

---

# **Introduction to Musical Corpus Studies**

***Release 0.0.1***

**Fabian C. Moss**

**Nov 09, 2020**



# CONTENT

<b>1 Organization</b>	<b>3</b>
1.1 About this course . . . . .	3
1.2 Overview . . . . .	3
1.3 Credits . . . . .	4
1.4 Deliverables and learning objectives . . . . .	4
<b>2 Folk Songs and the Melodic Arc</b>	<b>7</b>
2.1 Data . . . . .	8
2.2 Notes, Pitch Classes . . . . .	8
2.3 Melodic Arc . . . . .	9
2.4 Intervals . . . . .	9
<b>3 Solos in the Weimar Jazz Database</b>	<b>11</b>
<b>Bibliography</b>	<b>13</b>





**Warning:** This material is still (heavily) under construction and might change throughout the course!

You can help improving the course and [let me know](#) about any errors and inconsistencies that you find or suggest other ways of improving the course.

## Welcome!

These pages present the content of the course “Introduction to Musical Corpus Studies” at the [Institute of Musicology](#), given at [University of Cologne](#) in Fall 2020.

In the last two decades *Musical Corpus Studies* evolved from a niche discipline into a veritable research area. The growing availability of digital and digitized musical data as well as the application and development of modern methodologies from computer science, machine learning, and data science cast new light on old musicological questions and generate entirely novel approaches to empirical music research.

Moreover, the general methodological and epistemological approach of Musical Corpus Studies allows to transcend traditional intra-musicological boundaries between its sub-disciplines (historical/systematic/ethnological/...) without sacrificing the respective specific viewpoints and perspectives.

This course offers a fundamental and practical introduction into these topics. It demonstrates, explores, and critically reflects central thematic areas and methods by means of a number of case studies. In the engagement with these topics the course also introduces elementary methods from natural language and music processing, as well as statistics, data analysis and visualization.

The course is aimed at students at the undergraduate level who have little or no empirical background and are curious about quantitative approaches to musicology.



**ORGANIZATION**

## 1.1 About this course

This course aims at providing an example-based introduction to the rapidly developing field of Musical Corpus Studies (MCS). Introducing a field that relies equally on musicological domain knowledge and skills in computational and statistical methods faces obvious challenges: while most people interested in this field come with a background in either area, few people are versed in both, and it can take years to bridge the musicological-computational gap.

In particular, systematic introductions to programming or specific musicological topics can be at times quite arduous, even boring, because it takes a long time to proceed from learning basic concepts to actually interesting problems. The problems and “toy examples” that are presented to introduce the basic concepts are necessarily remote from real-world applications and challenging research problems.

This course takes an alternative route. It does not start with an introduction to the programming language `Python` (which will be used throughout to carry out the computational analyses) but rather showcases a number of recent corpus studies that take on musicological research questions. The focus thus lies in understanding how aspects of music can be studied with computational methods and by analyzing musical corpora.

If this sparks your interest, it will be much easier to pick up the basics for yourself, knowing what they are *for* and being motivated intrinsically. If you are not particularly interested in doing this kind of work yourself, you will still see a broad range of applications that are much more useful to you than learning (or not learning) programming basics.

## 1.2 Overview

This year’s course takes place on two weekends (13-14 November and 11-12 December 2020), comprising twelve sessions in total. The topics cover a broad range of musicological topics, from folk melodies and Jazz solos, over harmonies in Beethoven’s string quartets and 20th century Pop music, to Renaissance cadences and metric patterns in Malian drum music (see [Table 1.2](#)).

No.	Date	Time	Topics
1	Fr., 13.11.2020	16:00-17:20 Uhr	Introduction / Background
2		17:40-19:00 Uhr	Melody I: Pitches, intervals, folk song melodies
3	Sa., 14.11.2020	09:00-10:20 Uhr	Melody II: Jazz solos
4		10:40-12:00 Uhr	Harmony I: Beethoven's string quartets
		12:00-13:00 Uhr	<i>Lunch Break</i>
5		13:00-14:20 Uhr	Group work
6		14:40-16:00 Uhr	Harmony II: Pop charts (Billboard 100)
7	Fr., 11.12.2020	10:00-11:20 Uhr	Harmony III: Cadences in Renaissance polyphony (guest: <a href="#">Richard Freedman</a> )
8		11:40-13:00 Uhr	Harmony IV & Form: Brazilian Choro
9	Sa., 12.12.2020	09:00-10:20 Uhr	Rhythm & Meter: Malian percussion music
10		10:40-12:00 Uhr	Timbre: Electronic Music 1950-1990
		12:00-13:00 Uhr	<i>Lunch Break</i>
11		13:00-14:20 Uhr	Group work
12		14:40-16:00 Uhr	Recapitulation and conclusion

## 1.3 Credits

Active participation in this course is compensated with 3 credit points (CPs), equivalent to a work load of 90 hours. These are distributed as follows: 24 SWS (à 45 minutes) are allocated to presence in the block seminar. Additionally, 24 SWS are dedicated to the preparation and follow-up of the material. The remainder of 42 SWS goes to the reading of the relevant literature.

