ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

An ISO 3297: 2007 Certified Organization *Vol. 5, Special Issue 6, July 2016*

4th National Conference on Design Innovations for 3Cs "Compute-Communicate-Control"

Organized by

Dept. of ECE, MVJ College of Engineering, Bangalore-560067, India

Speech Based Emotion Recognition System

Akash S, Archana Krishnan, Abhijith M, Asst. Prof Ms. Shivalila

Dept. of ECE, MVJ College of Engineering, Bangalore, Karnataka, India

ABSTRACT: This paper presents a speech based emotion recognition system that can be installed on social robots. Three different classifiers namely: Support Vector Machine (SVM), Artificial Neural Network (ANN), and Bayes network were used to classify the different emotional states such as happy, sad, anger, surprise, and neutral. The classification was done using the features extracted from the speech sample. The various features extracted from the speech sample are pitch, intensity, Mel-frequency cepstral coefficients, harmonics to noise ratio(HNR), spectrogram and long term average spectrum(LTAS). Using these features the classification of different emotions are carried out.

KEYWORDS*:* Praat, Weka,SVM(Support Vector Machine), ANN(Artificial Neural Network).

I.INTRODUCTION

Recent years have been marked by the development of social robots, used for new educational technologies or for pure entertainment. The interactions with these machines are radically different from how humans interact with traditional computers. So far, humans have been learning very unnatural conventions and devices like keyboard and pop up windows to communicate to machines. In order to communicate in a natural way with humans, the robot should be able to recognize different expressions of emotions.Basically, Emotion recognition through speech can be explained as detection of emotion by feature extraction of voice conveyed by humans. Speech is a time varying signal, which represents the underlying patterns of emotions. There are various kinds of emotion present in a speech. Some of them are Anger, Happy, Sadness, Neutral, Fear and Surprise. Emotion detection is natural for humans but it is a very difficult task for a machine.A machine can detect who the speaker is and what the speaker says using speaker recognition and speech recognition techniques, but if weapplied emotion recognition we can get to know how the speaker is talking.Emotion plays an important role in the rational actions of humans. Therefore it is desirable for an intelligent machine-human interface for better human machine communication and decision making.

Speech emotion recognition has various applications in day-to-day life. Some of them are: Call center conversation, emotion analysis of telephonic conversation between two criminals, in aircraft cockpits, lie detection, psychiatric diagnosis, in enhancement in speech and speaker recognition, intelligent toys.This paper is based on the works presented in [1], which usestwo external tools – praat and weka. In this paper, the different emotional states are classified using three different classifiers Support Vector machine (SVM) [2], Artificial Neural Networks (ANN) [3] and Bayesian network. The pitch, intensity, Mel-frequency cepstral coefficients (MFCC), energy related features, formantsare some of the features used in the emotion recognition system.The rest of the paper is organized as follows;section two describes the emotion recognition system. In section three the details of the features extracted is discussed, finally, section four describes the conclusion.

II. SPEECH EMOTION RECOGNITION SYSTEM

During the past few years several techniques were used for emotion recognition. There are several types of classifiers which can be used for emotion recognition such as Hidden Markov Model (HMM) [4], Support Vector Machine (SVM), Artificial Neural Network (ANN), Gaussian Mixture Model (GMM) [5], k-nearest neighbour (KNN), decision trees, LDA. However, there are at least two different approaches: one is estimating the short-time signal parameters and modelling their changes, the other is to extract global features of the signal and applying various types of classifiers. Here, the second approach is chosen and each speech signal is analysed as a whole from which the global features are extracted and then used for classification.

Here, we use two external tools- the programs Praat [6] and Weka [7]. Praat is a free (Open source) program for phonetic analysis of speech. It can compute several parameters of speech: pitch, formants, spectrum, Mel frequency cepstral coefficients and many others. Weka 3 (Waikato Environment for Knowledge Analysis) is a suite of machine learning software written in java. It is freely available under GNU General Public License. It contains a collection of

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

An ISO 3297: 2007 Certified Organization Vol. 5, Special Issue 6, July 2016

4th National Conference on Design Innovations for 3Cs "Compute-Communicate-Control"

Organized by

Dept. of ECE, MVJ College of Engineering, Bangalore-560067, India

visualization tools and algorithms for classification, feature selection, regression modelling and filtration. The basic structure of the system is as illustrated in the figure 1. Basically there are two phases in the system's operation: learning phase and the evaluation phase.In the learning phase, feature selection and the training of classifier is carried out. Here, in this phase, various speech signals of different emotions are fed on to the classifier for training. In the evaluation phase, the trained system is used for emotion recognition.

