# Speech emotion recognition using real time database

[1]Supriya Jagtap, [2]Dr. K.R. Desai

[1]P. G Scholar, [2]Prof. and H.O.D. Dept. of Electronics and Telecommunication Engineering,
Bharati Vidyapeeth's college of Engineering, Kolhapur, (India)

***Abstract*: People express their emotions in a different way but speech through emotion recognition is a very difficult task. This is a primary challenge of emotion recognition are choosing the speech corpora and then feature extraction of a speech signal. Emotion recognition or affect detection from the natural speech is a challenging task in the field of human and machine interaction. In this paper, the performance of the Support Vector Machine (SVM) has been evaluated for two types of databases – simulated (acted) and real-time (natural) speech corpora. In this paper, the work is carried out with a reduction in the size of the feature using Principle Component Analysis (PCA) technique and performance evaluation is done based on emotion classification accuracy.**

***Keywords*: Database, Feature Extraction, PCA, SVM Classifier.**

## I. Introduction

Emotions are very important in human life that shows different roles of every person and the mental state of a human's perspective/feelings. There are some different ways to express emotions. People express their emotions through facial expression, body pose, and oral communication. A human can quickly identify the exact intention of the speaker and capture this emotion through this different way in mind, but machines can't understand or feel this emotion of humans. This is a difficult and challenging task in the field of human and machine interaction. But only a few human-machine interfaces being enforced presently able to reach that [1]. Every human has a different style to express their emotions because every person has a different spoken style. In this paper, we are using two types of database to recognize the emotion namely simulated database and natural database.

In this paper, we focused on speech signals and this speech signal contains a lot of different data related to speech like information related to the speaker, message of that person, language and actual intention of the speaker. For instance, once someone is angry, his/her tone is loud, his expression becomes serious or unsmiling and the content of his/her speech no longer remains agreeable [2]. Similarly, when a speaker is sad, he/she speaks in a normal tone and the content of speech is medium and nervous. At the instance of happy, he speaks in a musical and louder tone and the content of his speech is rather pleasant and glowing, feels satisfied. Similarly, when a speaker is a fear, his/her voice in a soft and down tone and the content of speech is no longer and feels horrified. Exactly based on these observations, in this work we comparatively studying effect of speech to recognize the different emotions like angry, happy, sad and fear, etc. speech-based this emotion recognition system is useful in various fields like call center, intelligent toys, e-learning, Healthcare centers, stress management apps, lie detection etc.

In this paper, the basic four emotional states such as happy, sad, angry and fear are identified. In this work, we are using two different databases such as simulated and natural database and then identify the emotions by using a multiclass support vector machine (SVM). In a feature extraction technique, speech signals find the different parameters of speech like pitch features, energy-related features, formant frequency and Mel-frequency cepstrum coefficients (MFCC) is a spectral feature which was used for the emotion recognition system. The performance evaluation rates of both databases were observed. The remaining paper is organized as follows: Section two describes some previous related work in a literature review. Section three describes the proposed work of this system one by one in detail including the database for emotion recognition system using a speech database, the describe various extracted features that were used in the emotion classification. Then describes the Principal component analysis (PCA) and emotion classification by using Support Vector Machine. Experimental results obtained during this study were discussed in section four. Section five is provided the Conclusion of this paper [3].

## II. Literature Review

Under this point, the focus is on the literature that is available for speech and emotions. There are some people they have done their work on speech-based emotion recognition by using different feature extraction technique and classification algorithm. The speech emotion recognition system has used various processes and some papers have been introduced here.

Dario Bertero et al. [4] during this analysis, centered on Sentiment Recognition by using real-time speech emotion and interactive dialogues system. By using the CNN model feature extracted and this approach achieves an average accuracy of 65.7% on six emotions. Sentiment analysis with CNN also done and it has an 82.5F-measure when trained from out-of-domain data.

S.S. Agrawal [5] the main target of this paper is using the Hindi language natural database to recognize emotion based on prosodic parameters as the F0, A0 and duration and phonetic includes these parameters as the MFCC and their derivatives.

Mr. Vaijanath. V. Yerger and Dr. L. K. Ragha [6] focused on speech emotion recognition using Marathi language, and Features like Mel-frequency Cepstrum coefficients (MFCC), pitch detection, formant frequency, zero-crossing detection (ZCD), jitter and energy are extracted after preprocessing stage. The paper surveys work done by researchers on speech emotion in different languages and will try to conclude the approach for analyzing emotions in the Marathi language.

