Lead Sheet Transcription

Pierre Lardet

4th Year Project Report
Computer Science and Mathematics
School of Informatics
University of Edinburgh

2024

Abstract

This skeleton demonstrates how to use the `infthesis` style for undergraduate dissertations in the School of Informatics. It also emphasises the page limit, and that you must not deviate from the required style. The file `skeleton.tex` generates this document and should be used as a starting point for your thesis. Replace this abstract text with a concise summary of your report.

Research Ethics Approval

This project was planned in accordance with the Informatics Research Ethics policy. It did not involve any aspects that required approval from the Informatics Research Ethics committee.

Declaration

I declare that this thesis was composed by myself, that the work contained herein is my own except where explicitly stated otherwise in the text, and that this work has not been submitted for any other degree or professional qualification except as specified.

(*Pierre Lardet*)

Acknowledgements

Thank you to my two lovely supervisors, Prof. Amos Storkey and Dr. Alec Wright, for allowing me to fulfil my desire for such a personally fulfilling project. They for provided me with invaluable guidance on both this project and research in general, which I will take forward to future research. I would also like to thank my friends and family for their unwavering support throughout the year, and for their patience and understanding when I was too busy to spend time with them.

Table of Contents

Chapter 1

Introduction

1.1 Motivation

1.2 Aims

The aims of this project are:

- To develop a model that can generate lead sheets from audio recordings of music.
- To implement and evaluate current state-of-the-art models for music transcription and generation.
- Investigate the use of different data representations for music transcription and generation.
- To evaluate the performance of the model on a variety of music genres and styles.
- To investigate the use of different models/datasets/data augmentation methods for automatic chord recognition, melody transcription and lyric transcription.

1.3 Outline

The report is structured as follows:

- **Chapter 2** provides background information on lead sheets, music transcription, and related work.
- **Chapter 3** describes the experiments conducted in this project, including the data, models, training, and evaluation.
- **Chapter 4** presents and analyses the results of the experiments.
- **Chapter 5** concludes the report and provides suggestions for future work.

Chapter 2

Background & Related Work

2.1 Background

2.1.1 Lead Sheets

Lead sheets are a form of musical notation that contain both the melody and harmony. Generally, the melody is a single line forming the dominant melody, which is often the vocal line of a song, written in standard musical notation on a stave. The harmony is represented by chord symbols, which are written above the melody. A lead sheet also contains the key, time signature and, optionally, the lyrics of a song. An example of a lead sheet is shown in Figure 2.1.

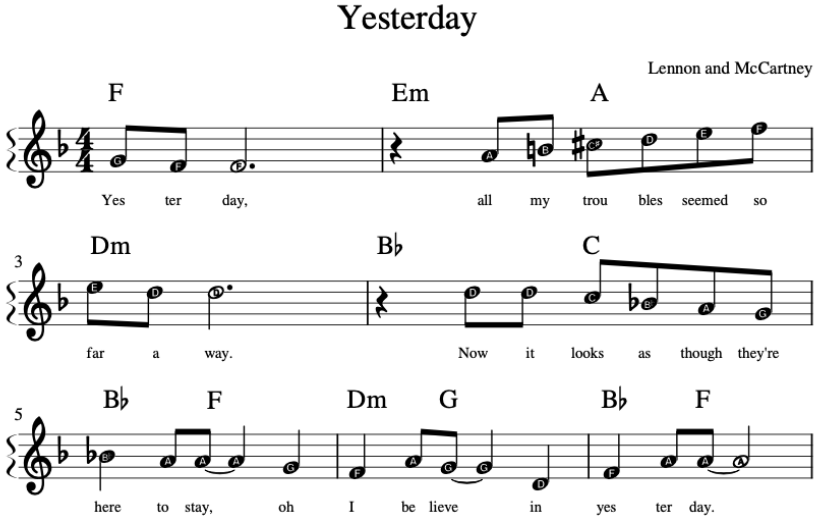

Figure 2.1: An example of a lead sheet for ‘Yesterday’ by the Beatles.

