Generating Rhythm

Mariella Daghfal, Kaede Johnson, Shayan Khajehnouri, & Tahel Singer
EPFL Course DH-501

I. Introduction

In this assignment, we seek to computationally generate a 'human-like' rhythm that emulates the rhythmic schemata underpinning a corpus of Swedish Slangpolska music. Slangpolska refers to a type of folk music that is usually played to accompany a dance of the same name, danced by grouping four dance steps into three beats, which is reflected in the meter of the music.

Professor Martin Rohrmeier introduced rhythm as "a pattern of event onsets and durations [and rests and accents] in reference to an underlying metrical structure." With this context, we try to recreate the same patterns of event onsets and durations typical of Slangpolska music, in a way that closely resembles what humans would require to follow its traditional dance steps.

We design a model that generates measures at random and a model that generates rhythm using duration and onset-informed bigrams, and compare results from the two by way of geographic disparity in their chronotonic chains. After calculating average distances between our corpus, our baseline rhythms, and our bigram-informed rhythms, we find that our generated rhythms can successfully mimic Slangpolska rhythm about a third of the time. We close with a few limitations of our model and suggestions for future work.

II. Representing Rhythm

In accordance with the initial definition provided to us by Professor Rohrmeier, we isolate onsets, durations, rests, and their positional mapping to underlying meter as the core tenets of rhythm. Duration patterns especially are key to the lens with which we study rhythm. Though we were introduced to other aspects of rhythm in class, notably syncopation and rubato, we do not consider these aspects relevant to a traditional, unperformed corpus.

Exploratory analysis using Python's music21 package [1] revealed that less than 10% of our corpus' scores contain a note with an articulation and just 1% of our corpus' notes are articulated, leading us to disregard articulation in rhythm representation. We also removed polyphonic scores, scores without a consistent ¾ meter, and scores containing less than 8 measures to simplify the application of our generative model and evaluation metric.

In an abstract sense, then, we conceptualize rhythm as the 'duration personality' of a series of events which consistently respect the 3/4 meter; it is primarily this 'personality' which informs our choice of generative model and evaluation metric.

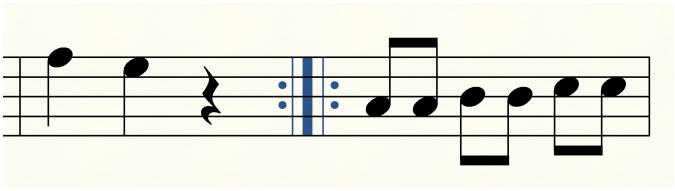

Fig. 1. Measures 4 and 5 of *Dahl polska efter Ola Olsson* as score

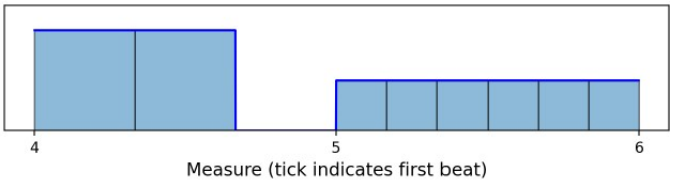

Fig. 2. Measures 4 and 5 of *Dahl polska efter Ola Olsson* as a chronotonic sequence

III. Evaluation Metric

While our reported definition of rhythm incorporates observable phenomena, we also believe a large part of rhythm classification is mental. Accordingly, we sought a metric which involved our corpus' quantifiable aspects while also respecting human judgment. Toussant et al. [2] and Beltran et al. [3] find transformation methods - broadly, those which measure how much effort it takes to mutate one rhythm into another - superior at matching human judgment than features-based methods, which compare the presence of pre-defined rhythmic qualities between two scores. Toussant also claims chronotonic distance is a particularly useful transformation-based method for describing the relationship between families of rhythms.[4] We select chronotonic distance as our evaluation metric for its status as a transformation-based method, its endorsement in the literature, and its ability to model what we consider the most important quantifiable aspects of rhythm: onsets, durations, and metrical structure.

