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, Issue 7, July 2016**

# Analysis of Affective Speech Signals for Emotion Extraction and Attitude Prediction

Punam M. Vitalkar[1], Prof. P.N. Bendre[2], Dr.S.M.Gulhane[3]

PG Student, Dept. of Digital Electronics Engineering, DBNCOET, Yavatmal , Maharashtra, India[1]

Professor, Dept. of Digital Electronics Engineering, DBNCOET, Yavatmal , Maharashtra, India[2]

HOD, Dept. of Electronics & Telecommunication Engineering, JDIET, Yavatmal , Maharashtra, India[3]

**ABSTRACT:** This paper deals with the development of a system for extraction of sentiments and attitude prediction from affective speech signals. As we know that, sentiment analysis for attitude prediction is one of most emerging and globally accepted technique used in business intelligence. In many business intelligence applications, the huge amount of telephone quality speech samples or telephone calls recorded by Business Process Organizations (BPOs) are processed for emotion extraction and attitude prediction. Through emotion extraction and attitude prediction, customer opinions, various trends are mined, which helps in decision making for corporate industries. In the proposed system, there is a mechanism provided for analysis of speech and extraction of various features viz. pitch, formants, short-time energy, Mel Frequency Cepstral Coefficients (MFCC), Spectral Flux and Zero Crossing Rate etc. By using these features and Support Vector Machine (SVM) as a classifier, the system tries to extract the emotion and predicts the attitude of the person whose voice is being analyzed.

**KEYWORDS:** Emotion extraction, sentiment analysis, speech signal, business intelligence.

## I. INTRODUCTION

In the modern age, the interest for emotional speech recognition has grown considerably during the past ten years [1]. Emotion processing is one of the main concept in NLP. Emotion can be easily extracted from facial expressions, from text data available. But extracting emotions from sound or speech has gaining major attention now a days. It is well known that emotional conditions such as anger, sadness and delight can have effect on speech sound. This effect can be observed mainly in the suprasegmental features, such as F0, intensity and temporal characteristics of speech. Since muscle tension may be raised in some emotional conditions, there is a possibility that some segmental features are also influenced by the speaker's emotional conditions [2].

Speech is a complex signal which contains information about the message, speaker, language and emotions. Emotion on other side is an individual mental state that arises spontaneously rather than through conscious effort. The database for the speech emotion recognition system is the emotional speech samples. Features for emotion recognition are extracted from these speech samples [4].

There are many ways of communication but the speech signal is one of the fastest and most natural methods of communications between humans. Therefore the speech can be the fast and efficient method of interaction between human and machine also Emotion recognition from the speaker's speech is very difficult because of the following reasons: In differentiating between various emotions which particular speech features are more useful is not clear. Because of the existence of the different sentences, speakers, speaking styles, speaking rates accosting variability was introduced, because of which speech features get directly affected. The same utterance may show different emotions. Each emotion may correspond to the different portions of the spoken utterance. Therefore it is very difficult to differentiate these portions of utterance. Another problem is that emotion expression is depending on the speaker and his or her culture and environment. As the culture and environment gets change the speaking style also gets change, which is another challenge in front of the speech emotion recognition system. There may be two or more types of emotions, long term emotion and transient one, so it is not clear which type of emotion the recognizer will detect [5].

Emotion recognition from the speech information may be the speaker dependent or speaker independent. The different classifiers available are k-nearest neighbors (KNN), Hidden Markov Model (HMM) and Support Vector Machine (SVM), Artificial Neural Network (ANN), Gaussian Mixtures Model (GMM). The application of the speech emotion recognition system include the psychiatric diagnosis, intelligent toys, lie detection, in the call centre conversations

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, Issue 7, July 2016**

which is the most important application for the automated recognition of emotions from the speech, in car board system where information of the mental state of the driver may provide to the system to start his/her safety[6].

In recent years the workings which requires human-machine interaction such as speech recognition, emotion recognition from speech recognition is increasing. Not only the speech recognition also the features during the conversation is studied like melody, emotion, pitch, emphasis. It has been proven with the research that it can be reached meaningful results using prosodic features of speech [2].

