Speech emotion recognition is a current topic of research since it has wide range of applications. Up to some extent, emotion recognition from the speech information is an easy task for the human beings because they have natural ability to analyze the speech information. But at the same time emotion detection by using the speech signal is a difficult task for the machine since machine does not have the sufficient intelligence to analyze the emotion from speech signal. Therefore machine does not have the ability to recognize the emotional state of the speaker [1]. By using speech recognition and speaker identification techniques machine can detect what is said and who has said it, but if it is equipped with speech emotion recognition system machine can also detect how it is said [2].

Speech emotion recognition can be defined as the extraction of the emotional state of the speaker from his or her speech signal. The most important application of the speech emotion recognition system is in the field of human machine interaction to make the human machine interface more efficient. Other applications of the speech emotion recognition system are in the psychiatric diagnosis, intelligent toys, lie detection, in call center and in the car board system. [3]

Many different systems have been proposed for the emotion extraction from the speech. These different systems were using the different features and classifiers. Spectral features and prosodic features can be used for speech emotion recognition because both of these features contain the emotional information. Linear predictive cepstral coefficients (LPCC) and Mel-frequency cepstral coefficients (MFCC) are some of the spectral features. Fundamental frequency, loudness, pitch and speech intensity and glottal parameters are the prosodic features used to model the different emotions [4]. The different classifiers available for the speech emotion recognition are k-nearest neighbors (KNN), Hidden Markov Model (HMM) and Support Vector Machine (SVM), Artificial Neural Network (ANN), Gaussian Mixtures Model (GMM). In schuller et al. emotion classification has been performed using HMM and obtained the recognition rate of 86%. But this study was limited only to the features related with the pitch and energy [5]. Shen et al. studied the emotion classification through speech using SVM classifier and performed an experiment on the Berlin emotional database. The overall recognition rate obtained was about 82.5% [6].

In this paper, Hidden Markov Model (HMM) and Support Vector Machine (SVM) classifier were used to classify the five emotional states such as anger, happiness, sadness, surprise, and neutral state which are known as the basic emotions as no distinct emotion is observed. The energy related features, Mel-frequency cepstral coefficients (MFCC), fundamental frequency are some of the features which were used for the speech emotion recognition system. The classification rate of both of this classifiers were observed.

The rest of the paper is organized as follows: Section two provides the details about speech emotion recognition system. The various features which were used in the emotion classification is described in the section three. The detailed information about the emotion classification by using Hidden Markov Model and Support Vector Machine is provided in the section four. Section five discussed about the experimental results obtained during this study. Conclusion of this paper is provided in the section six.

2. SPEECH EMOTION RECOGNITION SYSTEM

The structure of the speech emotion recognition system considered in this study is illustrated in Figure 1. The main elements of the speech emotion recognition system are same as any typical pattern recognition system. It consists of the emotional speech as input, feature extraction, classification of the emotion using HMM or SVM classifier and recognized emotion as the output.

Figure 1 Structure of Speech Emotion Recognition System

The efficiency of the speech emotion recognition system is highly depends upon the naturalness of database used in the system. The emotional speech input to the system may be the acted data or collection of the speech data from real world
Fourier transform representation of the log magnitude

The cepstral coefficients are the coefficients of the

shown in the following equation:

\[ \text{cc}(n) = \text{FT}^{-1}\{\log |\text{FT}(x(n))|\} \quad 3.1 \]

Frequency content of pure tones or that of speech signal does not follow a linear scale. Therefore for each tone with an actual frequency, \( f \) measured in Hz, a subjective pitch is measured on a scale which is referred to as the ‘mel’ scale [7,219]. The relation between the Mel frequency and real frequency is formulated as

\[ F_{\text{mel}} = 3233 \log_{10}(1 + \frac{f_{\text{Hz}}}{1000}) \quad 3.2 \]

The MFCC coefficients can be obtained, the Fourier transform is calculated and this spectrum is filtered by a filter bank in the mel domain. Then the logs of the powers at each of the mel frequencies are taken. Then the inverse Fourier transform is replaced by the cosine transform in order to simplify the computation and is used to obtain the mel frequency cepstrum coefficients [8].

4. EMOTION CLASSIFICATION

Another important aspect of the speech emotion recognition system is an emotion classification. Emotions can be classified from the given speech samples based on the different features extracted from that speech and by providing that features to the classifier. Out of the many classifiers specified in the introduction section, two classifiers Hidden Markov Model and Support Vector Machine were used in this study.

4.1 Hidden Markov Model

Hidden Markov Model is having the long history in the field of speech applications. The HMM consist of the first order markov chain whose states are hidden from the observer therefore the internal behavior of the model remains hidden. The hidden states of the model capture the temporal structure of the data [3]. Hidden Markov Models are statistical models that describe the sequences of events. HMM is having the advantage that the temporal dynamics of the speech features can be trapped due to the presence of the state transition matrix. During classification, a speech signal is taken and the probability for each speech signal provided to the model is calculated. An output of the classifier is based on the maximum probability that the model has generated this signal [8].

4.2 Support Vector Machine

Support Vector Machine is simple and efficient computation of machine learning algorithm, and used in the pattern recognition and classification issues. SVM is having the advantage that for the limited training data, it is having very good classification performance. The idea behind the SVM is to transform the original input set to a high dimensional feature space by using kernel function. Therefore non linear problems can be solved by doing this transformation [6]. Following figure 4.2.1 shows the support vector machine with kernel function, in which input space is consisting of input samples converted into high dimensional feature space and therefore input samples become linearly separable.

