ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

**International Journal of Innovative Research in Computer and Communication Engineering**

*(An ISO 3297: 2007 Certified Organization)*

*Website: www.ijircce.com*

**Vol. 5, Issue 3, March 2017**

# Real World Speech Emotion Recognition From Speech Spectrogram Using Gray-Level Co-Occurence Matrix

Lima Mathew[1], Sajin Salim[2]

PG Scholar, Dept. of ECE, TKM Institute of Technology, Kollam, Kerala, India[1]

Asst. Professor, Dept. of ECE, TKM Institute of Technology, Kollam, Kerala, India[2]

**ABSTRACT:** Recognition of human emotions from the speech samples recorded in a noisy environment is a changeling research area in the field of Human Computer Interaction (HCI).This novel algorithm aims to detect the emotions in real time speech samples using the information inside the spectrogram. The Least Mean Square Adaptive Filter is used for enhancing the speech samples. GLCM based Haralick feature extraction technique is used for extracting texture features from spectrogram. Support Vector Machine was used for being the classifier. This technique computes a Gray Level Co-occurrence matrix (GLCM) and extracts features such as energy, mean, entropy and standard deviation for feature classification. It uses adaptive filtering techniques for noise cancellation. The algorithm was tested in Real time speech signals and Berlin Emotional Database using MATLAB software. SER is applicable in man-machine interaction such as computer tutorial, automatic translation, mobile interaction, health care, children education, ATM security etc.

**KEYWORDS**: Speech Emotion Recognition (SER); Human Computer Interaction (HCI); Spectrogram; Haralick Feature Extraction; Support Vector Machine (SVM); Gray Level Co-occurrence matrix (GLCM).

## I. INTRODUCTION

Speech is a vocalized form of communication that arises from the human vocal tract. Emotion on other side is an individual mental state that arises spontaneously rather than through conscious effort. There are various kinds of emotions which are present in a speech. They are anger, happiness, sadness, surprise, fear, and neutral [2]. Speech Emotion Recognition system is a system which basically identifies the emotional as well as physical state of human being from his or her voice. Emotion recognition is gaining attention due to the widespread applications into various domains: detecting frustration, disappointment, surprise/amusement etc. There are many approaches towards automatic recognition of emotion in speech by using different feature vectors. A proper choice of feature vectors is one of the most important tasks. The speech feature vector can be classified as prosodic features and spectral features. The prosodic features such as pitch (F0), intensity and duration and spectral features such as Mel-Frequency Ceptral Coefficients (MFCC) or Linear Prediction Ceptral Coefficients (LPCC) [1]. The limitation of using the MFCC feature is that it is not superior in noisy environment. In real world application MFCC and LPC degrades rapidly because of noise.

Another feature extraction proposed here is by using a two dimensions magnitude spectrogram which is a graphical display of the squared magnitude of the time-frequency of speech. This magnitude spectrogram representation usually contains distinctive pattern that capture different characteristic of speech emotion signal. The purpose of using a two dimensions magnitude spectrogram is to analyse a spectrogram and capture all of characteristics of speech and preserve the important underlying dependency between different parameters [5]. The features from the spectrogram can be extracted by

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

# International Journal of Innovative Research in Computer and Communication Engineering

*(An ISO 3297: 2007 Certified Organization)*

*Website: www.ijircce.com*

**Vol. 5, Issue 3, March 2017**

using Haralick feature extraction technique. The most important task in Speech Emotion Recognition system is speech feature extraction and feature classification. The different types of classifiers which are used for classifying speech emotions are Support Vector Machine (SVM) [3] [4], Bayesian classifier, Artificial Neural Network (ANN) [6], Hidden Markov Model (HMM) and K-Nearest Neighbors (KNN) [7]. Among this SVM is used as a classifier to classify different emotional states such as anger, fear, joy, sad etc. SVM is simple and efficient algorithm having good classification performance compared to other classifiers.

## II. RELATED WORK

The goal of the speech emotion recognition system is for a machine to be able to "hear," understand," and "act upon" based on the emotions in the spoken information. The speech emotion recognition system may be viewed as working in a four stages: Analysis, Feature extraction, Feature classification and testing.