To define chronotonic distance, we must first define a chronotonic sequence. A rhythm's chronotonic sequence is a series of squares organized left to right on a 2-dimensional plane. Each square represents a note; a square's left edge corresponds to a note's onset location, while a square's height and width correspond to note duration. Empty space in a chronotonic sequence corresponds to a rest. Other elements (including pitch) are not measured. Figures 1 and 2 display the same two measures as musical score and as a chronotonic sequence respectively. Note that on a longer timeframe, onset patterns reveal the underlying metrical grid in a chronotonic sequence.

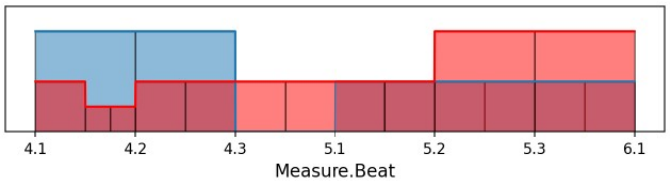

Fig. 3. Overlaid chronotonic sequences for measures 4 and 5 of *Dahl polska efter Ola Olsson* and *1814*.

We define the chronotonic **distance** between two scores as the sum of all geographic space occupied by one and only one chronotonic sequence when two chronotonic sequences are overlaid on the same graph. For example, Figure 3 overlays the chronotonic sequences of two scores, while Figure 4 illustrates their geographic disparity; the sum of the area shaded grey in Figure 4 is the chronotonic distance between measures four and five of the plotted scores. Under this framework, a "good" chronotonic distance corresponds to a lower number, which indicates similar onsets, durations, rests, and underlying metrical schemata. A "bad" chronotonic distance corresponds to a higher number, which indicates less overlap in the qualities listed above. While a distance of 0 corresponds to perfect rhythmic equality, it does not imply perfect equality in general, as certain unmeasured and irrelevant qualities (notably pitch) may still differ.

We may formalize the chronotonic distance between specific measures of score A and score B ($D_{A,B}$) as:

$$D_{A,B} = \sum_{m \in M} \sum_{i=1}^{m_\alpha} |C_{i,m,A} - C_{i,m,B}| \tag{1}$$

where *M* refers to the measures to be compared, m_α refers to the number of atomic beat moments in a given measure *m*, and $C_{i,m,A} - C_{i,m,B}$ refers to the area between score A and score B's chronotonic chains during atomic moment *i* of measure *m*. The set *M* within 1 is a random set of eight adjacent measures present in both score A and score B (to be clear, the starting measure of this set is a random odd number low enough to allow seven further measures to be pulled from each score, and in a cardinal sense, the same eight measures are pulled from score A and score B).

Letting A represent a score in our polska corpus, we calculate $D_{A,B}$ for all A and B where B can be:

1) One of 575 other polska scores.
2) One of 500 bigram-generated scores.
3) One of 500 randomly generated scores.

The results of these 907,200 comparisons are discussed in Section V below.

As a final comment on our evaluation metric, we note that chronotonic distance is blind to the type for geographic disparity measured. For example, the quarter rest just before measure five in Figure 1 is as geographically 'far away' from two eighth notes as the preceding quarter note; both have the same effect on distance. We discuss the implications of this impartial weighting in Section V as well.

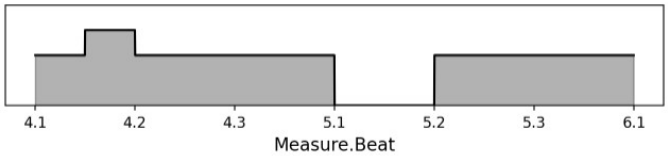

Fig. 4. Geographic difference in chronotonic sequences for measures 4 and 5 of *Dahl polska efter Ola Olsson* and *1814*. **The sum of this area is equal to distance.** In this case, the distance is 2.625 (note that it is the length of a quarter note which receives a geographic value of 1).

IV. Description of your model

We developed two models in this assignment: a Random model and a Generative model.

