Milestone-2 Report

Robust Music Genre Analysis & Audio Processing

1. Introduction

This report presents exploratory data analysis and a baseline machine learning model built on the Messy Mashup training dataset.

The objective was to:

- Analyze audio characteristics of genre-separated stems
- Compute signal-level statistics
- Investigate dataset integrity
- Build a simple Decision Tree classifier
- Evaluate performance using Macro F1 score

2. Dataset Description

The dataset consists of:

- **Genres:** 10 music genres : ['blues', 'classical', 'country', 'disco', 'hiphop', 'jazz', 'metal', 'pop', 'reggae', 'rock']
- **Stems per song:** drums.wav, vocals.wav, bass.wav, others.wav

Directory Structure:

- root/ ○ messy_mashup/ genres_stems/ genre/ song_folder/
  - drums.wav
  - vocals.wav
  - bass.wav
  - others.wav

3. Methodology

Audio Processing

Audio files were loaded using the Librosa library. A sampling rate of 22050 Hz was used to standardize processing across all files.

For feature extraction:

- Signals were converted into numerical arrays.
- Feature values were computed per file.
- Genre-level averages were calculated.

Spectral Centroid Analysis

The spectral centroid represents the “brightness” of a sound by measuring the center of mass of the frequency spectrum.

Steps performed:

1. For each genre, the harmonic stem was selected.
2. For each audio file, the spectral centroid was calculated..
3. Each genre's mean centroid values were determined.
4. It was determined which genre had the highest average spectral centroid.

This approach ensures fair comparison by analyzing the same stem type across all genres.

Result:
The genre with the highest mean spectral centroid is:

- hiphop

This suggests that the harmonic background of hip-hop tracks contains more powerful high-frequency elements.

Silence Detection in Audio Stems

To identify silent audio segments:

1. Only the first 0.5 seconds of each stem file were loaded.
2. The definition of silence was: amplitude < 1e-4
3. The maximum absolute amplitude of the segment was measured.
4. The number of files that met the silence condition was counted.

This method efficiently detects near-silent openings in audio stems.

Output produced by code:

- Total number of stem files analyzed
- Number of files containing silence in first 0.5 seconds: 349

4. Tools and Libraries

The following Python libraries were used:

- NumPy --- numerical computation
- Librosa --- audio processing and feature extraction
- OS module --- directory traversal

5. Key Observations

- Different genres exhibit different spectral energy distributions.
- Harmonic stems makes it easy to compare things across genre.
- Some stem files begin with near-silent segments, This is probably because of how they were recorded or mixed..

6. Conclusion

This suggests how audio signal processing can be used to look at musical traits in different genres. By extracting spectral features and identifying silence patterns, significant acoustic variations were noted.

The analysis shows that the spectral centroid is a good way to measure frequency distribution and gives us an idea of how different genres sound.