A. *Feature Extraction Methods*

Feature Extraction is the most important part of speech recognition since it plays an important role to separate one speech from other. Because every speech has different individual characteristics embedded in utterances. These characteristics can be extracted from a wide range of feature extraction techniques proposed and successfully exploited for speech recognition task.

i. **Mel Frequency Cepstral Coefficient (MFCC):** Mel-frequency cepstral coefficients (MFCCs) are a parametric representation of the speech signal, that is commonly used in automatic speech recognition, but they have proved to be successful for other purposes as well, among them speaker identification and emotion recognition. They are claimed to be robust of all the features for any speech tasks. A Mel is a unit of measure of perceived pitch or frequency of a tone. Through the mapping onto the Mel scale, which is an adaptation of the Hertz-scale for frequency to the human sense of hearing, MFCCs enable a signal representation that is closer to human perception. They are calculated by applying a Mel-scale filter bank to the Fourier transform of a windowed signal. Subsequently, a DCT (discrete cosine transforms) transforms the logarithmised spectrum into a cepstrum. Mel filter banks consist of overlapping triangular filters with the cut off frequencies determined by the centre frequencies of the two adjacent filters. The filters have linearly spaced centre frequencies and fixed band width on the Mel scale [5]. The logarithm has the effect of changing multiplication into addition. It converts the multiplication of the magnitude in the FT into addition. MFCC features are the most commonly used most popular and robust technique for feature extraction in currently available speech recognition systems. But the overall performance of MFCC features is not superior in noisy environment. In real world applications the performance of MFCC degrades rapidly because of the noise [2].

ii. **Linear Predictive Coding (LPC):** Linear prediction is a mathematical computational operation which is linear combination of several previous samples. LPC of speech has become the predominant technique for estimating the basic parameters of speech. It provides both an accurate estimate of the speech parameters and it is also an efficient computational model of speech. The basic idea behind LPC is that a speech sample can be approximated as a linear combination of past speech samples. Through minimizing the sum of squared differences (over a finite interval) between the actual speech samples and predicted values, a unique set of parameters or predictor coefficients can be determined. These coefficients form the basis for LPC of speech [2]. The principle behind the use of LPC is to minimize the sum of the squared differences between the original speech signal and the estimated speech signal over a finite duration. This could be used to give a unique set of predictor coefficients.

**B.** ***Classification Methods***

It is the most important task in speech recognition systems. It needs training vector and test vector. Classification problem grows in accordance with the dimension of the input vectors. The different classification techniques are given below:

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 3, March 2017

i. **Support Vector Machine (SVM):** Support Vector Machine is used for recognition of emotional states. Support Vector Machine (SVM) is a supervised machine learning algorithm which can be used for both classification and regression challenges. Classification is on the basis of finding the right hyper-plane that differentiates the two classes separable. SVM finds the particular hyper-plane having maximum margin of separation. It has good feature extraction but input features are processed separately. The SVM in particular defines the criterion to be looking for a decision surface that is maximally far away from any data point. This distance from the decision surface to the closest data point determines the *margin* of the classifier. This method of construction necessarily means that the decision function for an SVM is fully specified by a (usually small) subset of the data which defines the position. These points are referred to as the *support vectors* (in a vector space, a point can be thought of as a vector between the origin and that point).

$$\mathbf{x}_i \bullet \mathbf{w} + b \geq +1 \text{ when } y_i = +1 \qquad (1)$$

$$\mathbf{x}_i \bullet \mathbf{w} + b \leq -1 \text{ when } y_i = -1 \qquad (2)$$

ii. **Hidden Markov Model (HMM):** In hidden Markov model, the state is not directly visible, but the output dependent on the state is visible. The HMM needs to be trained on a set of seed sequences and generally requires a larger seed than the simple Markov models. The training involves repeated iterations of the Viterbi algorithm. The Viterbi algorithm is expensive, both in terms of memory and compute time.

iii. **Artificial Neural Network (ANN):** An artificial neural network (ANN), often just called a neural network (NN), is an interconnected group of artificial neurons that uses a mathematical model or computational model for information processing based on a connectionist approach to computation. In most cases an ANN is an adaptive system that changes its structure based on external or internal information that flows through the network. The particular model used in this technique can have many forms, such as multi-layer perceptions or radial basis functions. The MLP is a type of neural network that has grown popular over the past several years. A MLP with one input layer, one hidden layer, and one output layer. MLP’s are usually trained with an iterative gradient algorithm known as back propagation.

