International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 05 | May 2020 www.irjet.net p-ISSN: 2395-0072

EMOTION DETECTION USING SPEECH SIGNAL

Abhirami S[1], Sheetal D Kumar[2], Akash Bansal[3], Ankit Kumar[4],Kavita Guddad[5]

[1]UG Student, Dept. of Electronics and Communication Engineering, Dayananda Sagar College of Engineering, Karnataka, India
[2]UG Student, Dept. of Electronics and Communication Engineering, Dayananda Sagar College of Engineering, Karnataka, India
[3]UG Student, Dept. of Electronics and Communication Engineering, Dayananda Sagar College of Engineering, Karnataka, India
[4]UG Student, Dept. of Electronics and Communication Engineering, Dayananda Sagar College of Engineering, Karnataka, India
[5]Asst. Professor, Dept. of Electronics and Communication Engineering, Dayananda Sagar College of Engineering, Karnataka, India

---

Abstract - *Emotion is the complex occurrence of consciousness, sensation, and behavior reflecting the personal significance of a human. Emotion detection plays a vital role in the man-machine interface. It can also be used for identifying the psychological state of a person in lie detection. In this project, we are using MATLAB for recognizing different emotions like happy, sad, and angry using human voice signal. We will be adopting the Support Vector Machine (SVM), and Deep neural network (DNN) algorithm for emotion detection and compare their efficiency.*

Key Words: **Emotion Detection, Mel-frequency cepstral coefficient(MFCC), Support Vector Machine(SVM), Deep Neural Network(DNN), Berlin Database, Our Database**

1. INTRODUCTION

Communication is the key to every human relationship. This not only includes the relationship between two human beings but also with the newly emerging technology. We use commands to interact with the computer. During the emerging period of computers, we used to type commands to interact with the computer. The technology has evolved over the period, so much so that we can give commands to our computers just by using our voice. Applications like Google Assistant, Siri, Alexa use voice recognition with the help artificial intelligence (AI) to accept commands from their user. If the system could analyse and understand a person's state of mind, then it can provide better suggestions to the user.

Speech is one of the most prevalent modes of communication amongst humans. The realization of this fact has inspired data analysts, researchers, and programmers to integrate this feature to the pre-existing technology for the man-machine interface. This means the machine must have an adequate understanding and database for interpreting human voice. There are has been a lot of research in this field.

Emotion detection is becoming more prevalent in modern technological developments. Calls made to large corporations use voice detection for diverting calls to their respective departments. Emotion detection is a progression made in the field of voice detection and recognition.

Emotion detection using speech involves extraction of features. Mel-frequency cepstral coefficient (MFCC) is being used for feature extraction in this project. Once feature extraction is done algorithms such as Support Vector Machine(SVM) and Deep Neural Network(DNN).

Support Vector Machine is a supervised machine learning algorithm which can be used for classification as well as regression challenges. Deep Neural Network is a type of artificial neural network which can exhibit a temporary dynamic behavior.

First, this project was implemented using the Berlin Database and later, we developed our database and used it for training and testing for three emotions (Angry, Happy and Sad). For the implementation of this project MATLAB 2014 is being used. For developing the database Praat Software is used. The minimum system requirement is Intel i32.1 GHz with a memory of 4GB and hard disk of 40GB.

2. LITERATURE SURVEY

The categorization of emotions has long been an ardent subject in different fields like affective science, psychology, and emotion research. Two traditional approaches are: categorical (termed discrete) and dimensional (termed continuous). Many theorists have conducted studies to determine which emotions are basic [1]. A most popular example is Ekman [2] who proposed a list of six basic emotions, which are anger, disgust, fear, happiness, sadness, and surprise. He emphasizes that each emotion serves as a discrete category rather than an individual emotional state. In the second approach, emotion is a combination of several

psychical dimensions and identified by axis. Wilhelm Max Wundt proposed in 1897 that emotions can be described by three dimensions: (1) strain versus relaxation, (2) pleasurable versus unpleasurable, and (3) arousing versus subduing[3].PAD emotional state model is a three-dimensional approach by Albert Mehrabian and James Russell. PAD stands for pleasure, arousal, and dominance. Another conventional dimensional model was suggested by James Russell in 1977. Unlike the earlier three-dimensional models, Russell's model features only two dimensions which include (1) arousal (or activation) and (2) valence (or evaluation) [3]. The categorical approach is generally used in SER [4].