They date back to the mid 20th century, and were originally called ‘fake sheets’ because they were used by musicians to ‘fake’ their way through a song [4]. Lead sheets are most notably used in jazz music where their origins lie. Any jazz musician worth their salt owns a ‘real book’, so called to distinguish it from the fake books that were used

in the past. Real books contain lead sheets for hundreds of jazz standards, and are an essential tool for any jazz musician.

More recently they have served as a useful tool for musicians in other genres, such as pop and rock, who want to learn and perform songs quickly and easily. They allow easy communication of the important elements of a song, without the need for a full score, further allowing improvisation and personalisation of the song.

Lead sheets do not contain all the information of a full score, such as dynamics, articulation, or specific voicings of chords. Furthermore, they are not suitable for all types of music, such as classical music, where a full score is necessary to convey the composer's intentions, or rap music where the lyrics and beat are the most important elements. However, they are a useful tool for many musicians, and are a common way of representing music in the music industry.

2.1.2 Music Transcription

Musical Transcription is a field within MIR (Music Information Retrieval) that aims to convert audio into a symbolic representation. [1] provides a comprehensive review of different forms of transcription. They categorise transcription into frame-level, note-level, stream-level and notation-level transcription. In brief, frame-level transcription predicts musical features, such as pitch, for a short frame of audio, which can then be aggregated into note-level predictions for musical notes. Stream-level transcription involves looking at the large picture of a piece of music such as phrasing and structure, and notation level then assembles all of this information into a human-readable score.

For lead sheet transcription, we are interested in this highest level of transcription. We want to take an audio recording of a song and generate a lead sheet that contains the melody and harmony of the song. This is a challenging task, as it requires the model to isolate and transcribe a melody, transcribe the chordal information, and then combine these two elements into a coherent lead sheet where the melody and harmony are aligned in time/beat.

2.1.3 Music Features

Brief explanation of CQT and chroma vectors. GPT generated features [2].

2.2 Related Work

2.2.1 Lead Sheet Transcription

[2] is the only work I've found that does specific lead shet generation. However, they focused far more on melody transcription, and their model is only effective for 24s of audio. They also do not transcribe the lyrics. They also trained on HookTheory user-submitted data which is not ideal - perhaps biased to simpler transcriptions? This work uses well known, expert datasets.

They proposed the use of a generative pre-trained transformer to provide the features, rather than traditionally used CQT features.

2.2.2 Automatic Chord Recognition

JAAH,

Datasets: Isophonics, Mcgill, RWC, USPop, HookTheory

Evaluation: MIREX standard evaluation. Confusion matrices for chords. Qualitative evaluation.

Suffer from lack of variety and size, and distribution of chords. Attempted remedy by [6].

2.2.3 Melody Transcription

Transformer SOTA.

Datasets: MAPS, MDB, MedleyDB, WJazzD, RWC, USPop, HookTheory

Evaluation: MIREX standard evaluation. F1 scores, precision, recall. Qualitative evaluation.

2.2.4 Data Scarcity

Resampling: [6] Augmentation: Semi-supervised learning Synthetic data gen:

2.2.5 Music Generation

Jukebox [3]

Music generation for MIR: [5]

Chapter 3

Experiments

3.1 Data

3.2 Models

3.3 Training

3.4 Evaluation

Chapter 4

Results

4.1 Quantitative Results

4.2 Qualitative Results

4.3 Lead Sheet Generations

Chapter 5

Conclusions

Bibliography

[1] Bhuwan Bhattarai and Joonwhoan Lee. A comprehensive review on music transcription. *Applied Sciences*, 13:11882, 10 2023.

[2] John Thickstun Chris Donahue and Percy Liang. Melody transcription via generative pre-training, 2022.

[3] Prafulla Dhariwal, Heewoo Jun, Christine Payne, Jong Wook Kim, Alec Radford, and Ilya Sutskever. Jukebox: A generative model for music, 2020.

[4] 99The real book, 2021.

[5] Nadine Kroher, Helena Cuesta, and Aggelos Pikrakis. Can musicgen create training data for mir tasks?, 2023.

[6] Jeff Miller, Ken O'Hanlon, and Mark B. Sandler. Improving balance in automatic chord recognition with random forests. In *2022 30th European Signal Processing Conference (EUSIPCO)*, pages 244–248, 2022.

Appendix A

First appendix

A.1 First section