Sentiment Analysis is one of most emerging and globally accepted technique used in business intelligence. In many business intelligence applications, the huge amount of telephone quality speech samples or telephone calls recorded by Business Process Organizations (BPOs) are processed for emotion extraction and sentiment analysis. Through emotion extraction and sentiment analysis, customer opinions, various trends are mined, which helps in decision making for corporate industries.

## II. RELATED WORK

### *A. Literature Survey*

X. M. Cheng et al [7], analyze the feather of the time, amplitude, pitch and formant construction involved such four emotions as happiness, anger, surprise and sorrow in their paper. Through comparison with non-emotional quiet speech signal, they sum up the distribution law of emotional feather including different emotional speech. Nine emotional features were extracted from emotional speech for recognizing emotion. They introduce two emotional recognition methods based on principal component analysis and the results show that the method can provide an effective solution to emotional recognition.

D. Ververidis et al [8], introduce a more fine-grained yet robust set of spectral features: statistics of Mel-Frequency Cepstral Coefficients computed over three phoneme type classes of interest –stressed vowels, unstressed vowels and consonants in the utterance. They investigate performance of their features in the task of speaker-independent emotion recognition using two publicly available datasets. Their experimental results clearly indicate that indeed both the richer set of spectral features and the differentiation between phoneme type classes are beneficial for the task. Classification accuracies are consistently higher for their features compared to prosodic or utterance-level spectral features. They show that, while there is no significant dependence for utterance-level prosodic features, accuracy of emotion recognition using class- level spectral features increases with the utterance length.

Nwe et al [10], a text independent method of emotion classification of speech is proposed in their paper. The proposed method makes use of short time log frequency power coefficients (LFPC) to represent the speech signals and a discrete hidden Markov model (HMM) as the classifier. The emotions are classified into six categories. Performance of the LFPC feature parameters is compared with that of the linear prediction Cepstral coefficients (LPCC) and mel-frequency Cepstral coefficients (MFCC) feature parameters commonly used in speech recognition systems. Results show that the proposed system yields an average accuracy of 78%and the best accuracy of 96% in the classification of six emotions. Results also reveal that LFPC is a better choice as feature parameters for emotion classification than the traditional feature parameters.

A.R. Panat et al [11] presented research on affective speech analysis. The authors have carried out rigorous experimentation and statistical analysis to asses and analyze the perceptual parameters useful for emotion recognition and speaker verification. The continuing research addresses to the six emotions viz.: happiness, anger, sadness, surprise, disgust, and fear. The authors have successfully recognized nine emotions viz. Dreadful (El), Cheerful (E2), Calm (E3), Amorous (E4), Might (E5), Disgusting (E6), Fear (E7), Pathos (E8), and Surprise (E9). Each speaker has displayed special distinguishable features in expressing various emotions. These speaker based variations in features have been used for developing robust and precise automatic speaker verification system. The results were confirmed using statistical analysis based on human recognizers' feedback and automatic speaker verification systems performance analysis. Authors have designed and modeled adaptive neuro-fuzzy Information system based automatic agent for recognizing emotion and verifying speaker. Experimentation was done using Indian regional language 'Marathi' and 'English'.

### *B. **Feature Extraction***

In emotion extraction process, it is necessary to extract speech spectral features from the speech samples. Some of tge important features are

- Pitch (F0)
- Formant Frequencies (F1, F2,F3)

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, Issue 7, July 2016**

- Short Time Energy
- Mel Frequency Cepstral Coefficients (MFCC)
- Spectral Flux
- Zero Crossing Rate (ZCR)

A pitch is the Fundamental Frequency (F0) of the quasi-periodic speech signals. There are a number of algorithms to detect the Pitch of the quasi-periodic speech signals. Pitch is determined based on the cepstrum method. Basically the concept behind this is that if we consider that the log amplitude spectrum contains many regularly spaced harmonics, then the Fourier analysis of its spectrum will show a peak corresponding to the spacing between the harmonics: i.e. the fundamental frequency.

In speech signals, the Formant frequencies typically depict the resonance of the vocal tract. The Formant frequencies (F1,F2,F3) are found by using the poles of the AR model of the Vocal Tract.