![Figure 4.2.1 Support Vector Machine with Kernel Function](image)

5. EXPERIMENTAL STUDY

5.1 Emotional Speech Database

The emotional speech input to the system is given by recording the voice samples from the various speakers. The voice samples has recorded by the male as well as female speakers in five different emotions that is anger, happiness, sadness, surprise, and neutral state which is to be recognized using speech emotion recognition system. These voice samples are taken by using one common statement. The different voice samples were taken for the purpose training of the classifiers and testing of the system.

5.2 Experimental Results using HMM

For the emotion recognition using Hidden Markov Model (HMM), first the database is sort out according to the mode of classification, in this case 5 for five modes. Then the features from input waveform are extracted. These features were added to the database. The transition matrix and emission matrix has been made according to the modes, which generates the random sequence of states and
emissions from the model. Final is estimating the state sequence probability by using viterbi algorithm. From this probability of this HMM define the matching of mode with the database, from the outcome of hmm one can put the result tag as mode that is most match.

The emotion recognition rate by using HMM is shown in the Table 5.2.1. the test samples of the anger were correctly classified as anger at 83.33% whereas they were misclassified as happy at 16.67%. For the happy test sample one have got the recognition rate of 57.14% as happy, 14.29% as sad and 28.57% as surprise state. The sadness is correctly classified as 62.50% and also classified as surprise and neutral state as 12.50% and 25.00% respectively. The test samples of the surprise state were classified as surprise at 71.43% and also classified as anger at 28.57%. the test sample of the neutral state were correctly classified at 75.00% and misclassified as sadness at 25.00%. Therefore from this results which were obtained using hidden markov model one can observe that there was confusion between two or three emotional state.

Table 5.2.1 Emotion Recognition Rate using HMM

<table>
<thead>
<tr>
<th>Emotions</th>
<th>Anger</th>
<th>Happy</th>
<th>Sad</th>
<th>Surprise</th>
<th>Neutral</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anger</td>
<td>85.33</td>
<td>16.67</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Happy</td>
<td>0</td>
<td>57.14</td>
<td>14.29</td>
<td>28.57</td>
<td>0</td>
</tr>
<tr>
<td>Sad</td>
<td>0</td>
<td>0</td>
<td>62.50</td>
<td>15.50</td>
<td>25.00</td>
</tr>
<tr>
<td>Surprise</td>
<td>28.57</td>
<td>0</td>
<td>0</td>
<td>71.43</td>
<td>0</td>
</tr>
<tr>
<td>Neutral</td>
<td>0</td>
<td>0</td>
<td>25.00</td>
<td>0</td>
<td>75.00</td>
</tr>
</tbody>
</table>

5.3 Experimental Results using SVM

First of all the necessary features which are explained above are calculated. All the values that are calculated in the previous steps will get provide to the Support vector machines for the classifier training purpose. After completion of the training, test voice samples through which emotions has to be extracted is provided to the classifier. The Support vector machines again calculate the values of the features of the testing voice sample. Then on the basis of features extracted from the testing voice sample is then compared with the trained voice sample. During the comparison Support vector machines will find the minimum difference the trained voice sample and test voice sample. Using this difference SVM classifier will recognize emotion.

Table 5.2.2 shows the emotion recognition rate of the support vector machine.

Table 5.2.2 Emotion Recognition Rate using SVM

<table>
<thead>
<tr>
<th>Emotions</th>
<th>Anger</th>
<th>Happy</th>
<th>Sad</th>
<th>Surprise</th>
<th>Neutral</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anger</td>
<td>71.12</td>
<td>14.29</td>
<td>0</td>
<td>14.29</td>
<td>0</td>
</tr>
<tr>
<td>Happy</td>
<td>0</td>
<td>57.14</td>
<td>14.29</td>
<td>28.57</td>
<td>0</td>
</tr>
<tr>
<td>Sad</td>
<td>0</td>
<td>0</td>
<td>71.43</td>
<td>0</td>
<td>28.57</td>
</tr>
<tr>
<td>Surprise</td>
<td>22.19</td>
<td>14.29</td>
<td>0</td>
<td>63.33</td>
<td>0</td>
</tr>
<tr>
<td>Neutral</td>
<td>0</td>
<td>0</td>
<td>25.00</td>
<td>0</td>
<td>75.00</td>
</tr>
</tbody>
</table>

As shown in the table, test sample of the anger were classified as anger at 71.42% and also classified as happy and surprise state at 14.29%. the happy emotional test sample was classified as happy, sad and surprise state at 57.14%, 14.29% and 28.57% respectively. The test samples of the sad emotional state were classified as sad and neutral state at 71.43% and 28.57%. The test sample of the surprise state were classified as surprise at 63.33% and also classified as anger and happy state. Neutral test sample was classified as neutral state and sad state at 75.00% and 25.00%.

6. CONCLUSION

In this paper, two classification methods viz. hidden markov model and support vector machine were studied for speech emotion recognition system. Speech features were extracted from the emotional speech sample such as fundamental frequency, energy and mel frequency cepstrum coefficients. Both the classifiers obtained relatively similar accuracy for the emotion classification. The accuracy of the system, is highly depends on emotional speech database used in the system therefore it is necessary to record correct emotional speech database.

REFERENCES