IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661,p-ISSN: 2278-8727, Volume 17, Issue 4, Ver. I (July – Aug. 2015), PP 01-08
www.iosrjournals.org

Emotion Recognition using combination of MFCC and LPCC with Supply Vector Machine

Soma Bera[1], Shanthi Therese[2], Madhuri Gedam[3]
[1](Department of Computer Engineering, Shree L.R. Tiwari College of Engineering, India)
[2](Department of Information Technology, Thadomal Shahani Engineering College, India)
[3](Department of Computer Engineering, Shree L.R. Tiwari College of Engineering, India)

Abstract: *Speech is a medium through which emotions are expressed by human being. In this paper, a mixture of MFCC and LPCC has been proposed for audio feature extraction. One of the greatest advantage of MFCC is that it is capable of identifying features even in the existence of noise and henceforth it is combined with the advantage of LPCC which helps in extracting features in low acoustics. Two databases have been considered namely Berlin Emotional Database and a SAVEE database. SVM has been implemented for classification of seven emotions (Sadness, Joy, Fear, Anger, Boredom, Disgust and Neutral). Accuracy of the developed model is presented using confusion matrix. The Recognition outcome of the combined MFCC and LPCC extracted features are compared with the isolated results. The maximum accuracy rate that reaches by using the combination of feature extraction method is 88.59% for non linear RBF (Radial Basis Function) kernel SVM.*

Keywords: *Confusion matrix, Feature extraction, Linear SVM, LPCC, MFCC, RBF non linear SVM, Speech Emotion Recognition*

I. Introduction

Speech contains emotions which reflects the mood of a person that describes the physical state of mind. Human speech entails of various different emotions such as happiness, anger, sad, fear, disgust, boredom, compassion, surprise and neutral. A vocal communication through which people express ideas has ample amount of information that is understood discreetly. This information may be expressed in the intonation, pitch, volume and speed of the voice as well as through emotional state of mind. The speaker's emotional state is closely related to this information. The most common basic emotions are happiness (joy), anger, fear, boredom, sadness, disgust and neutral. Over few decades, the recognition of emotions has become a vital research area that has gained boundless interest.

II. Proposed Work

In our work, we combine the advantages of both the feature extraction techniques namely MFCC and LPCC. The extracted audio features of both the techniques are then supplied to SVM classifier for final emotion classification to reach an appropriate level of accuracy. The proposed method is based on the following four principals [1].

Feature Extraction: In this, the speech signal is elaborated in order to obtain a definite number of variables, called features, which resembles to be useful for speech emotion recognition.

Feature Selection: it chooses the most suitable features in order to reduce the computational load thereby reducing the time required to identify an emotion.

Database: it is the memory of the classifier; it contains sentences (audio clips) separated according to the emotions to be recognized.

Classification: this actually classifies the emotions by using the features selected by the Feature Selection block and the audio clips in the Database.

III. Emotional Database

In current research work, an ample amount of research work has been carried out in the field of Emotion Recognition. Various databases have been built for the same. In our case, we have used a Berlin Emotional Database (BED), which is a German corpus and established by Department of acoustics technology of Berlin Technical University [2]. Berlin Database plays a significant role in Speech Emotion Recognition. The entire database is easily accessible and is freely available. Moreover, it is one of the most important database for Emotion Recognition. Berlin Database consists of in total 800 utterances, out of which 535 utterances are successfully extracted in this project. The database consists of 7 emotions namely Joy, Anger, Sadness, Fear, Disgust, Boredom and Neutral. Another database that we are experimenting on is the one based on English corpus known as **Surrey Audio-Visual Expressed Emotion** "SAVEE" database. This database consists of 7 emotions namely joy, Surprise, anger, sadness, fear, disgust, and neutral. There are four male speakers recorded

for the SAVEE database. The database consists of audio WAV files sampled at a frequency of 44.1 kHz. There are 15 sentences for each of the seven emotion categories, 480 British English utterances in total [3].

IV. Feature Extraction

Various feature extraction techniques are available such as LPCC, MFCC, LPC, Perceptual Linear Predictive Coefficient (PLP), Power Spectral Analysis (FFT), Mel Scale Cepstral Analysis (MEL), etc. We have combined the advantages of MFCC and LPCC feature extraction in this project in order to reach a better accuracy level for emotion recognition.