The Random model builds measures by one 'event' (note or rest) at a time. It first selects among the dotted-half, half, quarter, eighth, or sixteenth duration at random. It then decides, again at random, whether this duration will apply to a note or a rest. Combining these two decisions, it adds the selected event to an empty measure with a 3/4 time signature. It then removes any durations from the aforementioned set that would extend the measure beyond three beats (for example, if the Random model's first choice is the quarter note, it may no longer choose the dotted-half note or rest). The Random model then proceeds to select a second duration at random, to decide if this duration will apply to a note or rest at random, to add this second event to the measure, and to filter the possible set of durations from its latest onset location. This process continues until the measure has been filled with events that total exactly 3 beats.

We built 500 Random scores by using the Random model to create and distribute 4000 measures into sequences of eight. These 500 random rhythms act as the baseline for our Generative model.

By contrast, our Generative model, a Bigram model, builds rhythm based on observed occurrences of duration-onset pairs in our corpus. The Generative model first splits each score in our corpus into 'tokens' - numerical pairs which indicate a note's duration and location (with regards to beat) within its respective measure. Our model then links neighboring tokens within the corpus and creates a distribution of 2-pair sequences observed across all scores. We remark that each note will appear twice in the Bigram distribution - once with the note succeeding it, and once with the note preceding it. Notes at the beginning of scores are linked to an added 'start' token to allow the Generative model a token which indicates the beginning of a score.

Our Bigram model architecture is modeled after traditional natural language processing application. To begin a new rhythm, the Generative model is fed a 'start' token, and the onset and duration of the first note of the first measure is selected according to the probability distribution of all the onset-duration pairs of score-starting notes in our corpus. Correspondingly, the end notes of each measure are concatenated with an added 'end' token. The onset and duration of the second note is similarly chosen according to the probability

distribution of onset-duration pairs that followed the preceding onset-duration pair in the corpus. As with the Random model, however, selected durations must adhere to the 3/4 meter. This process of looking backward one step for a probability distribution at the current onset persists even across the meter threshold.

In some instances, our Generative model can observe a preceding onset-duration pair that has no succeeding notes in the corpus, denying the model the probability distribution it needs to generate its next note. This can occur precisely because our onset measure in each token is *measure*-based rather than *score*-based; it may be that the model outputs at the end of the first measure a duration-onset pair which is only seen in the corpus at the very end of a score (and which therefore cannot have succeeding notes). In such cases, we cancel and reset the Generative model's attempt at creating eight measures.

After using our Generative model to create 500 Generated rhythms eight-bars at a time, we calculated pairwise chronotonic distances between our Random, Generated, and Actual rhythms, as discussed in the next section.

V. Results and Discussion

We conducted four sets of pairwise chronotonic distance calculations as defined by Formula 1 for $D_{A,B}$:

1) Actual-Random: distance between a score in our corpus and a score from the Random set
2) Actual-Generated: distance between a score in our corpus and a score from the Generated/Bigram set
3) Actual-Actual: distance between a score in our corpus and another score from our corpus
4) Generated-Generated: distance between a Bigram score and another Bigram score

Average distance values across all scores by comparison set along with 95% confidence intervals are displayed in V. Relative to other comparisons, the elevated average chronotonic distance between scores from the corpus and randomly-generated scores is immediately apparent. Meanwhile, there is no apparent distinction in average distance between corpus scores and Generated scores, corpus scores and corpus scores, and Generated scores against Generated scores.

Seeking to measure the power of these average differences, we performed one-sided t-tests on the average distances between set pairings for all 576 scores in our corpus V We find that the average distance from corpus score to corpus score is statistically lower than the average distance from corpus score to baseline score in all cases. Meanwhile, for about 33% of scores in our corpus, the average distance from the corpus score to Generated scores was not significantly different from the average distance to other corpus scores - indicating that our Generative model can successfully trace mimic aspects of rhythm captured by chronotonic sequences about a third of the time.

