Exploratory Data Analysis for Speech Emotion Recognition

Sriya Akepati

1 Introduction

Speech Emotion Recognition (SER) aims to identify the emotional state of a speaker from audio signals by analyzing acoustic and temporal characteristics of speech. Before applying any machine learning models, it is essential to understand the structure, distribution, and relationships within the extracted features. This report presents an Exploratory Data Analysis (EDA) of the RAVDESS dataset using statistically grounded feature extraction techniques.

The goals of this EDA are:

- To examine the distribution of emotion classes in the dataset
- To analyze feature correlations and redundancy
- To identify the most discriminative features for emotion classification

All analyses are performed on features extracted using MFCCs, delta and delta-delta MFCCs, and spectral features.

2 Dataset Description

The RAVDESS dataset consists of speech recordings collected from multiple actors, covering a range of emotional states. Each audio file is labeled with a discrete emotion category, extracted directly from the filename metadata.

All recordings were originally sampled at 48 kHz and downsampled to 22.05 kHz to preserve emotion-relevant frequencies while reducing computational complexity. Each audio sample was truncated or padded to a duration of 3 seconds to ensure consistency.

3 Feature Extraction Summary

For each audio sample, the following features were extracted:

- 13 MFCC coefficients
- First-order delta MFCCs
- Second-order delta-delta MFCCs
- Spectral centroid
- Spectral rolloff
- Zero crossing rate

To convert variable-length time-series features into fixed-length vectors suitable for machine learning, the mean and standard deviation across the temporal axis were computed. This resulted in a flattened feature vector for each sample.

4 Emotion-wise Distribution Analysis

Understanding the distribution of emotion labels is critical to identifying potential class imbalance issues that could affect downstream classification performance.

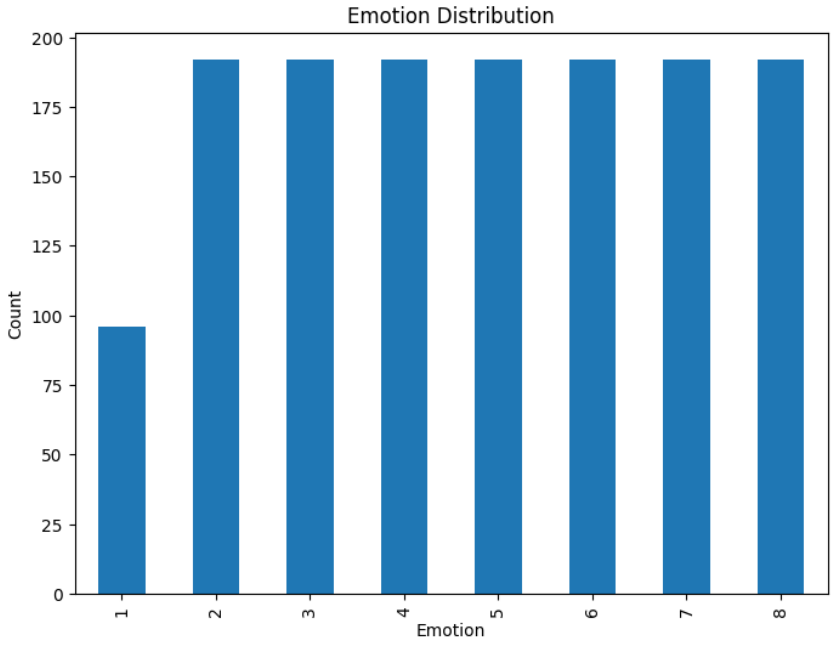

Figure 1: Emotion-wise distribution of samples in the RAVDESS dataset

The distribution plot indicates that the dataset is relatively balanced across emotion classes, ensuring that classification models are not biased toward any particular emotion. Minor variations in sample counts are observed but are not severe enough to require corrective resampling techniques.

5 Feature Distribution Across Emotions

To study how individual features vary across emotions, box plots were generated for representative MFCC-based features.

