteristics are distributed at different level of speech such as source level, system level & prosodic level. The performance of emotion reorganization is mostly dependent on the classifier after the future calculation & input unit is represented by feature vector & problem of emotional recognition can now be considered as data mining problem so, in principle any classification that can with high dimensional data can be used but static classification like support vector machine, Neural network and decimation in trees for global Features and hidden Markov model for short term features as a dynamic modeling techniques are most commonly found in literature on emotion speech recognition [26][27] .All these classifiers need training data to learn parameters & A direct comparison of static and dynamics classification is difficult since not the same features can be used so it difficult to say if just features have been chosen more favorable or if really the classifier has been superior dynamic classifier is very promising but currently for static classifier more features types can be exploited like jitter and shimmer to measure the voice quality so that the overall the performance better however when the features set is restricted to the same features type the quality of classifier can be determined in comparison to human rates in listening test or other classifier algorithms As a general tendency it can be observed that sophisticated classifier do achieve higher recognition rates than simple classifier but not much SVM is the popular and most often successfully applied algorithm [28] [29].

VI. DATABASES:

Performance of vocal emotion recognition is totally depend on the quality of the database generally research deals with database of acted by actor, induced or completely spontaneous emotions the complexity of the system increases with the naturalness, each database consist of corpus of human speech pronunciation under different emotional conditions with respect to authenticity there seems to be three types of databases [30].

Type 1 is acted emotion is obtained by professional actor to ask them to speak predefined sentences. In some experiment focusing on the production and perception of real and acted emotional speech supported the opinion that acted speech is not fact when spoken and perceives more strongly that real emotional speech e.g. Berlin database [31],Type II coming from real life systems like call center's e.g. Smartkom database ,Type III is elicited emotions database where emotions are provoked and self report is used for labeling control , generally researcher used type 1 database which is acted by actor and most simple to recognize ,whereas type IV is the Real life database which is very difficult to collect and processed as shown in the following figure The real life emotional speech in natural situations is difficult because of the poor acoustic recording conditions usually present. In sporting events competitions spectators and commentators alike often experience a range of emotions of quiet high intensity. Discrete emotion (arousal) is not differentiated by vocal uses [32].

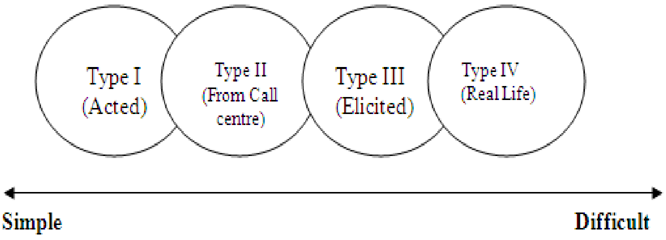

Figure 3: Types of databases used for emotion recognition and their difficulty

IV. APPLICATIONS:

Emotion in human and animal a new wave of interest has recently risen attracting both psychologists and artificial intelligence specialist there are several reasons for renaissance such as technological progress in recording storing and processing of audio and visual information the development of non intuitive sensors the advent of wearable computers urge to enrich human computer interface from point and click to sense and feel and invasion [33]. Emotion recognition could be used for Psychiatrics diagnosis in intelligent toys and lie detector alsoused for business in particular in call centre environment one potential application is the detection of the emotional state in telephone conversion and providing feedback to an operator supervisor for monitoring purposes another application in surfing voice mail messages according to the emotional expressed by the caller, One more challenging problem to use emotional content of the conversation for the operator performance evaluation Chongguo li al el made one chatting robot for communication with human in natural approach the main input of this chatting robot is speech from the user and the output is its response in speech with useful information It can be include information of emotion to other party in conventional video teleconferencing and web based teaching for added effects in distance learning if a student doesn’t understand what the teacher is saying the emotion may be detected on his face and in his/her speech These responses may have a direct and immediate influences to the teacher who would be turn try to explain the topic again the emotion based feedback is especially important in communicating with young children Possible applications includes intelligent speech based customer information systems ,human oriented HCI GUI interactive movies ,intelligent toys and games situated computer assisted speech training systems and supported medical system[34]. Voice analysis has often used to diagnose a variety of psychopathological states particularly depression and schizophrenia.

