International Journal of Applied Information Systems (IJAIS) – ISSN : 2249-0868
Foundation of Computer Science FCS, New York, USA
2nd National Conference on Innovative Paradigms in Engineering & Technology (NCIPET 2013) – www.ijais.org

Emotion Recognition through Speech

Akshay S. Utane
M.Tech Student
Dept. Of Electronics
and Telecommunications
Dr.Babasaheb Ambedkar
Technological
University,Lonere,India

S.L. Nalbalwar, Ph.D
Associate Professor
Dept. Of Electronics
and Telecommunications
Dr.Babasaheb Ambedkar
Technological
University,Lonere,India

ABSTRACT
In last two decades automatic emotion recognition based on speech become wide area of research for human-machine communication. Many systems has been implemented to recognize emotion in speech signals. In this paper previously implemented speech emotion recognition systems has been reviewed using various types of classifiers. The classifiers used to distinguish emotions such as neutral ,surprise ,anger ,happy, sad, fearful, disgust ,etc. emotional speech samples are used as database for emotion recognition from speech and extracted features from speech samples are prosodic and spectral features such as pitch, energy, formants, speech rate ,(MFCC) Mel frequency cepstrum coefficient and linear prediction cepstrum coefficient (LPCC).the performance of classifiers represented by extracted features. Advantages and performance of speech emotion recognition system using different types of classifiers are also discussed.

Keywords
Emotion recognition, Feature extraction, Feature Selection, spectral features , prosodic features, Classifier.

1. INTRODUCTION
SPEECH is one of the most fundamental and natural communication means of human beings in communication speech signal is one of the fastest methods of communications between humans. Therefore the speech can be more efficient and fast method of interaction between machine and human [1]. In advance human-machine interaction technology, a user-friendly interface is becoming more and more important for speech-oriented applications. Humans have the natural ability to use all their available senses for maximum awareness of the received message. Through all the available senses people easily sense emotional state of their communication partner. The emotion recognition is natural for humans but it is more difficult task for machine. Therefore the purpose of emotion recognition system is to improve machine and human communication using knowledge related emotion in such a way that human machine interface becomes more efficient. [2].

Speech emotion recognition has several applications in day-to-day life. Some of The application of the speech emotion recognition system includes:

1. in psychiatric diagnosis, lie detection
2. in Call center conversation may be used to analyze behavioral study of call attendants with the customers which helps to improve quality of service of a call attendant .
3. In aircraft cockpits, speech recognition systems trained to recognize stressed speech are used for better performance.
4. Emotion analysis of telephone conversation between criminals would help crime investigation department.
5. It is Useful for enhancing the naturalness in speech based human machine interaction
6. Interactive movie, storytelling & E-tutoring applications would be more practical, if they can adapt themselves to listeners or students emotional states.
7. Conversation with robotic pets and humanoid partners would be more realistic and enjoyable, if they are able to understand and express emotions like humans [1,3].

In speech emotion recognition, the emotions are found out from the speech of male or female speakers [1]. the past century some speech features were studied which involved fundamental frequencies, Spectral features have been widely used in speech emotion recognition , such as Mel frequency cepstrum coefficient (MFCC), linear prediction cepstrum coefficient (LPCC). another effective group of features are Prosodic features for describing emotional states such as energy , pitch , intensity , formants etc. [5]. Most traditional speech emotion systems have been focused on prosodic features or spectral features which form the basis for speech processing even today . In one of the research a correlation between spectral speech features and emotion were present. Further machine and humans emotion recognition rate was Compared, in which founded same recognition rates for both. In another research on real and acted emotional speech a spectrograms approach were studied and obtained similar recognition rate for both, which Recommended to use for the speech emotion recognition system. After this study on speech emotion recognition system using support vector machine was presented and achieved an accuracy of 73% for five different emotional states. In one of the research hidden markov model for emotion recognition from speech obtained 70% efficiency for seven different emotions [5,8]
In another research Gaussian mixture model studied The emotion recognition performance for certain set utterances has been observed to be 92.5%, 84.75% and 86% for male, female and combination of male & female speakers for four different emotional states respectively.[3] in recent research the Maximum Likelihood Parameter-based classifier obtained the best recognition performance on neutral emotion with an 72.40 percent efficiency. [7].The task Emotion recognition from the speaker‟s speech is very challenging for the following reasons: In this which particular features of speech are required to

differentiate between various emotions are not clear. Due to the existence of the different speaking styles, speakers, sentences ,languages, speaking rates introduces accosting variability , because of which speech features get directly affected. Each emotion may correspond to the different portions of the spoken utterance. The same utterance may show different emotions Therefore it is very difficult to differentiate these portions of utterances. Another problem is that Expression of emotion is depending on the speaker and their culture and environment. As there is change in culture and environment of different speakers their speaking style also gets changes. This is another challenge in front of the speech emotion recognition system. There may be two or more types of emotions transient and long term emotions so it is not clear the recognizer will detect which type of emotion [1].

