Gender Based Emotion Recognition using Speech Signals: A Review

Parvinder Kaur[1], Mandeep Kaur[2]
[1]*Department of Electronics and Communication Engineering, Punjabi University, Patiala, India*
[2]*Department of Electronics and Communication Engineering, Punjabi University, Patiala, India*
[1]parvinderkrm13@gmail.com
[2]ermandeep0@gmail.com

***Abstract*— Emotion recognition using speech signals has been a rising area in the recent years. The emotion recognition system discussed is gender based which is capable of recognizing six emotions (happiness, anger, surprise, disgust, sadness and fear) and neutral state. The above said system has two sub systems: 1) gender recognition 2) emotion recognition. If the gender of the speaker is known before finding the emotions of the speaker then it gives higher accuracy as mentioned in one of the papers. It also improves the human-computer interaction (HCI) which can be useful in giving feedback in real time applications. In this paper literature on emotion recognition through speech using different databases and different features is presented. Different models of classifiers are discussed here for their accuracy with respect to emotion recognition.**

***Keywords*— gender recognition, emotion recognition, pitch, Support Vector Machine,**

I. INTRODUCTION

Speech signal is the primary, natural and the fastest way of communication between human beings. Extracting emotions from speech is an efficient method for Human-Computer Intelligent Interaction (HCII) [1] and giving back appropriate feedback. Primary emotions depicted in the speech are: anger, feeling of repulsion, fear, happiness, sadness, astonished and neutral state. HCII is becoming in important in all circles of life viz. virtual reality, smart phones, smart offices and homes. The emotions can be recognized using either facial expressions or speech features. Speech signal is a non-stationary signal so the signal is divided into small fragments called frames each of length 20ms. Voice of human beings can be treated as stationary for interval of 20-40ms. Then features are extracted from each frame which can be local features or global features [1]

Emotion recognition through speech signals is done by first registering the audio signals and then classifying it on the basis of gender using pitch of the voice and then making decision on that speech emotions using classifier. Different classifiers used are Support Vector Machine (SVM), Gaussian Mixture Model (GMM), Hidden Markov Model (Hmm), Artificial Neural Network (Ann), K-Nearest Neighbour, Naive Bayes, Combined HMM-SVM, Hybrid Classifiers, Class Specific Multiple Classifiers.

II. LITERATURE SURVEY

A. *Features*

Different features of speech used for speech emotion recognition are:

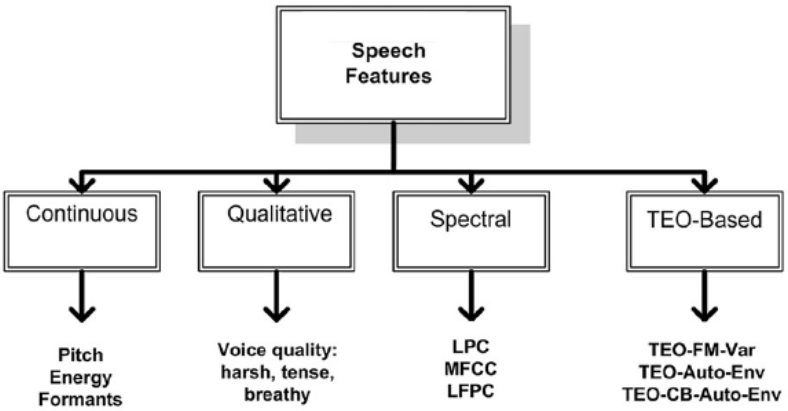

Fig. 1 Different Features of Speech

B. Databases

Different databases used are [1]:

1) *Reading-Leeds Database*: It includes natural speech rather than acted voices.
2) *Belfast Database*: It recognizes emotions from facial and voice features.
3) *CREST-ESP (Expressive Speech):* It has meaningful and natural database of speech
4) *Berlin Emotional Speech (BES*): It has a database of stimulated speeches
5) *LDC (Linguistic Data Consortium) speech emotion corpus*: It includes acted speeches.