Speech is a relevant communicational channel enriched with emotions: the voice in speech not only conveys a connotative message but also information about the emotional state of the person. Some significant voice feature vectors that have been chosen for research are fundamental frequency, MFCC, etc. Over the last years, an excessive investigation has been made to recognize emotions by using speech statistics.

3. BLOCK DIAGRAM AND WORKING PRINCIPLE

The block diagram shown in Fig -1 represents the working of our project. The steps followed in this project are: 1. Taking the input audio signal for a fixed period. 2. Pre-processing and feature extraction using MFCC is done. 3. Features are selected as per the requirements. 4. The database containing selected features are used for training using the following algorithms: a) Support Vector Machine (SVM) b) Deep Neural Network (DNN). 5. Once the training is done, we test using various database inputs and later using real-time input.

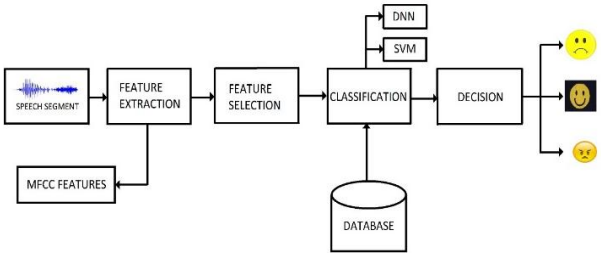

Fig - 1: Block Diagram

3.1 FEATURE EXTRACTION AND SELECTION

The first step to feature extraction is recording the speaker's voice. Many features can be extracted from a speech signal for emotion detection purpose. In this project, we are extracting Mel Frequency Cepstral Coefficient (MFCC) feature. The input audio signal is presented in Fig -2. Then we calculate the maximum and minimum frequency. The signal is then selected for a fixed duration as shown in Fig -3. Then we find the Fourier Transform using Hamming Window. The Fourier plot is shown in Fig -4. Cepstrum values are calculated and plotted as shown in Fig -5. Now the autocorrelation values are calculated and plotted as shown in Fig -6. Linear Predictive Coding(LPC) is used after resampling at 10,000/fs. Finally, the energy entropy is calculated and plotted as shown in Fig -7.

