International Conference - Innovation-2021-Innovation-2021
International Journal of Scientific Research in Computer Science, Engineering and Information Technology ISSN : 2456-3307 (www.ijsrcseit.com)
Volume 8, Issue 3, May-June-2021

Speech Emotion Recognition Based Patient Feedback for Hospitals

Rutuja Patil[1], Siddhi Salunke[1], Pournima Ubale[1], Mayur Talole[1], Prof. Ajita Mahapadi[2]
[1]UG Student, Department of Computer Engineering, Dr DY Patil School of Engineering, Pune, Maharashtra, India
[2]Assistant Professor, Department of Computer Engineering, Dr DY Patil School of Engineering, Pune, Maharashtra, India

ABSTRACT

This paper presents design and implementation of a patient feedback system based on speech emotion recognition (SER) for hospital purpose. Reviews are recorded through the microphone and based on that emotions are generated. The proposed system is implemented using Speech Features and Speech Transcriptions, which include Spectrogram, Mel-Frequency Cepstral Coefficients (MFCC) and TextBlob. Using this features and transcriptions different aspects of emotions are detected.

Keywords : Speech Emotion Recognition, SER, Speech Transcriptions, Speech Features

I. INTRODUCTION

Feedback is an event that occurs when the output of a system is used as input back into the system as part of a chain a of cause and effect. Feedback plays vital role for in almost every sector of an industry, which helps in adopting new knowledge and prevents any mistake. Since we know that traditional feedbacks are given manually through forms and online reviews, which can be very time consuming and irritating to fill out. So in today's world everything is just a few clicks away, so the people have become very impatient when it comes to giving feedbacks, so we have implemented a module which allows us to record a review through microphone which is when SER comes into the picture.

Speech Emotion Recognition (SER) systems can be defined as a collection of methodologies in which speech signals are processed and classified to detect the sentiments. The process of extracting of emotional state of the speaker from their speech is SER. SER is used in variety of applications such as caller agent conversation analysis, interactive voice based-assistant. For this we use speech features which are to be extracted.

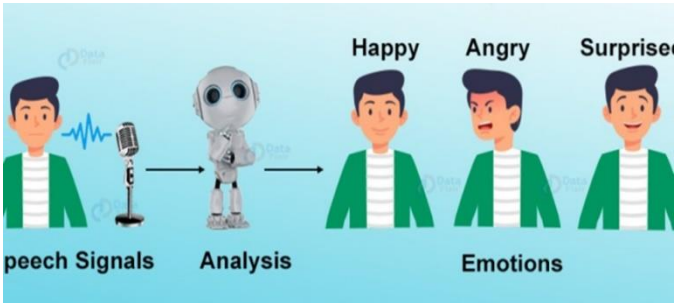

Fig1.1 Speech Emotion Recognition (SER)

Speech features are two types of that is **The temporal features** (time domain features) and **The spectral features** (frequency based features), that can be

obtained by converting the signals which are time based into the frequency domain using the Fourier Transform which are used to identify rhythm, pitch, notes etc. So a visual representation of this features is called as spectrogram.

A spectrogram displays signal strength over time at the various frequencies present in a waveform. Spectrograms can be two-dimensional graphs with a third variable represented by colour, or three-dimensional graphs with a fourth colour variable. When the audio file has been applied the spectrogram, then it is also called voiceprints, voicegrams or sonograms.

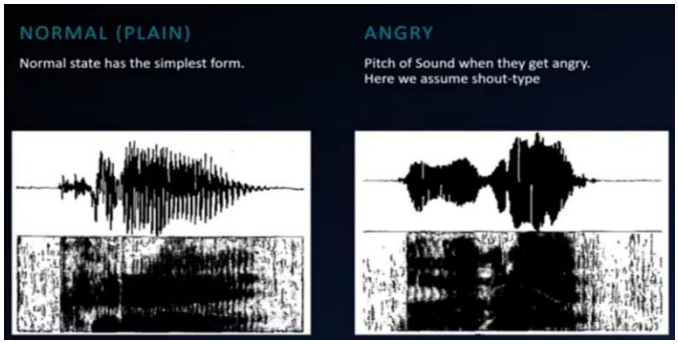

Fig 1.2 Normal, Angry Spectograms

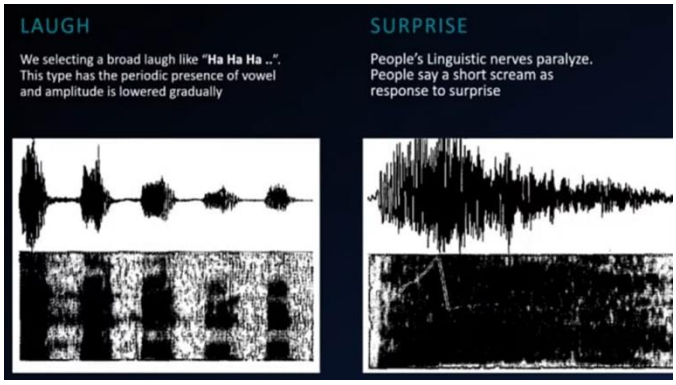

Fig 1.3 Laugh, Surprise Spectograms

The above pictures display's the different spectrogram for the different emotions such as normal, angry, laugh and surprise. The obtained Spectrogram magnitudes are then mapped to the Mel-scale to get Mel-spectrograms.

