

# **Introduction to musical corpus studies**

Fabian C. Moss

2023-02-13

# Table of contents

<b>Welcome</b>	<b>3</b>
About this course . . . . .	3
Learning objectives and deliverables . . . . .	4
Reading . . . . .	4
Group work . . . . .	4
Report . . . . .	5
Organization . . . . .	5
Credits . . . . .	6
<b>References</b>	<b>7</b>

# Welcome

These pages collect the materials for the course *Musikalische Korpusforschung*, taught in the spring term of 2023 at the Institut für Musikforschung at Julius-Maximilians-Universität Würzburg.

## Note

The course takes place on Wednesdays from 8 to 10 a.m. in room 107 (CIP Pool, Domerschulstr. 13). The materials will be updated weekly after the course.

## About this course

This course is designed to introduce students to the fascinating research area of *Musical Corpus Studies* (MCS), which involves analyzing large collections of musical data to gain new insights into the structure and style of music. Students will be exposed to various techniques for collecting and processing musical data, such as audio feature extraction, transcription, and musical annotation. Additionally, they will learn about different approaches to analyzing musical data, including machine learning, statistical modeling, and network analysis. The focus will be on understanding how music can be studied with computational methods and by analyzing musical corpora.

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. Rather than starting with an introduction to the programming language [Python](#), which will be used throughout the course to carry out the computational analyses, students will be introduced to a number of recent corpus studies that take on musicological research questions. The course aims to provide an example-based introduction to MCS, which will motivate students intrinsically and enable them to pick up the basics more easily. Throughout the course, students will work on hands-on projects to apply what they have learned and to gain a deeper

understanding of the field. 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.

## **Learning objectives and deliverables**

This course is suitable for students with a background in music, computer science, or related fields who are interested in exploring the practical applications of MCS, including the study of musical genre, style, and expression, as well as the use of musical corpora in musicology, music education, and music technology. By the end of the course, students will be equipped with the necessary skills and knowledge to tackle challenging research problems and contribute to the rapidly developing field of MCS.

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 workload 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 [WueCampus](#). 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). It is your task to write a report on this theme (8–10 pages long).

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 **21 August 2023, 23:59h** to [fabian.moss@uni-wuerzburg.de](mailto:fabian.moss@uni-wuerzburg.de) as a single PDF file per group, named `intro-corpusmus_<last_name>.pdf`, e.g. `intro-corpusmus_moss.pdf`.

## Organization

Classes will generally be structured as follows:

- Lecture (~25 min.)
- Q&A and discussion of reading material (~20 min.)
- Group work (~45 min.)

Table 1: Course overview. Dates marked with \* are taught online.

Week	Date	Topic
1	19. April 2023	Introduction to musical corpus studies
2	26. April 2023	Music as data: metadata & data collection
3	03. Mai 2023	Music as data: encoding formats
4	10. Mai 2023	Music as data: corpus building & digital editions
5	17. Mai 2023	Methods & techniques: the vector-space model
6	24. Mai 2023*	Methods & techniques: statistical methods
7	31. Mai 2023	Methods & techniques: machine-learning methods
8	07. Juni 2023	Methods & techniques: basics of visualization and sonification
9	14. Juni 2023	Analysis: melodic shapes in folk songs & Jazz solos
10	21. Juni 2023	Analysis: harmonic patterns in Mozart and Beethoven
11	28. Juni 2023	Analysis: metrical regularities in Malian Djembe music
12	05. Juli 2023	Analysis: character networks in opera libretti
13	12. Juli 2023	Analysis: authorship attribution in Renaissance polyphony
14	19. Juli 2023	Ethics and biases: representativity and canon; feedback

## Credits

For ungraded credit, active participation in the class discussions and group work are sufficient. For a graded credit, an additional report commenting on a particular corpus study on the background of the course contents (about 8-10 pages) is required.

## References