Digital Musicology (DH-401)
Assignment 3: Similarity

Group 6: Mickaël Achkar, Yinghui Jiang, Yichen Wang

May 4, 2022

1 Dataset preparation

In this project, we strive to implement a data-driven approach to compare the similarity between chorale incipits from a dataset of symbolically-encoded music. We were provided with a corpus comprised of 366 scores of the *Chorales* by Johann Sebastian Bach in CSV form. `Chorale043` was removed because the MuseScore file is corrupted, it only has one staff, and it does not include a closing fermata on or after the 4^{th} bar.
In fact, each incipit was taken until the fermata occurring on or after the 4^{th} bar.
For easier time series manipulation, we created a `global_onset` column by cumulatively summing the durations of previous notes of each staff (the first one initialized to zero) in each piece. The similarity between two pieces were measured with three key parameters, melody and melodic contour, rhythms, and harmony.

2 Approach

2.1 Rhythm

Rhythm consists of onsets and durations of musical events; thus, it is closely related to the metrical structure. The assumption is that metrical weights are proportional to the average across staves of note durations occurring at the same onsets. In this sense, onsets triggered simultaneously (ie. multiple voices) have more metrical weight than lone onsets which get smaller due to the averaging. The metrical weights of each incipit were computed. Since the length of incipits can be different, we rescaled them in 100 equally spaced bins to standardize the metrical weights and effectively use them for comparison.
To measure the similarity, the cross-correlation between two pieces was computed. The overlap between two signals becomes shorter as the lag increases, so a *Hann* window was applied to emphasize the middle of the resulting curve or smaller phase shifts, and in effect reward larger overlaps. After windowing the cross-correlation results, the L_∞ norm (the maximum cross-correlation value of the resulting signal, also plotted with a yellow bar) was used to get a single score for each pairs for similarity measurement. This measure provided results which were felt as much more similar than the ones we got with the $L_2 norm$. In some sense, the latter measure considers all other correlations, and may cumulate many insignificant rewards over time which become quite large.

2.2 Melodic Contour

Melodic contour generally describes a melody's pitch shape. Since the principal components of melodies are shaped like cosines, we proposed a low-dimensional representation of melodic contours using discrete cosine transform (DCT) inspired by [1]. We modelled all contours by taking the mean pitch across staves during simultaneous voices, or the current pitch if there was only one voice. Each incipit's minimum note duration is used for the fine-tune sampling and binning of onsets. Just like the provided CSV table, the representation ignores all rests by having non-zero pitch at all onsets. For standardization of the melodies, we sampled $N = 100$ equally spaced pitches from each modeled contour's distribution, thus all pieces are represented by vectors of 100 pitches of same length. We computed the first $D = 10$ cosine transform coefficients for each phase and projected contours on the results for a low-dimensional melody representation. This number was chosen to capture most of the explained variance while also generalizing the shape to allow a more abstract comparison between pairs of melodies. We discarded the first coefficient c_0 to make contour transposition invariant, which corresponds to a flat line and describes the overall pitch height of a contour.
For visualization and validation purpose of the method, we represented for each piece the cosine components

associated with coefficients c_1 and $-c_2$, for which the paper describes them as *descendingess* and *archedness*, respectively. From the plots, we can see that there is a large range of principal melody shapes.
We then computed the Euclidean distance (L_2 norm) between the D-dimensional coefficient representation of two pairs of melodies. A smaller distance indicates a larger similarity in melody.

2.3 Harmony

Harmony is defined as the interplay of note pitches occuring at the same onset. We grouped the pitches by global onset for each incipit in the dataset. Most pieces in the dataset have four staves, so there are six possible intervals between two notes. As the table shows, intervals can be categorized into 7 classes of two sound groups according to the difference of chromatic pitch-class (cpc) values ordering in semitones.

Interval class and *cpc* distance	0: (0)	1: (1, 11)	2: (2, 10)	3: (3, 9)	4: (4, 8)	5: (5, 7)	6: (6)
Consonances	P1/P8			m3/M6	M3/m6	P4/P5	
Dissonances		m2/M7	M2/m7				4+/5°

The interval classes is the shortest distance between two notes measured in semitones, and they are pitch transposition invariant. Thus, we calculated the intervals by subtracting the `midi` of two notes then converted them to cpc ordering in semitones by $\bmod 12$. In this way, we mapped the intervals to the interval class. We used a vector representation in $\mathbb{R}^9$ for harmony, which consists the frequencies of seven interval classes, as well as two additional classes: the frequencies of consonances and dissonances. Then the L_2 norm Euclidean distance was used to score the similarity between two incipits. Smaller distances indicates larger similarity in harmony.

2.4 Evaluation

Given the difficulty of an objective evaluation of our metrics, we conducted a limited and simplified subjective quality evaluation among us three by manually checking the best and worst few results to see if one is close to another within the pairs. Although possibly influenced by other attributes, we found the results are very satisfactory for subjective perception.
Also, some incipits are almost the same in scores and they simultaneously appear in the top few results of every measure that is independently implemented, which validates our approach's accuracy.
For example, for melodic contour, the best pair `Chorale009` and `Chorale361` are exactly the same in incipits, which also appears in top best results of harmony; few more examples are the combination between `Chorale130`, `Chorale320`, and `Chorale358`. In rhythm measurement, since we only used the incipits to extract metrical weight, short incipits with fewer notes can have good correlation score with many others.
Based on our perceptual experience, melodic contour tend to capture our hearings best, since it's assumed that human brains are sensitive to pitches. Rhythm is the second easy to perceive with the help of metronome. Although harmony may need professional training, consonance and dissonance are obvious to our hearing since melodies and harmony all based on interval compositions.

3 Conclusion

In this assignment, we adopted data-driven approach to measure the similarity between pieces by three parameters. We can conclude that each of our measures considered important invariances for each similarity feature. For rhythm, our measure was invariant to phase shift of the beat, to a certain extent. In fact, by windowing the signal, more importance is given to smaller relative phase shifts done in the cross-correlation (ie. giving more importance to larger overlaps). For melodic contour, our measure was pitch transposition-invariant and low-dimensional to capture a general shape that must be comparable. Finally, for harmony, our measure was pitch transposition-invariant (by using cpc and 7 grouped interval classes), as well as time-invariant (it does not consider the timing of the chords). A possible improvement would be to combine these parameters to create a compound measure that could reflect our intuition better. This would require human cognitive knowledge to model parameters with different weights according to hearing sensitivity. In our study, melody would be weighted most and follows by rhythm and harmony, which could be determined by previous subjective experiments results.

References

[1] Bas Cornelissen, Willem Zuidema, John Ashley Burgoyne, et al. “Cosine Contours: a Multipurpose Representation for Melodies”. In: *ISMIR* (2021).