From the Audio data we have extracted three key features which have been used are MFCC (Mel Frequency Cepstral Coefficients), Mel Spectrogram and Chroma. coefficients that collectively make up an MFC. They are derived from a type of cepstral representation of the audio clip (a nonlinear "spectrum-of-a-spectrum"). MFFC are commonly used as features in speech recognition system, such as the system which can automatically recognize the voice spoken through the microphone into the system.

Speech Transcriptions are also used along with speech features. In Speech transcription the audio file is transcripted from the spoken audio into the text and it returns a block of text for each portion of the audio which is transcribed. For this TextBlob has been used.

TextBlob is a Python library for processing textual data. It provides a simple API for diving into common natural language processing (NLP) task such as part-of-speech tagging, translations noun phrase extraction, sentiment analysis, classifications, sentiment analysis.

II. METHODOLOGY

The design of the Feedback System for patients using Speech Emotion Recognition involves the incorporation of the following steps:

1) Initially the speaker uses the microphone for giving the review using the record a review button.
2) The voice is recorded as a wav file and then analyzed and the preprocessing part is done. Preprocessing can include normalization, noise removal, cleaning of the audio file.
3) After preprocessing all the speech features and transcriptions are extracted such as Pitch, tone, spectrograms and passed to the classifier.
4) From the Training Dictionary also the speech features and transcriptions are extracted and sent to the classifier. We have used RAVDESS dataset.

5) After receiving the the features and transcription the MLP classifier classifies the emotions and TextBlob generates the transcriptions.
6) Based on the results the accuracy score and emotions are generated.
7) Generated emotions are mapped into a review which can be either Excellent, Satisfied, Not satisfied, Needs Improvement.

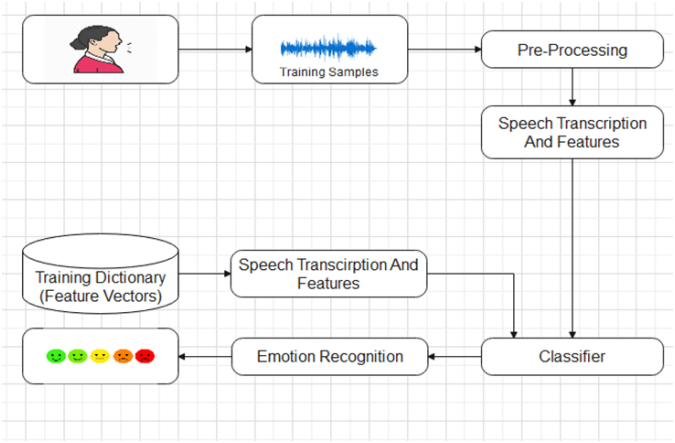

Fig 2.1 Block Diagram of the Total System

III.DESIGN AND IMPLEMENTATION OF THE PROPOSED SYSTEM

The Python implementation of Librosa package, Soundfile package, numpy, Scikit were used.

Librosa is a python package for music and audio analysis. It provides the building blocks necessary to create music information retrieval systems. **PyAudio** is a set of Python bindings for PortAudio, a cross-platform C++ library interfacing with audio drivers. Soundfile is an audio library based on libsndfile, CFFI and NumPy. SoundFile can read and write sound files.Scikit learn is a free software machine learning library for the Python programming language.

We have used **Multi-layer Perceptron (MLP) Classifier to classify the emotions from the given wave signal, which makes the choice of learning rate to be adaptive.**

Multi-layer Perceptron (MLP) Classifier is a neural network algorithm which is suitable for classification prediction problem where inputs are assigned a class as a label.

A. Module I

1 Start the microphone for recording the review
2 All the preprocessing of the recorded review will be done such as cleansing of the audio file will be done, silence removal, noise cancellation, handling the missing data.
3 After this the framing of audio file is done and converted into a standardized format.
4 Now the file is ready for further processing.
5 Stop

B. Module II

Module II consist of TextBlob.

1. Start the sentiment analysis using TextBlob sentiment analyser, in this Naïve Bayes analyser predicts the sentiment.
2. In return it will generate positive and negative scores.
3. Based on this score we will predict the emotions.
4. Stop

C. Module III

Module III consist of Mel-Frequency Cepstral Coefficients (MFCC)

1 Start
2 Import libraries such as sklearn, librosa, numpy, soundfile.
3 Load the data from dataset (RAVDESS) and extract the features.
4 Train the data on the MLP Classifier
5 Test the data using model.fit().
6 calculate the accuracy and predict the emotions
7 Stop

We compare the module I and module II and based in that result and score emotions are detected.

