Introduction to musical corpus studies

Fabian C. Moss

2023-04-11

Table of contents

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.

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

1 Introduction

2 On corpus studies in musicology

Our in-class research has produced the following important aspects for music corpus studies.

Corpus studies is mostly associated with computational linguistics. Linguistics have earlier adopted computational methods in their research. This is possibly also due to the fact that there was and is a separation between linguistics and literature studies, the latter only lately opening up to computation in the context of digital humanities. In music research, there is no such distinction.

Part of the definition of corpus studies is the adherence to quantitative methods and to working with machine-readable data, especially when using automated methods for analysis. Corpus studies can be seen as a sub-field of digital musicology.

Several aspects play an important role for corpus research in music, e.g. corpus compilation, that is, assembling scores,recordings, or other musical data that are sometimes deemed to be representative of a certain style or genre.

Another type of data are annotations (or tags) and metadata. They can be useful for applications like querying the corpus. For this, several programming languages or software tools can be used. Corpus studies moreover frequently employ statistical methods in order to identify patterns. A central distinction with respect to other data analytical studies with music is that corpus studies are, in general, motivated by musicological or music theoretical research questions.

Homework

- Read (Pugin 2015)

Recommended reading

1. Temperley and VanHandel (2013); VanHandel and Temperley (2014)
2. Neuwirth and Rohrmeier (2016)

Part I

Music as data

3 Metadata and data collection

3.1 Dataset collections

Tip

From the following two overview sites of music datasets, try to get an overview of what is listed. Try to find good descriptors for the different entries.

- ISMIR datasets
- Musical Corpora Register

After you have a good overview, choose one or two items and have a closer look at them. Collaborate with a colleague on a more detailed description of the dataset, what it contains, and its potential usage in music research.

3.2 Metadata

Tip

The Répertoire International des Sources Musicales is probably the largest music database worldwide. Perform a couple of queries on their interface, note down what you can find out about pieces, which information is contained in the database, what other information you would be interested in, and how you would use it to do a corpus study.

References

Neuwirth, Markus, and Martin Rohrmeier. 2016. "Wie Wissenschaftlich Muss Musiktheorie Sein? Chancen Und Herausforderungen Musikalischer Korpusforschung." *Zeitschrift Der Gesellschaft Für Musiktheorie [Journal of the German-Speaking Society of Music Theory]* 13 (2): 171–93. https://doi.org/10.31751/915.

Pugin, Laurent. 2015. "The Challenge of Data in Digital Musicology." *Frontiers in Digital Humanities* 2 (August): 1–3. https://doi.org/10.3389/fdigh.2015.00004.

Temperley, David, and Leigh VanHandel. 2013. "Introduction to the Special Issues on Corpus Methods." *Music Perception: An Interdisciplinary Journal* 31 (1): 1–3. https://doi.org/10.1525/MP.2013.31.1.1.

VanHandel, Leigh, and David Temperley. 2014. "Introduction to the Second Special Issue on Corpus Methods." *Music Perception: An Interdisciplinary Journal* 31 (3): 191–91. https://doi.org/10.1525/mp.2014.31.3.191.