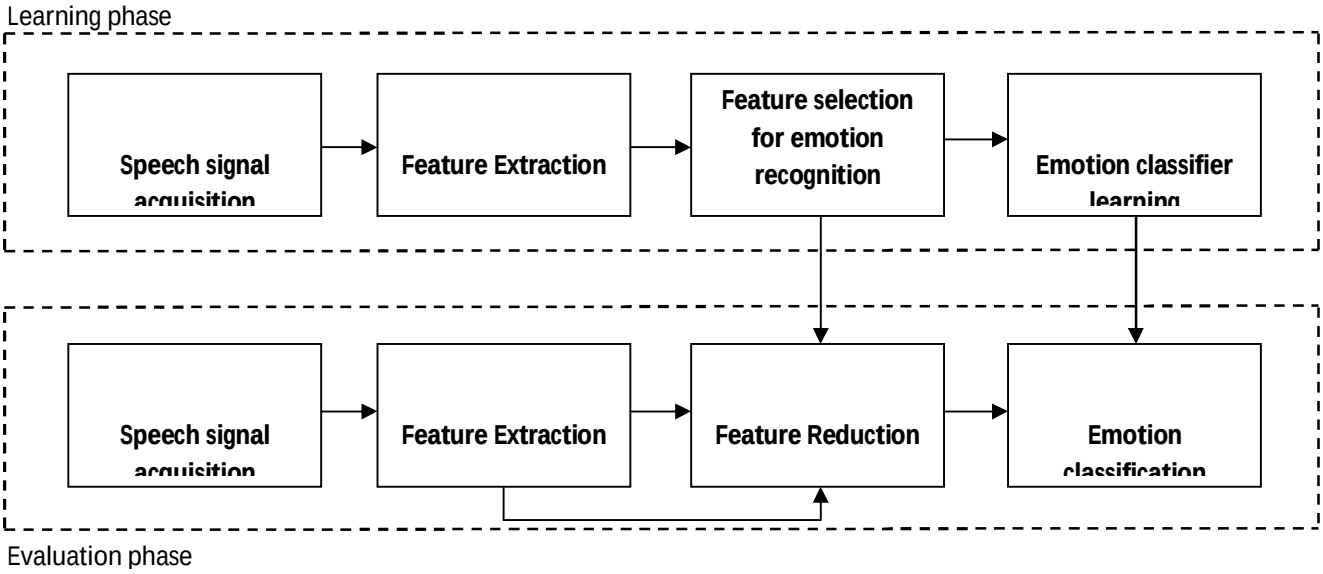

Figure 1: block diagram of speech emotion recognition system.

Emotion recognition from speaker's speech is a difficult task because of the following reasons: because of the existence of different sentences, speakers, speaking style, speaking rates variability is introduced, because of which speech features get affected directly. The same utterance may show different emotions. Another problem is that emotion expression depends on the speaker, his or her culture and environment. As the culture and environment gets changed, the speaking style also changes, which is another challenge.

One can also use a high pass filter [3] before feature extraction process to increase the accuracy and reduce noise in the speech sample. All the features obtained in the feature extraction process are not used for classification, instead some of the features are selected and based on those features classifier learning and classification is done. This is illustrated in the block diagram as feature selection for emotion recognition and feature reduction.

III.FEATURES EXTRACTED AND CLASSIFIERS USED

Various parameters of the speech signal can be computed by the program Praat. Some of the parameters that are computed and used in this system for emotion recognition are:

a) Pitch: is a fundamental frequency of the speech. It is produced by vocal cords. Various algorithms can be used to estimate pitch. Since the pitch exist only the voiced parts of the signal, the resulting waveform would be discontinuous, but smoothing and interpolation functions can be used to overcome this drawback.

b) Spectrogram: here the speech signal is split into 16ms frame with a 10ms step. Each frame is then fourier transformed to compute the magnitude of the frequency spectrum. The phase is neglected. Then logarithm is taken from the magnitude of the spectrum, and the result is appended to a matrix and this resulting matrix is called spectrogram.

c) Intensity: it is the instantaneous sound pressure value in dB.