VII. CHALLENGES:

The challenges involved in the development of human emotion recognition the nature of the human expressions are complex and lack of ability to classify and label them precisely, in the speech recognition, there are layers of information with absolute labels of words and meanings. However in the area of expressive speech recognition these higher levels do not exist, at least not in such an absolute and obvious form. Some of the main challenges which are inherent to the expressions domain are listed and more technical oriented challenges are briefly discussed.

i) Obscure information

Human expressions are conscious, unconscious and intentional expressions reveal mental states, emotions and attitudes. They also reveal intentions, speech acts like greeting, apologizing, describing, asking and dialogue acts such as statement, acknowledgement, question and answer. In addition, human expressions are affected by physiological states like discomfort, and environmental contexts, starting from back-ground noise that affects the loudness or intensity of speech to social contexts that may affect the display rules. All this information is carried by the same basic cues that are also shared by the verbal speech.

ii) Neutral Vocal expression

Another implication is that there is no 'neutral' expression. Because of the assumption of prevailing moods and mental states, there is no real definition of neutral. Therefore, there could be no requirement to initiate a system using a certain 'neutral' expression according to which all the other expressions would be calibrated. The definition of 'neutral' as small variations of the same expressions may still exist, as well as neutral as opposed to opinionated.

iii) Lexical information in speech

Lexical definitions of emotions, mental states, speech acts and communication cues depend on language and cultural background. The difference in language and culture can mean that certain labels of expressions cannot be analyzed or translated from one language to another, because the vocabulary has no parallel notions, so an expression may be interpreted in a slightly different manner .Many subtle nuances of daily expressions have no name. Analyzing and labeling mixtures of expressions and their underlying meanings are even more complicated. People recognize many of the underlying expressions, but oftened them very difficult to name or label [20, 21]. Even a small subset of definitions, for example *anger*, which is considered one of the basic emotions [26], or the more elaborate set of *cold anger* and *warm anger*, cover a whole variety of nuances, among which we can distinguish by context, vocal cues and sometimes also lexically.

iv) Disruptive behavior of Speech

The probability of a person repeating precisely the same expression, with the same text and with the same underlying meaning or meanings, is very low. Different people respond in different manners to the same scenarios, and have different ways of expressing themselves according to their personality and background.

v) Context in speech

Every utterance has a certain context, and analysis of human expressions is related to it. It may be the interaction with the audience or environmental parameters. These parameters influence the way the utterance is expressed, and the way it should be analyzed. The interaction point of view influences the design in two major ways, the first is the time domain, and the second is the context domain A complication may occur when the target of the speech, either audience or medium changes.

vi) Adaptive and asynchronous changes in time

Mixtures of vocal expressions occur at all the time. These durations and occurrences of those expressions changes asynchronously during an interaction and carry their own impact. Some people do not speak continuously and endlessly, but rather in time segments whose duration and timing are set according to the personality and the nature of the interaction. An analysis of a continuous interaction should integrate the analysis of the current speech segment with the analysis of the previous transitions among speech segments. Dynamic tracking of changes is significant for analysis of speech tendencies, such as escalations of situations, and emotions. An ideal system would track both long term states and gradual changes in addition to sudden changes. The manner of response to these changes depends on the specific application.

vi) Ambiguity resolution in speech

Spoken language is filled with ambiguities at the lexical level. Words always have more than one meaning, and listeners must resolve this ambiguity to access the appropriate meaning of a given lexical item. Research in which this issue has been investigated has focused primarily on determining at what *point* during the course of spoken language processing semantic and sentential context serve to disambiguate lexical items

For example, for subdued some vocal expressions perceived meaning can be tired, bored, sad, can be interpreted as related to an introvert personality. Additional cues during sustained interactions can help to adjust and to refine the initial estimations.

vii) Dynamic changes in spoken utterance

Another kind of dynamic behavior can be found within an utterance. This level is related more to the technical side of the inference system. It includes various intonation and energy patterns. Short-term variations within an utterance can change our perception of the uttered expression in a way that statistics over the whole utterance do not reveal. Implications for design for all these reasons, the required solutions for both inference and continuous synthesis should enable both the estimation of single utterance expressions and an analysis at the interaction level.

VIII. CONCLUSION:

Dealing with the vocal emotions is one of the challenges for speech processing technologies. Whereas the research on automated facial emotions recognition has been quite extensive, that focusing on speech modality, both for automated production and recognition by machines, has been active only in recent years and has mostly focused on English. Possible applications include intelligent speech-based customer information systems, human oriented human-computer interaction GUIs, interactive movies, intelligent toys and games, situated computer-assisted speech training systems and supported medical instruments. The selection of a feature set is a critical issue for all recognition systems. In the conventional approach to emotion classification of speech signals, the features typically employed are the fundamental frequency, energy contour, duration of silence and voice quality. The effect of any emotion on speech will depend on the attention and cognitive conflict between the speaker's emotional response and the focus of speech; physiological changes such as different breathing styles.for instant, when one is in a state of anger fear or joy the sympathetic nervous system is aroused, the heart rate and blood pressure increase the mouth becomes dry and there are occasional muscle tremors speech is then loud fast enunciated with strong high frequency energy. when one is bored or sad the parasympathetic nervous system is aroused, the heart rate and blood pressure decrease and salivation increases which results in slow pitched with little high frequency energy, In vocal transmission characteristics visual contact between the speaker and the listener is not necessary, sound can be used to attract attention as listeners always have their listening "turn on" and open to sound coming from any source. Vocal channel should be more suitable for the signaling of certain emotions than of others. Fear and alarm are the most clearly vocally expressed emotions in long distance communication whereas disgust is the bad one for the long distance.