4.1 MFCC

The reason for MFCC being most widely used is its capability of easy calculation, noble ability of the peculiarity, anti-noise and other various such advantages. MFCC in the low frequency region has a noble resolution of frequency, and the robustness to noise is also very good, but the accuracy of the coefficient in the high frequency is not satisfactory. The process of calculating MFCC is shown in Fig (1).

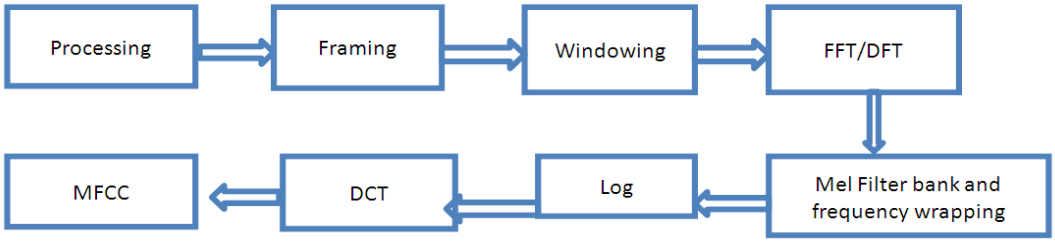

Fig (1) Process of Calculating MFCC

In the pre-processing stage first each signal is de-noised and are then divided into frames using a Hamming window. In framing, the speech samples are segmented into the small frames with the time length within the range of 20-40ms. In our case, we keep the frame length to be of 20ms. Next, in windowing phase, each individual frame is windowed in order to minimize the signal discontinuities at the beginning and end of each frame. The next phase known as FFT transforms each frame of N samples from the time domain into the frequency domain. Mel Filter Bank consists of overlapping triangular filters in order to filter out any noise in the speech sample. It is calculated by using the following equation (1) [4] [5]:

$$B(f) = 2595.\log(1 + f/700) \qquad \ldots\ldots\ldots (1)$$

where f resembles the frequency denoted on a linear scale axis.

Taking logarithm simply converts the multiplication of the magnitude in the Fourier transform into addition. DCT converts the frequency domain back to time domain and finally the MFCC features of the speech samples are extracted successfully.

4.2 LPCC

LPCC represents the characteristics of certain speech channel, and the same person with dissimilar emotional speech will have dissimilar channel features, thereby extracting these feature coefficients to classify the emotions contained in speech. The computational process of LPCC is usually a repetition of computing the linear prediction coefficients (LPC) [6] means the process of calculating LPCC features remain the same as that of calculating LPC features.

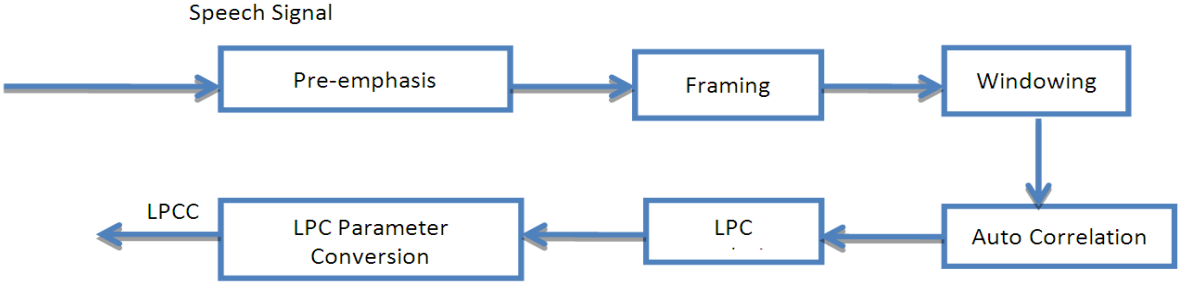

Fig (2) Process of calculating LPCC

The signal s[n] is predicted by a linear combination of its past values. The predictor equation (2) is defined as

$$s^{\sim}[n] = \sum_{k-1}^{p} ak\, s[n-k] \qquad \ldots\ldots\ldots\ldots\ldots\ldots (2)$$

Here s[n] is the signal, a_k is the predictor coefficients and s[n-k] is the predicted signal. The predicted error signal in equation (3), is defined as

$$E[n] = s[n] - s\square\,[n] \qquad \ldots\ldots\ldots\ldots\ldots\ldots (3)$$

the linear prediction coefficients are transformed into cepstral coefficients using a recursive relation as illustrated in equation (4) [8].

$$IDFT(\log(|DFT(x)|)) \qquad \ldots\ldots\ldots\ldots\ldots\ldots (4)$$

If x is LPC, the cepstral coefficients are known as linear prediction cepstral coefficients (LPCC).