## III. Proposed work

The flow diagram of the emotion recognition system through speech considered in this study is illustrated in Figure 1. The emotion recognition system through speech is really challenging. The main issue in the evaluation of the Emotion recognition system through speech is two types of the database are used where one is simulated type and another is real-time (natural) database is used. It consists of emotional speech as input, then preprocessing is done. In this work proposed system is based on prosodic and spectral features of speech by using this method features are extracted, and then the classification of Emotional state using multiclass SVM classifier and detection of emotion as the output is done.

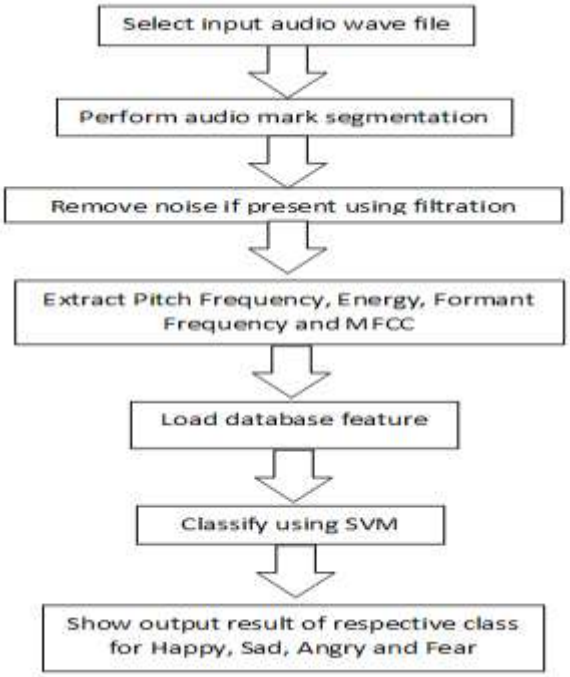



Fig 1. Shows flowchart of speech emotion recognition system

The emotional speech input to the system may contain the collection of the acted speech data and the real-world speech data. After collection of the database containing short Utterances of emotional speech sample which was considered as the training samples, proper and necessary features such as prosodic and spectral features were extracted from the speech signal. In this technique number of features, vectors are evaluated after the feature extraction. PCA algorithmic rule is used to optimize and reduce the dimensionality of the feature vector and select the best, strong feature vector. These feature values were provided to Support Vector Machine for the training of the classifiers. And remaining emotional speech samples presented to the classifier as test input. Then the classifier classifies the test sample into one of the emotions from the above mentioned four emotions and gives output as recognized emotion. [3]

### i. Database

The emotional database is divided into two types that is simulated database and real-time (natural) database. In this work, these two databases are used for emotion recognition techniques. In the simulated database is created by actor/ artist with emotional linguistic speech, an actor records his/ her speech in a recording room in certain emotions. Whereas in real databases, speech databases are obtained by recording conversations in real-life situations like in talk shows and call centers. But there is a difference between the features of acted and real emotional speeches. [7]

**A. Berlin database:** is a type of simulated database which is created by using the German language with 10 actors. It contains 800 sentences with 7(Happiness, anger, disgust, fear, sadness, surprise and neutral) emotions represented by 10 actors with 10 sentences. This standard Berlin Database of Emotional Speech (EMO-DB) is used in most of the researches for emotion recognition [8].

**B. Real-time database:** We have constructed our own database with 5 subjects for different emotions. The conditioning of the environmental is kept natural without considering any acoustic factor as like living room environment. The recording is done on Motorola mic without any gain control mechanism and also without any noise-canceling filters. Intentionally environmental conditions are not artificially arranged and are kept naturally. To evaluate the performance of the system in normal conditions. At the same time, the care has been taken that while recording speech other environmental sounds are sufficiently low in intensity and are not masking any speech signal. They're by keeping speech clarity sufficiently audible and understandable. In this database, we have used the English language.

**ii. Feature extraction**

Feature extraction is the most important stage of speech emotion recognition. This feature extraction involves the additional information of the speech signal and finds a number of variables called features. In this speech-based emotion recognition system, it is not clear which feature is most powerful for distinguishing the emotion. This technique reduces the computational complexity of the approach and finds a number of parameters from this audio signal. So, in this work, we are using four types of features namely Pitch Frequency, Formant Frequency, Energy and Mel frequency cepstrum coefficient (MFCC).