Speech Emotion recognition are of two or more types may be the speaker independent or speaker dependent. The different classifiers available are Bayes classifier, Hidden Markov Model (HMM) and Support Vector Machine (SVM), Gaussian Mixtures Model (GMM), Artificial Neural Network (ANN), k-nearest neighbors (KNN). Maximum Likelihood bayesian classifier This paper gives reviews on mentioned classifiers [5, 6, 7].

2. SPEECH EMOTION RECOGNITION SYSTEM

Speech emotion recognition system is typical pattern recognition system. This shows that the stages involved in the pattern recognition system are also present in the Speech emotion recognition system. There are five main modules in speech emotion recognition system consist emotional speech input, feature extraction, feature selection, classification and recognized emotional output [2].

Speech emotion recognition system is as shown in Figure 1.

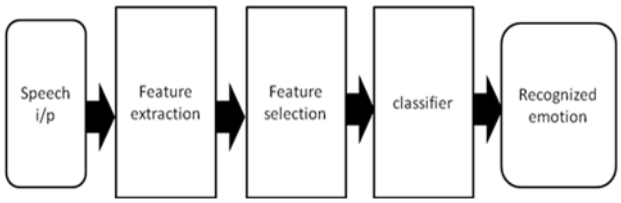

Fig 1 Structure of the Speech Emotion Recognition System.

There is requirement to find out a set of the significant emotions to be classified by an automatic emotion recognizer is a main concern in speech emotion recognition system. A typical set of emotions contains 300 emotional states. Therefore to classify such a great number of emotions is very complicated. According to "Palette theory" any emotion can be decomposed into primary emotions similar to the way that any color is a combination of some basic colors. Primary emotions are neutral, joy, anger, disgust, fear, sadness, stressed and surprise [1]

In speech emotion recognition system database is critically important to consider. The evaluation of emotion recognition system is depends on the naturalness and efficiency of the database which is used as an input to the speech emotion recognition system. If the improper database is given as an input to the system then incorrect conclusion may be drawn. The input as database feed to the speech emotion recognition system may contain the real world emotions or the acted ones. It is more practical to use database that is collected from the real life situations [1] most of the researcher used Berlin emotional speech database is a simulated speech database contains are totally about 500 acted emotional speech samples. Which are simulated by professional actors.[9]

3. FEATURE EXTRACTION

The important task to extract proper features for emotion recognition. emotion from the speech is represented by the large number of parameters which is contained in the speech due to change in these parameters will result in corresponding change in emotions. Therefore an extraction of these speech features which represents emotions is an important factor in speech emotion recognition system [12 ,14]. The features of speech can be categories as short term and long term features. The important issue in feature extraction is the region of analysis of the speech signal used which is to be considering in the feature extraction. The speech signal is non stationary hence divided into the small intervals which are called as frames [1].

The spectral and prosodic features are known as the primary indicator of the speakers emotional states. Research on emotion recognition showed important features are speech rate, energy, pitch, intensity, duration, formant, linear prediction cepstrum coefficient (LPCC) and Mel frequency cepstrum coefficient (MFCC) [8,9,14]. With the different emotional state, corresponding changes occurs in the pitch, energy, speak rate and spectrum. Typically anger has a higher mean value of energy, mean value and variance of pitch, having high frequency formants and there has improvement of spectrum in high frequency components as per the feature of fear has a high mean value and variation range of pitch ,better intensity and higher energy. On the other hand sadness has decreases in the mean value, variation range and variance of pitch, speak rate is slow, low intensity and spectrum of high frequency components decreases , also the energy is weak. Therefore to recognize emotions from speech statistics of energy, formants , pitch, and some spectrum feature can be extracted [8, 9 ,11,14].

The pitch signal is also called as the glottal wave form which is one more main feature which detected emotion in speech. The pitch signal produced from the vibration rate of the vocal cords , and it is depends on the tension of the vocal folds and sub glottal air pressure. Number of harmonics present in the spectrum is directly get affected by the pitch frequency. The pitch signal is characterize by the two features that is glottal air velocity at the vocal fold opening time instant and pitch frequency. Another main features of speech which showed emotion is energy and the study of energy is depends on short term average amplitude and short term energy [9]. As the arousal level of emotions affects the energy of speech signal therefore it can be used in the field of emotion recognition. [10]. Formants are the efficient and meaningful frequency components of human speech. it provides purely quantitative frequency content of the vowel and characteristic of speech signal. it is very advantageous to use of formant frequencies for emotion recognition. Mel frequency cepstrum coefficient (MFCC) is well known features and widely used in speech emotion recognition and speech recognition systems and the recognition rate of the using MFCC is very high. Mel frequency cepstrum is an illustration of short term power spectrum of sound [5,11,14]. In the low frequency region better frequency resolution and robustness to noise could be achieved with the help of MFCC rather than that for high frequency region. Linear prediction cepstrum coefficient (LPCC) the

