**A Minor Project Synopsis**

**on**

**Speech Emotion Recognition**

Submitted to Manipal University Jaipur

towards the partial fulfillment for the award of the degree of

**Bachelor of Technology**

**In Information Technology**

By

**Name of the Student – Dev Beniwal**

Registration Number - 209302060

Section **- F**

Under the Guidance of

**Name of the Supervisor** – Ms. Neha V Sharma



**Department of Information Technology**

**School of Information Technology**

**Manipal University Jaipur**

**2022-2023**

**Synopsis**

**1. Introduction**

For several years now, the growth in the field of Artificial Intelligence (AI) has been accelerated. AI, which was once a subject understood by computer scientists only, has now reached the house of a common man in the form of intelligent systems. The advancements of AI have engendered to several technologies involving Human-Computer Interaction (HCI). Aiming to develop and improve HCI methods is of paramount importance because HCI is the front-end of AI which millions of users experience. Some of the existing HCI methods involve communication through touch, movement, hand gestures, voice and facial gestures. Among the different methods, the voice-based intelligent devices are gaining popularity in a wide range of applications. In a voice-based system, a computer agent is required to completely comprehend the human’s speech percept in order to accurately pick up the commands given to it. This field of study is termed as Speech Processing and consists of three components:

• Speaker Identification

• Speech Recognition

• Speech Emotion Detection

Speech Emotion Detection is challenging to implement among the other components due to its complexity. Furthermore, the definition of an intelligent computer system requires the system to mimic human behavior. A striking nature unique to humans is the ability to alter conversations based on the emotional state of the speaker and the listener. Speech emotion detection can be built as a classification problem solved using several machine learning algorithms. This project discusses in detail the various methods and experiments carried out as part of implementing a Speech Emotion Detection system.

**2. Motivation**

There are several motivations for building a speech emotion recognition project, including:

* Improving human-computer interaction: By recognizing and responding to the emotional state of the speaker, machines can provide more natural and intuitive interfaces, making it easier for people to communicate with computers and devices.
* Enhancing user experience: By recognizing and responding to the emotional state of the user, machines can personalize their responses and tailor their interactions to the user's needs, preferences, and emotional state.
* Improving accessibility: For individuals with communication difficulties or disorders, such as autism, recognizing emotions from speech signals can help them express their emotions more effectively and communicate more easily.
* Advancing research in psychology and neuroscience: Speech emotion recognition can provide insights into the nature of human emotions, the relationship between emotions and language, and the neural mechanisms involved in emotion processing.
* Enabling better analysis of media content: Recognizing emotions from speech signals can help analyze the emotional content of media, such as movies, music, and social media, for applications such as sentiment analysis and recommendation systems.

Overall, the motivation for building a speech emotion recognition project is to create more intelligent and responsive machines that can understand and respond to human emotions, which can improve communication, accessibility, and user experience, and advance research in psychology and neuroscience.

**3. Project Objectives**

The primary objective of developing a speech emotion recognition project is to accurately detect and classify the emotional state of a speaker from their speech signals. This involves analyzing various acoustic features of the speech, such as pitch, intensity, tempo, and spectral features, to identify emotional states such as happiness, sadness, anger, fear, and so on.

Other objectives of a speech emotion recognition project may include:

* Developing more natural and intuitive human-computer interfaces that respond to the emotional state of the user.
* Improving the accuracy of human-computer interaction by making it more personalized and context-aware.
* Providing better support for individuals with communication difficulties or disorders, such as autism, by enabling them to express their emotions more effectively.
* Enhancing the emotional content of speech-based applications, such as virtual assistants, gaming, and entertainment, by enabling them to respond appropriately to the user's emotional state.
* Enabling better analysis of emotional content in media, such as movies, music, and social media, for applications such as sentiment analysis and recommendation systems.

Overall, the development of a speech emotion recognition project has the potential to significantly improve the quality of human-computer interaction and communication, as well as open up new avenues for research and development in various fields.