V. SVM Classification

SVM, a binary classifier is a simple and effective computation of machine learning algorithms, and is generally used for pattern identification and classification problems, and in case of restricted training data, it can have a very noble classification performance as compared to other classifiers [7]. The different SVM that are available for classification are linear and nonlinear SVM. Linear are the ones that classify the data elements which have a linear hyperplane, meaning classes separated by a straight line. And the nonlinear are the ones wherein the separating hyperplane are nonlinear that is not a straight line. The function of SVM is to transform the original input set to a high dimensional feature space by making use of kernel function. Hence, nonlinear problems can be solved by doing this transformation. Different SVM techniques are available such as One vs One, One vs All, Polykernel SVM, Gaussian kernel SVM, Radial Basis Function (RBF), etc. RBF kernel is implemented in our case for nonlinear classification.

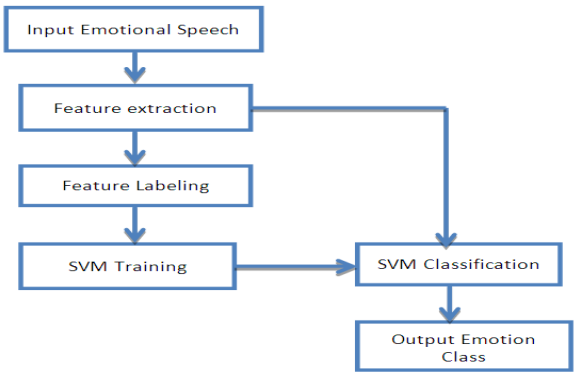

Fig (3) Block Diagram of Speech Emotion Recognition System using SVM Classification

5.1 RBF (RADIAL BASIS FUNCTION) KERNEL SVM

The standard form for implementing RBF function [9] is given by the following function as shown in equation (5):

$$h(x) = \sum_{n=1}^{N} Wn\ \exp(-\gamma\ ||X - Xn||^{2}) \qquad \ldots\ldots\ldots\ldots\ldots\ldots (5)$$

where h(x) = hypothesis of the given data element x, Wn = weight of the given samples, $\exp(-\gamma\ ||X\text{-}Xn||^2)$ = the basis function, wherein $||X\text{-}Xn||^2$ = the radial distance

VI. Experimental Results And Analysis

The data is trained by extracting the voice features MFCC and LPCC. The performance of speech emotion recognition system is subjective to many factors, mainly the quality of the speech samples, the extracted features and classification algorithm that has been used. The following figure illustrates the extracted MFCC and LPCC features of various emotions respectively.