Figure 6 displays the chronotonic sequence of one such successful Generated rhythm (red) along with the chronotonic sequence of a Slangpolska score from our corpus. (See 3 and 4

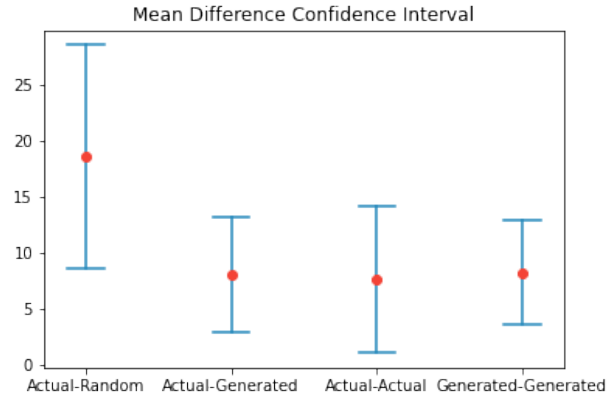

Fig. 5. Average chronotonic distances between different set pairings. Actual refers to the Slangpolska corpus, Random refers to the baseline model, and Generated refers to the Bigram model.

Comparison 1	Comparison 2	% of p-values <.05
Actual-Actual	Actual-Random	100%
Actual-Actual	Actual-Generated	67%
Actual-Generated	Actual-Random	100%

TABLE I
One-sided Student t-tist results. Average chronotonic distance between corpus scores and Generated scores are not significantly different for 33% of scores in the corpus.

for Actual and Generated sheet music respectively). In addition to the perfect overlay of measures 2, 6, and 8, geographic discrepancies in measures 1 and 7 are extremely minimal. This example showcases the ability of our Generative model to not only match patterns in onset and duration, but also to recover from slight discrepancies in duration.

However, our Generative model clearly cannot perform this well universally, as is exemplified in Figure 7. In this case, the Generative model appears incapable of matching rhythm characteristics. In fact, from measures 3 to 4 and 6 to 8, the two rhythms appear to be alternating in their note durations. Onsets too more often diverge than converge between the two scores. With our Generative model unable to sufficiently 'blend-in' with corpus rhythms for nearly 70% of corpus scores, it is clear our model is missing some of the signal it needs to rectify such mismatches in rhythm and onset patterns - a failing which surely bleeds into human perception of rhythm given the success of transformation-based distance metrics at modeling this phenominon.

One limitation to our model is we do not enforce any features on our Generated rythms. We could have, for example, studied the most common patterns of notes to occur in our

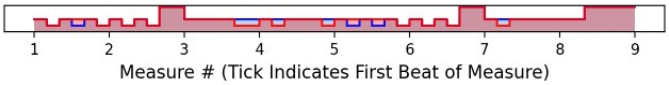

Fig. 6. 'Good' rhythm example: overlaid chronotonic sequences for a Slängpolska and bigram sequence.

corpus, and designed a model that was forced to conform these patterns more often, especially when they began approaching such patterns due to probabilisitc behavior.

As mentioned in the evaluation metric section, chronotonic distance as a metric poses some issues for our ability to claim we have matched (or failed to match) the corpus' rhythm. A rest may cause the same increase in chronotonic distance as a note, but only one registers a beat on a metrical grid. If our measure cannot capture the complexities of rhythm satisfactorily enough, we risk some measurement error in our results.

Future work might augment our transformation-based distance metric with feature-based metrics, extend our bigrams into trigrams (or longer n-grams), exchange our n-gram model for an LSTM neural network or Transformer-based neural network, or explore tree-based methods for rhythm creation. It is plain to see that a model with more memory could better approximate rhythm; metrical grids operate on the scale of measures, not notes, and our bigrams have but residual ability to preserve the grid through even short distances.

VI. Conclusion

Using bigrams to model rhythm and chronotonic distances to quantify our results, we successfully mimic the onset and duration behavior of a Slangpolska corpus about 33% of the time. We have learned that bigrams are a weakly sufficient tool for generating human-like rhythm, and that when they work well, it is thanks to their ability to stay in the neighborhood of common note patterns. A model which is more stringent about evoking these patterns and capable of longer memory might improve its performance as measured by chronotonic distance.