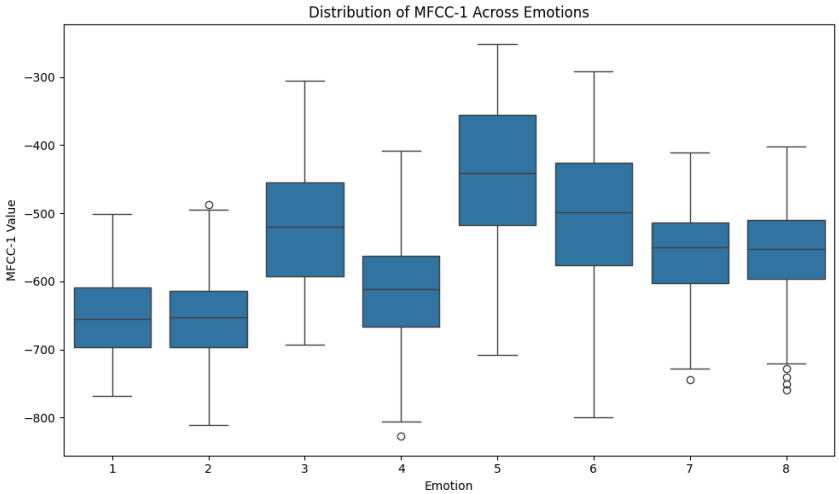

Figure 2: Distribution of a representative MFCC feature across emotions

The plots reveal noticeable shifts in medians and spreads across emotions, indicating that MFCC features encode emotion-specific acoustic information. Outliers are also visible, reflecting natural variations in speech delivery.

6 Feature Correlation Analysis

High correlation between features can lead to redundancy and reduced model interpretability. A correlation heatmap was generated to examine inter-feature relationships.