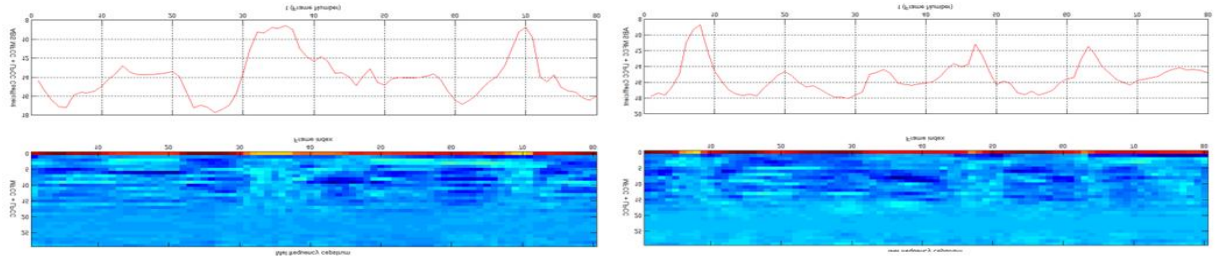

Fig (4) Combination of MFCC and LPCC features for Joy and Fear

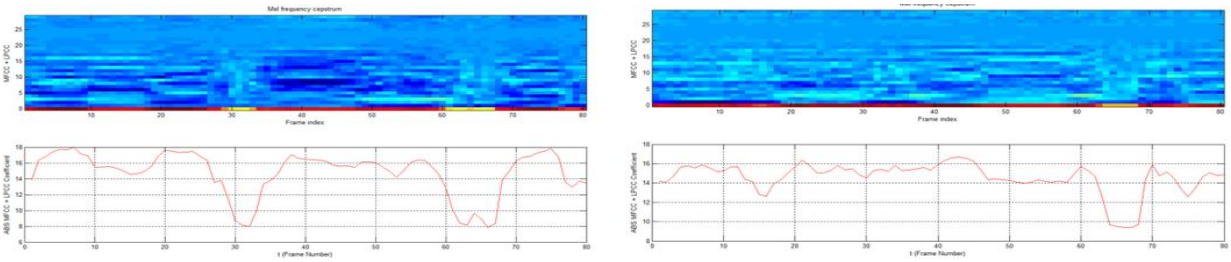

Fig (5) Combination of MFCC and LPCC features for Anger and Sadness

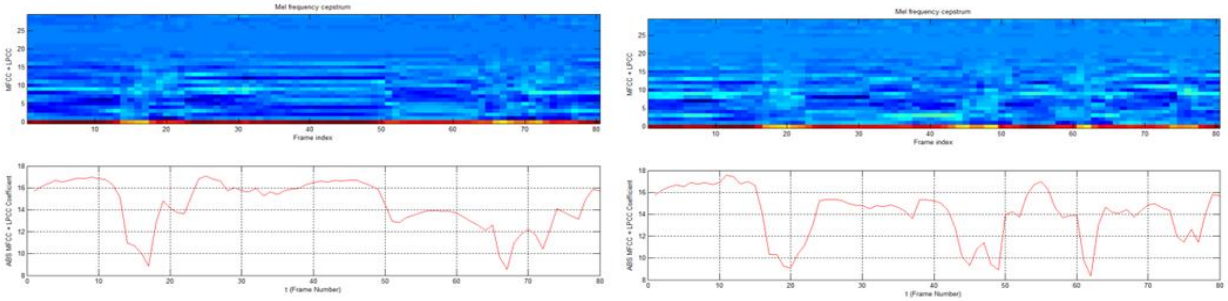

Fig (6) Combination of MFCC and LPCC features for Boredom and Neutral

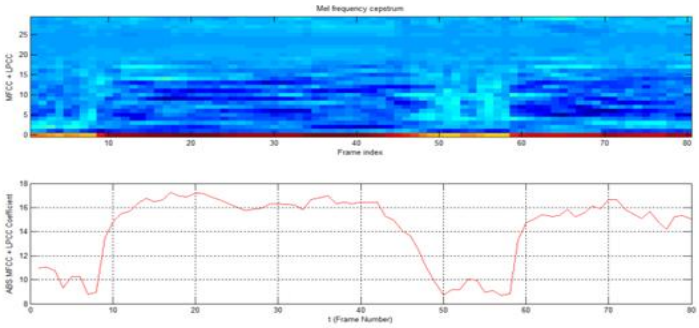

Fig (7) Combination of MFCC and LPCC features for Disgust

VII. Results For Berlin Emotional Database

The following tables illustrate the Recognition rate (%) of all the emotions (starting with only 2 emotions and then gradually increasing the number of emotions) classified by SVM.

Table 1.1: Recognition rate (%) of only 2 emotions (Anger & Sad)

Recognition (%)	Linear SVM	Non linear SVM
MFCC	87.83	**100**
LPCC	91.01	**100**
MFCC+LPCC	90.48	**100**

Table 1.2: Recognition rate (%) of only 3 emotions (Anger & Sad & Joy)

Recognition (%)	Linear SVM	Nonlinear SVM
MFCC	53.85	**100**
LPCC	53.46	**100**
MFCC+LPCC	**61.54**	**100**

Table 1.3: Recognition rate (%) of only 4 emotions (Anger & Sad & Joy & Fear)

Recognition (%)	Linear SVM	Nonlinear SVM
MFCC	46.06	85.45
LPCC	53.33	87.88
MFCC+LPCC	**56.06**	86.36

