Milestone-1 Report

Robust Music Genre Analysis & Audio Processing

1. Introduction

This milestone focuses on understanding the dataset and establishing a baseline performance through exploratory data analysis (EDA) and simple heuristic-based methods using librosa.

The objectives were:

- Validate dataset integrity and filter corrupted files.
- Perform silence detection and analysis.
- Generate a mixed sample from stems.
- Compute RMS amplitude and apply peak normalization.

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

Q1 Result

Total corrupted (<4KB) + files < 5.0491MB:

(1 MB = 1024 × 1024 bytes)

- Final Answer: Corrupted (<4KB): 0 and Files < 5.0491MB: 1256 so we get **1256**

Q2 Result

Absolute difference between:

(Here, we took two threshold upper and lower to calculate the difference)

- Total sounds > 5.0493MB : 184
- Total sounds < 5.0491MB : 1256

- Final Answer: **1072**

Q3 Result

Absolute difference between:

- Training reggae drum samples : 83
- Validation country vocal samples : 17

- Final Answer: **66**

4. Silence Analysis (Q4–Q9)

Silence detection was performed using:

librosa.effects.split()

The function identifies non-silent intervals based on top_db threshold.

Silence Cases Handled:

- Long and very long Silence
- Stem-specific silence
- Genre-specific silence
- Silence in middle

Q4

Total files with silence ≥ 5 seconds:

- **678**

Q5

Total vocal tracks with silence ≥ 5 seconds:

- **315**

Q6

Average Silence Length in Vocals (seconds):

- **12.78**

Q7

Jazz drum tracks with silence ≥ 5 seconds:

- **20**

Q8

Jazz drum tracks with silence ≥ 5 seconds AND silence only in middle:

(The track included start and end and middle is not alone)

- **0**

Q9

Jazz drum tracks with silence ≥ 5 seconds AND Max_Silence_Sec ≥ 10:

- **7**

5. Audio Mixing & Signal Processing (Q10–Q12)

For the first song in the **rock** genre:

- All 4 stems were loaded
- Duration fixed to 5 seconds
- Stems stacked and summed element-wise
- RMS amplitude computed manually
- Peak normalization applied

Q10

Mix Sample Length (Q10)

Length of mix samples are :

- Final Answer**: <u>110250</u>**

Q11

RMS Amplitude (Q11)

RMS was computed using:

RMS=1N∑x2RMS = \sqrt{\frac{1}{N} \sum x^2}RMS=N1∑x2

- RMS Value: **<u>0.16697016</u>**

Q12

Peak Normalization (Q12)

Normalization performed as:

mixnorm=mixrawmax⁡(|mixraw|)mix_{norm} = \frac{mix_{raw}}{\max(|mix_{raw}|)}mixnorm =max(|mixraw|)mixraw

Maximum absolute value after normalization:

- Max Value**: <u>1.0</u>**

6. Observations

- Vocal stems tend to have longer silence segments compared to drums.
- Silence ≥ 10 seconds is relatively rare.

- RMS amplitude reflects energy content of the mixed signal.
- Peak normalization ensures consistent amplitude scaling without clipping.

7. Challenges Faced

- Handling fully silent audio files.
- Implementing logic.
- Correctly converting MB to bytes using 1024 multiplier.
- Managing floating threshold precision (5.0491 vs 5.0493 MB).
- Ensuring all stems had equal length before summation.

8. Conclusion

This milestone strengthened understanding of:

- Audio data preprocessing
- Silence pattern analysis
- Multi-track signal mixing
- Energy-based audio metrics