most effective representation of speech signal provide the details information about the characteristics of particular channel of any individual person and due to accordance of different emotions this channel characteristic will get change , so emotions in speech can extract by using these features. The advantage of using the LPCC is that its algorithm is more efficient ,it involves less computation and it could describe the vowels in efficient manner[9].

In feature extraction all of the basic feature of speech extracted may not be essential and helpful for speech emotion recognition system. If all the extracted features gives as an input to the classifier this would not guarantee about accurate performance of system which signifies that there may needed the removal of such a unused features from the base features. Therefore there is a requirement of proper and sufficient feature selection to reduce these features. to select the significant feature subset Forward selection(FS) feature selection method could be used. In the first stage forward selection initializes with the single best feature from total feature set. And further The remaining features are added which increases the classification efficiency. the selection process should stop If the added number of features attained the preset number.[13,14]

4. CLASSIFIER

An important task in speech emotion recognition system is selection of classifier. after calculation of Speech features, the appropriate features are provided to the classifier. A classifier detects the emotion from speech utterances of different speaker's. To perform emotion recognition from speech Various types of classifier have been proposed. Hidden Markov Model (HMM), Bayes classifier , Support Vector Machine (SVM), Gaussian Mixtures Model (GMM) ,k-nearest neighbors (KNN), Artificial Neural Network (ANN) and. Maximum Likelihood Bayesian classifier etc. are the classifiers used in the speech emotion recognition system. Each classifier having some advantages and limitations over each others.

Only Gaussian Mixture Model is more efficient over global features are extracted from the training utterances are suitable for emotion recognition from speech. Its based on expectation-maximization algorithm or Maximum A Posteriori (MAP) Parameter Estimation. All the training and testing equations are based on the assumption that all vectors are independent therefore GMM cannot form temporal structure of the training data. GMM achieved maximum efficiency of 78.77% using the accurate features of speech signal . In speaker dependent system calculated 89.12% for recognition performance using GMM and obtained typical performance of 75% using speaker independent recognition system. [1,3].

Speech emotion recognition systems are similar to speech recognition, speech segmentation and isolated word recognition. hidden markov model is widely used classifier for speech application the main reason behind is its physical interconnection with the production mechanism of speech signals. HMM has achieved high accuracy for modeling acoustic and temporal information in the spectrum of speech In speech emotion recognition system. The HMM is doubly stochastic process
consist of first order markov chain whose states are buried from the observer[1]. generally for each emotion a single HMM is trained and an random sample is classified according to the model which illustrate for best feature sequence derived For speech emotion recognition [4]. HMM having advantage that the temporal dynamics features of speech can be taken as second accessibility procedure established for optimizing the recognition framework. The process for features selection occurred as main problem in building the HMM based recognition model. Because features carries information is not enough about the emotional states, but it must be significant for the HMM structure as well. In speech emotion recognition system HMM provides higher classification accuracy for emotion recognition as compared to other classifiers [8]. HMM classifiers using spectral features having higher recall rates than that for prosody and formant features using HMM classifier. [12]. The efficiency for speech emotion recognition by using HMM classifier for the speaker independent system is observed 64.77% and for the speaker dependent it was 76.12% [1]. In one of the research **B**ayes classifier is adapted with genetic algorithm and sequential floating feature selection that employed ability of accurate classification. It achieved probability of correct classification in first stage related to spectral and prosodic feature with efficiency at an average rate of 67% for surprise and happiness emotional utterances [15].