Table 1.4: Recognition rate (%) of 5 emotions (Anger & Sad & Joy & Fear & Disgust)

Recognition (%)	Linear SVM	Nonlinear SVM
MFCC	46.67	84.53
LPCC	45.60	85.87
MFCC+LPCC	**47.20**	**86.13**

Table 1.5: Recognition rate (%) of 5 emotions (Anger & Sad & Joy & Fear & Disgust)

Recognition (%)	Linear SVM	Nonlinear SVM
MFCC	46.67	84.53
LPCC	45.60	85.87
MFCC+LPCC	**47.20**	**86.13**

Table 1.6: Recognition rate (%) of 6 emotions (Anger & Sad & Joy & Fear & Disgust & Boredom)

Recognition (%)	Linear SVM	Nonlinear SVM
MFCC	42.76	83.33
LPCC	41.23	82.24
MFCC+LPCC	**46.27**	**87.50**

Table 1.7: Recognition rate (%) of 7 emotions (Anger & Sad & Joy & Fear & Disgust & Boredom & Neutral)

Recognition (%)	Linear SVM	Nonlinear SVM
MFCC	35.70	84.30
LPCC	38.88	60.00
MFCC+LPCC	**40.00**	**88.04**

Table 1.8 Confusion Matrix of MFCC+LPCC feature extracted using RBF kernel SVM:

	Anger	Joy	Sadness	Fear	Disgust	Boredom	Neutral
Anger	94.30	0	0	4.06	0	0	1.62
Joy	0	90.47	0	9.52	0	0	0
Sadness	0	0	95.00	0	0	0	5.00
Fear	19.69	0	0	75.75	1.51	3.03	0
Disgust	0	0	0	4.87	95.12	0	0
Boredom	0	0	2.46	0	0	97.53	0
Neutral	3.89	0	1.29	0	0	0	94.80

Table 1.9 Confusion Matrix of MFCC+LPCC feature extracted using Linear SVM:

	Anger	Joy	Sadness	Fear	Disgust	Boredom	Neutral
Anger	79.67	0	4.06	2.43	0	9.75	4.06
Joy	68.85	0	8.19	6.55	0	14.75	1.63
Sadness	28.57	0	41.07	1.78	0	10.71	17.85
Fear	31.74	0	6.34	44.44	0	4.76	12.69
Disgust	64.10	0	15.38	10.25	5.12	0	5.12
Boredom	31.94	0	5.55	2.77	0	43.05	16.66
Neutral	32.85	0	2.85	7.14	1.42	10.00	45.71

VIII. Results For Savee Emotional Database

The SAVEE emotion speech database has been considered with 7 different emotions such as happiness, angry, sad, surprise, fear, disgust and neutral. This database has been recorded as a necessity for the improvement of an automatic emotion recognition system. This database comprises of recordings from 4 male actors in 7 different emotions. The 4 different speakers are Darren Cosker (DC - 01), James Edge (JE - 02), Joe Kilner (JK - 03) and Kevin Lithgow (KL -04). The following tables illustrate the emotion recognition rate for 7 different emotions.

Table 1.10: Emotion Recognition rate (%) of Speaker 01 – DC

Recognition (%)	Linear SVM	Nonlinear SVM
MFCC	59.17	91.67
LPCC	59.17	86.67
MFCC+LPCC	**73.33**	**92.50**

Table 1.11: Emotion Recognition rate (%) of Speaker 02 - JE

Recognition (%)	Linear SVM	Nonlinear SVM
MFCC	71.67	93.33
LPCC	76.67	78.33
MFCC+LPCC	50.83	88.33

Table 1.12: Emotion Recognition rate (%) of Speaker 03 - JK

Recognition (%)	Linear SVM	Nonlinear SVM
MFCC	51.67	77.50
LPCC	51.67	75.83
MFCC+LPCC	**62.50**	**81.67**

Table 1.13: Emotion Recognition rate (%) of Speaker 04 - KL

Recognition (%)	Linear SVM	Nonlinear SVM
MFCC	36.67	97.50
LPCC	51.67	80.83
MFCC+LPCC	35.00	84.17

The following table 1.14 illustrates the Confusion Matrix for all 7 emotions of SAVEE Database for Speaker 01 - DC using linear kernel.

