# Speech Emotion Recognition Using Recurrent Neural Network

D.P.C.L. Gayathri[1], M. Sushmitha[2], G. Durga Prasad[3], M. Tharun Jamal Kumar[4],
Dr.CH. Surya Kiran[5]
[1, 2, 3,4]Student, Department of CSE, NRI Institute of Technology, Vijayawada, A.P., India.
[5]Professor, Department of CSE, NRI Institute of Technology, Vijayawada, A.P., India.

***Abstract***

Speech Emotion Recognition is a project that identifies the emotion of a person based on his/her voice. This project is based on the Recurrent Neural Network(RNN), which uses different modules for emotion recognition. The classifiers are used to identify various emotions namely happiness, anger, sadness, disgust, fear, neutral state and surprise.

In this there will be a recorded voice of a person and our system identifies the emotion from that recorded audio. Various features are extracted from the voice using the LIBROSA package of language python. Our system uses the same phenomenon that animals like horses and dogs use to understand human emotion. It analyzes the audio files in WAV format and returns the intended outcome i.e. the emotion.

## *I. Introduction*

This project is an act of analyzing the emotion of a person based on their speech. Emotional state of a person influences the interaction with another person. Since speech is one of the important expressions of emotion, our aim is to develop a system which predicts the emotion based on the speech. There are various applications of Speech Emotion Recognition. The emotion of a person plays an important role in decision making.

The relation between people mainly depends upon emotions. There are many expressions to exhibit human emotion. One of those expressions is speech. Our system SER detects emotion based on speech.

Our system takes the speech and extracts various features such as tone, pitch, frequency etc. These features are extracted by using a package of python named LIBROSA. Later, by using a deep learning algorithm namely Recurrent Neural Network those features are classified into various emotions. We are taking a dataset which consists of a number of speeches of various emotions to train our model.

This system SER helps in recognizing the emotion of a person and act accordingly. This system is used in various fields such as call centers etc.

## *II. Existing System*

The existing system is identifying the emotions of a person based on the text. It is being used in some social media apps to identify the emotion of the person based on comments they make.

Since the emotion is identified through text, the accuracy is low. This system identifies some particular words from the text and analyzes the emotion. Apps like facebook, twitter use text based sentiment analysis to identify the emotion of the writer and make computational upgrades to their respective apps.

## *III. Propose System*

We are using an advanced deep learning model, Recurrent Neural Network(RNN). This RNN enables it to capture the frequencies of speech in an efficient way. When speech is collected in real form, it may contain lots of noise. So, we perform a noise reduction phase. The speech exhibiting a maximum number of features regarding an emotion, then that emotion is our intended outcome. Our proposed system has the advantages of high accuracy, high computational processing.

The most efficient way of analyzing speech signals is converting them to 2D spectrograms. The commonly used method for this is Time-frequency analysis. After the preprocessing of the audio signals, the audio signal is converted to 2D using Short Time Fourier Transform. This 2D audio is analyzed through RNN. By traversing through sequentially constructed networks, the result is produced based on the sum of probabilities.

- **Recurrent Neural Network:** Our project needs sequential data information to predict

the emotion. RNN is a deep learning model that is capable of dealing with sequential data. The other neural networks such as CNN deal with the data individually and all the input signals should be independent to each other. The traditional neural networks use different parameters in each layer, but RNN shares the parameters within all the sequential steps. Since our project deals with sequential data we use the Recurrent Neural Network. The system architecture of SER is as follows:

There are various steps involved in Speech Emotion Recognition. A speech input is taken from the database. The audio is processed through librosa, which extracts the features from the audio. After feature extraction, the features are classified using the Recurrent Neural Network. After classification the emotion of the audio is displayed, which is the intended outcome. The steps involved in Speech Emotion Recognition are as follows:

### *IV. Implementation (Modules)*

In this SER project, we use the libraries like sklearn, soundfile, and librosa to build a model using an MLPClassifier. To implement the deep learning model we are using python module keras. This enables our system to recognize emotions from audio files. We load the audio files, extract features from it. After that dataset is split into training and testing sets .Then, MLPClassifier will be initialized and train the model. Finally, we'll calculate the accuracy of our model which provides the intended outcome.

### *V. Sample Screens*

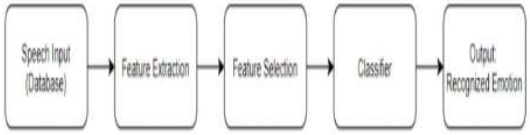
```
"""
This file can be used to try a live prediction.
"""

import keras
import numpy as np
import librosa

class livePredictions:
    """
    Main class of the application.
    """

    def __init__(self, path, file):
        """
        Init method is used to initialize the main parameters.
        """
        self.path = path
        self.file = file

    def load_model(self):
        """
        Method to load the chosen model.
        :param path: path to your h5 model.
        :return: summary of the model with the .summary() function.
        """
        model=keras.models.load_model(self.path)
        self.loaded_model = keras.models.load_model(self.path)
        #return self.loaded_model.summary()
```



### *VIII.* References

[1] https://www.analyticsinsight.net/speech-emotion-recognition-ser-through-machine-learning/

[2] http://scholarworks.sjsu.edu/cgi/viewcontent

.cgi?article=1647&context=etd_projects

[3] https://core.ac.uk/download/pdf/15986611.pdf

[4] https://data-flair.training/blogs/python-mini-project-speech-emotion-recognition/

[5] https://en.wikipedia.org/wiki/Recurrent_neural_network

[6] https://www.researchgate.net/publication/299185942_Human_speech_emotion_recognition

[7] https://core.ac.uk/download/pdf/15986611.pdf