**Energy, Pitch and Related Features:**

Energy and Pitch are basic features of speech signals. To obtain the energy feature from a speech a short-term function is used to obtain the value of energy in each of the speech frames. This we can obtain by calculating mean value, local maxima, local minima, variance, the difference between local extreme and variance ranges to obtain the energy feature in the speech signal. [9]

**Mel Frequency Cepstrum Coefficient (MFCC):**

MFCC is one of the spectral features and is computed on the basis of human hearing ability. Mostly in speech synthesis for feature extraction, the MFCC method is used because of its low complexity and better ability to extract features. It is a very efficient technique and reduces noise from speech signals. Speech signals are continuously varying in a pre-emphasis high pass filter that is applied to increase signals energy. In framing this continuous signal are divided into N number of samples and this samples segmented into 20 to 40ms. Windowing is used to reduce the signals discontinuities at the start and end of each frame and those frames are shifted with 10ms span.

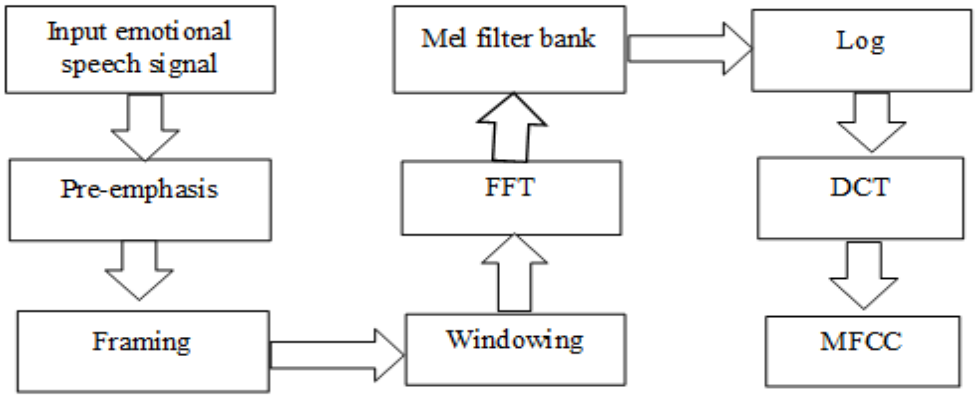



Figure 2. Block diagram for MFCC feature extraction

Fast Fourier Transform algorithm is used for converting this sample time domain to frequency domain. In this step of the Mel Scale Filter, it identifies how much energy exists in a particular frame and converts the frequency Hz into Mel Scale Frequency. And log energy computation, this is inspired by human hearing ability. A human does not listen to loud volume on linear scale. It gives those features for which humans can listen clearly.[12]

**Formant frequency:**

Formants are nothing but the spectral peaks of sound which is created from the human's vocal tract. It measures the amplitude peaks in the frequency spectrum of sound. In this work using LPC based formant frequency which estimate formants, remove their effects from the speech signal and estimate intensity and frequency of remaining buzz. And calculating mean value, local maxima, local minima, variance, the difference between local extreme and variance range to obtain the formant frequency feature in speech signal [13].

**iii. Principle Component Analysis (PCA)**

PCA collects a feature data like pitch related, energy related, MFCC related and formant related these are extracted from berlin and real-time database. During the training phase, lots of parameters/ features taken for the model building then generated model get confused at that time. PCA tries to reduce of overfitting problem and removes the irrelevant information. Then in this work overall system also ultimately reduces the processing speed.

**iv. Multiclass SVM**

SVM is the supervised machine learning algorithm which is mostly used for classification. Classifying data is a common task in machine learning. In this work, we are using multiclass SVM for classing four emotions like Happy, Sad, Angry and Fear. SVM is an easier and effective computation technique of machine learning algorithms, and under the conditions of limited training data, it is widely used for classification and pattern recognition issues. SVM provides better classification performance over the limited training data. It is one of the advantages of the SVM classifier.