Table 1.14: Confusion Matrix of MFCC + LPCC feature extracted for Speaker 01 - DC using Linear Kernel

Emotions	Anger	Joy	Sadness	Fear	Disgust	Boredom	Neutral
Anger	75	0	0	0	0	16.66	8.33
Joy	0	**100**	0	0	0	0	0
Sadness	0	0	0	0	0	11.11	88.88
Fear	6.66	0	0	85.71	0	0	14.28
Disgust	0	0	0	0	66.66	0	33.33
Boredom	0	0	0	0	6.66	93.33	0
Neutral	0	0	0	0	0	0	**100**

The following table 1.15 illustrates the Confusion Matrix for all 7 emotions of SAVEE Database for Speaker 01 - DC using Non Linear kernel.

Table 1.15: Confusion Matrix of MFCC + LPCC feature extracted for Speaker 01 - DC using Non Linear Kernel

	Anger	Joy	Sadness	Fear	Disgust	Boredom	Neutral
Anger	64.28	0	0	0	7.14	7.14	21.42
Joy	0	**100**	0	0	0	0	0
Sadness	0	0	**100**	0	0	0	0
Fear	0	0	0	**100**	0	0	0
Disgust	0	0	0	0	**100**	0	0
Boredom	0	0	0	0	0	**100**	0
Neutral	0	0	0	0	0	0	**100**

Table 1.16: Confusion Matrix of MFCC + LPCC feature extracted for Speaker 02 - JE using Linear Kernel

	Anger	**Joy**	**Sadness**	**Fear**	**Disgust**	**Boredom**	**Neutral**
Anger	25	0	0	0	0	0	75
Joy	0	78.57	0	0	0	0	21.42
Sadness	0	0	**100**	0	0	0	0
Fear	0	0	0	20	10	0	70
Disgust	0	10	40	0	**0**	0	50
Boredom	0	0	0	0	0	11.11	88.88
Neutral	0	0	0	0	0	0	**100**

Table 1.17: Confusion Matrix of MFCC + LPCC feature extracted for Speaker 02 - JE using Non Linear Kernel

	Anger	**Joy**	**Sadness**	**Fear**	**Disgust**	**Boredom**	**Neutral**
Anger	72.72	0	0	0	0	0	27.27
Joy	0	69.23	0	0	0	0	30.76
Sadness	0	0	**100**	0	0	0	0
Fear	0	0	0	**100**	0	0	0
Disgust	0	0	0	0	**100**	0	0
Boredom	0	0	0	0	0	**100**	0
Neutral	0	0	0	0	0	0	**100**

Table 1.18: Confusion Matrix of MFCC + LPCC feature extracted for Speaker 03 - JK using Linear Kernel

	Anger	**Joy**	**Sadness**	**Fear**	**Disgust**	**Boredom**	**Neutral**
Anger	76.92	0	0	0	0	0	23.07
Joy	13.33	80.00	0	0	0	0	6.66
Sadness	0	0	7.14	0	0	0	92.85
Fear	0	0	0	53.84	0	7.69	38.46
Disgust	6.66	0	0	0	20	0	80
Boredom	13.33	0	0	0	0	86.66	0
Neutral	0	0	0	0	0	0	**100**

Table 1.19: Confusion Matrix of MFCC + LPCC feature extracted for Speaker 03 - JK using Non Linear Kernel

Emotions	**Anger**	**Joy**	**Sadness**	**Fear**	**Disgust**	**Boredom**	**Neutral**
Anger	86.66	0	0	0	0	0	13.33
Joy	6.66	93.33	0	0	0	0	0
Sadness	0	0	**100**	0	0	0	0
Fear	0	0	0	64.28	0	0	35.17
Disgust	0	0	0	14.28	85.71	0	0
Boredom	0	0	0	0	0	38.46	61.53
Neutral	0	0	0	0	0	0	**100**

Table 1.20: Confusion Matrix of MFCC + LPCC feature extracted for Speaker 04 - KL using Linear Kernel