Cepstrum is a spectrum of spectrum. Mel Frequency Cepstral coefficients are computed by taking Discrete Cosine Transform (DCT) of log energy. Multiple filter banks are used to compute MFCC.

The Short Time Energy is used find the energy of a signal in predefined frame or window size.

The Zero Crossing Rate (ZCR) is a measure of the count that how many times the signal crosses the axis at zero point.

## III. PROPOSED WORK

The sentiment analysis using emotion extraction is proposed in this paper. The given speech signal is pre-processed by using applying appropriate filter to remove noise from it. After preprocessing end points are detection algorithm is applied. Feature extraction process comprises of extracting various features viz. Pitch (F0) which is also called as fundamental frequency, Formant frequencies (F1,F2,F3), Short Time Energy, MFCC, Spectral Flux and Zero Crossing Rate (ZCR) etc. The feature vector of all these values is formed. The feature vector is then given as input to the emotion classifier. In our proposed system we have used Support Vector Machine (SVM) classifier for classification. The model is shown below:

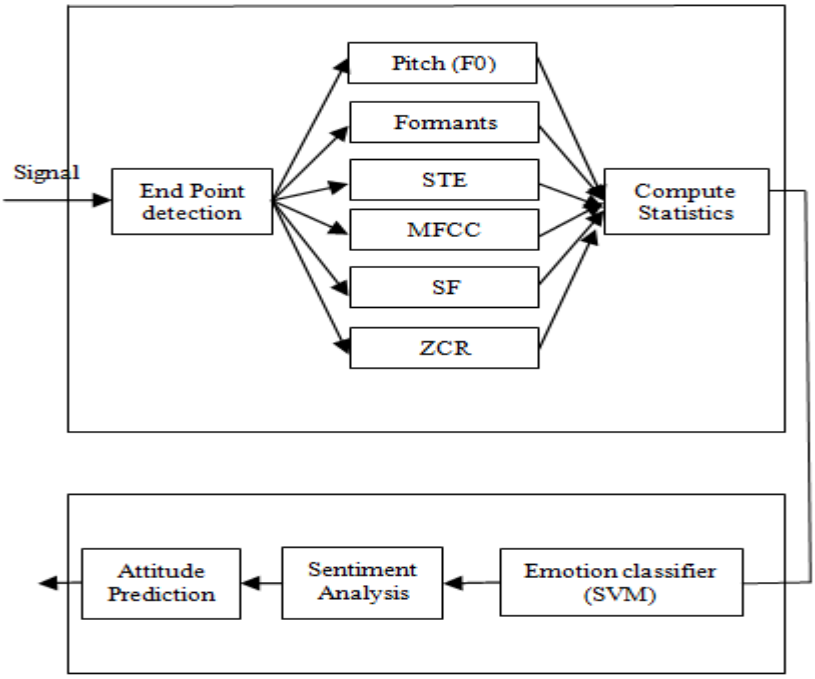


Figure 1 : Emotion recognition system

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, Issue 7, July 2016**

In the proposed system the speech signal is to be first loaded, then the signal is utilized for analysis and sentiment extraction. The chosen signal can be displayed in the form of time-amplitude waveform as shown in Figure 2.

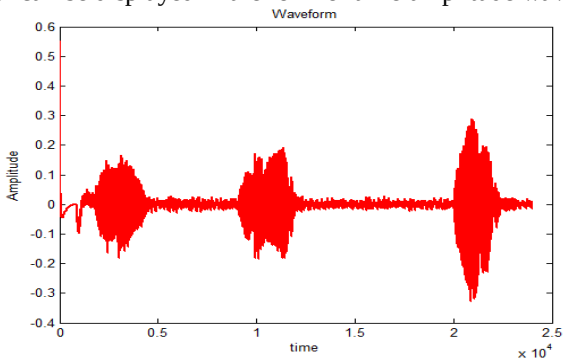


Figure 2 : Time-Amplitude waveform

*1) Pitch (F0)*

Pitch is one of most fundamental parameter or feature of any speech signal. It is also known as fundamental frequency. We have determined the fundamental frequency by auto-correlation.