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

An ISO 3297: 2007 Certified Organization Vol. 5, Special Issue 6, July 2016

4th National Conference on Design Innovations for 3Cs "Compute-Communicate-Control"

Organized by

Dept. of ECE, MVJ College of Engineering, Bangalore-560067, India

d) Mel-frequency cepstral coefficients (MFCC): they are used for parameterisation of the speech.
e) Harmonics to noise ratio (HNR): is the energy of the harmonics parts of the signal related to the energy of noise parts and is expressed in dB.
f) Long-term average spectrum (LTAS): is the averaged logarithmic power spectral density of the voiced parts of the signal with an influence of the pitch corrected away.

Additional vectors are also derived for more information. After generating the various features, the numbers of features are then reduced using weka's ATTRIBUTESELECTION function in the feature selection process.

The various classifiers used here are:

1) Support Vector Machine (SVM): Since it was first proposed, SVM has attracted a high degree of interest in the machine learning research community. SVM is used for pattern recognition and for classification of patterns and it is efficient and simplecomputation for machine learning algorithm.

SVM are a set of related supervisor learning methods for classification and regression. Given a set of training examples, each marked as belonging to one of the two categories, an SVM training algorithm builds a model that predicts whether a new example falls into one category or the other.

2) Artificial Neural Networks (ANN): Artificial representation of human brain that tries to stimulate the learning process of human brain is known as neural networks. It is inspired by biological systems.

ANN's are used to model complex relationships between inputs and outputs and to find patterns in data. ANN is made of interconnected artificial neurons. It learns by example as humans. It is configured through a learning process for a specific application like pattern recognition or data classification. In biological systems, for learning, connections are adjusted that are between neurons. It is true for ANN as well.Neural networks are based on biological brains parallel architecture. Large number of interconnected neurons works in union to solve problems.

3) Bayes Network: A Bayesian network is a probabilistic graphical model that represents a set of random variables and their conditional independencies via a directed acyclic graph.For example, a Bayesian network could represent the probabilistic relationships between diseases and symptoms. Given the symptoms the network can be used to compute the probabilities of presence of various diseases.

All these classifiers were borrowed from weka. Using these classifiers various emotions are recognised. An example of feature extraction of a speech signal using praat is shown in figure 2.

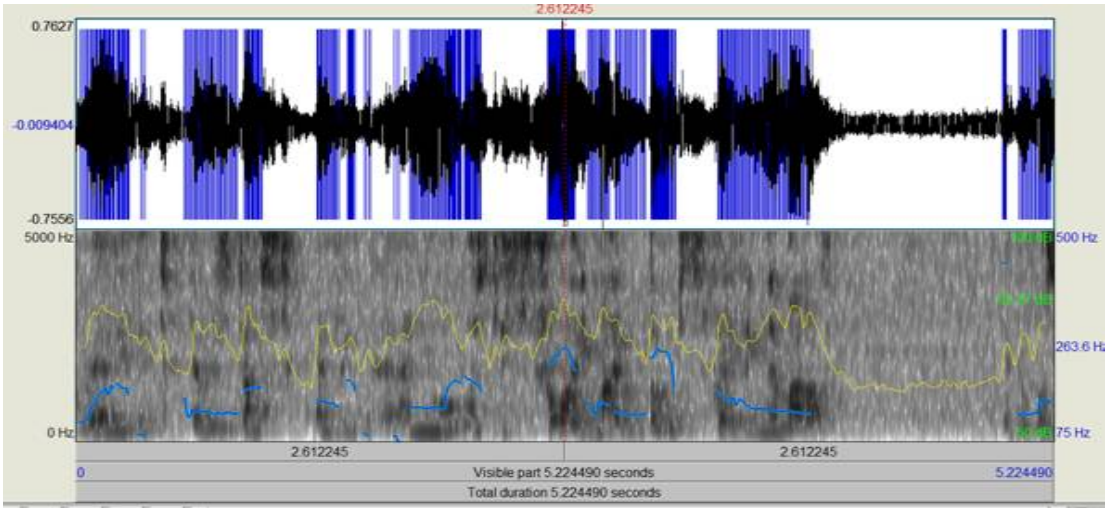

figure 2, feature extraction of an example speech signal using praat.