The basic 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 [10][11]. In two-dimensional space, this hyperplane is

a line dividing a plane into two parts wherein each class lay on either side. The learning of the hyperplane in linear SVM is done by transforming the problem using some linear algebra. This is where the kernel plays a role. For linear kernel the equation for prediction for new input using the dot product between the input (x) and each support vector (xi) is calculated as follows:
F(x)= B (0) +sum (ai*(x, xi)) (4.5.1)

This is an equation that involves conniving the inner products of latest input vectors (x) with all support vectors in training information. The coefficient B0and ai (for each input) should be calculable from the training information by the learning rule. The support vector machine with kernel perform, within which input area is consisting of input samples regenerate into high dimensional feature space and thus input samples become linearly severable [3].

## v. Results

This is the results evaluated for the simulated database and real-time database with PCA algorithmic rule. This work evaluates the overall emotion recognition in terms of parameters like sensitivity, specificity, correct rate, error rate and accuracy of the system is presented. The recognized emotions are anger, fear, sad and happy. The results are reported in two states the first part shows the performance of the berlin database system with PCA and without PCA algorithmic rule. The second part of the result provides the performance of the real-time database system with and without the PCA rule. The experimental results highlight that the PCA based results allow incrementing accuracy of emotion recognition system on average. And as shown below the results of Before and after PCA Obtained values of accuracy tables followed by Graphs of corresponding tables.

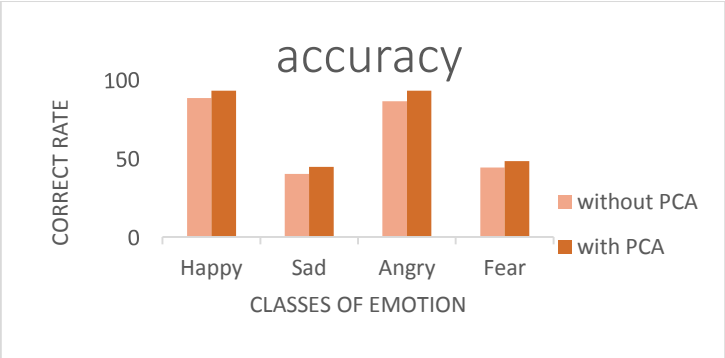
**Table: 1 Performance evaluation of berlin database for Emotion Recognition with and without PCA**

| parameters | Happy | | Sad | | Angry | | Fear | | Average | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Without PCA | With PCA | Without PCA | With PCA | Without PCA | With PCA | Without PCA | With PCA | Without PCA | With PCA |
| **Sensitivity** | 85 | 91 | 10 | 0 | 85 | 91 | 12 | 1 | 48 | 46 |
| **Specificity** | 82 | 92 | 44 | 58 | 82 | 91 | 66 | 55 | 68.50 | 74 |
| **Correct rate** | 81.19 | 92 | 43.59 | 44 | 61.19 | 74 | 39 | 42 | 56.24 | 63 |
| **Error rate** | 12 | 7 | 60 | 55 | 14 | 7 | 66 | 57 | 38 | 31.50 |
| **Accuracy (%)** | 88 | 92.59 | 40 | 44.44 | 86 | 92.59 | 44 | 48 | 64 | 68 |

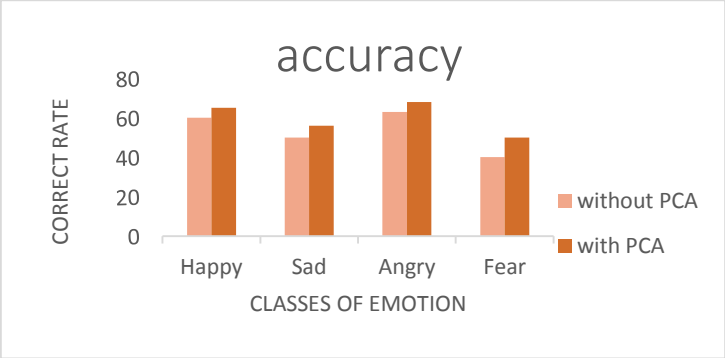
**Table: 2 Performance evaluation of real time database for Emotion Recognition with and without PCA**

| Parameters | Happy | | Sad | | Angry | | Fear | | Average | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Without PCA | With PCA | Without PCA | With PCA | Without PCA | With PCA | Without PCA | With PCA | Without PCA | With PCA |
| **Sensitivity** | 80 | 83 | 67 | 75 | 72 | 83 | 65 | 75 | 73.5 | 79 |
| **Specificity** | 40 | 38 | 40 | 38 | 50 | 43 | 17 | 33 | 37 | 38 |
| **Correct rate** | 57 | 67 | 40 | 54 | 67 | 71 | 38 | 43 | 51 | 59 |
| **Error rate** | 40 | 35 | 50 | 44 | 37 | 32 | 60 | 50 | 47 | 40 |
| **Accuracy in (%)** | **60** | **65** | **50** | **56** | **63** | **68** | **40** | **50** | **53** | **60** |

The bar graphs given shows number of emotions recognized with PCA and without PCA and with considering two database Berlin and Real- time database with accuracy in percentage.




