Automated Chorus Detection for Enhanced Music Discovery and Annotation Workflows

Project proposal by Dennis Dang

Problem Statement

High-quality, annotated music data have the potential to enhance music discovery processes, recommendation systems, marketing strategies, and other machine learning pipelines in Music Information Retrieval (MIR) workflows. However, the labor-intensive process of manually annotating musical data limits the efficiency, scalability, and scope of data-driven music endeavors across the industry spectrum, from streaming platforms and record labels to live performers. To address these challenges, we propose the development of an automated music labeling and annotation system, with a particular emphasis on chorus detection. Such a system would significantly improve the way music data is processed, analyzed, and utilized, ultimately leading to enhanced user experiences, optimized business processes, and more efficient resource allocation within the music industry.

Context

The proliferation of music streaming services has led to an exponential growth in the volume of music data generated and consumed. This abundance of data presents both opportunities and challenges for the industry. Automated music segmentation and structural annotation, in particular, are relevant tasks in MIR, with applications in music summarization, playlist generation, and content-based music retrieval. However, traditional approaches that rely on manual annotation by music experts are time-consuming, expensive, and struggle to keep pace with the ever-growing volume of music data. To overcome these challenges, there is a pressing need for automated methods that can efficiently and accurately detect and label chorus sections in music data, leveraging recent advancements in digital signal processing and machine learning.

Criteria for Success

The success of the automated chorus detection system will be evaluated based on the following criteria:

1. **Accuracy/Precision**:
    a. The system should achieve a chorus detection accuracy of at least 90% on a homogenous dataset of music genres, 70% on a diverse dataset.
    b. The system should maintain a precision of over 85% in identifying the exact start and end times of chorus sections within a tolerance of ±0.5 seconds.

2. **Integration:**
    a. The system should offer a flexible and modular architecture that allows for easy integration with third-party APIs, such as YouTube's Data API or Spotify's Web API, enabling the development of standalone web applications that leverage the chorus detection functionality alongside the rich music metadata and content provided by these platforms.
    b. The system's output should be compatible with widely used music tagging formats, such as ID3v2, XML, and MP4 metadata, ensuring easy adoption by DJ software, music management tools, and other industry-standard applications.

Scope of Solution Space

The proposed automated chorus detection system will focus on the following key areas:

1. **Machine Learning and Deep Learning:**
    - A range of supervised and unsupervised deep learning algorithms will be explored and evaluated for their performance in chorus detection tasks, including but not limited to:
        - CRNN
        - Clustering
        - HMM
        - Similarity Matrices/Pattern detection
2. **Parameter and Hyperparameter Tuning:**
    - A systematic approach to parameter and hyperparameter tuning will be employed to optimize the performance of the selected deep learning models.
3. **Digital Signal Processing Methods:**
    - Various digital signal processing techniques will be applied to preprocess and transform the raw audio data into a suitable format for machine learning models.
    - A/B testing will be conducted to compare the performance of different signal processing methods and identify the most effective techniques for chorus detection in electronic music.
4. A**udio Feature Selection and Engineering:**
    - A comprehensive set of audio features will be extracted from the preprocessed audio data to capture the relevant characteristics for chorus detection.
    - Domain knowledge and musical insights will be incorporated to engineer novel features specifically tailored to the characteristics of chorus sections in electronic music.
5. **Focus on Electronic Music Genres:**
    - In this stage, the project will primarily focus on developing and evaluating chorus detection models for electronic music genres, rather than aiming for immediate generalization across all music genres.

Constraints

One significant constraint is the limited availability of large-scale, high-quality annotated music datasets, particularly for diverse music genres. The current focus on electronic music genres, while essential for establishing a proof-of-concept, may limit the system's immediate generalizability. Additionally, the computational resources required for processing and analyzing large volumes of audio data may pose challenges in terms of scalability and efficiency.

Stakeholders

The primary stakeholders in the development of an automated chorus detection system include music streaming platforms, record labels, music producers, and artists. These stakeholders stand to benefit from improved music data annotation and enhanced music discovery processes. Music streaming platforms can leverage the system to optimize their recommendation algorithms and user experiences, while record labels and artists can utilize the technology for more targeted marketing and promotional strategies. Music researchers and academics also have a stake in the project, as advancements in chorus detection can contribute to the broader field of Music Information Retrieval (MIR).

Data Sources

The current dataset consists of 332 electronically produced songs that have been manually labeled according to a comprehensive data annotation guide. The guide provides a clear definition of a chorus, outlining the criteria for chorus labeling based on distinctiveness, thematic representation, and structural alignment. The annotation process involves data preprocessing, ensuring consistent formatting and sampling rates, and the use of specialized tools like Serato DJ for accurate labeling. The guide also addresses ambiguities and edge cases, ensuring high-quality annotations. While this dataset serves as a valuable starting point, future efforts should aim to expand the data sources to encompass a wider range of music genres, enhancing the system's generalizability.