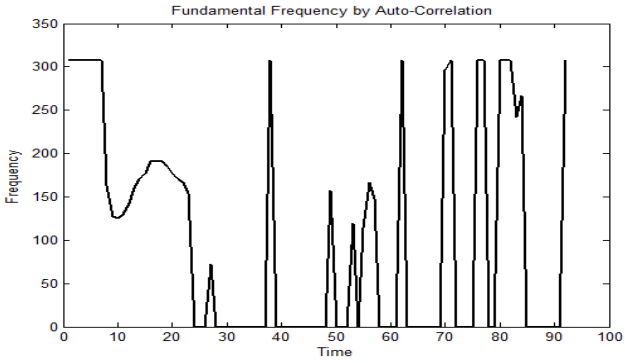


Figure 3 : Pitch (F0)

*2) Short –Time Energy*

It is used to predict the degree of energy in the human speech. It is calculated by using FFT algorithm

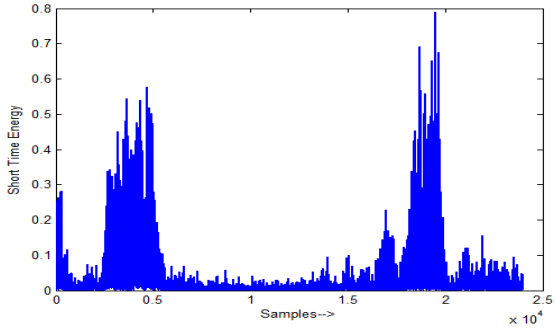


Figure 4 : Short_Time Energy

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, Issue 7, July 2016**

## IV.IMPLEMENTATION AND EXPERIMENTAL RESULTS

We have conducted rigorous experimentation on the speech corpus. Various speech features are extracted. The spectrogram of different speech signals are examined. Following are some of the spectrograms of the speech samples containing various emotions viz. *amorous, surprise, cheeeful, pathos, might, dreadful ,fear, disgusting, calm* etc.

