Assignment-1

ASLP-SongEval: Exploratory Data Analysis Report

Ahsan Adil - 523193
Muhammad Hamza Malik - 503376
Abdullah Azmat - 507975
Ayesha Kashaf Aslam - 515191

BSCS-14-C

1 Introduction and Motivation

For this project, our group wanted to work on something that was both technically challenging and achievable within the given time frame. Music is an essential part of most people's lives today, and the ability to analyze and categorize songs in a meaningful way seemed both practical and fascinating. Beyond just being an engaging dataset, this project also provides the opportunity to explore multimodal data (audio, text, and metadata) and apply signal processing, statistical analysis, and data visualization methods to a real-world domain.

Our motivation stemmed from wanting to understand how various musical and acoustic features correlate with listener experience and annotated categories such as emotion or perceived quality. The ASLP-SongEval dataset presented the right balance between complexity, scale, and accessibility, allowing us to apply core machine learning and audio analysis techniques while learning about music data representation.

2 Dataset Overview

We obtained the dataset by cloning it directly from Hugging Face using a simple Python script that utilized the `snapshot_download` method from the `huggingface_hub` library. The full dataset is approximately 10 GB in size and contains around 2400 MP3 files. Most of the songs are in either English or Chinese.

A metadata file accompanies the audio data. It includes information about each song's singer gender and provides four different annotation sets for five metrics, all collected from professional audio evaluators. These annotations represent subjective assessments of specific aspects of each recording, which makes the dataset valuable for studying human perception of musical quality.

Example structure:

- Audio files: 2400 MP3 songs (average duration $\approx$ 250–300 seconds)
- Metadata: singer gender, annotation scores, and category labels
- Languages: English and Chinese

3 Domain-Specific EDA Results

For the exploratory data analysis, we agreed to use a consistent subset of audio files so that comparisons across analyses would be meaningful. The main tools used were `librosa`, `matplotlib`, and `numpy`.

Part 1: Waveforms

Waveforms provide a direct visualization of amplitude over time, essentially showing how loudness varies as the song progresses. The patterns were fairly straightforward, with most songs beginning with a low amplitude section (an intro) and showing one or two strong peaks around the middle of the track, corresponding to chorus or instrumental sections.

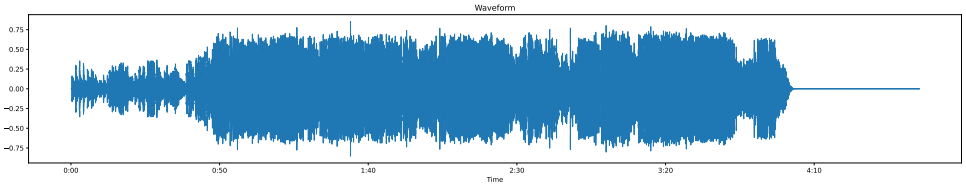

Figure 1: Waveform comparison plot

Part 2: Spectrograms

Spectrograms provided deeper insight into the audio, as they reveal how frequency content evolves over time. We initially used a relative decibel scale (`ref=np.max`) for clearer single-file visualization. While this normalization method highlights the structure within a track, it is less suitable for comparing across different files, as the absolute loudness information is lost. Future analyses will use an absolute reference scale for fairer comparison.

Examining the spectrogram of `21.mp3`, we observed that both pitch and loudness start relatively low, rise to a peak midway through the track, and then taper off towards the end—consistent with the waveform analysis.