C. Classifiers

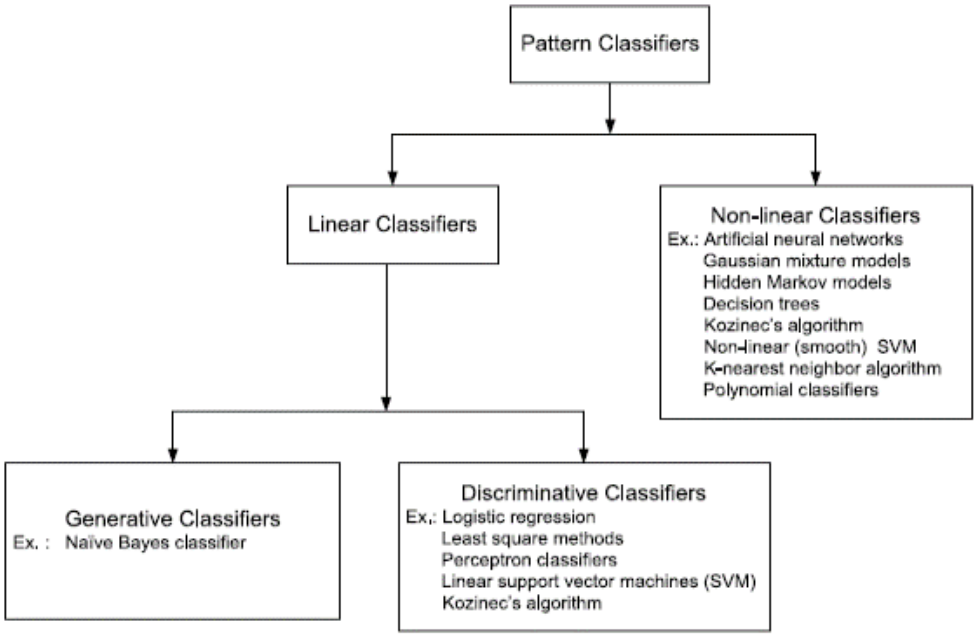

Fig.2 Different types of classifiers

Some of the classifiers are described below in context of emotion recognition:

1) *Gaussian Mixture Model [4*]: GMM-based speech emotion recognition has two phases. It is combination of emotion model training and emotion recognition phases. Firstly, the speech signal is passed through a pre-processing system that standardizes the amplitude, reduces the noise, and extracts features. A model that has the maximum log-likelihood is determined to be the result of recognition. The parameters of the Gaussian model (the mean and the variance of Gaussian distribution) are calculated using the algorithm called Expectation- Maximization (EM)

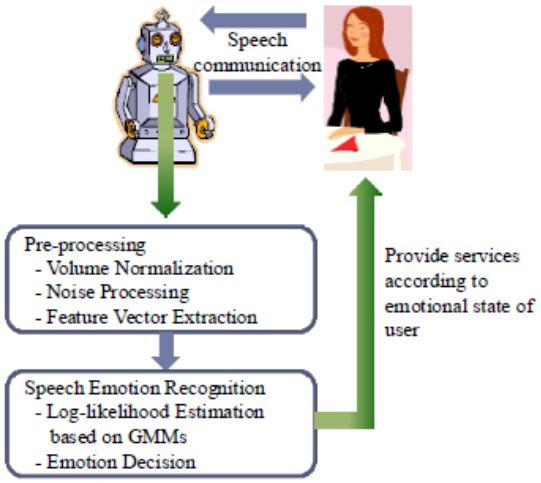

Fig 3 speech emotion recognition procedure using GMM [4]

2) *Class-Specific Multiple Classifiers [5]:* The classifiers can be combined in three ways: hierarchical, serial and parallel. In the first combination i.e. hierarchical, classifiers are connected in a tree fashion. At the end node only one class remains after decision. In the second combination of classifiers i.e. serial, classifiers are arranged in a line fashion and the number of classes reduces for the next classifier at each stage. In the third combination i.e. parallel, all the classifiers are independent of each other and the result of each of these is combined at last using some algorithm.

The classifiers for the given system are chosen based on different features and suitable classifier according to its performance for a group of emotions. The result of the different class specific emotion classifiers are combined by an algorithm to make a final decision on the detected emotions.