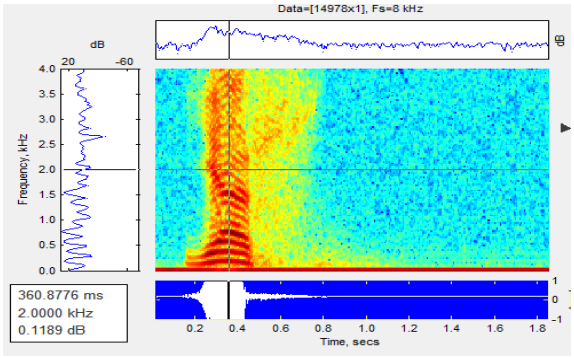


Figure 5 : Spectrogram of “amorous” emotion

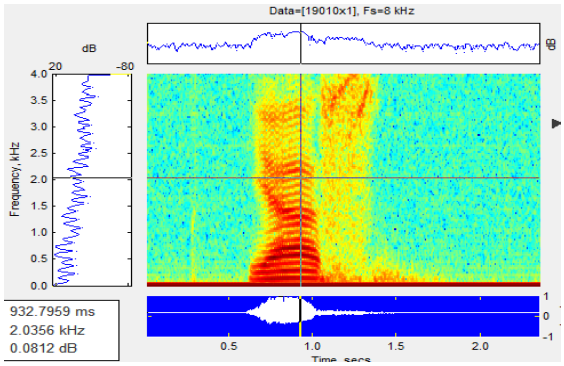


Figure 6 : Spectrogram of “surprise” emotion

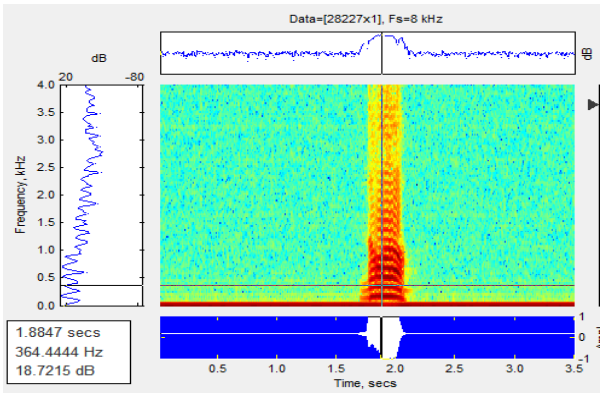


Figure 7 : Spectrogram of “Cheerful” emotion

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, Issue 7, July 2016**

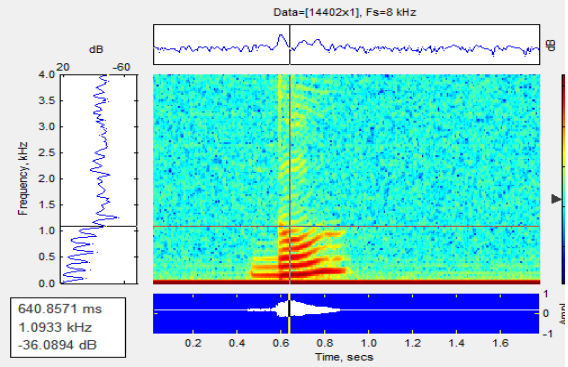


Figure 8 : Spectrogram for “pathos” emotion

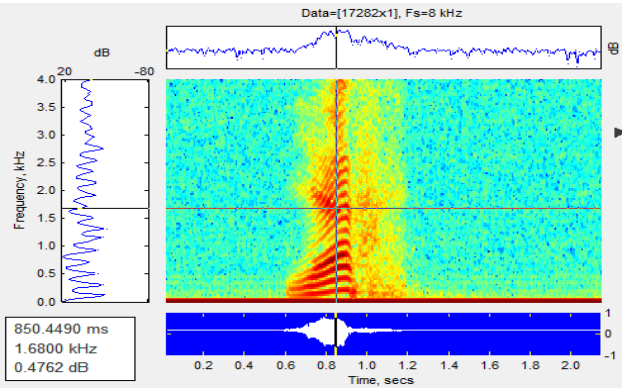


Figure 9 : Spectrogram of “might” emotion

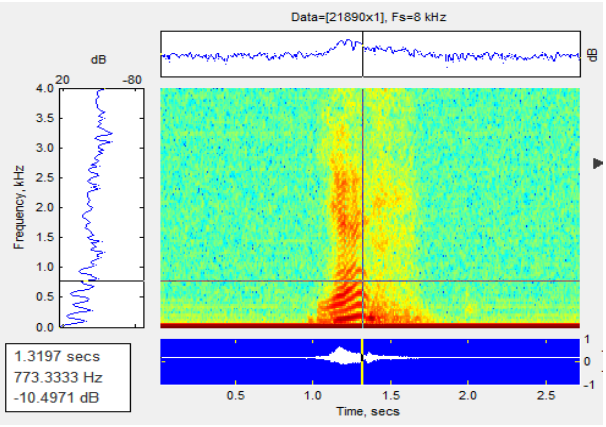


Figure 10 : Spectrogram of the “dreadful” emotion

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, Issue 7, July 2016**

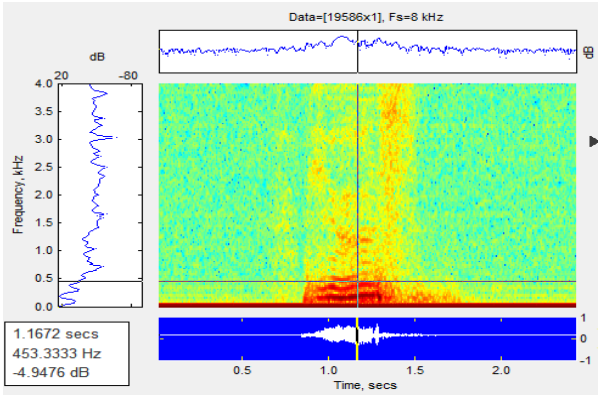


Figure 11 : Spectrogram for the “fear” emotion

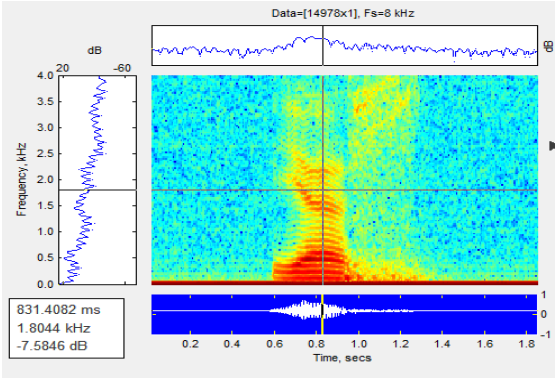


Figure 12 : Spectrogram for the “disgusting” emotion

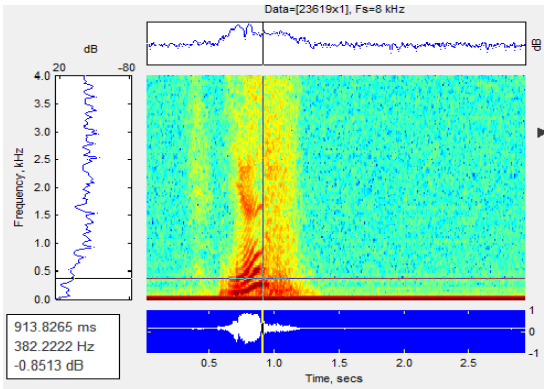


Figure 13 : Spectrogram for the “calm” emotion

The extracted features are combined to form feature vector. In Table I, features of speech samples containing various emotions are displayed.

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, Issue 7, July 2016**

Table I : Speech Features

| Emotion | F0 (kHz) | Short_time_energy (dB) |
|---|---|---|
| Amorous | 2.000 | 0.1189 |
| Surprise | 2.035 | 0.0812 |
| Cheerful | 0.364 | 18.72 |
| Pathos | 0.293 | -36.08 |
| Might | 1.680 | 0.4762 |
| Dreadful | 0.777 | -10.497 |
| Fear | 0.453 | -4.94 |
| Disgusting | 1.804 | -7.58 |
| Calm | 0.388 | -0.8573 |

These values are used for classification of speech samples using Support Vector Machines (SVM) technique, which is discussed below.

*3) Support Vector Machines (SVM)*

