HMR Institute of Technology and Management
Plot No. 326, Hamidpur, Delhi-110036

Synopsis of Minor Project

Date:

Minor Project Title: **Speech Emotion Recognition**

Name of Guide(s):

Programme- **B. Tech (CSE)**		Year/Semester- **7th Sem**	
S.No.	Enrollment No.	Name	Signature
1.	41713302717	Kanishk Gupta	
2.	41013302717	Parth	
3.	41113302717	Abhishek Sharma	

Signature of student

Signature of Guide(s)

Signature of Proctor:
Name:

Approval by Board of Faculty

Member	Signature	Remark (Approved/Not Approved)

Contents

1. Abstract

Speech Emotion Recognition is a technology that extracts emotional features from speech signals by computer and contrasts and analyses the characteristic parameters and the emotional change acquired. Many systems have been developed to identify the emotions from the speech signal. Emotion recognition is based on the technologies which use different classifiers for emotion recognition is reviewed. The classifiers differentiate emotions such as anger, happiness, sadness, surprise, neutral state, etc. The dataset for the speech emotion recognition system is the emotional speech samples and the features extracted from these speech samples. There are many voice products developed like Amazon Alex, Google Home, Apple HomePod, which functions mainly on voice-based commands. It is evident that voice will be the better medium for communicating with the machines.

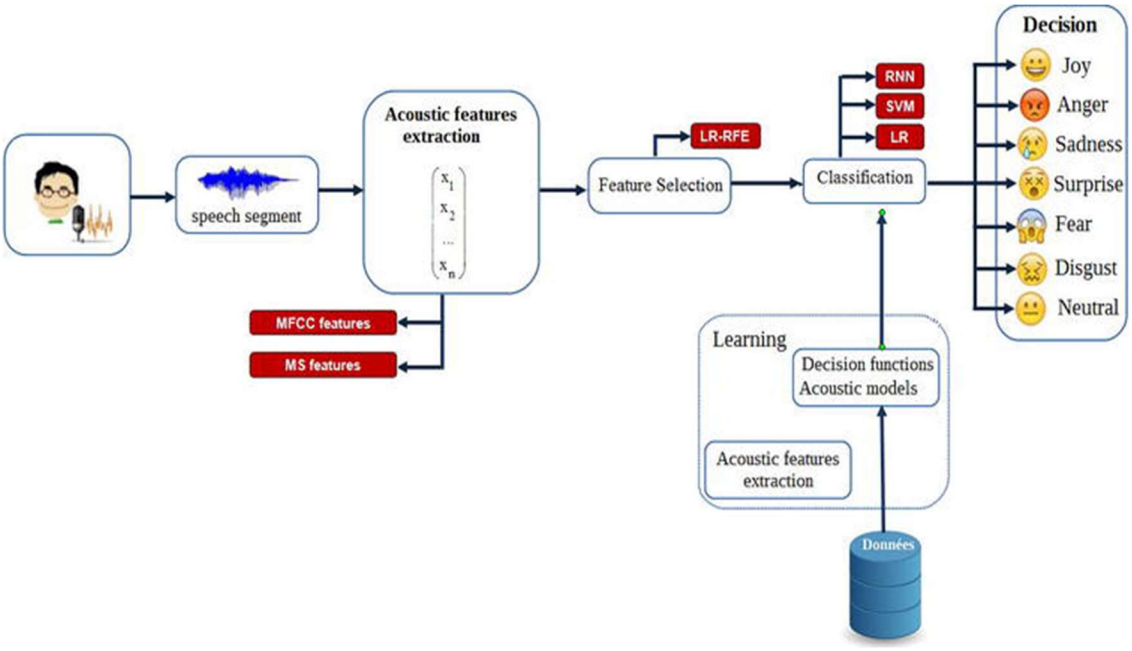
2. Introduction

Emotion plays a significant role in daily interpersonal human interactions. This is essential to our rational as well as intelligent decisions. It helps us match and understand others' feelings by conveying our feelings and giving feedback to others. Research has revealed the powerful role that emotion play in shaping human social interaction. Emotional displays convey considerable information about the mental state of an individual. In prior studies, several modalities have been explored to recognize the emotional states such as facial expressions, speech, physiological signals, etc. Several inherent advantages make speech signals a good source for affective computing.

SER aims to recognize the underlying emotional state of a speaker from her voice. The area has received increasing research interest all through current years. There are many applications of detecting the emotion of the persons like in the interface with robots, audio surveillance, web-based E-learning, commercial applications, clinical studies, entertainment, banking, call centres, cardboard systems, computer games, etc. For classroom orchestration or E-learning, information about students' emotional state can help enhance teaching quality.

Three key issues need to be addressed for successful SER system, namely,

1. choice of a good emotional speech database,
2. extracting effective features
3. designing reliable classifiers using machine learning algorithms.



3. Objectives of Project:

To build an end-to-end hardware-software solution that extracts features from human voice such as pitch, loudness, spectrum, and speech rate that contain crucial information on the emotional state of the speaker. And, Recognizes the emotions of a person based on his voice in real-time and provide a suitable response.

Solution Overview

There are two types of features in a speech namely, the lexical features (the vocabulary used) and the acoustic features (sound properties like pitch, tone, jitter, etc.).

