Project Update Report:
Measuring music quality using JETSING
(JET Serialism Is Not Good)

Jonathan Sabini **Ethan Fleming** **Taylor Noah**

Portland State University, 2022

1 Introduction

Music is inherently repetitive, reusing motifs to form a compelling melody that's self-coherent throughout the piece. Many current models for music generation often have no notion of a long-term structure in musical composition and sound, as though the computer musician is wandering around unsure how to compose a consistent rhythm. The Music Transformer model (Huang et al., 2018) uses close attention to overcome this issue, maintaining long-range coherence throughout its generated musical pieces.

Although the music sounds drastically better than models that don't use relative self-attention, the negative log-likelihood metric's numerical results don't demonstrate this difference well. Many metrics, including the negative log-likelihood, can only be utilized on a machine learning model and wouldn't be able to judge a stand-alone music piece. In general, from what we have found, most metrics used to evaluate music generation models have little foundation on musical theory.

2 Motivation

Our group has an interest and background in music, so we wanted to incorporate this knowledge into our machine learning research. There are many models for music generation, but many popularly used metrics to judge them result in dramatically different qualities of musical pieces being considered relatively the same. Many music papers in artificial intelligence, to our knowledge, are often forced to resort to qualitatively comparing generated pieces when numerical metrics fail to show the contrast. We want a metric that will more effectively ascertain how tasteful a composition sounds or perhaps how distasteful. Listening to the sample music generated by the relative self-attention transformer model inspired us to do something on this topic.

We propose a metric based on the rules of serialism in western music theory, which we call the JETSING metric. In western music, which most of us are accustomed to, there is a central tonality from which notes and keys are chosen to sound "good." Conversely, serialism has no central tonality. Instead, it is based on the concept of a "row." A row is a random ordering of the 12 unique half-step pitches in western music theory (C - B on the piano). In strict serialism, you may not repeat a pitch until all 12 pitches in a row have been played, the row must always appear in the same order, and notes are allowed to be in any octave. Because there is no central tonality in this method, music based on serialism tends to sound creepy, confusing, lost, and seemingly aimless.

Many models that audibly perform worse tend to have characteristics of serialism. They wander aimlessly, don't repeat motifs, and lack a sense of central tonality. Conversely, the models that perform well tend to have central tonality and can repeat motifs and create a long-term sense of direction in the piece.

Our JETSING metric is based on the rules of serialism and captures how well a piece performs with regard to those rules. In other words, we measure how serial a piece is. The less serial a piece is, the better it adheres to western music theory and likely sounds "good" to our ears. JETSING is comprised of two sub metrics JETSING-ROW and JETSING-OCTAVE. The former metric measures how well a piece adheres to the first two rules and the latter metric measures the average octave difference between each pair of melodic notes.

3 Related Work

Countless papers have been written on the topic of music generation, but many of them only use an understanding of musical theory to improve

the models themselves. (Huang et al. 2018) worked with Vaswani to utilize the self-attention mechanisms of their original transformer model from (Vaswani et al., 2017), but adjusting it for music generation. (Kotecha et al., 2018) used a bi-axial LSTM to generate polyphonic music aligned with musical rules. Finally, (Zhao et al., 2020) using an extensive knowledge of musical theory worked to create a lightweight variational auto-encoder model. While many papers have used musical theory extensively to improve music generation models, we hope to create a metric that will simplify the process of assessing these models numerically.

4 Experiments

We re-run experiments done by Huang et.al. with their music transformer and generate pieces. We then apply both their metrics and our novel serial metric to rate the performances. As a baseline, we apply JETSING to music created by humans and qualitatively analyze its measurements. Then we will compare baseline results to machine-generated results and the standard deviations of each metric.

4.1 Datasets

For these experiments we utilize datasets gathered from our referenced papers. These include Bach-Midis, Piano-e-Competition, JSBChorale, and the MAESTRO dataset.

4.2 Experiment Method

If possible we will train one or two other models to generate music and be scored as well. We hope to use the LSTM from Kotecha et.al. and we are searching for a transformer that does not use attention (similar to the setup of Huang et.al). This way we can score models of fairly different generative strengths to see how much the scores contrast.

Lastly, we will examine the results of both NLL and JETSING to see which shows more contrast. We find the average and standard deviation of both metrics across all scores. A higher standard deviation indicates that a metric shows more contrast. A lower standard deviation would indicate that the metric shows less contrast.

4.3 Definition of Serialism

The rules of serialism are as follows:

1. No note should be repeated until all 12 notes of a note row have been played.

2. The order of the row remains consistent throughout the piece.

3. Notes can be played at any octave.

5 JETSING

Based on the rules we create two sub-metrics of JETSING: JETSING-ROW and JETSING-OCTAVE.

5.1 JETSING-ROW

JESTSING-ROW will be responsible for the first two rules. It begins at 0 and increments for each break of rules 1 and 2.

Here are examples and valid and invalid note orderings using just 5 tones:

- Valid Row: A-B-C-D-E
- Invalid (Violates 1): A-A-B-C-D-E
  - All notes are accounted for, but a note is repeated before all notes are played.
- Invalid (Violates 2): A-B-D-C-E
  - All notes are accounted for without repetition, but the original order is not maintained.

5.2 JETSING-OCTAVE

This metric handles the third rule of serialism. Since the rule simply allows a pitch to be of any octave it doesn't have a binary rule break. In light of this we separated the rule into its own metric JETSING-OCTAVE.

In this metric we measure the octave difference between notes p_t and p_{t+1}, where p is the pitch at time t. If these two notes are in the same octave the metric reports a 0. If the second note is at least one octave higher or lower the metric reports a 1 and so on. This metric reports the difference in octaves from one note to the next.

We define the JETSING-OCTAVE below:

- $\alpha_t = \Omega_{p_{t+1}} - \Omega_{p_t}$
- $A = \frac{\sum_{t=0}^{n-1} \alpha_t}{n-1}$

α_t represents the octave, Ω, difference of pitches p_{t+1} and p_t. We then sum over α at each time step, to get the total octave differences in the piece, and then divide by one less than the number of notes in the piece, n, to get the average octave difference over the piece, A.

For each octave difference between notes n_t and n_{t+1} the metric reports the octave difference. We take the sum of octaves differences and divide by number of pitches -1 in the piece.

References

Cheng-Zhi Anna Huang, Ashish Vaswani, Jakob Uszkoreit, Noam Shazeer, Ian Simon, Curtis Hawthorne, Andrew M. Dai, Matthew D. Hoffman, Monica Dinculescu, and Douglas Eck. 2018. Music transformer.

Nikhil Kotecha and Paul Young. 2018. Generating music using an lstm network.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need.

Yizhou Zhao, Liang Qiu, Wensi Ai, Feng Shi, and Song-Chun Zhu. 2020. Vertical-horizontal structured attention for generating music with chords.