Human Emotion Detection from Voice: Project Report

June 18, 2025

Abstract

This project develops a system to detect human emotions from voice using audio signal processing and machine learning. By leveraging the RAVDESS dataset, the system extracts audio features such as MFCC and Chroma, trains a Random Forest classifier, and predicts emotions in real-time through a Streamlit-based user interface. The system achieves reliable emotion classification and includes a graphical representation of emotional trends, with session data storage for future analysis. This report outlines the tools, methodology, and outcomes of the project.

1 Introduction

Emotion detection from voice is a critical component in human-computer interaction, enabling applications in mental health monitoring, customer service, and interactive systems. This project aims to build an end-to-end system that processes audio inputs to classify emotions such as happiness, sadness, anger, and others. Using the RAVDESS dataset, the system extracts meaningful audio features, trains a machine learning model, and provides a user-friendly interface for real-time emotion prediction. The project integrates audio recording, model inference, and visualization, with session data saved for further analysis.

2 Tools Used

The following tools and libraries were employed:

- **Python**: Core programming language for data processing and model development.
- **Librosa**: Audio signal processing for feature extraction (MFCC, Chroma, etc.).
- **Scikit-learn**: Machine learning library for training Random Forest classifier.
- **Streamlit**: Web framework for building the interactive user interface.
- **Sounddevice**: Library for real-time audio recording.
- **Pandas/NumPy**: Data manipulation and numerical computations.

3 Steps Involved in Building the Project

The project was developed through the following steps:

1. *Dataset Preparation*: The RAVDESS dataset, containing 1440 audio files with eight labeled emotions, was preprocessed to ensure consistent sampling rates and file formats.
2. *Feature Extraction*: Audio features, including Mel-Frequency Cepstral Coefficients (MFCC), Chroma, and spectral contrast, were extracted using Librosa to represent emotional characteristics.

3. *Model Training*: A Random Forest classifier was trained on the extracted features, with data split into 80% training and 20% testing sets. Label encoding was applied to emotion categories.
4. *Model Evaluation*: The model was evaluated using accuracy and confusion matrix, achieving satisfactory performance on the test set.
5. *UI Development*: A Streamlit interface was built, allowing users to record audio, trigger predictions, and view results. A line graph was added to display emotional trends over time.
6. *Real-Time Prediction*: The system processes live audio inputs, extracts features, and predicts emotions using the trained model.
7. *Data Storage*: Session data, including audio features and predicted emotions, were saved as CSV files for future analysis.

4 Conclusion

The Human Emotion Detection from Voice project successfully demonstrates the integration of audio processing, machine learning, and web development to classify emotions in real-time. The system achieves robust performance using the RAVDESS dataset and a Random Forest classifier, with an intuitive Streamlit interface enhancing user interaction. The addition of emotional trend visualization and session data storage provides valuable insights for further research. Future improvements could include expanding the dataset, incorporating deep learning models, and optimizing real-time processing for deployment in practical applications.