References

[1] Michael Scott Cuthbert and Christopher Ariza. Music21: A toolkit for computer-aided musicology and symbolic music data. In J. Stephen Downie and Remco C. Veltkamp, editors, *ISMIR*, pages 637–642. International Society for Music Information Retrieval, 2010.

[2] Godfried T Toussaint, Luke Matthews, Malcolm Campbell, and Naor Brown. Measuring musical rhythm similarity: Transformation versus feature-based methods. *Journal of Interdisciplinary Music Studies*, 6(1), 2012.

[3] Juan Beltran, Xiaohua Liu, Nishant Mohanchandra, and Godfried Toussaint. Measuring musical rhythm similarity: Statistical features versus transformation methods. *International Journal of Pattern Recognition and Artificial Intelligence*, 29:1550009, 03 2015.

[4] Godfried Toussaint. A comparison of rhythmic similarity measures. 01 2004.

Appendix

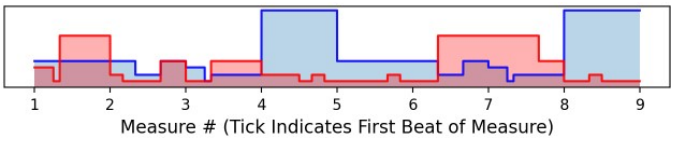

Fig. 7. 'Bad' rhythm example: overlaid chronotonic sequences for a Slängpolska and bigram sequence.

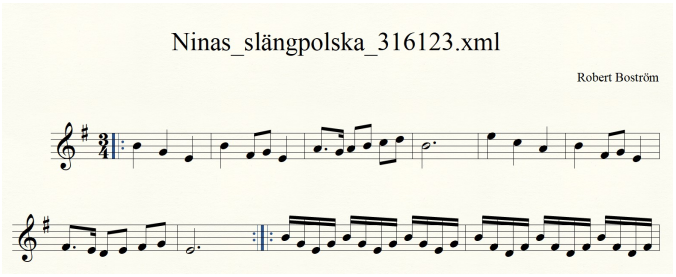

Fig. 1. Slängpolska score from the 'bad' rhythm example (see measures 1-8)

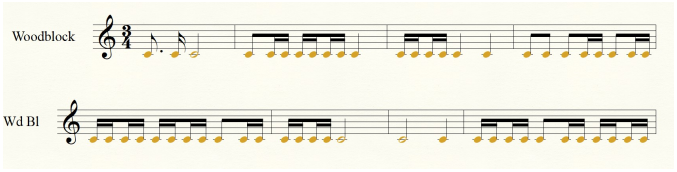

Fig. 2. Bigram-generated score #35; from the 'bad' rhythm example

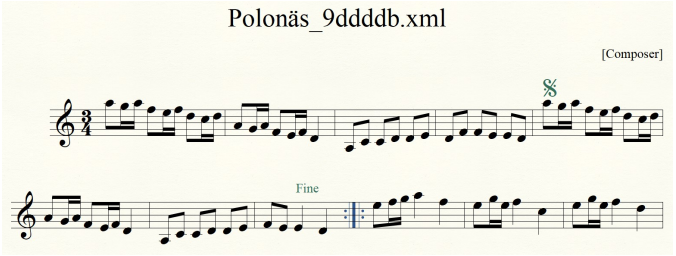

Fig. 3. Slängpolska score from the 'good' rhythm example (see measures 1-8)

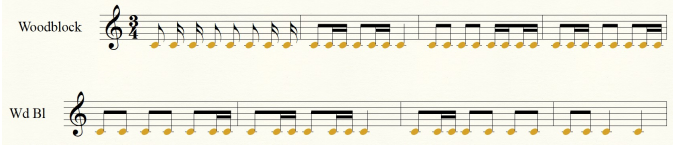

Fig. 4. Bigram-generated score #35; from the 'good' rhythm example