DMRN+14: DIGITAL MUSIC RESEARCH NETWORK
ONE-DAY WORKSHOP 2019

*QUEEN MARY UNIVERSITY OF LONDON*
*TUE 17 DECEMBER 2019*

Digital Music
Research Network

# Automatic Music Accompaniment with a Chroma-based Music Data Representation

Lele Liu[1] and Emmanouil Benetos[1]

[1]School of Electronic Engineering and Computer Science, Queen Mary University of London, UK,
lele.liu@qmul.ac.uk

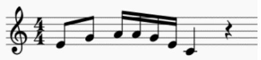
***Abstract*—We propose a model to generate melodic music accompaniment for classical piano music using Long-Short Time Memory (LSTM) networks. We design a chroma-based music data representation that combines music knowledge including pitch chroma, pitch height, onsets, tempo, and modes. A subjective listening test shows that by using the musical features in the music data representation, the LSTM model can generate better accompaniment compared to using simple MIDI pitch numbers in the data representation.**

## I. INTRODUCTION

Automatic music generation, aiming at generating music using machine algorithms, has attracted a lot of interest in recent years due to machine and in particular deep learning. Many music generation systems use MIDI-like encodings [1]. We pay special attention to how different musical features can influence the performance of the music data representation. By taking musical features into account, we describe pitch by pitch chroma and pitch height. This operation greatly decreases the dimensionality of the data representation. We also include beat starts and music modes in the input features. This data representation allows the LSTM network to learn easier and generate more harmonious music accompaniments compared to a MIDI-pitch representation. A subjective user study shows that the musical features (especially the chroma feature) used in our proposed data representation increase the users' overall evaluation on the music accompaniments.

## II. METHODOLOGY

Although it is common to use MIDI pitch numbers in music data representations, people perceive pitch as having two dimensions, where pitch chroma plays a role in representing melodies and pitch height helps in separating music streams and sound sources [2]. Thus, we use chroma features and separate MIDI pitch into 12 pitch chromas and 11 pitch heights. This encoding describes pitch in a more musical way and decreases the dimension of a pitch encoding from 128 to 23. We divide the data representation for a monophonic music piece into two parallel sequences -- a melody sequence *M* and a music annotation sequence *A*. We build the *M sequence* using note pitch chroma, pitch height, note sustain and rest, and the *A sequence* using beat information and music mode. The two sequences are sampled in parallel along the music pieces based on a musically relevant time step. An example of encoding the *M sequence* for a monophonic music piece is shown in Figure 1.



Symbolic *M sequence*: [E4 - G4 - A4 A4 G4 E4 C4 - - - 0 0 0 0]

Figure 1. Encoding *M sequence* for a monophonic music piece.

We assume that all music pieces are within the style of Western tonal music in major/minor modes and contain a monophonic melody and a monophonic accompaniment. The model is based on a recurrent neural network architecture, generating music accompaniment from left to right and predicting one note (or rest/sustain) at each musically relevant time step. The model is composed of two components - a *pitch chroma component* and a *pitch height component*. Each component contains one embedding layer, one LSTM layer and a SoftMax activation output layer. We define both components to solve a categorical classification problem. The final system predicts the accompaniment pitch chroma part at first and uses the chroma part result to predict pitch height.

## III. EXPERIMENT RESULTS

We experiment on different data representations within the LSTM network structure and evaluated their performance in a user study. In the user study, 20 listeners (average 5.85 years of musical training) are asked to rate the overall quality of the generated accompaniment. The statistical data of the listeners' ratings are shown in Table 1. According to the ratings, our proposed data representation (No.5) showed the best performance. The t-values and p-values are calculated from t-tests between the specific data representation and the MIDI-pitch data representation.

TABLE I. STATISTICAL DATA OF PARTICIPANTS' RATINGS.

| ***No*** | ***Data Rep*** | ***Mean*** | ***Median*** | ***t-Value*** | ***p-Value*** |
|---|---|---|---|---|---|
| 1 | MIDI-pitch | 5.698 | 6 | -- | -- |
| 2 | chroma-only | 6.368 | **7** | 2.278 | **0.02384** |
| 3 | chroma+beat | 5.823 | 6 | 0.409 | 0.68271 |
| 4 | chroma+mode | 5.674 | 6 | -0.078 | 0.93807 |
| 5 | proposed | **6.847** | **7** | 3.872 | **0.00015** |

## REFERENCES

[1] F. T. Liang et al., Automatic stylistic composition of bach chorales with deep lstm. In ISMIR, pages 449–456, 2017

[2] J. D. Warren et al., Separating pitch chroma and pitch height in the human brain. Proceedings of the National Academy of Sciences of the United States of America, 100 17:10038–42, 2003.W.-K. Smith, *Linear Networks and Systems* (Book style).Belmont, CA: Wadsworth, 1993, pp. 123–135.