

# Prediction of musical phrase endings

Shamkhal Guliyev

December 2024

The subject of this project is to prepare data, explore it, and build and compare predictive models for musical phrase endings.

## 1 Introduction

Disclaimer: I am not a musicologist. I describe the field in my own words with a very simplified view of a domain that is far more complex than this brief description.

Music can be studied in a way similar to language. This leads musicologists to sometimes use the same vocabulary in music as in literature or linguistics, even though many differences arise. A musical piece can be played or written, just as a text can be spoken or written. Like a text, a musical piece is structured. This structure can be described as a sequence of phrases. With a little attentive listening, we can identify these musical phrases, even if they are not explicitly noted in the writing of the piece.

Several factors contribute to identifying the end of a phrase. Of course, the end of the piece is also a phrase ending. However, these endings can appear in many places within the piece and are sometimes difficult to identify. Indicators of these endings may include, for example, the sequence of notes played, rhythm, etc. While some rules could be provided, they are not always valid or precise. This is where machine learning can help. If we have annotated corpora, we can implement algorithms to learn these phrase endings. That is the task you will undertake.

## 2 Data

The data at your disposal come from the Meertens Tune Collections. This is a corpus of melodies from the Netherlands, written (we are not working with sounds but with symbolic textual representations of music).

The data has been prepared by specifying a set of attributes. Each musical piece is essentially viewed as a JSON file. Some attributes are metadata describing the piece (type, year, etc.) and are not relevant to the intended task. Others describe the melody as sequences of values: the sequence of notes, the

sequence of durations, etc. These useful attributes are grouped under a **feature** attribute.

The package described at <https://pvankranenburg.github.io/MTCFeatures> provides access to the data and an API for manipulation.

The *playground.ipynb* notebook gives a basic example of using this API.

### 3 Task 1: Preparation of data

The first step is to prepare the data for prediction. We could use different types of models, but we will restrict ourselves to those studied in the course this semester. To do this, you will need to construct matrices representing descriptions of parts of melodies, some corresponding to phrase endings and others not.

The technique involves constructing subsequences of a certain length from all sequences in all melodies. Of course, it would not be wise to construct all possible subsequences. You may try to understand why.

Also, it will not be optimal to consider all feature attributes. An objective of this first task is to select the best attributes.

### 4 Task 2: Learning

We expect:

- The comparison of different learning methods,
- The tuning of various hyperparameters,
- Identifying the optimal set of attributes for predicting musical phrase endings.

Iterations between this task and Task 1 (data preparation) will likely be necessary.

An essential aspect of this task is choosing evaluation methods for your results and clearly justify these choices.

The critical part of this work is to demonstrate the methodology you followed. The evaluation of your work will primarily focus on this demonstration. In other words, achieving the highest score in the group does not guarantee the best grade!

### 5 Task 3: Interpretation

This entire study should allow you to draw insights from your findings and attempt to make sense of them. It is essential to adopt a critical perspective and interpret the results.

## 6 Organization and expectations

The project must be carried out in pairs. You must submit a written synthesis document with one or more codes, either as Python script files or notebooks. The quality of the synthesis in terms of structure, grammar, and clarity, as well as the clarity of the code and the presence of informative comments, will significantly influence the final grade.

Links:

- Meertens Tune Collections: <https://www.liederenbank.nl/mtc/>
- Attributes: <https://pvankranenburg.github.io/MTCFeatures/melodyrepresentation.html>