According to the language very short term changes have been observed, in fundamental frequency and expressed emotion in tone language than in indo American languages. Difficulties in vocal Emotion recognition is also due to human comprehension of speech compared to emotion recognition human can predict the words with the idioms, in computer based emotion recognition used some kind of statistical model to improve the prediction. It is difficult to model the word knowledge, knowledge of the speaker and encyclopedic knowledge, body language, noise ,spoken language is dialog oriented,and with dissiliences like hesitations repetitions,changes of subject in the middle of an utterance,slip of tongue, continuous speech, channel variability,speaker variability.

Characteristics of speech samples changes with respect to speaking style, gender Anatomy of vocal tract, speed of speech and also changes as per social and regional dialect, natural language has an inherent ambiguities like homophones word boundary ambiguities [35]. Further improvement and expansion may be achieved according to the following suggestions: The set of the most efficient features for emotion recognition is still vague. A possible approach to extracting non-textual information to identify emotional states in speech is to apply all known feature extraction methods. Thus, we may try to incorporate the information of different features into our system to improve the accuracy of emotion recognition. Recognizing emotion translation in real human communication is also a challenge. Thus, it will be worthwhile to determine the points where emotion transitions occur.

REFERENCES:

1. Murray, I.R.; Arnott, J.L," Synthesizing emotions in speech: is it time to get excited"Fourth International Conference on ICSLP 96.
2. Banse, R., & Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression.*Journal of Personality and Social Psychology*, *70*(3),614–636.
3. Laukka, P. (2004). *Vocal expression of emotion—discrete-emotion and dimensional accounts*. Comprehensive Summaries of Uppsala Dissertations from the Faculty of Social Sciences 141, ACTA UniversitatisUpsaliensis, Uppsala. Experiments.
4. Burkhart, F., van Ballegooy, M., Englet, R., Huber, R. *An emotion aware voice portal.*Proc. Electronic Speech Signal Processing ESSP. 2005.
5. Burkhardt F., Ajmera J., Englert R., Stegmann J., Burleson W. *Detecting anger in automated voice portal dialogs.* Proc. INTERSPEECH'2006, Pittsburgh, 2006.
6. Montero J. M., Gutierrez-Arriola J., Cordoba R., Enriquez E., Pardo J.M. *Spanish emotional speech: towards concatenative synthesis* COST 258,1998.
7. M. Lee, S. Narayanan. R. Pieraccini Recognition of Negative Emotions from the Speech Signal Publication Year: 2001 , Page(s): 240 – 243.
8. T sang long paoyutechenMandarin emotion recognition in speechIEEE workshop ASRU 2003 pp 227 230.
9. Lin, Yi-Lin, and Gang Wei. "Speech emotion recognition based on HMM and SVM." Machine Learning and Cybernetics, 2005.Proceedings of 2005 International Conference on.Vol. 8.IEEE, 2005.
10. Fu, Liqin; Wang, Changjiang; Zhang, Yongmei; Relative Speech Emotion Recognition Based Artificial Neural Network ICSPS, 2010 Volume: 1.
11. Thurid Vogt, Elisabeth Andr´e, and Johannes WagnerAutomatic Recognition of Emotions from Speech: A Review of the Literature and Recommendations for Practical RealisationSpringer-Verlag Berlin Heidelberg 2008.
12. LingHe; Lech, M.; Maddage, N.; Emotion Recognition in Spontaneous Speech within Work and Family Environments 3rd International Conference onBioinformatics and Biomedical Engineering , 2009. ICBBE 2009.pp 1-4.
13. *Mingli Song, Chun Chen, Mingyu You* Audio-Visual Based Emotion Recognition Using Tripled Hidden Markov Model Icassp 2004 pp 876 - 880.
14. SCHELLURE, B.; Rigoll, G.; Lang, M.; Emotional Feature Extraction Based On Phoneme Information for Speech Emotion Recognition Acoustics, Speech, and Signal Processing, 2003. Proceedings. (ICASSP '03). 2003.
