Digital Musicology (DH-401)
Assignment 2: Pitch and keys

Group 6: Mickaël Achkar, Yinghui Jiang, Yichen Wang

April 6, 2022

1 Dataset preparation

In this project, we strive to implement a data-driven approach to infer the key (tonic and mode) of musical passages from a dataset of symbolically-encoded music. We were provided with a corpus comprised of scores from the *Piano Sonatas* by Wolfgang Amadeus Mozart, the *String Quartets* by Ludwig van Beethoven, and some movements from the *Trio Sonatas* by Arcangelo Corelli in CSV form.
We created additional columns with the chromatic pitch-class (`cpc`) for each note of the dataset, both ordered in fifths and in semitones. For the labeled dataset, we also create the `cpc` for each global key, and for each local key. This conversion is done for ease of manipulation and standardization of the data. To get the ordering by fifths, we use $\texttt{cpc\_f} = \texttt{tpc} \mod 12$, and to get the ordering by semitones, we use $\texttt{cpc\_s} = \texttt{midi} \mod 12$. Since `midi` is not available for `globalkey_tpc` and `localkey_relativetpc`, we first convert these columns to `cpc` ordered in fifths, then use an empirical *lambda* function called `fifths_to_semitones` to order them into semitones.
For local keys, `cpc` values are obtained by adding the `localkey_relativetpc` values with each piece's `globalkey_tpc` value and applying the $\mod$ and *lambda* functions.
For the data-driven mode prototype creation, we have also transposed all the pieces' `cpc` values to their tonic reference given as ground truth by subtracting `globalkey_cpc` and applying the $\mod 12$ function to make sure they still lie in $[0; 11]$.

2 Approach

2.1 Constructing mode prototypes through data

Each key is identified by a tonic and a mode. To infer the tonality of a piece, a training subset of the labeled dataset was used to generate two prototypes of the hierarchical importance of different pitch-classes: one for the major mode, and one for the minor mode. To create them, the duration of each pitch-class of the major and minor training pieces was summed to estimate the importance of notes, and normalized by the total count of pitch-classes (the L1-norm), since the key of a piece does not depend on the frequencies of pitch-classes [1].
For the two prototypes, we found that the dominant, tonic, supertonic, mediant, subdominant, and submediant are significant in the hierarchies for modes. Conforming to the musical theory, the major mode has an emphasis on the leading tone and the minor mode has an emphasis on the subtonic. Besides that, the leading tone also appeared important in minor mode, which confirms the common $\sharp 7$ accidentals employed by musicians playing in harmonic or melodic minor.
We used a vector representation of pieces and prototypes in $\mathbb{R}^{12}$ inspired by [1] as the basis of the mathematical space in which pieces that have similar pitch-class counts are close to one another. To estimate if a piece is close to a certain prototype, we use in a similar fashion to [1] the key distance $d_{\mathbb{K}}$ (a Euclidian distance) (Eq. 1) between its vector representation and the two mode prototype vectors transposed to each of their 12 possible transpositions of the chromatic pitch space $\left((\sigma_i(p))_k = p_{k-i \mod 12} \text{ for } k \in \mathbb{Z}^{12}\right)$. The minimum of these key distances will characterize the best corresponding prototype (tonic and mode) for a certain piece (Eq. 2). From the plots, it is obvious that notes weigh differently in different keys, which provides profiles for key estimation.

$$d_{\mathbb{K}}(p,q) = \left\| \frac{p}{\sum_i p_i} - \frac{q}{\sum_i q_i} \right\|_2 \tag{1}$$

$$d_{\mathbb{M}}(p,q) = \min_{i \in \mathbb{Z}_{12}} d_{\mathbb{K}}(\sigma_i(p), q) = \min_{i \in \mathbb{Z}_{12}} d_{\mathbb{K}}(p, \sigma_i(q)) \tag{2}$$

2.2 Global key estimation

The approaches mentioned in the previous section constitutes our classifier. First, we took 70% of labeled pieces as the training set and separated them into the major and minor subsets. Then, we processed the two sets by transposing them to their tonic reference given as ground truth to generate the prototypes for the two modes. We computed the key distance between each piece in the testing set with the 12 transposed hierarchies of the two prototypes. The prototype which has the least distance with the test piece would be the predicted global key.

2.3 Local key modulation identification

We analyzed a few random pieces and found that key modulations usually happens at the beginning of a bar. We chose to split the pieces in the corpus by bars to better identify the changes in local key. To identify the local key for every n-bar sized window of a piece, we compared its distribution of pitch-classes with the 12 transpositions of the two prototypes derived from global key training data. The challenge was to find an appropriate window size for the splitting of a piece. We implemented a cross-validation for our local key classifier by varying the window size and found the optimal size of 16 bars.

2.4 Model evaluation

To estimate the performance of both global and local key classifiers, the dataset was randomly split and shuffled 50 times into training (70%) and testing (30%) data.

For global key estimation, the accuracy after 50 folds of our model was on average 93.38% with a low variance of 0.00092 ($\sigma = 3.0\%$), which implied close prediction for most pieces in the corpus. Overall, it performed well in global key estimation. For the few pieces that may not be inferred correctly, it is possibly due to ambiguity in closeness to prototypes in their vectorized key spaces.

For local key modulation identification, the accuracy after 50 folds of our 16-bar window model was on average 69.91% with a low variance of 0.00035 ($\sigma = 1.9\%$). The cross-validated window size parameter increased the average accuracy of the model. Compared with the global key classifier accuracy, the lower accuracy for the local key classifier may be explained due to the usage of a constant window size for all the pieces as well as human expert annotation disagreements. In fact, their could hardly be any consensus in a modulation and its exact boundaries due to human expert biases and ambiguity of perception.

2.5 Classification on unlabeled pieces

For the rest of the 20% unlabeled data for global keys and local keys we repeated the previous procedures to predict these pieces. Results were exported to `PredictedKeys.csv`. Considering the previous satisfactory performance and low variance of our classifiers, we expected that they will have around 93% accuracy in inferring global keys, while around 70% accuracy in identifying local keys.

3 Conclusion

In this assignment, we adopted data-driven geometrical models based on duration-weighted pitch-class distributions. This study not only highlighted the high certainty in predicting a piece's global key but also the possible challenges when investigating local key modulations. The prototypes derived by global key training data may not provide sufficient insights for regional tonality. A possible improvement would be to optimize the window size for each piece. In fact, a one-size-fits-all approach seems rather naive given the diversity of the corpus. Another improvement would be to capture a modulation via another detection mechanism, based on sudden pattern irregularities. But more importantly, even humans do not share a consensus on the ground truth.

References

[1] Daniel Harasim et al. “Exploring the foundations of tonality: Statistical cognitive modeling of modes in the history of Western classical music”. In: *Humanities and Social Sciences Communications* 8.1 (2021), pp. 1–11.