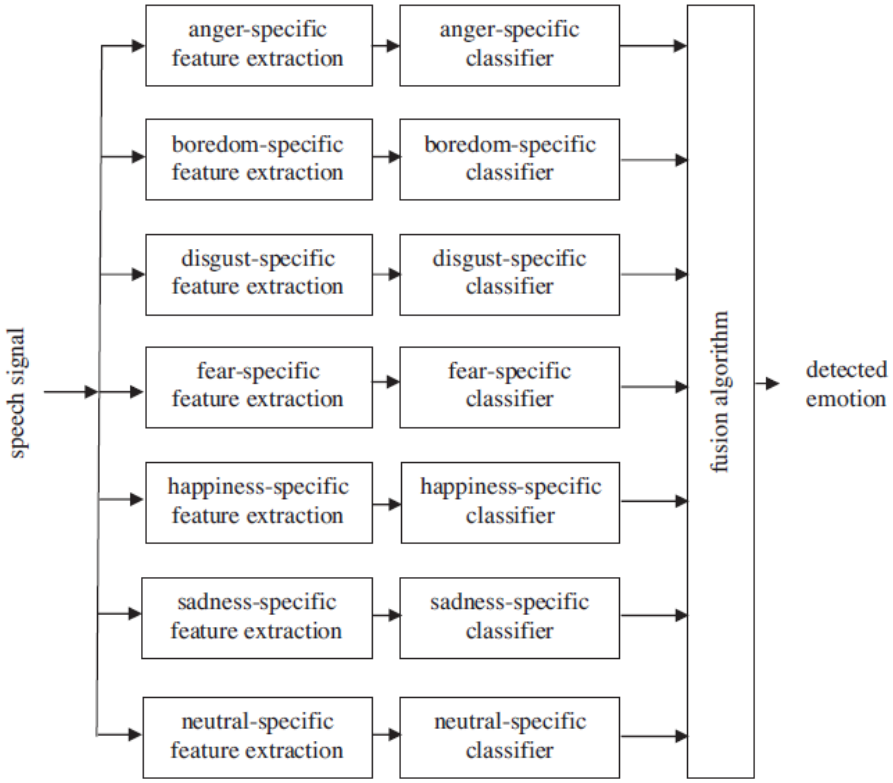

Fig 4 class-specific multiple classifiers (parallel) scheme for seven emotions [5]

3) *Support Vector Machine (Svm) Classifier [1]:* It works on the kernel functions which transform the original features into high dimensional area. Feature separation is obtained by placing the hyperplane precisely between the two different classes. The classifier has two phases first the training phase and the second is testing phase. The model is made to be based on each emotion and train it against all other emotions. Then the decision is made for in favor of that class to whichever the distance is the highest.

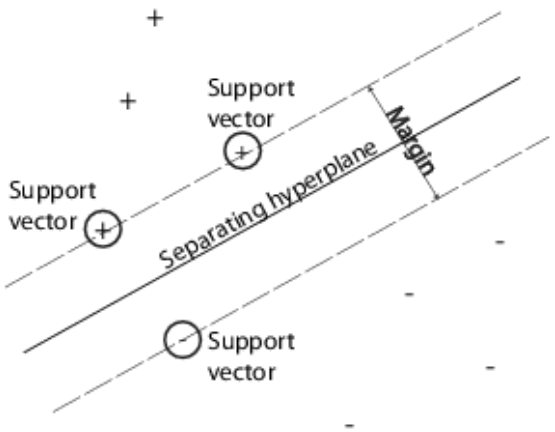

Fig 5 Separating hyper plane showing support vectors in classification

Four principal parts of recognizing emotions from speech are [1]:

1) *Feature extraction* : elaboration of speech signal to get features useful for emotion recognition

2) *Feature selection* : selecting more appropriate features to reduce computational load

3) *Classification*: assigning labels representing an emotion using the information provided by above three parts.

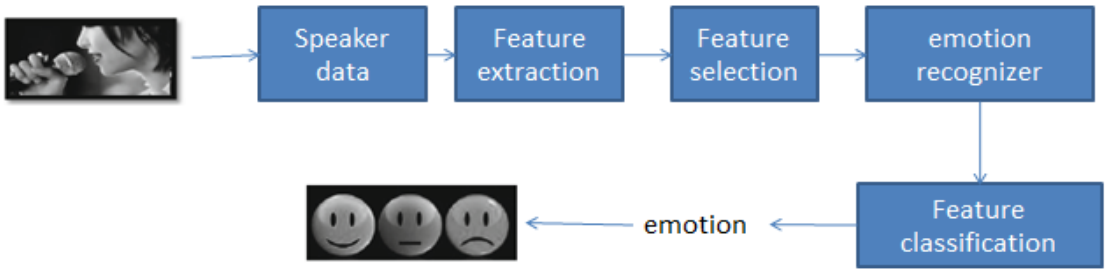

Fig.6 Block diagram of emotion recognizer.

It consists of two subsystem gender recognition based on pitch extraction and emotion recognition based on SVM (Support Vector Machine) classifier.

