📘 **PROJECT REPORT**

VOICE EMOTION RECOGNITION USING MATLAB

---

1. INTRODUCTION

Human speech carries not only linguistic information but also emotional cues such as happiness, anger, sadness, and neutrality. Automatic recognition of emotions from speech plays an important role in applications such as human–computer interaction, mental health analysis, call-center analytics, and assistive technologies.

This project presents a **Voice Emotion Recognition (VER) system** implemented using **MATLAB**, which classifies human emotions from speech signals using **Digital Signal Processing (DSP)** and **Machine Learning techniques**. The system extracts meaningful features from speech using **Mel Frequency Cepstral Coefficients (MFCCs)** and classifies emotions using a supervised learning model.

---

2. OBJECTIVES

The objectives of this project are:

- To analyze speech signals in time and frequency domains
- To preprocess audio signals for noise robustness
- To extract MFCC and delta features from speech
- To build an emotion classification dataset
- To train and test a machine learning classifier
- To recognize emotions from unseen speech samples

---

3. DATASET USED

The project uses the **RAVDESS (Ryerson Audio-Visual Database of Emotional Speech and Song)** dataset.

Dataset Characteristics:

- Professional actors

- Clean studio recordings
- Multiple emotional categories
- WAV audio format

Emotions Considered:

- Neutral
- Happy
- Sad
- Angry

---

4. SYSTEM OVERVIEW

The Voice Emotion Recognition system follows a structured pipeline:

1. Audio dataset preparation
2. Signal analysis
3. Preprocessing
4. Framing and windowing
5. MFCC feature extraction
6. Delta feature computation
7. Dataset feature vector creation
8. Machine learning classification
9. Emotion prediction for test audio

The entire workflow is implemented in **modular MATLAB scripts**.

---

5. SOFTWARE TOOLS USED

- **MATLAB R2023**
- Signal Processing Toolbox
- Statistics and Machine Learning Toolbox

---

6. SIGNAL ANALYSIS

Speech signals are first analyzed in:

- **Time domain** to observe amplitude variations
- **Frequency domain (FFT)** to analyze spectral content
- **Spectrogram** to visualize time–frequency characteristics

This analysis helps in understanding emotional variations in speech patterns.

---

7. PREPROCESSING

Preprocessing improves recognition accuracy by:

- Resampling audio to **16 kHz**
- Normalizing signal amplitude
- Removing silence and DC offset
- Applying pre-emphasis filter

This ensures consistent feature extraction across samples.

---

8. FEATURE EXTRACTION

8.1 MFCC Features

Mel Frequency Cepstral Coefficients (MFCCs) are extracted as they closely model the human auditory perception.

Steps involved:

- Framing
- Windowing (Hamming window)
- FFT
- Mel filter bank
- Log compression

- Discrete Cosine Transform (DCT)

8.2 Delta Features

First-order delta MFCCs are computed to capture temporal variations in speech.

---

9. FEATURE DATASET CREATION

MFCC and delta features are combined to form a single feature vector per audio sample. These vectors are stored in a structured dataset and saved as:

features.mat

Each feature vector is associated with its corresponding emotion label.

---

10. CLASSIFICATION MODEL

A **supervised machine learning classifier** is trained using the extracted features.

Model Characteristics:

- Trained on labeled MFCC feature vectors
- Learns emotion-specific patterns
- Saved as a trained model file:

svm_model.mat

---

11. TESTING WITH UNSEEN AUDIO

The trained model is tested using an unseen speech sample:

- Test audio is preprocessed
- Features are extracted
- Emotion is predicted using the trained classifier

This validates the generalization capability of the system.

---

12. RESULTS AND OBSERVATIONS

- Emotions were classified accurately for clean speech samples
- MFCC features provided strong emotion discrimination
- Delta features improved classification stability
- The system performed well for controlled dataset conditions

---

13. APPLICATIONS

- Human–Computer Interaction
- Emotion-aware virtual assistants
- Call center sentiment analysis
- Mental health monitoring
- Speech analytics systems

---

14. ADVANTAGES

- Modular MATLAB implementation
- Uses well-established DSP techniques
- Scalable to more emotions
- Easy to extend with deep learning models

---

15. LIMITATIONS

- Performance depends on dataset quality
- Limited robustness to background noise
- Emotion overlap in speech can reduce accuracy
- Real-time deployment not implemented

---

16. FUTURE ENHANCEMENTS

- Use deep learning (CNN / LSTM)

- Add noise-robust feature extraction
- Real-time emotion recognition
- Support multilingual datasets
- Deploy as standalone application

---

17. CONCLUSION

The Voice Emotion Recognition system successfully demonstrates how **speech signal processing and machine learning** can be combined to classify human emotions. By extracting MFCC-based features and training a supervised classifier, the system achieves reliable emotion recognition for clean speech samples. This project highlights the practical application of DSP and ML techniques in modern intelligent systems.

---

18. REFERENCES

1. RAVDESS Dataset Documentation
2. MATLAB Signal Processing Toolbox
3. Rabiner & Schafer – Digital Processing of Speech Signals
4. Emotion Recognition from Speech – IEEE Papers

---

19. AUTHOR

Sudeep J Elahole
B.Tech Electronics and Communication Engineering
DSP | MATLAB | Machine Learning