## 1.4 Deliverables and learning objectives

Apart from attending and following the presentations by the lecturer, course work consists of three main parts: preparing the relevant literature (reading), completing the assigned exercises (group work), and critically engaging with the course materials in the form of a report written together with your group (report).

These deliverables will broaden your knowledge and understanding of current musicological research, enhance your organizational and social skills, and help you to develop efficient work-load management strategies. Finally, compiling a report will advance your communication and writing abilities.

### Reading

For each session, the relevant literature is cited in the text and provided on [ILIAS](#). Careful preparation of the reading material is required in order to be able to follow the content of the course. Because the course will mainly talk about methods and general points of musical corpus research, the content (and musical topic) will mainly be introduced by the literature.

I am aware that the reading workload is relatively high since the course will be taught as a block seminar and doesn't spread out over the entire semester. I hope that the fact that the course is finished before the end of the year compensates for this.

### Group work

At the beginning of the course, you will be randomly assigned to a group. Together with your group (which will stay fixed for the entire semester), you will work on a number of exercises during the course, e.g. in Zoom breakout rooms. You will collaborate on specific tasks related to the topic at hand, discuss methodological questions, and help each other in the understanding of some of the concepts that are introduced in the course.

### Report

After the course has ended, your group will be randomly assigned a course topic (one of the twelve sessions in Table 1.2). It is your task to write a report on this theme. Questions that you could address are: What did you learn? Which concepts are not clear? Which methods did you (not) understand? What is missing? How can the textual descriptions be improved? Who in your group did what? Was the presentation of the material adequate? If not, what was missing? Please also write about the organization of your group, challenges and benefits.

#### Recommended structure for the report

1. **Introduction:** general description and summary of the course and your assigned session in particular.
2. **Discussion:** summarize the main discussion, open questions, and how you would continue this line of research.
3. **Issues:** describe in detail what was crucial for your understanding of the topic, what was missing, etc.
4. **Various:** anything that you would like to write in the report
5. **Author contributions:** describe briefly how each of you specifically contributed to the report.

---

**Important:** Submit your report by **31 January 2021, 23:59h** to [fabian.moss@epfl.ch](mailto:fabian.moss@epfl.ch) as a single PDF file per group, named *intro\_corpusmus\_<group\_number>.pdf*, e.g. *intro\_corpusmus\_1.pdf*.