## III. METHODOLOGY

To recognize the emotion from speech signal, speech emotion recognition system using speech spectrogram is introduced. The whole processes for speech emotion recognition system can be shown in Fig.1. The input speech signal undergoes pre-processing such as making the signal mean to zero, pre-emphasis, least mean square adaptive filtering and amplitude normalisation. The spectrogram of the speech signal is generated and extracts the texture features from it using Gray-Level Co-Occurence based Haralick feature extraction techniques. These extracted features such as mean, entropy and standard deviation is given to SVM classifier.

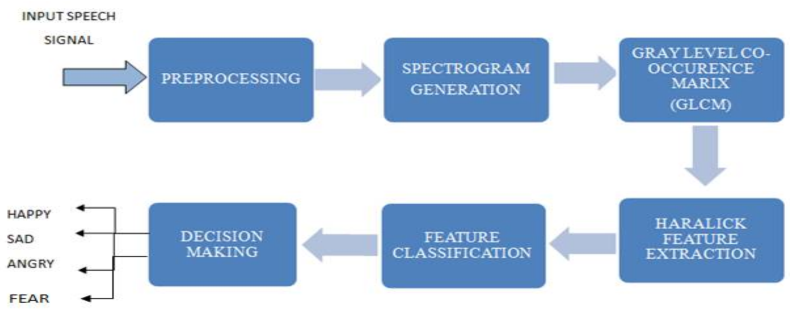



Fig 1: Speech emotion recognition of the system

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

**International Journal of Innovative Research in Computer and Communication Engineering**

*(An ISO 3297: 2007 Certified Organization)*

*Website: www.ijircce.com*

**Vol. 5, Issue 3, March 2017**

*A. Emotional speech input*

The emotional speech dataset of real time speech samples collected from different peoples in a noisy environment is selected for the proposed framework. The Berlin Emotional dataset is also used for comparing accuracy of the given system.

*B. Pre-processing*

The continuous time signal is selected. At first, zero mean is done. This process will make the speech signal mean to be zero for easy to process. Actually, the speech amplitude level in each person will different. So it needs to adjust this speech signal before processing. The pre-emphasis for flattening speech spectrum is processed and normalize the amplitude of each speech to -1 to 1. For a real world application this pre-processing steps are not sufficient because audio recording involves 4 major types of noise: hiss, rumble, crackle and hum. Other commonly occurring noises are acoustic noise, white noise and environmental noises. So in order to reduce the effect of these noises

Adaptive Filtering is used. Adaptive filters are variable filters whose filter coefficients are adjustable or modifiable automatically to improve its performance in accordance with some criterion, allowing the filter to adapt the changes in the input signal characteristics. Because of their self-adjusting performance and in-built flexibility, adaptive filters have found use in many diverse applications such as telephone echo cancelling, radar signal processing, and noise cancelling and biomedical signal enhancement. Adaptive filters are the alternate method for recovering desired speech from the noise. Several algorithms have been proposed in earlier days to detect the desired signal. Least mean square (LMS) algorithm was the most efficient method in terms of computation and storage requirements.

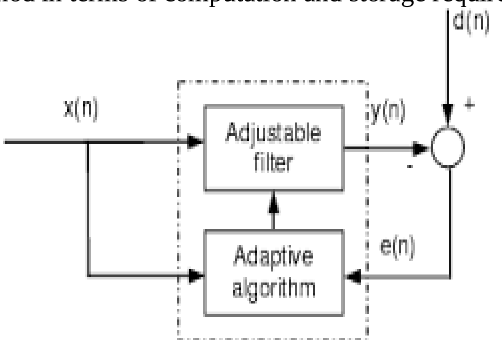



Fig.2. Block diagram of Adaptive Filter.

