Digital Musicology (DH-401)
Assignment 1: Meter and time signatures

Group 6: Mickaël Achkar, Yinghui Jiang, Yichen Wang

March 23, 2022

1 Dataset preparation

For this project, we were provided with a corpus comprised of scores from the *Piano Sonatas* by Wolfgang Amadeus Mozart and the *String Quartets* by Ludwig van Beethoven, in CSV form. Relevant columns were selected, including `piece`, `staff`, `timesig`, `voice`, `global_onset`, `duration`, and `tied`. We assumed that grace notes do not have a significant impact on inferring time signature: they are sparse, require more data manipulation as their duration is only specified in `nominal_duration`, and their ties are not clearly encoded in the `tied` column. The unit of time of `global_onset` is in quarter notes: $\frac{1}{4}$.

1.1 Tie aggregation

To assess whether tied notes impact determination of meter, we aggregated them by duration: we sum the duration of tied notes and assumed that they are equivalent to a longer note occurring at the `global_onset` of the first tied note in the series. We sorted our dataset by `piece`, `staff`, `voice`, `global_onset` to allow relevant tie aggregation. One can freely choose between the tied and untied dataframes and assess their classification impact.

2 Approach

2.1 From musical score to metrical weights

For a preliminary analysis, we selected a random piece and staff pair to see if there is a pattern related to the time signature. Since time signatures might change per staff, both piece and staff information were taken into consideration. Even if there are cases of time signature changing in the *same* staff, we assumed in our analysis that they do not, since these cases are sparse and add complexity in the analysis. The minimum note duration in the selected piece and staff is found and used for fine-tune sampling and binning of onsets. We assumed that metrical weights are directly proportional to the sum of note durations occurring at the same `global_onset` and hence computed these sums (Eq. 1). From the plot, it is obvious that some onsets have higher metrical weights than others, which implies possible periodicities.

$$w(t) \propto \sum_{n|onset(n)=t} duration(n) \tag{1}$$

2.2 From metrical weights to periodicities

2.2.1 Discrete Fourier Transform

This project is based on the hypothesis that metrical grids could reveal the patterns of periodicities. However, it is difficult to infer the time signature from metrical grids computationally without some pre-processing. If we consider metrical grid as a discrete signal, we could identify the most salient frequencies though a Discrete Fourier Transform (via FFT algorithm). As a good practice, we use a Hann window to reduce spectral leakage (hence reducing large spikes) of the signal when computing the FFT. We identify the frequency peaks with a minimum inter-frequency distance of $7\% \cdot N_{samples}$.

2.2.2 Autocorrelation function

Furthermore, we computed the autocorrelation function (ACF) of the metrical grid: this plots the correlation of the original signal with a lagged copy of itself as a function of lag. A high correlation peak at a corresponding lag indicate high recurrence of pattern along the metrical grid. The confidence interval of the ACF decreases with lag (the overlap with the copied signal becomes shorter), hence we only plot and select the ACF coefficients corresponding to a lag range with the length of the maximum duration of a bar found in the whole dataset (i.e. $\frac{9}{4}$).

In an effort to distinguish between *duple* and *triple* meters, we used a modified criterion M (Eq. 2) inspired from [1]. The more positive, the more it is representative of a duple meter; the more negative, the more it represents a triple meter. For each piece and staff, we computed this criterion 3 times, for each possible beat unit [$\frac{1}{2}$, $\frac{1}{4}$, $\frac{1}{8}$].

$$M_{beat\ unit} = \frac{acf(2b) + acf(4b) + acf(8b)}{3} - \frac{acf(3b) + acf(6b) + acf(12b)}{3} \tag{2}$$

2.3 From periodicities to time signature: supervised classification pipeline

2.3.1 Chosen features and labels

We have used a data-driven approach, a supervised classification pipeline to infer the time signature of 388 piece-staff pairs from 16 distinct features: the top 8 positive autocorrelation coefficients in the chosen lag window, the top 5 signal frequencies, and 3 duple/triple decision criteria (one per beat unit). If there were not enough features, we replaced the missing ones with a 0. This was later found to be better performing than replacing with the average of features. The labels are the first occurring time signature in the piece-staff pair.

2.3.2 Classifier: Random Forest and evaluation of model

Random Forest was chosen as our main classifier. The whole dataset was randomly split into training (75%) and testing (25%) data.
For the untied dataset, the accuracy of our model was 67% on average, sometimes performing as high as 78%. It performed well in inferring simple and compound common meters such as $\mathbf{\frac{2}{2}}$, $\mathbf{\frac{2}{4}}$, $\mathbf{\frac{3}{4}}$, $\mathbf{\frac{3}{8}}$, and $\mathbf{\frac{6}{8}}$. The most common confusions were between time signatures with similar structures, such as ($\mathbf{\frac{2}{2}}$, $\mathbf{\frac{4}{4}}$), ($\mathbf{\frac{2}{2}}$, $\mathbf{\frac{2}{4}}$), ($\mathbf{\frac{2}{4}}$, $\mathbf{\frac{4}{4}}$), ($\mathbf{\frac{2}{4}}$, $\mathbf{\frac{3}{4}}$), and ($\mathbf{\frac{3}{4}}$, $\mathbf{\frac{6}{8}}$). Some time signatures such as $\mathbf{\frac{3}{2}}$, $\mathbf{\frac{9}{8}}$, and $\mathbf{\frac{12}{8}}$ were not always correctly predicted, possibly due to their sparsity in the dataset, not allowing the model to effectively learn their structure.
The addition of the duple/triple decision criteria increased the performance of classification as it may capture well the multiplicity of beat structure.
The cross-validation curve for a 10-fold cross-validation shows a bit of overfitting in the variance of testing score while the training score is almost 100%. This is possibly due to having a small dataset of 388 datapoints.

For the tied dataset, the classification accuracy was lower, at around 55%. A possible explanation would be that the sum of durations at an onset is overly high for tied notes relative to the rest of the onsets, or that aggregating tied notes may remove an implicit link between rhythm and meter.

3 Conclusion

In this assignment, we realized that we had to make multiple assumptions and decisions regarding data selection and manipulation. A possible improvement would be to better capture the multiple levels of beat, with carefuly crafted tools such as Eq. 2. In fact, we have not used the individual autocorrelation coefficients or frequency amplitudes as features, but rather the top positive n lags in a specified window or the top n frequencies, which do not capture their relative importance. These continuous features may provide more details about the structure of the mapping.
Moreover, this study highlights the possible challenges at stake when manipulating musical corpora or other cultural datasets. The computational tools used may not reflect the cognitive aspects that are inherently human in the perception of meter and time signature.

References

[1] Fabien Gouyon and Perfecto Herrera. "Determination of the meter of musical audio signals: Seeking recurrences in beat segment descriptors". In: *Audio Engineering Society Convention 114*. Audio Engineering Society. 2003.