---

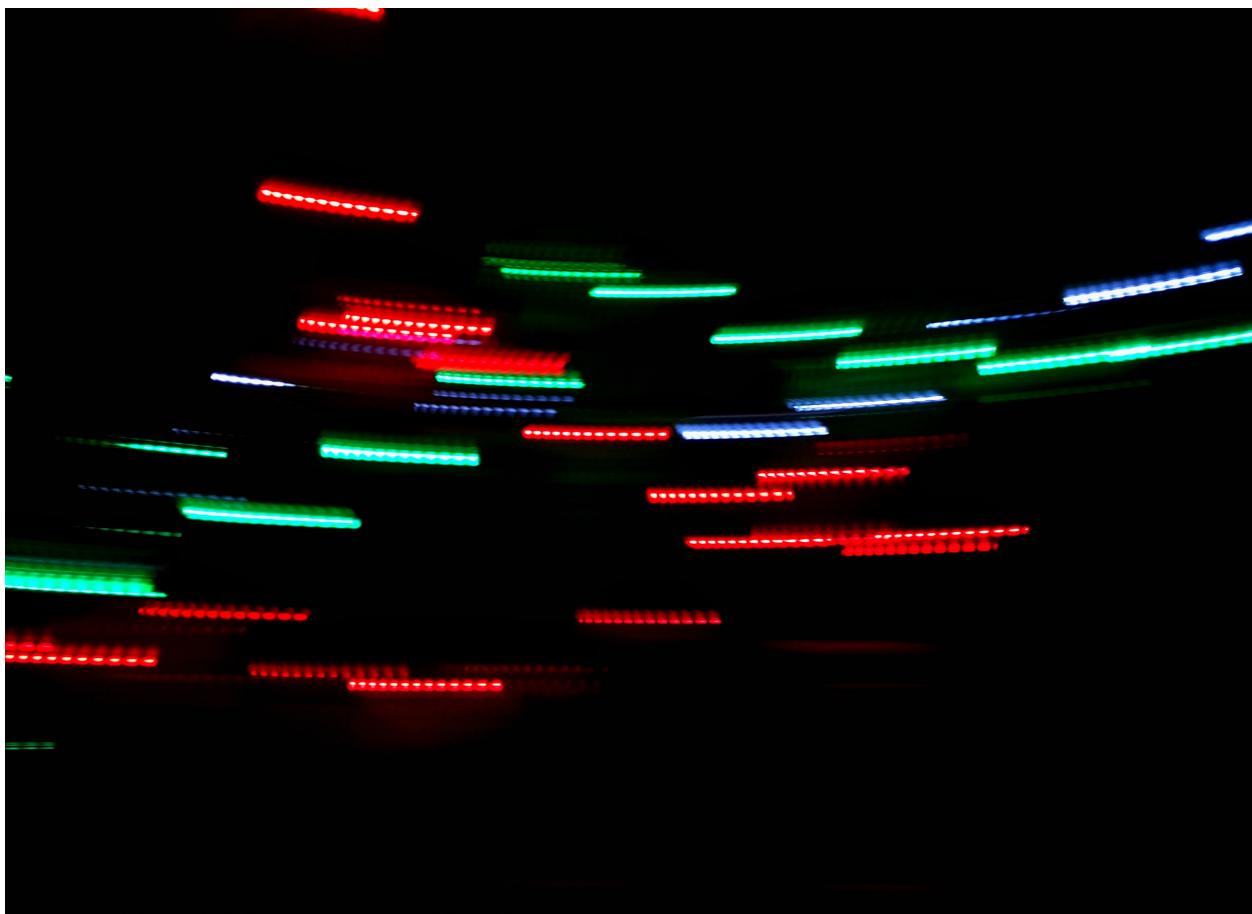


---

CHAPTER  
TWO

---

FOLK SONGS AND THE MELODIC ARC



Tones are among the basic elements of music. Most musical styles combine tones in different ways to create songs, chants, instrumental pieces, or other elaborate compositions. In this chapter, we will analyze some basic aspects of songs by studying distributions of tones and intervals.

Huron... / MusThe Tutorial

## 2.1 Data

<http://kern.humdrum.org/data?f=zip&l=/essen> or <http://kern.humdrum.org/help/data/>

Open and read *README.txt*

Essen Folksong Collection

- keine Texte!

### Die plappernden Junggesellen



Fig. 2.1: German song *Die plappernden Junggesellen* from the Essen Folksong Collection.

Analysis:

- AABA' form
- ascending / descending motives (local level) but also overall
- A' part elaboration of A by insertion of passing notes
- B part movement from ^3 to ^2

## 2.2 Notes, Pitch Classes

[https://github.com/DCMLab/DigitalMusicologyExercises/tree/master/tone\\_profiles](https://github.com/DCMLab/DigitalMusicologyExercises/tree/master/tone_profiles)

means, variance

also multidimensional (for later)

## Shengsi liangxianglian



Fig. 2.2: Chinese song *Shengsi liangxianglian* from the Essen Folksong Collection.

### 2.3 Melodic Arc

Melodic arc was studied first by [Hur06].

### 2.4 Intervals

[https://github.com/DCMLab/DigitalMusicologyExercises/tree/master/interval\\_bigrams](https://github.com/DCMLab/DigitalMusicologyExercises/tree/master/interval_bigrams)

maybe extend with Hansen and Pearce (2014) (but data not available?)

