Final Project PROPOSAL

Title: MUSICAL CLIMAX DETECTION

Submitted by Lily Qin and Jason Yang

Team Members:
Lily Qin
Jason Yang

1. Motivation and Problem Statement

A. PROBLEM STATEMENT (CLEAR AND CONCISE)

The problem we are looking to solve is to identify the climactic and most intense part of a music song.

B. MOTIVATION

Both Lily and Jason enjoy listening to music, but sometimes it is difficult to find the most well-known part of the song when showing a friend. Often the climax is the most recognizable song, and it would be very useful if a song would signify to the user which part is the most intense.

This problem is important, as a lot of people such as Jason and Lily who both go to the gym, it is very common to play the most intense and high energy part of the song when lifting weights. By doing this, we will be able to solve this frequent problem and allow people to quickly find the climax without having to scrub through and look through each song. People do not remember the exact timestamp, and especially since the same song on different platforms might be slightly different lengths, having a feature to find the exact time of the high intensity part of the song is very useful.

C. EXISTING SOLUTIONS

Currently, there are solutions to categorize songs into different genres, which is used by Spotify. Spotify does display the lyrics of a song, however it does not display the climax of the music.

The closest solution is Google's YouTube, as YouTube videos show the "most watched" part, and it also shows the audio profile of the video. Sometimes the "most watched" part of a video is the climax, as it is the part that most people want to rewatch. This is not sufficient to solve the problem though, because not only does it require enough people to have watched the video for the algorithm to choose this part, but sometimes the most watched part is not actually the climax.

As for showing the audio profile, it is also not sufficient to solve our problem, as there may be multiple spikes in audio. Sometimes too, the high intensity is not necessarily from the rise in volume but it may be also unique from other elements by its speed or other parts.

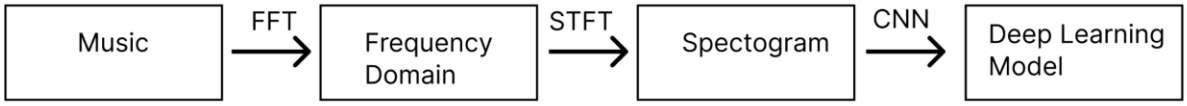

Figure A: Schematic Design

2. Proposed Methodology

A. SIGNAL PROCESSING TECHNIQUES

Firstly, we will use FFT to convert the song from time domain into frequency domain. Then, we will use sampling techniques to improve the efficiency of processing and compute. We will construct a spectrogram using STFT. The spectrogram will show how the energy of different frequencies changes over time. we will use STFT to identify spectral changes

B. MACHINE LEARNING METHOD

We will use CNN (convolutional neural networks) to identify the climax in the songs, as it is unsupervised and we do not need to manually label. Additionally, using CNN will allow us to use the spectrograms themselves as the input, and the model will be trained to learn and identify the climax itself.

C. DATASETS

The dataset we will likely use is http://millionsongdataset.com/, which has a lot of free songs (over 280GB), and there is a dataset of 10,000 songs for just 1.8GB and free. Additionally, there is documentation about how to use and read the files.

The other dataset we are considering to use is the GTZAN dataset found on Kaggle (https://www.kaggle.com/datasets/andradaolteanu/gtzan-dataset-music-genre-classification). However, this seems to be over multiple genres and the audio files are relatively short (30 seconds). The climax will most likely not be in the 30 seconds provided for every song. Finally, there is also https://freemusicarchive.org/home, however we would need to manually download each piece which is time consuming.

3. Expected Outcomes and Deliverables

A. EXPECTED OUTCOMES

Climax Detection Model: An algorithm uses signal processing and machine learning models to identify the climax or most intense section of a song.

Model Performance Metrics: Use F1 score, precision, recall, accuracy to evaluate the model performance. Our accuracy goal is at least 80% success rate for targeting the actual climax.

Spectrogram Visualizations: Use spectrogram highlighting the detected climax segment.

Database Analysis: Evaluation and preprocessing steps for the chosen datasets.

B. DELIVERABLES

Source Code: Well-structured code including signal processing and machine learning models for targeting the climax of a song.

Technical Report: A detailed report explaining the problem statement, proposed methodology that contains signal processing and machine learning models, implementation, experiments, evaluation, and conclusion.

Presentation: A project presentation summarizing the objectives, methods, experiments, and conclusions.

Gantt Chart: A clear and detailed timeline outlining the tasks, responsibilities, and milestones.

4. Challenges and Milestones

A. CHALLENGES

Definition of Climax: It's hard to identify the definition of climax in a song, especially on the signal side.

Data Quality and Availability: Songs may have inconsistent formats, and free datasets may contain songs that are hard to identify their climax.

Spectrogram Construction: Data may contain noise and show noise in spectrogram.

Model Training and Performance: Difficulty in training CNN to reliably identify climax.

B. MILESTONES

Week 5: Project Planning and Dataset Exploration
Week 6: Signal Processing and Spectrogram Analysis
- Preprocess audio data and create initial spectrograms.
- Implement FFT and STFT for spectrogram generation.
- Analyze frequency components and refine spectrogram visualization.

Week 7: Model Development
- Build and train the CNN model.
- Test model performance.

Week 8: Model Evaluation and Optimization
- Improve detection accuracy through data augmentation and model adjustments.

Week 9: Documentation and Final Report
Week 10: Final Presentation and Submission

Gantt Chart

Task	Week 5	Week 6	Week 7	Week 8	Week 9	Week 10
Project Planning & Dataset Exploration	✅					
Signal Processing & Spectrogram Analysis		✅				
Model Development			✅			
Model Evaluation & Optimization				✅		
Documentation & Final Report					✅	
Final Presentation & Submission						✅

5. References

https://brianmcfee.net/papers/ismir2014_spectral.pdf
http://millionsongdataset.com/,
https://www.kaggle.com/datasets/andradaolteanu/gtzan-dataset-music-genre-classification

https://freemusicarchive.org/home

TECHIN 513: MUSICAL CLIMAX DETECTION

Lily Qin, Jason Yang

Objective

The problem we are looking to solve is to identify the climactic and most intense part of a music song.

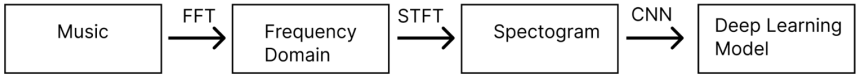

Figure A: Schematic Design

Approach

1. Find open source musical songs
2. FFT to convert music into frequency domain
3. STFT to create energy spectrogram
4. Use CNN to train on the spectrogram images
5. Validate and test CNN model
   - Evaluate with F1 score, Recall, Precision, Accuracy

Expected Results and Milestone

Task	Week 5	Week 6	Week 7	Week 8	Week 9	Week 10
Project Planning & Dataset Exploration	✅					
Signal Processing & Spectrogram Analysis		✅				
Model Development			✅			
Model Evaluation & Optimization				✅		
Documentation & Final Report					✅	
Final Presentation & Submission						✅

Climax Detection Model: identifies the climax or most intense section of a song.

Model Performance Metrics: F1 score, accuracy, precision, recall