**4. Methodology/ Planning of work:**

The speech emotion detection system is implemented as a Machine Learning (ML) model. The steps of implementation are comparable to any other ML project, with additional fine-tuning procedures to make the model function better. The flowchart represents a pictorial overview of the process. The first step is data collection, which is of prime importance. The model being developed will learn from the data provided to it and all the decisions and results that a developed model will produce is guided by the data. The second step, called feature engineering, is a collection of several machine learning tasks that are executed over the collected data. These procedures address the several data representation and data quality issues. The third step is often considered the core of an ML project where an algorithmic based model is developed. This model uses an ML algorithm to learn about the data and train itself to respond to any new data it is exposed to. The final step is to evaluate the functioning of the built model. Very often, developers repeat the steps of developing a model and evaluating it to compare the performance of different algorithms. Comparison results help to choose the appropriate ML algorithm most relevant to the problem.

**5. Facilities required for proposed work:**

Hardware Requirement

Laptop or PC

* Windows 7 or higher
* I3 processor system or higher
* 4 GB RAM or higher
* 100 GB ROM or higher

Software Requirement

* Python 3
* Librosa library for audio feature extraction
* Scikit-learn library for machine learning algorithms
* Jupyter Notebook for code development and analysis

**6. Gantt Chart**

|  |  |  |  |
| --- | --- | --- | --- |
| TASK | START DATE | END DATE | DURATION |
| Data Collection | 01-Feb-2023 | 15-Feb-2023 | 2 Weeks |
| Data Pre-Processing | 16-Feb-2023 | 22-Feb-2023 | 1 Week |
| Data Analysis | 23-Feb-2023 | 21-March-2023 | 4 Weeks |
| Model Development | 22-March-2023 | 10-April-2023 | 2.5 Weeks |
| UI Development | 11-April-2023 | 22-April-2023 | 1.5 Weeks |
| Model Testing and Validation | 23-April-2023 | 30-April-2023 | 1 Week |

**Bibliography/References**

* *https://ieeexplore.ieee.org/abstract/document/8805181*
* *http://ieeexplore.ieee.org/abstract/document/7002390/*
* *http://ieeexplore.ieee.org/document/9383000*
* [*https://ieeexplore.ieee.org/document/4555476*](https://ieeexplore.ieee.org/document/4555476)
* [*https://www.analyticsinsight.net/speech-emotion-recognition-ser-through-machine-learning/*](https://www.analyticsinsight.net/speech-emotion-recognition-ser-through-machine-learning/)
* [*https://www.researchgate.net/publication/330383935\_A\_Literature\_Review\_on\_Emotion\_Recognition\_Using\_Various\_Methods*](https://www.researchgate.net/publication/330383935_A_Literature_Review_on_Emotion_Recognition_Using_Various_Methods)
* [*https://www.researchgate.net/publication/225359641\_Towards\_Emotion\_Recognition\_from\_Speech\_Definition\_Problems\_and\_the\_Materials\_of\_Research*](https://www.researchgate.net/publication/225359641_Towards_Emotion_Recognition_from_Speech_Definition_Problems_and_the_Materials_of_Research)
* *Liberman, Mark, et al. Emotional Prosody Speech and Transcripts LDC2002S28. Web Download. Philadelphia: Linguistic Data Consortium, 2002.*
* *IEEE Research Paper 12 https://ieeexplore.ieee.org/document/7344793/?reload=true*
* *M.M.H.E. Ayadi, M.S. Kamel, F. Karray, "Survey on speech emotion recognition: Features classification schemes and databases", Pattern Recognition, pp. 572-587, 2011.*
* [*https://www.kaggle.com/datasets/uwrfkaggler/ravdess-emotional-speech-audio?resource=download*](https://www.kaggle.com/datasets/uwrfkaggler/ravdess-emotional-speech-audio?resource=download)