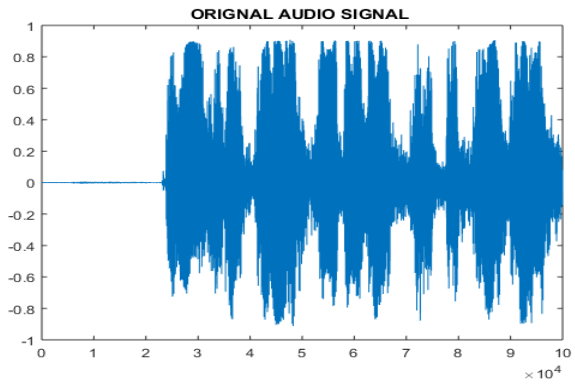

Fig -2: Original Audio Signal

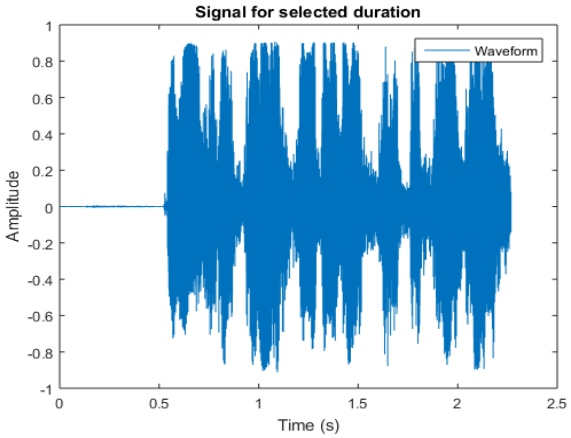

Fig -3: Signal for selected duration

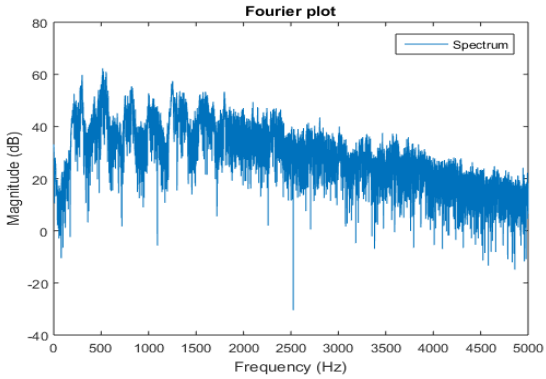

Fig -4: Fourier Plot of the signal

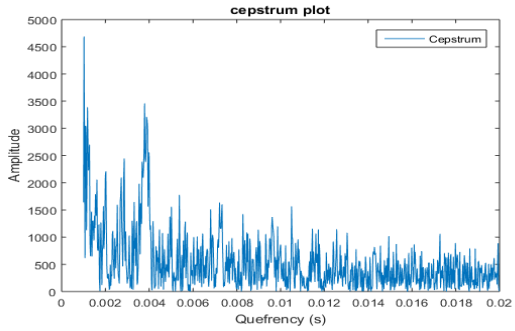

Fig -5: Cepstrum plot of the signal

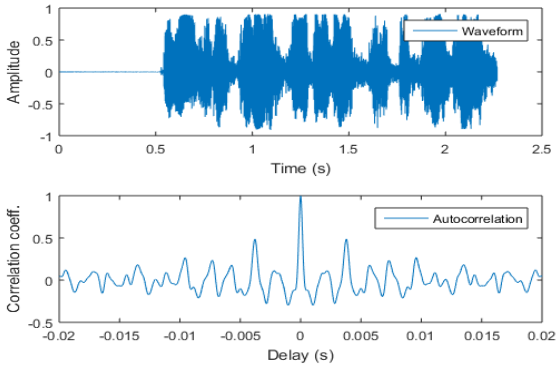

Fig -6: Signal and its autocorrelation

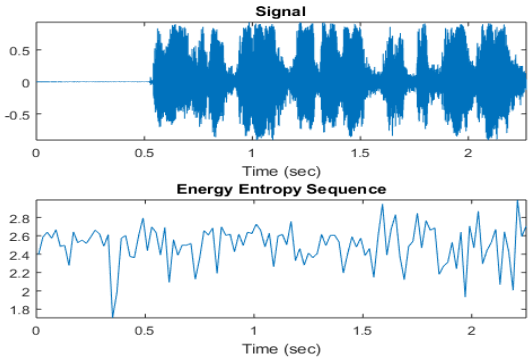

Fig -7: Signal and its energy entropy sequence

3.2 CLASSIFICATION

For classification, the signal is trained and tested using two algorithms:

3.2.1 SUPPORT VECTOR MACHINE

First, we train using Berlin Database and then using Our Database for three different kernels. The kernels used are Quadratic, Linear and Polynomial. But testing is done for the model using Quadratic kernel only.

The training was done using three kernels to observe the differences in plots. The plots for Angry emotion are shown in Chart -1. Similar plots were obtained for other emotions.