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

An ISO 3297: 2007 Certified Organization *Vol. 5, Special Issue 6, July 2016*

4th National Conference on Design Innovations for 3Cs "Compute-Communicate-Control"

Organized by

Dept. of ECE, MVJ College of Engineering, Bangalore-560067, India

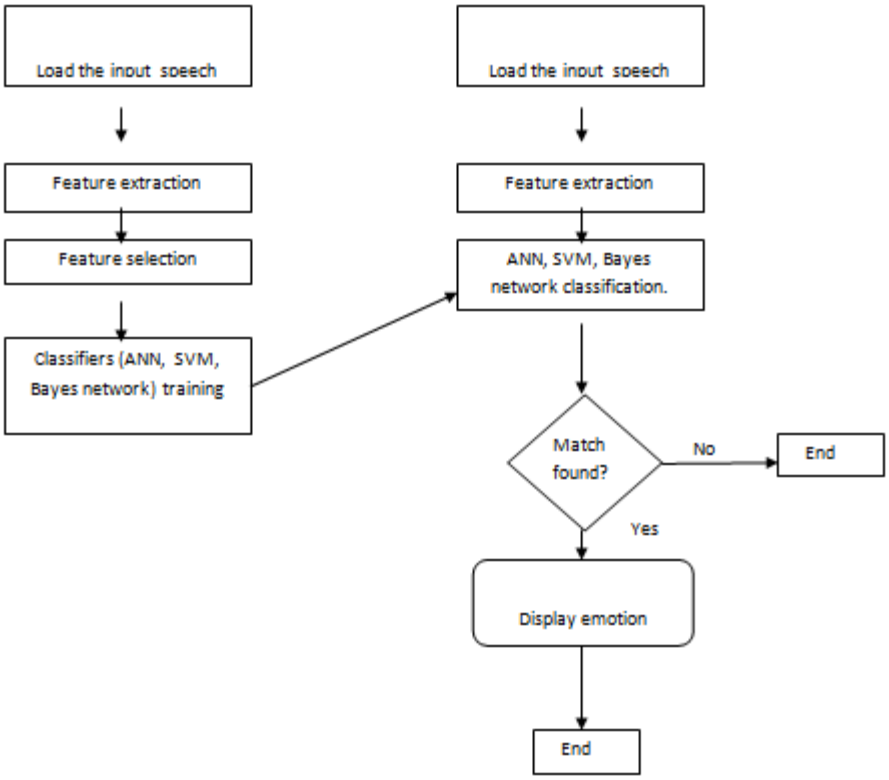

Figure 4, flowchart of the emotion recognition system

IV.CONCLUSION

This paper presents a speech emotion recognition system which can be installed onto a social robot. For the evaluation of the system, three classifiers are used: Bayes network, Support Vector Machine (SVM), Artificial Neural Networks (ANN). Six different features are extracted in the feature extraction process which is loaded on to the classifiers for learning and classification. The expected results are increase in accuracy of emotion detection than in the previous methods and classifiers. Further increase in accuracy may be brought by including filters before the feature extraction process to remove noise. Further enhancements can be brought by the inclusion of speaker and gender based emotion recognition.

REFERENCES

[1] Lukasz Juszkiewicz, Wroclaw University of Technology, "Improving Speech Emotion Recognition System for a Social Robot with Speaker Recognition", 2014 IEEE.
[2]Vaishali M. Chavan and V.V. Gohokar, "Speech Emotion Recognition by using SVM-Classifier" International Journal of Engineering and Advanced Technology (IJEAT) Volume-1, Issue-5, June 2012.
[3]Arti Rawat and Pawan Kumar Mishra, "Emotion Recognition through Speech Using Neural Networks" International Journal of Advanced Research in Computer Science and Software Engineering (IJARCSSE) Volume 5, Issue 5, May 2015.
[4]Prasad Reddy P.V.G.D, Prasad A, Srinivas Y, Brahmaiah P," Gender Based Emotion Recognition
System for Telugu Rural Dialects Using Hidden Markov Models" Journal of Computing, Volume 2, Issue 6, June 2010.
[5]Nitin Thapliyal and Gargi Amoli, "Speech based Emotion Recognition with Gaussian Mixture Model" International Journal of Advanced Research in Computer Engineering and Technology Volume 1, Issue 5, July 2012.
[6] A brief introduction by Pranav Jawale, "Speech Analysis using Praat".
[7]Zdravko Markov and Ingrid Russell, "An Introduction to the WEKA Data Mining System".