One of the most successful adaptive algorithms is the LMS algorithm. Instead of computing in one go as in wiener–Hopf equation, in LMS the coefficients are adjusted from sample to sample in such or way as to minimize the MSE[13] . The LMS algorithm is based on the steepest descent algorithm where the weight vector is updated from sample to sample as follows

$$w_{k+1} = w_k - \mu \nabla_k$$

Where $w_k$ and $\nabla_k$ are the weight and the true gradient vectors respectively at the $k^{th}$ sampling instant. $\mu$ controls the stability and rate of convergence. The steepest descent algorithm in the above equation still requires knowledge of R and P. since is obtained by evaluating the equation.

$$\nabla = \frac{d\varepsilon}{dw} = 0 - 2P + 2RW$$

The LMS algorithm is a practical method of obtaining estimates of the filter weights $w_k$ in real time without matrix inversion in the equation $W_{opt} = R^{-1}P$ or the direct computation of the auto correlation and cross correlation.

$$\nabla = 0 - 2P + 2RW$$

In the LMS algorithm, instantaneous estimates are used for. Thus

$$\nabla = 0 - 2P + 2RW$$

$$\nabla_k = -2x_k y_k + 2x_k x_k^T w_k$$

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

**International Journal of Innovative Research in Computer and Communication Engineering**

*(An ISO 3297: 2007 Certified Organization)*

*Website: www.ijircce.com*

**Vol. 5, Issue 3, March 2017**

$$\nabla_k = -2x_k e_k$$

Where $e_k = y_k - x_k^T w_k$ , On replacing the value of $\nabla_k$ in steepest descent algorithm yields

$$w_{k+1} = w_k - \mu\nabla_k$$

$$w_{k+1} = w_k - \mu(-2x_k e_k)$$

$$w_{k+1} = w_k + \mu 2x_k e_k)$$

Clearly, the LMS algorithm above doesn't require prior knowledge of the signal statistics (that is the correlations R and P), but instead uses their instantaneous estimates as shown above. The weights obtained by the LMS algorithm are only estimates, but these estimates improve gradually with time as weights are adjusted and the filter learns the characteristics of the signals. Eventually, the weights converge [14].

The condition of convergence is

$0 < \mu > 1/\lambda_{max}$, where $\lambda_{max}$ is the maximum Eigen value of the input data covariance matrix.

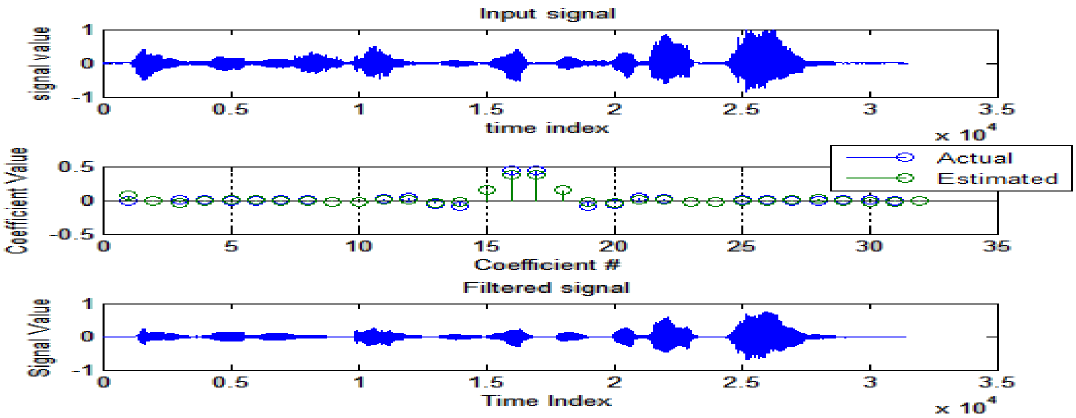



Fig.3. LMS Adaptive Filter response

*C. Spectrogram Generation*

Traditional representations of speech are derived from short-time segments of the signal and result in time-frequency distributions of energy such as the short-time Fourier transform and spectrogram. Speech-signal models of such representations have had utility in a variety of applications such as speech analysis, recognition, and synthesis. Nonetheless, they do not capture spectral, temporal, and joint spectro-temporal energy fluctuations (or "modulations") present in local time-frequency regions of the time-frequency distribution. Inspired by principles from image processing and evidence from auditory neurophysiologic models, a variety of two-dimensional (2-D) processing techniques have been explored in the literature as alternative representations of speech. This thesis develops speech-signal models for a particular 2-D processing approach in which 2-D Fourier transforms are computed on local time-frequency regions of the canonical narrowband or wideband spectrogram. The spectrogram is a 2-D graphical form of the time-frequency representation of speech signal. A common format is a graph with two geometric dimensions: the horizontal axis represents time or rpm, the vertical axis is frequency; a third dimension indicating the amplitude of a particular frequency at a particular time is represented by the intensity or colour of each point in the image. This graph is used for speech emotion classification. This is shown in the above Fig. 5.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