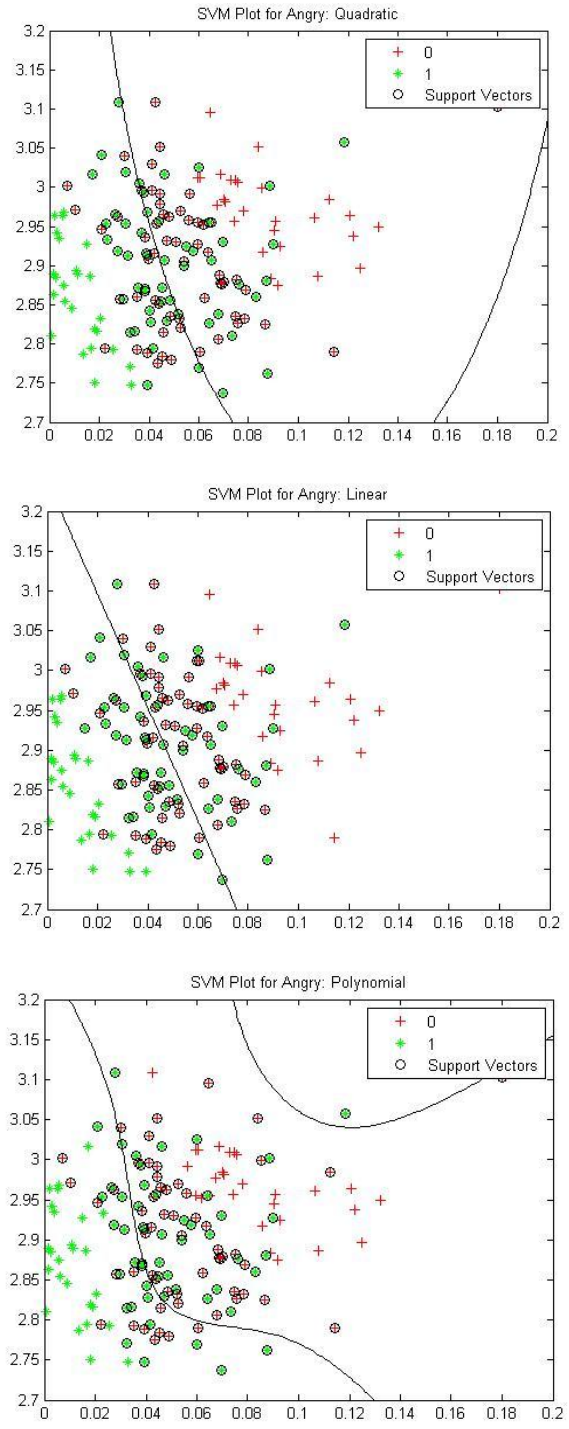

Chart -1: SVM plot for Angry Emotion Using
a) Quadratic Kernel b) Linear Kernel
c) Polynomial Kernel

3.2.2 DEEP NEURAL NETWORK

Like SVM, the training was done using Berlin and our database. Deep Neural Network Training has three main layers, Input Layer, Hidden Layer and Output Layer as shown in Fig - 8.

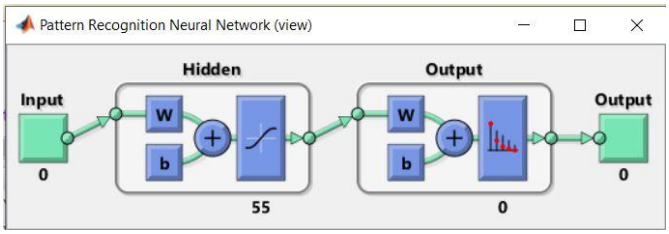

Fig -8: Model

The confusion matrix is a table to visualize the performance of the algorithm. The Confusion Matrix for DNN is shown in Table - 1. The error histogram is shown in Chart -2. The best Validation Point at 22 epochs as shown in Chart -3. The Receiver operating characteristic plot is shown in Chart -4.