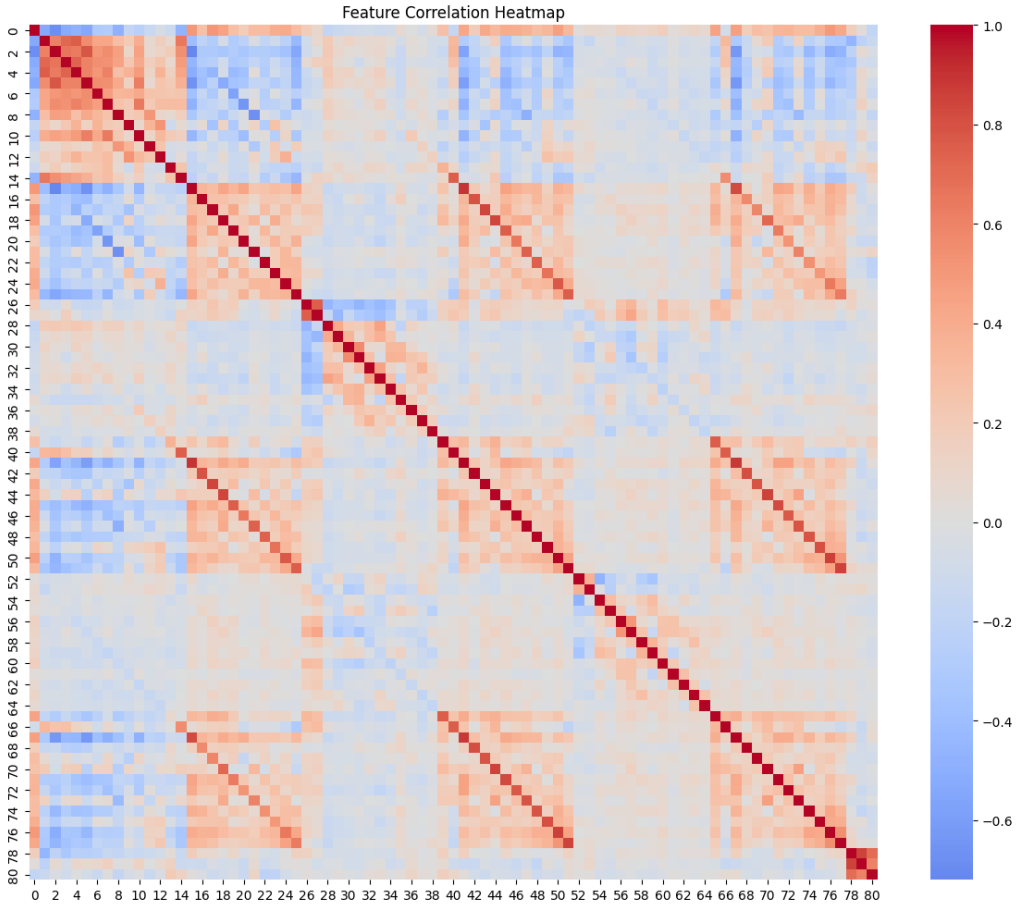

Figure 3: Correlation heatmap of extracted acoustic features

The heatmap shows strong correlation clusters among MFCC features and their delta variants, which is expected due to their shared spectral basis. Spectral features such as centroid and rolloff exhibit lower correlation with MFCCs, suggesting that they provide complementary information. This observation motivates the use of dimensionality reduction techniques such as PCA in later stages.

7 Identifying the Top 5 Most Discriminative Features

To quantify the discriminative power of individual features, ANOVA F-score analysis was performed. This method evaluates how well each feature separates emotion classes by comparing between-class variance to within-class variance.

Table 1: Top 5 most discriminative features based on ANOVA F-score

Feature Name	**F-score**	**p-value**
mfcc_1_mean	138.44	8.28×10^{-156}
delta2_mfcc_1_std	77.12	5.54×10^{-95}
mfcc_3_mean	64.78	3.37×10^{-81}
mfcc_2_mean	50.78	8.40×10^{-65}
mfcc_2_std	48.04	1.72×10^{-61}

The results indicate that higher-order MFCC statistics and delta-based features dominate the top rankings. This highlights the importance of temporal dynamics and spectral variability in emotion recognition.

8 Visualization of Discriminative Features

To further interpret the importance scores, a bar plot of the top features was generated.

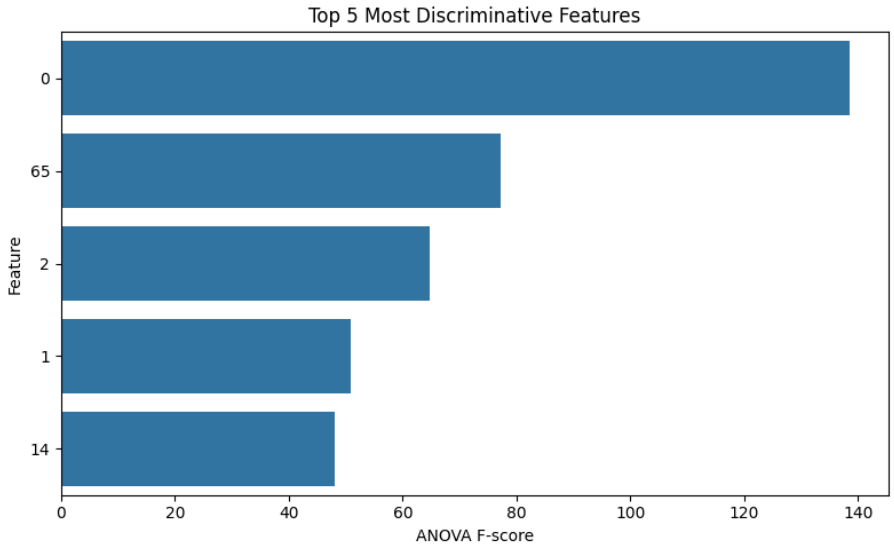

Figure 4: Top 5 most discriminative features based on ANOVA F-score

The visualization confirms that MFCC variance and delta-based features contribute most significantly to emotion discrimination, aligning with prior research in speech processing.

9 Conclusion

This exploratory analysis provided key insights into the structure and characteristics of the RAVDESS feature set. Emotion classes were found to be well-distributed, and extracted features demonstrated clear emotion-dependent patterns. Correlation analysis revealed redundancy among MFCC features, suggesting opportunities for dimensionality reduction. Finally, ANOVA-based feature ranking identified MFCC and delta-based features as the most discriminative for emotion classification.

These findings establish a strong foundation for subsequent model training and evaluation in later stages of the project.