**Graph1. Performance evaluation of Berlin speech database Recognize Emotion with PCA and without PCA with different parameters.**




**Graph2. Performance evaluation of Real-time speech database Recognize Emotion with PCA and without PCA with different parameters.**

## VI. Conclusion

In this paper, we have shown speech emotion recognition using a real-time database to compare with the standard database. The performance evaluation shows that the method of emotion recognition that we have used out Performance in terms of accuracy of emotion detection. It has also been found that during experimentation. If speech energy increases due to the change in distance between the source of speech and mic. Then estimated energy gets affected irrespective of emotion type. Hence our approach of a combination of various features is capable of decreasing such effects also system performs well on the English language database that we have generated.

## References

[1] S. G. Koolagudi, S. Maity, V. A. Kumar, S. Chakrabarti, and K. S. Rao, IITKGP-SESC : Speech Database for Emotion Analysis. Communications in Computer and Information Science, JIIT University, Noida, India: Springer, issn: 1865-0929 ed., August 17-19 2009.

[2] K. Sreenivasa Rao, Tummala Pavan Kumar, Kusam Anusha, Bathina Leela, Ingilela Bhavana and Singavarapu V.S.K. Gowtham, "Emotion Recognition from Speech", K. Sreenivasa Rao et al, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 3 (2) , 2012,3603-3607

[3] Akshay S. Utane, Dr. S. L. Nalbalwar, "Emotion Recognition through Speech Using Gaussian Mixture Model and Support Vector Machine", International Journal of Scientific & Engineering Research, Volume 4, Issue 5, May-2013, ISSN 2229-5518

[4] DarioBertero, FarhadBinSiddique ,Chien-ShengWu, YanWan, RickyHoYinChan and PascaleFung, "Real-TimeSpeechEmotionandSentimentRecognitionforInteractive DialogueSystems", Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, pages 1042–1047, Austin, Texas, November 1-5, 2016. c2016 Association for Computational Linguistics

[5] S.S. Agrawal, "EMOTIONS IN HINDI SPEECH- ANALYSIS, PERCEPTION AND RECOGNITION", IEEE2011

[6] Mr. Vaijanath. V. Yerigeri, Dr. L. K. Ragha, " Marathi speech emotion detection: A retrospective analysis", IEEE – 40222, 8th ICCCNT 2017 July 3-5, 2017, IIT Delhi, Delhi, India

[7] Monalisha Patro, Dr. Kanhu Charan Bhuyan, "EMOTION RECOGNITION FROM SPEECH SIGNAL USING SPECTRAL FEATURES", Proceedings of International Interdisciplinary Conference On Engineering Science & Management Held on 17th - 18th December 2016, in Goa, India. ISBN: 9788193137383

[8] Supriya B. Jagtap, Dr.K.R.Desai, Ms.J.K.patil, " A Survey on Speech Emotion Recognition Using MFCC and Different classifier", 8th National conference on emerging trends in engineering and technollogy(NCETET-2018),ISBN:978-93-87793-03-3

[9] Vinay, Shilpi Gupta, Anu Mehra, "Gender Specific Emotion Recognition Through Speech Signals", 2014 International Conference on Signal Processing and Integrated Networks (SPIN)
[10] Ashish B. Ingale, D. S. Chaudhari "Speech Emotion Recognition" International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-2, Issue-1, March 2012
[11] Rabiner L. R. and Juang, B., 'Fundamentals of Speech Recognition', Pearson Education Press, Singapore, 2nd edition, 2005.
[12] Ritu D. Shah, Dr. Anil C. Suthar, "speech emotion recognition based on SVM using MATLAB", International Journal of Innovative Research in Computer and Communication Engineering, Vol.4,Issue 3, March 2016, ISSN 2320-9798
[13] A.Khulage and Prof. B.V.Pathak, "Analysis of speech under stress using linear techniques and nonlinear techniques for emotion recognition system"