Another classifier used for emotion recognition is k-nearest neighbor classifier (k-NN).it is typical form of nearest neighbor technique based classifier for random samples. which appointed an utterance to an emotional conditions According to the k utterances for emotional states. The classifier can classify all the utterances in the design set properly, if "k" equals to 1, however there will be decrease in performance on the test set. K-NN classifier will Succeed in achieving classification rate of 64% for four emotional states by Utilizing the information of energy contours, pitch and formants etc [13]. the basic idea behind the support vector machine (SVM) classifier is basically Transforming the original set of feature to a higher dimensional feature space by using the kernel function , which required to get optimum classification in this new feature space. SVM classifier are generally used as important applications such as classification problems and pattern recognition The kernel functions such as linear, polynomial, radial basis function (RBF) can be used in SVM model for large extent. Hence SVM is gives better classification performance over the other classifiers and due to which it is used in the speech emotion recognition system[1, 5]. The support vectors is well known for measurement vectors which defines the boundaries of the margin. This margin is nothing but the width of the largest tube without any utterances, which can *get around* the decision boundary. emotional states can be separated to huge margin by using SVM classifier. An typical SVM classifier was implemented for two class problems, but it can be use for more classes. Because of the structural risk minimization oriented training SVM is having high generalization capability. SVM has the efficiency for the speaker dependent classification are above 80% and speaker independent classification are 75% respectively[1,10].Other classifier proposed for the classification of emotion is an artificial neural network (ANN), which is having ability to find nonlinear boundaries for separating the emotional states[1].in speech emotion recognition Multilayer perceptron layer neural networks are commonly used because it has well defined training algorithm as it is relatively easy to implement. Most frequently used feed forward neural network for purpose of speech emotion recognition. The classification rate achieved by ANN based classifiers for speaker dependent recognition with accuracy of 51.19% and for speaker independent recognition with 52.87% accuracy [10].

5. CONCLUSION

Automatic speech emotion recognition is emerging field are increasing now a day which it achieved results in the better human and machine interaction. in this depending on the various classifiers illustrated the Speech emotion recognition systems there are certain issues regarding speech emotion recognition system are the need of significant features extraction using signal processing unit from available speech signal, it is important to select appropriate feature selection. And another is a classifier model which used for emotions recognition from the speech signal. Most of the classifiers the average precision for speaker dependent recognition system is higher than speaker independent system. To enhance performance of the emotion recognition process multiple classifier or hybrid classifiers methods used to derive the classification. Also by extracting more effective and accurate speech features efficiency of the speech emotion recognition system can be enhanced.

.

6. REFERENCES

[1] M. E. Ayadi , M. S. Kamel , F. Karray, "Survey on Speech Emotion Recognition: Features, Classification Schemes, And Databases", Pattern Recognition 44, PP.572-587, 2011.

[2] I. Chiriacescu , "Automatic Emotion Analysis Based On Speech" , M.Sc. THESIS Delft University of Technology, 2009.

[3] Nitin Thapliyal , Gargi Amoli "Speech based Emotion Recognition with Gaussian Mixture Model" international Journal of Advanced Research in Computer Engineering & Technology Volume 1, Issue 5, July 2012

[4] T.Vogt, E. Andre and J. Wagner, "Automatic Recognition of Emotions from Speech: A review of the literature and Recommendations for practical realization", LNCS 4868, PP.75-91, 2008.

[5] S. Emerich, E. Lupu, A. Apatean, "Emotions Recognitions by Speech and Facial Expressions Analysis", 17th European Signal Processing Conference, 2009.

[6] Shashidhar G. Koolagudi, K. Sreenivasa Rao "Emotion recognition from speech: a review" Int Journal of Speech Technol (2012).

[7] Chung-Hsien Wu, and Wei-Bin Liang "Emotion Recognition of Affective Speech Based on Multiple Classifiers Using Acoustic-Prosodic Information and Semantic Labels " IEEE TRANSACTIONS ON AFFECTIVE COMPUTING, VOL. 2, NO. 1, JANUARY-MARCH 2011.

[8] A. Nogueiras, A. Moreno, A. Bonafonte, Jose B. Marino, "Speech Emotion Recognition Using Hidden Markov Model", Eurospeech, 2001.

[9] P.Shen, Z. Changjun, X. Chen, "Automatic Speech Emotion Recognition Using Support Vector Machine", International Conference On Electronic And Mechanical Engineering And Information Technology, 2011.

[10] D. Ververidis and C. Kotropoulos, "Emotional Speech Recognition: Resources, Features and Methods", Elsevier Speech communication, vol. 48, no. 9, pp. 1162-1181, September, 2006.

[11] Z. Ciota, "Feature Extraction of Spoken Dialogs for Emotion Detection", ICSP, 2006.

[12] E. Bozkurt, E, Erzin, C. E. Erdem, A. Tanju Erdem, "Formant Position Based Weighted Spectral Features for Emotion Recognition", Science Direct Speech Communication, 2011.

[13] C. M. Lee, S. S. Narayanan, "Towards detecting emotions in spoken dialogs", IEEE transactions on speech and audio processing, Vol. 13, No. 2, March 2005.

[14] Yu Zhou , Yanqing Sun , Jianping Zhang and Yonghong Yan ,"Speech emotion recognition using both spectral and prosodic features" ieee conference 19-20 dec 2009 published in information engineering & computer science.

[15] Mohammad H. sedaaghi , Constantine Kotropoulos and Dimitrios Ververidis " Using Adaptive Genetic Algorithms To Improve Speech Emotion Recognition " ieee conference 2007.