COSE474-2021F: Final Project Proposal
Chroma Feature Extraction and Harmonic Classification with Deep Learning Techniques

WonJae Gye

1. Introduction

Recent development of involving deep learning techniques in the domain of audio signal processing, has demonstrated various advantages over the conventional methods. Research efforts have prioritised human speech, and natural languages, due to the wide range of applications such advancements promise. Nevertheless, the capability of learning high-level representations from complex data that deep learning techniques provide is just as useful for various Music Information Retrieval (MIR) tasks.

Harmonic classification is the task of analysing and understanding the relations of multiple sound frequencies in a musical context. When multiple pitches (or tones) are played simultaneously (referred to as a chord), relationships are formed between each and every other pitch. Harmony refers to the complex interaction of these relationships of pitches. While harmony is the mathematical ratios and combinations of the frequencies of various sound waves, music is capable of expressing and conveying emotions far more abstract and ambiguous. Harmony's versatile capability of expressing joy and serenity, to nostalgia and melancholy, with our mapping of such high-level concepts to a complex pattern of frequencies, are the foundation of music: a distinctly human communication, much like language.

The project will attempt to train a model which extracts chroma features from sampled audio signals. Chroma features are analysed by the classifier and the samples are classified into the appropriate class of harmony. Such a model will have the potential for various uses in the music industry when developed further. Much like the field of NLP, which aims to bridge the gap between the human and computer's understanding of natural languages for their more seamless interactions, advancements in MIR may lead to a similar interaction, at an emotional level too.

2. Problem definition & challenges

The goal of this project is described by two objectives. The first objective is chroma feature extraction with a sample of a music audio file as input. The second objective is harmonic classification given the chroma features extracted from the first task.

Chroma feature refers to the representation of audio signals as their pitch classes. In Western music, twelve pitch classes with equal temperament is typical, where an octave (an interval between two pitches whose frequencies are 2:1) is divided into 12 intervals (classes) of equal size.

Audio signal will be sampled and preprocessed to extract its chroma feature. The chroma feature will be used to train a classifier that will label the sample with a tag which best describes the sample's harmonic structure.

Harmonic classification is a difficult task which can take trained musicians multiple years for proficiency. This can be attributed to the difficulty in identifying individual pitch components in a chord, as well the task's non-deterministic nature.

Harmonic classification requires more than correctly identifying all the individual pitch components in a chord, as it requires understanding the relations among the pitches. While there is a guideline for naming (classifying) each chord, there are multiple possible interpretations which rely on context. In other words, the same group of pitches can correctly and simultaneously belong in more than one classes of harmonic structure, and the name of the chord is often determined by the musical context. Therefore, the model will have to predict (rather than compute) the most likely class for each sample. At the time of proposal of this project, it is uncertain whether musical context, especially temporally related sound samples, are necessary to train a successful classifier or whether isolated samples are sufficient.

The purpose of this project is to explore the challenges that lie in improving the currently available models by giving it the capabilities of identifying more complex (7th, extended chords etc.) harmonic structures, with or without context.

3. Related Works

3.1. Traditional Machine Learning for Pitch Detection

Drugman's study (2018) compares multiple machine learning methods for pitch detection. While the study focuses on the speech processing aspect, the F0 contour estimation may relate to this project's task of finding the key note of the audio sample for determining its harmonic structure. If found to be useful, this can potentially eliminate some of the ambiguity that the classification model has to deal with.

3.2. Harmonic Classification with Enhancing Music Using Deep Learning Techniques

Tang (2021) considers automatic extraction of features of harmonic information. Their model is able to extract chord sequence with the goal of determining the key of the music. It is a model based on the CNN architecture that relies on temporal context. Despite this, Tang's acoustic model for feature extraction can be a reference point for this project.

4. Datasets

International Society for Music Information Retrieval (ISMIR) maintains a list of datasets (Lerch), from which the datasets used in this project are sourced. The primary candidate for training data is the JAAH dataset (Demirel). Jazz Audio-Aligned Harmony (JAAH) contains many jazz records' chroma features extracted and with their harmonic information annotated. This dataset may be used for supervised learning when harmonic structure classifier is trained.

5. State-of-the-art methods and baselines

Zhou (Zhou & Lerch, 2015) implemented various models capable of chord classification on a pre-processed input audio signal. Their models were successful with triad (harmonic structure defined by their three most defining tonalities) recognition, but does not further distinguish the more complex harmonies. This may due to the difficulty in training a model of much more thorough harmonic classification, or the training dataset which consisted of pop music with a limited use of extended harmonies.

This project will explore the possibility and challenges of extending the capabilities of current harmonic structure classifier models. By training with datasets involving the more complex and numerous harmonic structures, classification of extended chords can be tested.

6. Schedule

- Week 1 : chroma feature extraction; comparisons of various audio signal sampling methods and preprocessing methods.

- Week 2 : test implementation based on previous models of triad classifiers

- Week 3 : implementation of extended classification capabilities

- Week 4 : tune model and experiment with other dataset if required

References

Demirel, Eremenko, P. Jazz audio-aligned harmony (jaah) dataset. URL https://github.com/MTG/JAAH?search=1.

Drugman, T., Huybrechts, G., Klimkov, V., and Moinet, A. Traditional machine learning for pitch detection. *IEEE Signal Processing Letters*, 25(11):1745–1749, nov 2018. doi: 10.1109/lsp.2018.2874155. URL https://doi.org/10.1109%2Flsp.2018.2874155.

Lerch, A. mir-datasets. URL https://www.ismir.net/resources/datasets/.

Tang, W., Gu, L., and Selistean, D. Harmonic Classification with Enhancing Music Using Deep Learning Techniques. *Complexity*, 2021:1–10, September 2021. doi: 10.1155/2021/5590996. URL https://ideas.repec.org/a/hin/complx/5590996.html.

Zhou, X. and Lerch, A. Chord detection using deep learning. 01 2015.