**International Journal of Innovative Research in Computer and Communication Engineering**

*(An ISO 3297: 2007 Certified Organization)*

*Website: www.ijircce.com*

**Vol. 5, Issue 3, March 2017**

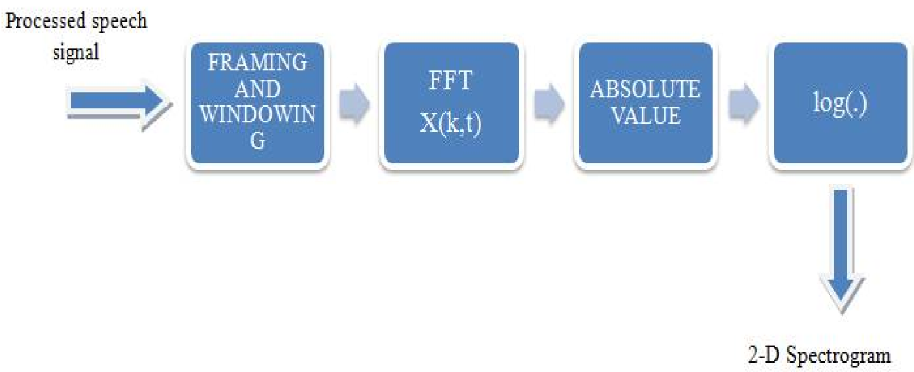
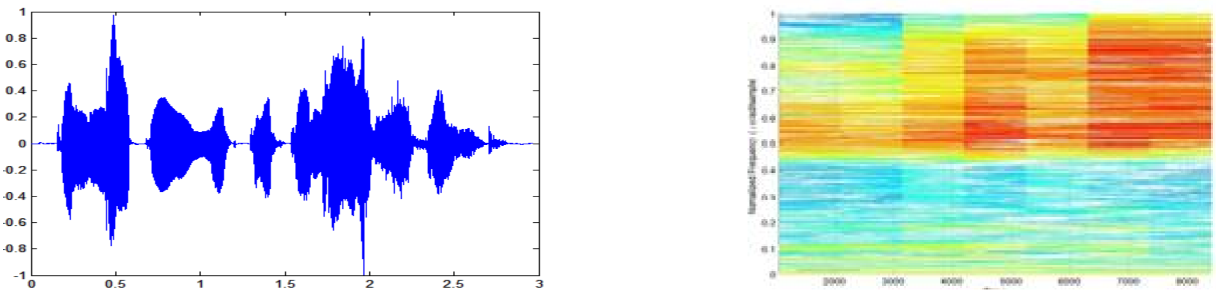



Fig.6. Steps for 2-D Spectrogram generation.

In order to create spectrogram of speech signals the following steps has to follow: Framing, Windowing, and FFT. The Fig.4 represents the steps in generating the two-dimensional colour spectrogram of each speech emotion signals.



(a) (b)

Fig. 5 Examples of the samples of fear emotion produced by a female speaker (a) Speech signal (b) Spectrogram.

***i) Framing:*** Segment the speech signal into small frames each having duration of 20-25ms.

**ii)** ***Windowing:*** It removes the discontinuity at the edge of each frame by using Hamming window. The Hamming window for the speech signal is formulated by using the following equation.

$W[n] = 0.54 - 0.46 \cos(2n\pi/(N-1))$, where, $N$ is the length of the window, w[n] is Hamming window.

**iii)** ***FFT:*** The FFT is an efficient algorithm for transforming the signals from the time domain to frequency domain. Normally the speech signal is in time domain format. For easy to speech comparison time domain speech is transformed to frequency domain.