The problem of speech emotion recognition can be solved by analyzing one or more of these features. Choosing to follow the lexical features would require a transcript of the speech which would further require an additional step of text extraction from speech if one wants to predict emotions from real-time audio. The analysis of the acoustic features can be done in real-time while the conversation is taking place as we'd just need the audio data for accomplishing our task. Hence, we choose to analyze the acoustic features in this work.

Furthermore, the representation of emotions can be done in two ways:

- Discrete Classification: Classifying emotions in discrete labels like anger, happiness, boredom, etc.
- Dimensional Representation: Representing emotions with dimensions such as Valence (on a negative to positive scale), Activation, Energy (on a low to high scale) and Dominance (on an active to passive scale)

We will be following the Discrete Classification for this project.

ARDUINO APPROACH:-

Arduino UNO will be used to give the user a suitable response according to the emotion detected from the speech. We will be using Arduino for the result generation of our SER model. Arduino will be interfaced with an LCD to display the result. The display units are very important in communication between the human world and the machine world.
The display unit works on the same principle, it does not depend on the size of the display; it may be big or the small. We are working with the simple display 16×2 unit. The liquid crystal display uses the property of light monitoring of liquid crystal and they do not emit the light directly.

LCDs are also used in embedded system applications for displaying various parameters interfaced with Arduino UNO. LCD 16x2 is a 16-pin device that has 2 rows that can accommodate 16 characters each. It can be used in 4-bit mode or 8-bit mode and also possible to create custom characters. The type of emotion detected will be displayed by this 16x2 LCD display by communicating with Serial port.

4. Tools/Materials

- **JupyterLab:** JupyterLab is an open-source, web-based UI for Project Jupyter and it has all the basic functionalities of the Jupyter Notebook, like notebooks, terminals, text editors, file browsers, rich outputs, and more. However, it also provides improved support for third party extensions.

- **Pandas:** Pandas is a fast, powerful, flexible and easy to use open-source data analysis and manipulation tool, built on top of the Python programming language.

- **TensorFlow:** TensorFlow is an end-to-end open-source platform for machine learning. It has a comprehensive, flexible ecosystem of tools, libraries and community resources that lets researchers push the state-of-the-art in ML and developers easily build and deploy ML-powered applications.

- **Librosa:** librosa is a Python library for analyzing audio and music. It has a flatter package layout, standardized interfaces and names, backward compatibility, modular functions, and readable code.

- **Keras:** Keras is a deep learning API written in Python, running on top of the machine learning platform TensorFlow. It was developed with a focus on enabling fast experimentation.

- **Tkinter:** Tkinter is the standard GUI library for Python. Python, when combined with Tkinter, provides a fast and easy way to create GUI applications. Tkinter provides a powerful object-oriented interface to the Tk GUI toolkit.
- **Numpy:** NumPy is a library for the Python programming language, adding support for large, multi-dimensional arrays and matrices, along with a large collection of high-level mathematical functions to operate on these arrays.

- **ReactJS**: It is a JavaScript framework. It will be used to design simple views for each state in our application and will help in efficiently updating and rendering just the right components when our model runs.

- **Dataset:** For this project, we'll use the RAVDESS dataset; this is the Ryerson Audio-Visual Database of Emotional Speech and Song dataset and is free to download. This dataset has 7356 files rated by 247 individuals 10 times on emotional validity, intensity, and genuineness.

- **Arduino IDE:** The Arduino Integrated Development Environment (IDE) is a cross-platform application that is written in functions from C and C++. It is used to write and upload programs to Arduino compatible boards, but also, with the help of third-party cores, other vendor development boards.

HARDWARE

- Arduino or Genuino Board
- LCD Screen (compatible with Hitachi HD44780 driver)
- pin headers to solder to the LCD display pins
- 220-ohm resistor
- Jumper wires
- Breadboard

4. REFERENCES

Research Papers:

[1] Panuwit Nantasri ; Ekachai Phaisangittisagul ; Jessada Karnjana ; Surasak Boonkla , "A Light-Weight Artificial Neural Network for Speech Emotion Recognition using Average Values of MFCCs and Their Derivatives" , 2020 17th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON), IEEE

[2] Kannan Venkataramanan; Haresh Rengaraj Rajamohan; , "Emotion Recognition from Speech", 22 Dec 2019, https://arxiv.org/pdf/1912.10458v1.pdf

Other References

[3] https://www.intechopen.com/books/social-media-and-machine-learning/automatic-speech-emotion-recognition-using-machine-learning

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

[5] https://towardsdatascience.com/speech-emotion-recognition-with-convolution-neural-network-1e6bb7130ce3

[6] http://tomas.pfister.fi/files/pfister10dissertation.pdf

[7] https://reader.elsevier.com/reader/sd/pii/S0167639319302262?token=2BFF3509ECDF5FDE55EDA7013E9A32A03C676AC3A96BDCFC4AFD2064FA381FE5F9E20EC8D71D2547429B2436386C64BA

[8] https://www.ijeat.org/wp-content/uploads/papers/v9i1s5/A10681291S52019.pdf

[9] https://www.arduino.cc/en/Tutorial/LiquidCrystalDisplay#:~:text=Before%20wiring%20the%20LCD%20screen,pin%20to%20digital%20pin%2012