15. KyungHakHyun; Eun Ho Kim; Yoon KeunKwakPerformance analysis of spectral andprosodicfeatures and their fusion for emotion recognition in speech ICASSP.2003.
16. *Manish Gaurav*Performance Analysis Of Spectral And Prosodic Features And Their Fusion For Emotion Recognition In Speech SLT 2008 pp 313-317 .
17. CarlosBusso; Zhigang Deng; Michael Grimm; ; IEEE Transactions on Audio, Speech, and Language Processing, Volume: 15 Issue: 3 2007 , Page(s): 1075 – 1086.
18. Koolagudi, Shashidhar G.; Reddy, Ramu; Rao, K. SreenivasaEmotion Recognition from Speech Signal using Epoch Parameters International Conference on Signal Processing and Communications (SPCOM), 2010 pp 1-5.
19. B. Schuller: *"Towards intuitive speech interaction by the integration of emotional aspects,"* IEEE Int. Conf. SMC 2002,
20. Kandali, A. B., Routray, A., &Basu, T. K. (2008a). *Emotion recognition from speeches of some native languages of Assam independent of text and speaker*.National Seminar on Devices, Circuits and Communication, Department of E.C.E., B.I.T.Mesra, Ranchi, Jharkhand, India, 6–7 Nov.
21. L.R. Rabiner and M.R. Sambur, "An Algorithm for Determining the Endpoints for Isolated Utterances", *The Bell System Technical Journal*, Vol. 54, No. 2, Feb. 1975, pp. 297-315
22. J.L. Shen, J.W. Hung and L.S. Lee, "Robust Entropy-based Endpoint Detection for Speech Recognition in Noisy Environments," *1998 International Conference on Spoken Language Processing* Sydney, Australia, Nov-Dec 1998.
23. Li Y. and Zhao Y. *Recognizing emotions in speech using short-term and long-term features.* Proc. of the international conference on speech and language processing.pp. 2255-2258, 1998.
24. J. H. L. Hansen and B. Womack, "Feature analysis and neural network based classification of speech under stress,"*IEEE Trans. Speech Audio Process.*, vol. 4, no. 4, pp. 307–313, Jul. 1996.
25. Patil, H. A., Dutta, P. K., &Basu, T. K. (2006). The wavelet packet based cepstral features for open set speaker classification in Marathi. In M. Spiliopoulou et al. (Eds.), *Studies in classification, data analysis, and knowledge organization* (pp. 134–141).
26. B. Schuller, G. Rigoll, and M. Lang, "Hidden Markov Model-based Sueech Emotion Recomition," Proceedings of IEEE-ICASSP, pp. 401-405,2003.
27. Nogueiras, A., Moreno, A., Bonafonte, A., and Mariño, J.B., "Speech Emotion Recognition Using Hidden Markov Models", *EUROSPEECH 2001*, Scandinavia.
28. A. Razak, A. H. M. Isa and R. Komiya, "A Neural Network Approach for Emotion Recognition in Speech",*Proc. 2nd Int. Conf. Art. Intell.In Engineering &Technology*, Aug 3-5, 2004, Kota Kinabalu, Sabah, Malaysia.
29. Kandali, A. B., Routray, A., & Basu, T. K. (2008b). Emotion recognition from Assamese speeches using MFCC features and GMM classifier.In *Proc. IEEE region 10 conference TENCON 2008*.
30. Murray and J.L. Amott, "Towards the Simulation of emotion in Synthetic Speech A review of **the** Literature on Human Vocal Emotion," Journal of the Acoustic Society of America, pp. 1097-1108, 1993.
31. Burkhardt, F., Paeschke, A., Rolfes, M., Sendlmeier, W., Weiss, B.*A database of German Emotional Speech.*Proc. Interspeech 2005, ISCA, pp 1517-1520, Lisbon, Portugal, 2005.
32. Montero J. M., Gutierrez-Arriola J., Cordoba R., Enriquez E., Pardo J.M. *Spanish emotional speech: towards concatenative synthesis* COST 258,1998
33. valery A pettrushin,"Emotion Recognition in Speech signal Experimental Study ,Devolopment ,and application sixth international conference on spoken language processing 2000
34. *S* **.**Yacoub, **S.** Simske, X. Lin, J. Burns, "Recognition of Emotions in Interactive Voice Response Systems," Eurospeech, HPL-2003-136,2003.
35. Scherer, K., "A Cross-Cultural Investigation of Emotion Inferences from voice and Speech: Implications for Speech Technology", *ICSLP 2000*, Beijing.