	Anger	**Joy**	**Sadness**	**Fear**	**Disgust**	**Boredom**	**Neutral**
Anger	0	0	0	0	0	0	100
Joy	0	100	0	0	0	0	0
Sadness	0	0	0	0	0	0	100
Fear	0	0	0	0	0	0	100
Disgust	0	0	0	0	14.28	0	85
Boredom	0	0	0	0	0	0	100
Neutral	0	0	0	0	0	0	**100**

Table 1.21: Confusion Matrix of MFCC + LPCC feature extracted for Speaker 04 - KL using Non Linear Kernel

Emotions	**Anger**	**Joy**	**Sadness**	**Fear**	**Disgust**	**Boredom**	**Neutral**
Anger	33.33	0	0	0	0	0	66.6
Joy	0	**100**	0	0	0	0	0
Sadness	0	0	**100**	0	0	0	0
Fear	0	0	0	**100**	0	0	0
Disgust	0	0	0	0	**100**	0	0
Boredom	0	0	0	0	0	**100**	0
Neutral	0	0	0	0	0	0	**100**

IX. Conclusion

In conclusion, experiments show that the highest accuracy rate for Speech emotion Recognition has been achieved using RBF kernel. The table noticeably depicts that emotion is highly recognized accurately where nonlinear SVM (RBF kernel) is used as compared to that of linear SVM. It provides 100 % accuracy when only 2 to 3 emotions are supplied to nonlinear SVM. But as an when the number of emotions goes on increasing, the accuracy level slowly and gradually goes down. In case of Linear SVM, even 2 emotions results in fair accuracy and when more number of emotions are added unto it, the recognition rate yet goes down to less than even 50 percent.

Accuracy of the developed model is presented using Confusion Matrix. The Recognition result of the combined MFCC and LPCC extracted features are compared with the isolated results. The maximum accuracy rate that reaches by using the combination of feature extraction method is **88.59%** for nonlinear SVM and **37.38 %** for linear SVM. From the above analysis, it is clear that the proposed system gives different accuracy level for both the databases. The system works fine for the Berlin Database as compared to SAVEE database. The only drawback of this system is that the Emotion Recognition rate of Linear SVM turns out to be too low. The future task that can be done is to improve the accuracy rate when using a linear SVM.

References

[1]. Igor Bisio, Alessandro Delfino, Fabio Lavagetto, Mario Marchese, And Andrea Sciarrone, "Gender-Driven Emotion Recognition Through Speech Signals For Ambient Intelligence Applications", Ieee Transactions On Emerging Topics In Computing, Digital Object Identifier 10.1109/Tetc.2013.2274797, 21 January 2014.

[2]. http://www.expressive-speech.net/, Berlin emotional Speech database

[3]. http://personal.ee.surrey.ac.uk/Personal/P.Jackson/SAVEE/, SAVEE Database

[4]. Nitin Thapliyal, Gargi Amoli, 'Speech based Emotion Recognition with Gaussian Mixture Model', International Journal of Advanced Research in Computer Engineering & Technology, Volume 1, Issue 5, July 2012, ISSN: 2278 1323.

[5]. Inma Mohino-Herranz, Roberto Gil-Pita, Sagrario Alonso-Diaz and Manuel Rosa-Zurera, "MFCC Based Enlargement Of The Training Set For Emotion Recognition In Speech", Signal & Image Processing : An International Journal (SIPIJ) Vol.5, No.1, February 2014.

[6]. Yixiong Pan, Peipei Shen and Liping Shen, "Speech Emotion Recognition Using Support Vector Machine", International Journal of Smart Home, Vol. 6, No. 2, April, 2012.

[7]. T.-L. Pao, Y.-T. Chen, J.-H. Yeh, P.-J. Li, "Mandarin emotional speech recognition based on SVM and NN", Proceedings of the 18th International Conference on Pattern Recognition (ICPR?06), vol. 1, pp. 1096-1100,September 2006.

[8]. Showkat Ahmad Dar, Zahid Khaki ' Emotion Recognition based on Audio Speech', IOSR Journal of Computer Engineering (IOSR-JCE)),e-ISSN: 2278-0661, p-ISSN: 2278-8727, Volume 11, Issue 6 (May - Jun 2013), PP 46-50.

[9]. Yaser Abu-Mostafa, "Radial Basis Function", Learning from data: Introductory Machine Learning, Hameetman Auditorium, Caltech, May 24, 2012.