*D. Feature extraction*

The feature extraction is the changeling step in speech emotion recognition. In this step the feature for speech signal classification is extracted from the spectrogram of corresponding speech emotions such as anger, fear, joy and sad. The spectrogram is used for feature extraction in this paper on the basics of the concept that the spectrograms of different emotion are different in their texture appearance. The spectrogram of fear and anger emotions has different color combinations [12]. The features from the spectrogram can be extracted by using GLCM based Haralick feature extraction technique and extracts the haralick texture features such as energy, mean, entropy and standard deviation. GLCM is one of the better texture feature extraction technique which provides a better feature vector than while using Law's mask based texture feature extraction technique. In GLCM matrix the number of rows and columns is equal to

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

*(An ISO 3297: 2007 Certified Organization)*

*Website: www.ijircce.com*

**Vol. 5, Issue 3, March 2017**

the number of gray levels in the image. The GLCM matrix tabulation is shown in the Fig.6. Mathematically a co-occurrence matrix C is defined over an n x m.

i) GLCM Algorithm:

1. Quantize the image data. Each sample on the spectrogram is treated as a single image pixel and the value of the sample is the intensity of that pixel. These intensities are then further quantized into a specified number of discrete gray levels as specified under **Quantization**.

2. Create the GLCM. It will be a square matrix *N* x *N* in size where *N* is the **Number of levels** specified under **Quantization**. The matrix is created as follows:

a. Let *S* be the sample under consideration for the calculation.

b. Let *W* be the set of samples surrounding sample *S* which fall within a window centered upon sample s of the size specified under **Window Size**.

c. Considering only the samples in the set *W*, define each element *i, j* of the GLCM as the number of times two samples of intensities *i* and *j* occur in specified **spatial relationship**. The sum of all the elements *i, j* of the GLCM will be the total number of times the specified spatial relationship occurs in *W*.

d. Make the GLCM symmetric:

i. Make a transposed copy of the GLCM

ii. Add this copy to the GLCM itself

This produces a symmetric matrix in which the relationship *i* to *j* is indistinguishable for the relationship *j* to *i* (for any two intensities *i* and *j*). As a consequence the sum of all the elements *i, j* of the GLCM will now be twice the total number of times the specified spatial relationship occurs in *W* (once where the sample with intensity *i* is the reference sample and once where the sample with intensity *j* is the reference sample), and for any given *i*, the sum of all the elements *i, j* with the given *i* will be the total number of times a sample of intensity *i* appears in the specified spatial relationship with another sample.

3. Normalize the GLCM:

i. Divide each element by the sum of all elements,
the elements of the GLCM may now be considered probabilities of finding the relationship *i, j* (or *j, i*) in *W*.

$$P_{i,j} = \frac{W_{i,j}}{\sum_{i,j=0}^{N-1} W_{i,j}}$$

where , i is the row number and j is the column number.

*E) SVM Classifier:* The identification of emotion related speech features is extremely challenging task. Support Vector Machine is used as a classifier to classify different emotional states such as anger, sadness, fear and joy. SVM has a better classification performance on a small amount of training samples. The major principle of SVM is to establish a hyperplane as the decision surface maximizing the margin of separation between negative and positive samples. Thus SVM is designed for two class pattern classification. Multiple pattern classification problems can be solved using a combination of binary support vector machine.

## IV. RESULTS AND DISCUSSIONS

In this paper, two different speech data set were used which are Berlin Emotion Database and Real time speech samples collected in noisy environment. The Berlin Emotional Database is a standard database. The main emotions such as happy or joy, anger and sad are classified for a particular incoming speech signal using MATLAB software. In our experiments, we selected 4 speech emotions of about 50 males and females are recorded in noisy environment which are anger, sadness, happy and fear. These emotions data can be enhanced by using adaptive filtering with LMS

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

# International Journal of Innovative Research in Computer and Communication Engineering

*(An ISO 3297: 2007 Certified Organization)*

*Website: www.ijircce.com*

**Vol. 5, Issue 3, March 2017**

algorithm. The spectrogram of the each emotion speech signal is computed. This shows the obviously different spectrogram. The textures features from each emotion spectrogram were selected using the GLCM texture feature extraction techniques and classified using Multiclass SVM. Table. I below list the number of samples in Berlin Emotional database for testing with the proposed algorithm. A message box will displays the result specifying whether the given input is fear, anger, happy or sad on the basics of the training vector that is used for training the support vector machine.