D. RAVDESS dataset

(The Ryerson Audio-Visual Database of Emotional Speech and Song)

The RAVDESS dataset is being used which is in English language.

It contains 7,356 files (total size: 24.8 GB) out of which speech audio-only files are 1440.

There are 24 Actors: 12 male and 12 female. 7 different emotion classes (calm, happy, sad, angry, fearful, surprise, and disgust expressions).

IV. EXPERIMENTAL SETUP

System Feature 1:

In our system the first action is performed by microphone, microphone will record the voice of the patient who is giving the feedback about the hospital. The recorded voice will go through CNN algorithm in order to find attributes.

System Features 2:

In our system the second action is performed by SER is extracting speech features and speech transcriptions. It is then compared with the features of trained dataset.

System Feature 3:

After Comparison with the trained dataset, the best fitted emotion is generated.

User Interface:

In our system user interface will be a webpage where patients will record their feedback using inbuilt microphone and they can see their ratings and reviews with the help of monitor, by using html, CSS, JavaScript a webpage is designed.

Hardware Interface:

Hardware interfaces is an inbuilt microphone to record the feedback given by the patients, a monitor to display the ratings and reviews.

Software Interface:

Implementation is done using python programming language.

Webpage is developed using html, CSS, JavaScript.

V. RESULT AND DISCUSSION

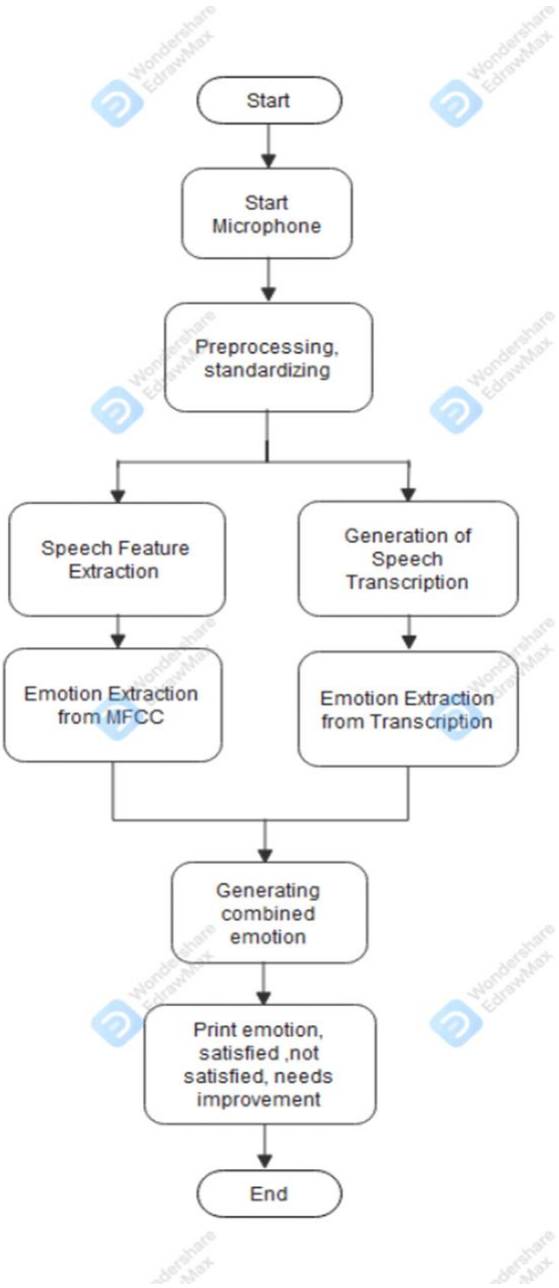

Fig 5.1 Flowchart Of The System

Fig 5.2 Result of the emotion generation through voice

VI. CONCLUSION

This System therefore helps us to record a review with the help of the Speech Emotion Recognition, which indeed reduces the times and effort needed to give a feedback. Feedbacks have always been ignored or neglected, because it's an annoying process. Therefore with the help of this system, feedback can be given in just one click and helps us from the lengthy process of filling out forms or typing a review.

VII. REFERENCES

[1]. Suraj Tripathi1, Abhay Kumar1*, Abhiram Ramesh1*, Chirag Singh1*, Promod Yenigalla1, "Deep Learning based Emotion Recognition System Using Speech Features and Transcriptions, Samsung R&D Institute India – Bangalore , arXiv.org, 2019.

[2]. Nithya Roopa S., Prabhakaran M, Betty.P, Nov 2018. "Speech Emotion Recognition using Deep Learning". International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-7 Issue-4S, November 2018

[3]. k Ashok Kumar, J L Mazher Iqbal . "Machine Learning Based Emotion Recognition using Speech Signal". International Journal of Engineering and Advanced Technology (IJEAT) ISSN: 2249 – 8958, Volume-9 Issue-1S5, December, 2019

[4]. https://www.geeksforgeeks.org/speech-recognition-in-python-using-google-speech-api/

[5]. https://www.tutorialspoint.com/artificial_intelligence_with_python/artificial_intelligence_with_python_speech_recognition.html