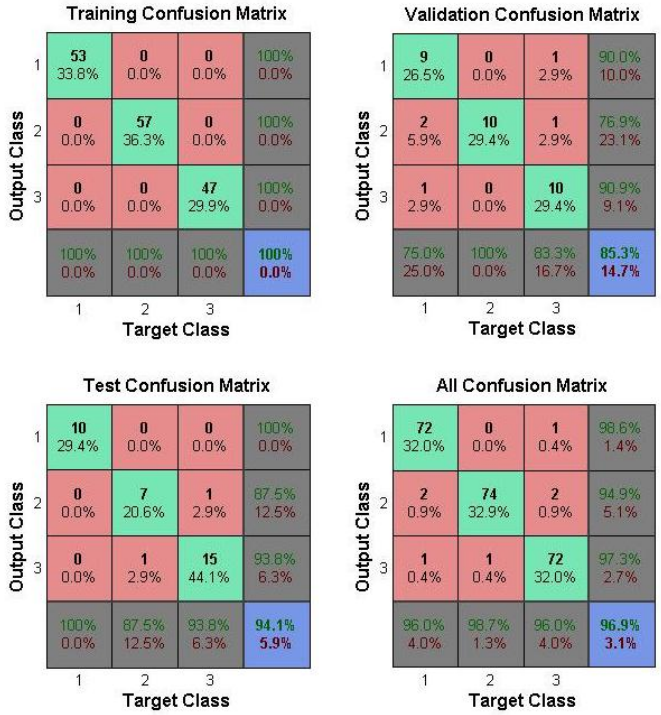

Table -1: Confusion Matrix

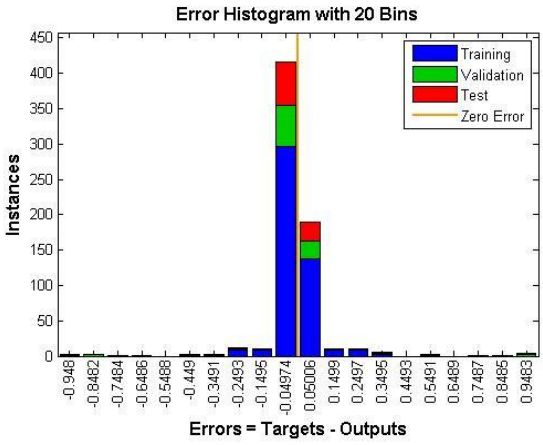

Chart -2: Error Histogram

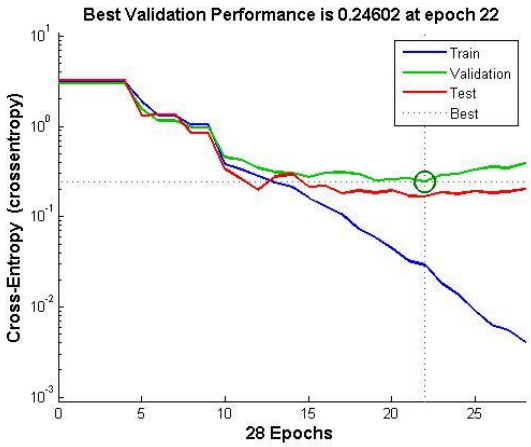

Chart -3: Best Validation Performance

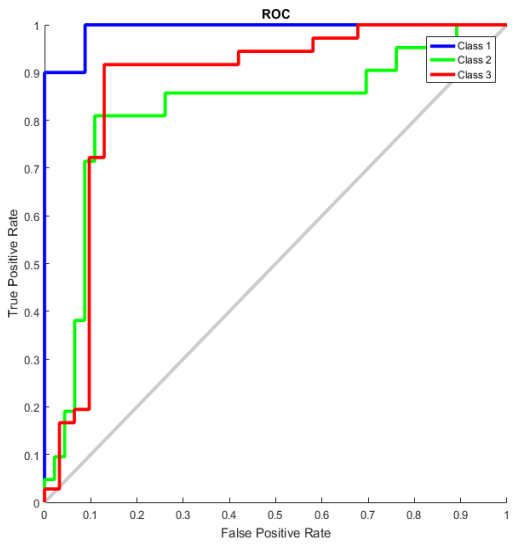

Chart -4: Receiver operating characteristic

4. RESULTS

Samples were taken for testing output for Happy, Sad, and Angry and determining the efficiency of each algorithm in both databases, i.e. for Berlin Database and Our Database.

4.1 BERLIN DATABASE:

The number of samples taken for training is 163 and for testing 30 clips has been taken. For each emotion 10 clips were selected for testing.

4.1.1 SVM:

Table -2: Confusion Matrix

Number of Samples= 30		PREDICTED		
		ANGRY	HAPPY	SAD
ACTUAL	ANGRY	7	0	3
	HAPPY	1	6	3
	SAD	2	2	6

4.1.2 DNN:

Table -3: Confusion Matrix

Number of Samples= 30		PREDICTED		
		ANGRY	HAPPY	SAD
ACTUAL	ANGRY	9	1	0
	HAPPY	4	2	4
	SAD	0	0	10

4.2 OUR DATABASE

The number of samples taken for training is 225 and for testing 60 clips has been taken. For each emotion 20 clips were selected for testing.