TABLE I
NUMBER OF SAMPLES IN BERLIN DATABASE

| Emotions | Female | Male |
|---|---|---|
| Anger | 67 | 60 |
| Fear | 32 | 37 |
| Joy | 44 | 27 |
| Sad | 37 | 25 |
| Neutral | 30 | 28 |

TABLE II
NUMBER OF SAMPLES IN REAL TIME DATABASE

| Emotions | Female | Male |
|---|---|---|
| Anger | 52 | 45 |
| Fear | 29 | 26 |
| Joy | 56 | 45 |
| Sad | 27 | 25 |
| Neutral | 32 | 20 |

The real time speech signals of different emotions such as anger, fear, sad, joy and neutral are recorded from many males and females at different situations. Table II below lists the number of male and female speech emotion samples to be tested with the proposed algorithm. The experimental results of Real time database shows that the accuracy in Real time database which is the real speech with represent emotion of common peoples recorded in a noisy environment has higher accuracy compared to other researcher's methods. Real time database (female) gives an accuracy of 88.4% for anger, 89.18% for sad, 84.09% for joy, 78.60% for neutral and 92.4% for fear. The experimental result of Real time database is shown in Table.III and Table.IV.

TABLE III
CONFUSION MATRIX OF THE PROPOSED ALGORITHMS (FEMALE) WITH REAL TIME EMOTIONAL DATABASE

| Emotions | Angry | Sad | Joy | Neutral | Fear |
|---|---|---|---|---|---|
| Anger | 88.46 | 0.00 | 3.84 | 0.00 | 7.69 |
| Sad | 2.7 | 89.18 | 0.00 | 2.7 | 5.4 |
| Joy | 9.0 | 4.5 | 84.09 | 2.27 | 0.00 |
| Neutral | 1.4 | 7.8 | 3.2 | 78.60 | 9.00 |
| Fear | 2.7 | 1.5 | 2.8 | 1.2 | 92.4 |

TABLE IV
CONFUSION MATRIX OF THE PROPOSED ALGORITHMS (MALE) WITH REAL TIME EMOTIONAL DATABASE

| Emotions | Angry | Sad | Joy | Neutral | Fear |
|---|---|---|---|---|---|
| Anger | 92.46 | 1.50 | 4.82 | 1.02 | 1.98 |
| Sad | 5.7 | 84.18 | 2.3 | 7.82 | 5.4 |
| Joy | 2.12 | 4.5 | 88.09 | 3.27 | 2.02 |
| Neutral | 8.91 | 4.28 | 1.2 | 71.43 | 14.18 |
| Fear | 6.7 | 1.5 | 3.8 | 1.2 | 88.02 |

The Berlin emotional database is also an real speech with long sentences of a real world speech. For this reason in most case and earlier methods Berlin emotion database has lower accuracy compared to Audiovisual Thai emotion database. But using this proposed algorithm the Berlin emotion database produce better accuracy compared to other methods. The experimental results of Berlin emotion database in shown in Table.V and Table.VI.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 3, March 2017

TABLE V

CONFUSION MATRIX OF THE PROPOSED ALGORITHMS (FEMALE) WITH REAL TIME EMOTIONAL DATABASE

| Emotions | Angry | Sad | Joy | Neutral | Fear |
|---|---|---|---|---|---|
| Anger | 96.46 | 1.20 | .82 | . 02 | 1.50 |
| Sad | 2.7 | 90.82 | 3.3 | 2.4 | .88 |
| Joy | 2.12 | 2.5 | 89.16 | 1.4 | 4.82 |
| Neutral | 3.91 | 2.28 | 1.2 | 86.43 | 6.18 |
| Fear | 1.3 | 1.02 | .81 | 1.2 | 96.02 |

TABLE VI

CONFUSION MATRIX OF THE PROPOSED ALGORITHMS (FEMALE WITH REAL TIME EMOTIONAL DATABASE

| Emotions | Angry | Sad | Joy | Neutral | Fear |
|---|---|---|---|---|---|
| Anger | 94.58 | .98 | 1.2 | 2.13 | 1.11 |
| Sad | 1.7 | 88.82 | 2.3 | 4.4 | 2.78 |
| Joy | 5.42 | 2.5 | 87.16 | 1.4 | 3.82 |
| Neutral | 1.91 | 1.28 | 5.2 | 88.43 | 4.18 |
| Fear | 3.4 | 2.50 | 1.05 | 1.2 | 91.64 |

## V. CONCLUSION

Real world speech emotion recognition from speech spectrogram using gray-level co-occurence matrix presents a novel algorithm for detecting the different emotions in speech by using speech spectrogram. The two databases used are Real time database and Berlin emotion database. The Real time database is recorded in a noisy environment. The noisy signals are enhanced by using LMS adaptive filtering. The texture features from the spectrogram of the speech samples are computed for feature extraction. Features are extracted from spectrogram using GLCM based Haralick feature extraction technique. The features are classified by using SVM. The experimental results show that the proposed framework can efficiently find the correct speech emotion compared to using the traditional method in most case.