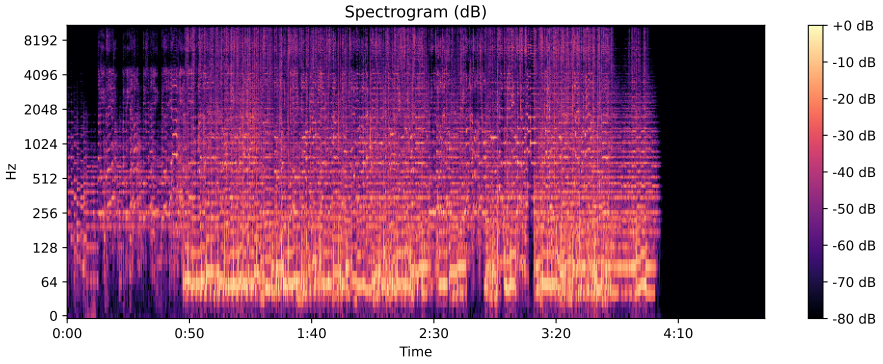

Figure 2: Spectrogram of 21.mp3

Part 3: Duration Distribution

We limited the computation to the first 30 audio files to reduce runtime. The duration histogram shows that most songs cluster around 290 seconds (approximately 4.8 minutes), with a small number extending beyond that. The duration spread appears reasonable, confirming data consistency.

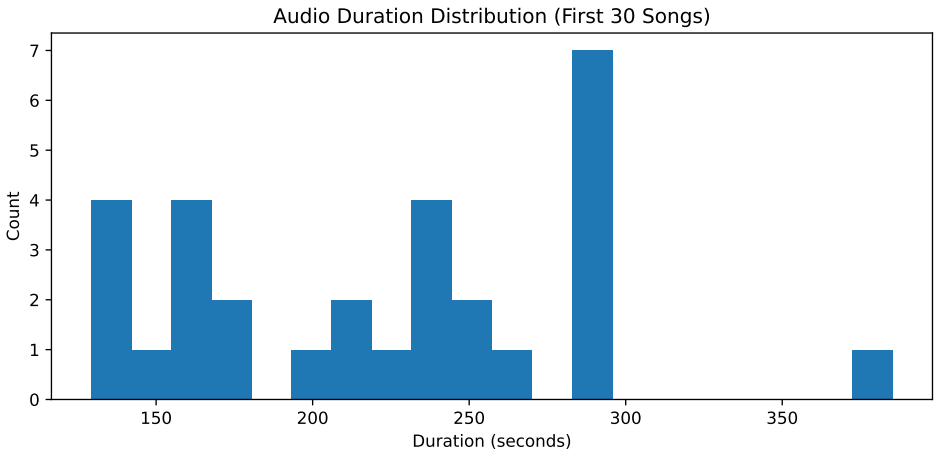

Figure 3: Duration Histogram of first 30 songs

4 Problem Formulation and Metrics

The primary goal of the ASLP-SongEval project is to explore how different modalities (audio, text, and metadata) can be used to model song-level attributes such as emotion, quality, or listener engagement. In this phase, we focused on understanding the dataset and identifying meaningful patterns and challenges.

Problem formulation: Given a set of songs with corresponding audio signals and annotation-based labels, the problem can be formulated as a supervised learning task—either regression (predicting continuous quality scores) or classification (predicting categorical labels such as emotion type or gender). Before training any models, exploratory analysis helps verify data balance, quality, and modality alignment.

Evaluation metrics: For classification tasks, we would typically use accuracy, precision, recall, and F1-score. For regression-style predictions, appropriate metrics include Mean Squared Error (MSE), Mean Absolute Error (MAE), and Pearson correlation with ground truth ratings. In the current EDA stage, metrics are limited to statistical summaries (mean, variance) and distribution-based visualization to assess data consistency.

5 Task Division Table

Each group member contributed to a specific part of the exploratory data analysis. The following table summarizes individual contributions:

Member	Modality / Focus Area	Main Tasks	Figures / Outputs
Hamza	Audio – Waveform Analysis	Implemented waveform plotting and interpreted loudness variations across time.	Figure 1: Waveform comparison
Ahsan	Audio – Spectrograms	Generated spectrograms using both relative and absolute scales; analyzed pitch–loudness correlation.	Figure 2: Spectrogram
Abdullah	Statistical EDA	Computed song durations, created histograms, and summarized trends across the first 30 samples.	Figure 3: Duration histogram
Ayesha	Report Compilation & Coordination	Structured LaTeX report, handled dataset setup, and integrated team analyses.	Final report and dataset summary

Table 1: Division of tasks among the four team members.