Support Vector Machine is a natural extension of LDCs which provides good generalization properties even for a large feature vector. *Support vector machine (SVM) is an effective approach for pattern recognition. Here, the basics of the SVM briefly will be presented. These basics of the SVM can be found in references Cristianini and Shawe-Taylor (2000), Fernandez Pierna, Baeten, Michotte Renier, Cogdill, and Dardenne (2004), Huang and Wang (2006), Kecman (2001) and Scholkopf and Smola (2000), Burges (1998). In SVM approach, the main aim of an SVM classifier is obtaining a function f(x), which determines the decision boundary or hyperplane (Fernandez Pierna et al., 2004). This hyper-plane optimally separates two classes of input data points. This hyperplane is shown in Fig. 1.Where M is margin, which is the distance from the hyperplane to the closest point for both classes of data points (Fernandez Pierna et al., 2004; Gunn et al., 1998).*

In SVM, the data points can be separated two types: linearly separable and non-linearly separable (Fernandez Pier-na et al., 2004). For a linearly separable data points, a training set of instance label pairs ($x_k$,$y_k$), where $k$=1,2,3,……..,$t$ $xk \in R^n$, and $y_k \in \{+1, -1\}$, the data points can be classified as $\langle w.x_k \rangle + b_0 \geq 1 \quad \forall y_k = 1$

$$\langle w.x_k \rangle + b_0 \leq -1 \quad \forall y_k = -1 \qquad (1)$$

Where $\langle w.x_k \rangle$ shows the inner product of $w$ and $x_k$. The inequality can be combined as,

$$y_k \left[ \langle w.x_k \rangle + b \right] - 1 \geq 0, \qquad \forall k = 1, ....., t \qquad (2)$$

The SVM classifier places the boundary by using maximal margin among all possible hyper planes.