## REFERENCES

[1] Sathit Prasomphan, "Detecting Human Emotion via Speech Recognition by Using Speech Spectrogram", *16th IEEE International journal on Human Communication Recognition, January, Thailand, .2016.*

[2] M.M.H.E. Ayadi, M.S., Kamel and F. Karray, "Survey on speech emotion recognition: Features, classification schemes and databases", *Pattern Recognition, pp. 572–587, 2011.*

[3] E.H. Kim, K.H. Hyu, S.H. Kim and Y.K.Kwak, "Speech Emotion Recognition Using Eigen-FFT in Clean and Noisy Environments", *16th IEEE International Conference on Robot Human Interactive Communication August, Jeju, Korea, 2007.*

[4] Akash Shaw, Rohan Kumar Vardhan ," Emotion Recognition and Classification in Speech using Artificial Neural Networks", *International Journal of Computer Applications (0975 – 8887) Volume 145 – No.8, July 2016* .

[5] Shambhavi S.S, Dr. V.N Nitnaware,"Emotion Speech Recognition using MFCC and SVM", *International Journel of Engineering & Technology (ISSN:2278-0181) Vol.4 Issue 06,June-2015.*

[6] L. He, M. Lech, M.C. Maddage and N.B. Allen," Time-frequency feature extraction from spectrogram and wavelet packets with application to automatic stress and emotion classification in speech Information", *7th International Conference on Communications and Signal Processing*, ICICS 2009, Macau, pp. 1–5, 2009, .

[7] A. Harimi, A. Shahzadi, A.R. Ahmadyfard and K. Yaghmaie, "Classification of emotional speech through spectral pattern features", *Journal of AI and Data Mining*, vol.2, no.1, pp.53–61, 2014, .

[8] Japanese Emotion Database (Japanese emotional speech).

[9] M.M. Javidi and E.F. Roshan. ,"Speech Emotion Recognition by Using Combinations of C5.0, Neural Network (NN) and Support Vector Machines (SVM) Classification Methods". *Journal of mathematics and computer Science*. vol. 6, pp. 191-200, Apr. 2013.

[10] Mihir Narayan Mohanty, A Routray.,"Machine Learing Approach for Emotional Speech Classification",*SEMCO-14*, Book chapter, Springer /Verlag Berlin Heidelberg, 2014.

[11] K.Z. Mao et.al.,"Probabilistic Neural-Network Structure Determination for Pattern Classification". *IEEE Transactions on neural networks*.,vol. 11, no. 4, pp. 1009-1016. Jul. 2000.

[12] Chuang, Z.J., Wu, C.H. "Emotion recognition using acoustic features and textual content", *In Proc. ICME, Taipei, Taiwan*. pp. 53- 56, 2004.

[13] H.Kaur and R.Talwar, "Performance and Convergence Analysis of LMS Algorithm," IEEE ICCIC, Dec.2012.

[14] J.Gorriz and J.Ramrez, "A Novel LMS Algorithm Applied to Adaptive Noise Cancellation," *IEEE Signal Process Letters*, vol. 16, no. 1, Jan. 2009.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

*(An ISO 3297: 2007 Certified Organization)*

*Website: www.ijircce.com*

**Vol. 5, Issue 3, March 2017**

## BIOGRAPHY

**Lima Mathew** is a PG scholar in Electronics and Communication Department, TKM Institute of Technology, Kollam, Kerala, India. She received Bachelor of Technology (B Tech) degree in 2015 TKM Institute of Technology, Kollam, Kerala, India.

**Sajin Salim** is working as Assistant Professor in Electronics and Communication Department, TKM Institute of Technology, Kollam, Kerala, India. He received Master of Engineering (ME) in Communication System at Sethu Institute of Technology, Madurai under Anna University in 2012 and Bachelor of Technology (B Tech) degree in 2007 from Travancore college of Engineering under Kerala University,