4.2.1 SVM

Table -4: Number Samples Predicted Correctly using SVM

EMOTION	NUMBER SAMPLES PREDICTED CORRECTLY OUT OF 20 SAMPLES
ANGRY	8
HAPPY	16
SAD	19

4.2.2 DNN

Table -5: Confusion Matrix

Number of Samples= 60		PREDICTED		
		ANGRY	HAPPY	SAD
ACTUAL	ANGRY	7	8	5
	HAPPY	9	11	0
	SAD	0	0	20

4.3 COMPARISON

The efficiency of SVM and DNN for both databases is shown in the table below.

Table 6: Efficiency of SVM and DNN for both databases

	SVM	DNN
BERLIN DATABASE	63.33%	70%
OUR DATABASE	71.66%	63.33%

The efficiency is affected by various parameters like the quality of audio clips in the database, sampling rate of audio clips, presence background noise in the clips, size of database per class, etc. We observe that DNN has higher efficiency than SVM for Berlin database. Berlin database is prepared in an ideal environment for research purpose. In case of Our database, DNN has a lower efficiency than SVM. But it is observed that the SVM model predicts multiple outputs. Hence, DNN algorithm will be a better choice for this application. Another point to be noted is Our Database is not created in an ideal isolated environment unlike the Berlin Database and hence results will be affected by the noise.

5. APPLICATIONS

Implementation of emotion detection on machines will take human-machine interaction to the next level. For the daily improving technologies field like Alexa, Siri, etc. this addition will enrich their features. Humans need psychological support in their lives, if the machines can detect the emotion and interact with human giving psychological assistance then the machines can act as virtual companions. There are a variety of applications for emotion detection like Improves man-machine interface. It is used to monitor the psycho-physiological state of a person. It can be used in a lie detector. It also finds its application in medicine and forensics, interfaces with robots, Audio surveillance, Web-based E-learning, Commercial applications, Clinical studies, Entertainment, Banking, Call centers, and Computer games.

6. CONCLUSION & FUTURE SCOPE

The proposed model of Emotion Detection Using Speech Signal with Real-Time input signal for which we have used two algorithms (SVM and DNN) for classification was implemented using MATLAB. The efficiency is affected by various parameters like the quality of audio clips in the database, sampling rate of audio clips, presence background noise in the clips, size of database per class, etc. Improving the database quality would help in improving the efficiency of the models. Future scope for this project would be implementing it using various other to identify the best-fit algorithm for this application. Increasing the size of the database is also a factor for improving efficiency.

Recent discoveries in the field of neurosciences and physio-psychology, together with the extensions of notions like emotional intelligence and multilevel intelligence, has led to the new framework called “affective computing”, according to which, the main objective of this framework is to build machines that recognize, express, model, communicate and respond to users emotion indicators. In future, the model can be optimized by adding more appropriate samples and increasing the number of samples per class to get better efficiency.

REFERENCES

[1] L. Kerkeni, Y. Serrestou, M. Mbarki, K. Raoof and M. A. Mahjoub, "A review on speech emotion recognition: Case of pedagogical interaction in classroom," 2017 International Conference on Advanced Technologies for Signal and Image Processing (ATSIP), Fez, 2017, pp. 1-7, doi: 10.1109/ATSIP.2017.8075575.

[2] Ekman, P. (1992). An argument for basic emotions. *Cognition and Emotion, 6*(3-4), 169–200. https://doi.org/10.1080/02699939208411068

[3] Matilda's. Emotion recognition: A survey. International Journal of Advanced Computer Research. 2015;3(1):14-19

[4] Koolagudi SG, Rao KS. Emotion recognition from speech: A review. International Journal of Speech Technology. 2012;15(2):99-117

[5] Ali H, Hariharan M, Yaacob S, Adom AH. Facial emotion recognition using empirical mode decomposition. Expert Systems with Applications. 2015.

[6] Liu ZT, Wu M, Cao WH, Mao JW, Xu JP, Tan GZ. Speech emotion recognition based on feature selection and extreme learning machine decision tree. Neurocomputing. 2018.

[7] Ragot M, Martin N, Em S, Pallamin N, Diverrez JM. Emotion recognition using physiological signals: Laboratory vs. wearable sensors. In: International Conference on Applied Human Factors and Ergonomics. Springer; 2017.

[8] Surabhi V, Saurabh M. Speech emotion recognition: A review. International Research Journal of Engineering and Technology (IRJET). 2016.