Next step to emotion classification using SVM is the sentiment analysis. In sentiment analysis, most dominating emotions are mined and depending the pattern, sentiment can be predicted. e.g. if for a set of samples of some person, dominating emotions are *disgusting, dreadful* then the person could be “negative” thinker. If dominating emotions are *cheerful*, *amorous* then person is “positive” etc.

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, Issue 7, July 2016**

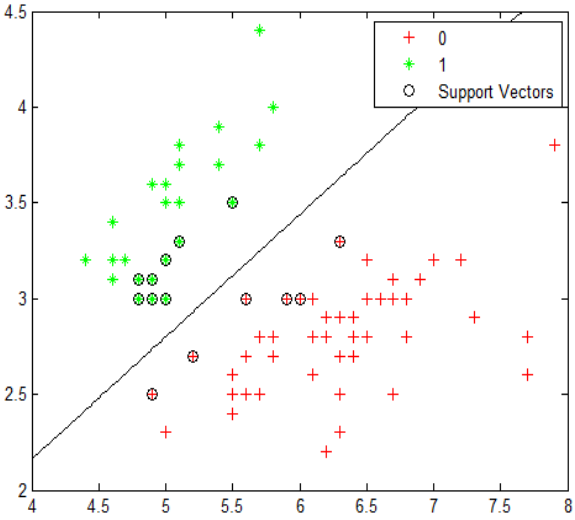


Figure 14 : SVM Classification

## V. CONCLUSION

Although it is difficult to get a accurate result, but we can show the variations that occur when emotion changes. By using MFCC algorithm feature is extracted from which we can observe how changes occur in pitch, frequency and other features when emotion changes .We have done Frame blocking and windowing steps of MFCC algorithm for a same voice and a same sentence in two different emotions and showed difference in pitch with change in emotion. By using some classifying algorithm we can classify different emotions. Some classifying algorithms are SVM,K-MEAN etc. We propose to create a speech corpora to make a database which will be used later in classifying algorithm. We will use SVM to classify some emotions.

## REFERENCES

[1] Sierro Herve, "Extracting from Speech signal : State of the art", Seminar paper, University of Fribourg, Switzerland .pp. 1-4
[2] S. Demicran and H. Kahramanli, "Feature Extraction from Speech Data for Emotion Recognition", A Journal of Advances in Computer Networks, Vol .2 , No. 1, March 2014, pp. 28-30.
[3] Swati Bhutekar, M.B. Chandak, "Corpus Based Emotion Extraction to implement prosody feature in speech sysnthesis sysyem", International Journal of Computer and Electronics Research, Vol 1, Issue 2, August 2012. pp. 67-75.
[4] Shivani Goel, "Extracting MFCC Features for Emotion Recognition from Audio Speech Signal", International Jouranl Advances in Science and Technology (IJAST) . Vol 2, Issue 3, 2014.
[5] Ashish Ingale and D.S Chaudhari, "Speech Emotion Recognition", International Jouranl of Soft Computing and Engineering.(IJSCE). Vol.2, Issue 1, 2012.
[6] 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.
[7] X. M. Cheng, P. Y. Cheng, and L. Zhao, "A study on emotional feature analysis and recognition in speech signal," in *Proc. International Conference on Measuring Technology and Mechatronics Automation*, 2009, IEEE, pp. 418-420.
[8] D. Ververidis and C. Kotropoulos, "Emotional speech recognition: Resources, features, and methods," *Speech Communication*, vol. 48, no. 9, pp. 1162–1181, Sep. 2006.
[9] Xu Zhe, David John and Anthony C. Boucouvalas, " Text-to-Emotion Engine for real time internet communication" Multimedia Communications Research Group, Bournemouth University.
[10] T. L. Nwe, S. W. Foo, and L. C. De Silva, "Speech emotion recognition using hidden Markov models," *Speech Communication*, vol. 41, no. 4, pp. 603–623, 2003.
[11] A. R. Panat and V. T. Ingole, "Affective State Analysis of Speech for Speaker Verification : Experimental Study, Design and Development". Conference on Computational and Multimedia Applications 2007 pp 255-161