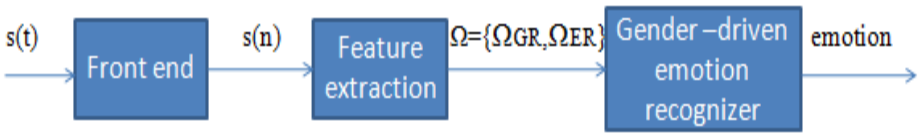

Fig 7 Block diagram of gender driven emotion recognizer

s(t) is continuous input audio signal . s(t) is input at the front end and sampling is made at frequency Fs=16 KHz to get discrete signal s(n)Features vector Ω are extracted from sampled sequence s(n) which contain both gender (ΩGR), and emotion features (ΩER). Output of gender recognition is used as input for emotion recognition. In the first step to emotion recognition gender recognition is done based on pitch frequency estimation. For the emotion recognition feature selection and reduction algorithm Principal Component Analysis (PCA) is used. It also employs two classifiers (two support vector machine), among which one is trained by the male speech signals and the other one is trained by female speech signals. This of the two classifier is to be chosen depend on gender information.

4) *Artificial Neural Network (Ann) [6]:* It has units called neurons and are arranged in layers which convert the input vector into output. Each neuron takes an input and an activation function (which is generally non-linear) is applied and the output is passed to the next layer. ANN are more advantageous than HMM and GMM because they can be used for non-linear functions also.
5) *Naive Bayes Classifier:* Naive Bayes method is a set of supervised learning algorithms based on Bayes' theorem which plays important role in probabilistic learning and classification. It uses the prior probability of an item if no description about that item is given. It also includes two phases, training and testing. Training is very fast and easy because it requires considering each attribute in each class whereas in Testing it is looking up tables and calculating the conditional probability.
6) *Hybrid Classifier*: It is the combination of different classifier such as HMM-SVM

D. Major Issues

Some of the major challenges are:

1) *Acoustic variability*: It is due to different speakers, different styles and rates of speaking because all of these decide speech energy and pitch of the voice [2]. More than one emotion can also be drawn in the same speech.
2) *Environment*: Emotions also depend on speaker's culture and environment. While recognizing emotions it is assumed that there is no cultural difference [2].
3) *Single classifier*: The work on emotion recognition is done based on single classifier model only [3]. Use of hybrid classifier may also improve recognition system.

4) *Acted voices*: Also most of the work is done on simulated or acted speech or audio voices stored in databases which has limited application to real time applications.

E. Application Areas

Application areas include Web movies and computer applications, car board system for driver assistance, Diagnostic tool, Automatic translation systems, Service robots, Forensic data analysis, Audio based surveillance system [2]

III. CONCLUSION

Speech emotion recognition is helpful in human machine interaction like service robots. Gender driven emotion recognition system recognizes the emotional state by first recognizing gender of the speaker and then recognizing emotions and this gives increased recognition accuracy. By reducing number of features using selection algorithm, number of operations required and the time required can be decreased. Different databases consisting of recorded utterances can be used to recognize emotions. Different classifiers such as SVM, GMM, Multiple classifiers are used to classify emotions using different features of speech signal. Accuracy of each classifier is different depending on the database and environment.

REFERENCES

[1] Igior Bisio, Alexander D., Fabio Lavagetto, Mario Marchese And Andrea S., "Gender driven emotion recognition through speech signals for ambient intelligence application", jan 2014
[2] Moataz El Ayadi , M. S. Kamel , F. Karray , "Survey on speech emotion recognition", vol. 44, 2011
[3] Shashidhar G. Koolagudi · K. Sreenivasa Rao , "Emotion recognition from speech" , 2012
[4] Jeong-sik P., Ji-hwan K., Y. hwan Oh, "Feature Vector Classification based Speech Emotion Recognition" , IEEE, Vol. 55, Aug 2009
[5] A. Milton , S. T. Selvi , "Class-specific multiple classifiers to recognize emotions from speech signals", vol 28, 2014
[6] Pravina L., "Hybrid classifiers for emotion recognition", IJSR, Jan 2015
[7] Puneet S., Mandeep K., "Study of the circuit partitioning using genetic algorithm", IJETAE, Vol. 3, Apr 2013
[8] Avninder K., Mandeep K., "Optimization of linear filters using particle swarm optimization (pso)", IJAREEIE, Vol. 5, Feb 2016
[9] Pankaj Gupta, Mandeep K., "Optimization of Fir Filters using Multiple Constant Multiplication (MCM) and common sub-expression elimination (CSE) Techniques", IJISET, Vol. 3, Aug 2016