---

**Note:** In this chapter we covered the following musical terms:

- a
  - b
  - c
-

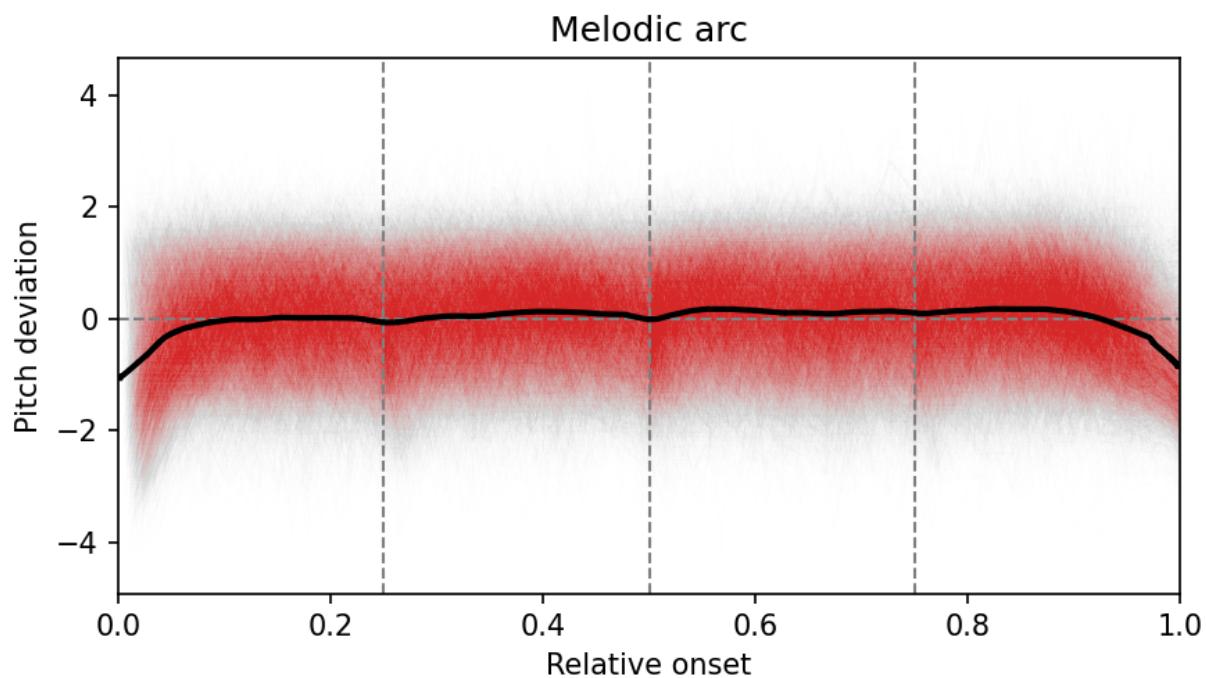


Fig. 2.3: Melodic arc. The red lines in the background show the melodic profile of each song in the Essen Folksong Corpus; the thick black line shows the melodic arc that was obtained by using *Locally Weighted Scatterplot Smoothing* (LOWESS) [CD88]. The dashed horizontal line marks the mean standardized pitch, and the vertical dashed lines mark the quartiles of the songs, showing that most songs also have an arc-like shape at local formal levels.

## SOLOS IN THE WEIMAR JAZZ DATABASE



Fig. 3.1: Photo by Janine Robinson on Unsplash

The first project we will have a look at is the Jazzomat project. Transcriptions of Jazz solos [PFA+17]. The *Weimar Jazz Database* (WJD) consists of 456 transcriptions of Jazz solos from diverse substyles. As all the corpora that we deal with here, it is freely available on the internet.<sup>1</sup>

The project is described in [PFA+17].

The WJD contains a number of tables:

---

<sup>1</sup> <https://jazzomat.hfm-weimar.de/dbformat/dboverview.html>

Table 3.1: Tables in the *Weimar Jazz Database*

Table name	Description
beats	Table for beat annotation of WJD melodies, referenced by melody (melid)
composition_info	Infos regarding the underlying composition of a WJD solo, referenced by melody (melid)
db_info	Information regarding the distributed database file like version information, license, etc
esac_info	EsAC infos for EsAC melodies, referenced by melody (melid)
melody	Main table for all melody events
melody_type	Indicated type of melody: WJD solos or EsAC (Folk songs using Essen Associative Code), referenced by melody (melid)
popsong_info	Pop song infos, referenced by melody (melid)
record_info	Infos regarding the specific audio recording of a WJD solo was taken from, referenced by melody (melid)
sections	All sections (phrase, chorus, form, chords, etc.), referenced by melody (melid)
solo_info	Solo infos for WJD solos, referenced by melody (melid)
track_info	Information specific to a track on a record (or CD)
transcription_info	Transcription infos for WJD solos, referenced by melody (melid)

Here, we focus on the main table `melody`. First, we download the entire database from <https://jazzomat.hfm-weimar.de/download/download.html> (under “Weimar Jazz Database”) and save it as the file `wjazz.db`.

```
import sqlite3 # for working with databases
import pandas as pd # for working with tabular data

# create connection to database
conn = sqlite3.connect("wjazzd.db")

# read all entries of the 'melody' table into a pandas DataFrame
df = pd.read_sql("SELECT * FROM melody", con=conn)

df.head()

>>> df.head()
output
```

The part of the code `SELECT * FROM melody` reads “Select all entries from the table ‘melody’”.

## BIBLIOGRAPHY

- [CD88] William S Cleveland and Susan J Devlin. Locally Weighted Regression: An Approach to Regression Analysis by Local Fitting. *Journal of the American Statistical Association*, 83(403):596–610, 1988.
- [Hur06] David Huron. *Sweet anticipation: Music and the psychology of expectation*. MIT Press, 2006.
- [PFA+17] M Pfleiderer, K Frieler, J Abeßer, W G Zaddach, and Burkhart B, editors. *Inside the Jazzomat: New Perspectives for Jazz Research*. Schott Campus, Mainz, Germany, 2017.