

---

# **Introduction to Musical Corpus Studies**

***Release 0.0.1***

**Fabian C. Moss**

**Aug 24, 2020**



# CONTENT

<b>1 Organization</b>	<b>3</b>
<b>2 Introduction / Background</b>	<b>7</b>
<b>3 Folk Songs and the Melodic Arc</b>	<b>9</b>
<b>4 Solos in the <i>Weimar Jazz Database</i></b>	<b>11</b>
<b>5 Harmony in Beethoven's String Quartets</b>	<b>13</b>
<b>6 Billboard Pop Charts</b>	<b>15</b>
<b>7 Brazilian Choro</b>	<b>17</b>
<b>8 History of Tonality</b>	<b>19</b>
<b>9 Malian Percussion Music</b>	<b>21</b>
<b>10 Electronic Music 1950-1990</b>	<b>23</b>
<b>11 Conclusion</b>	<b>25</b>
<b>Bibliography</b>	<b>27</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”, 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. Among the contents are:



---

**CHAPTER  
ONE**

---

**ORGANIZATION**

## **1.1 Schedule**

The following table outlines the schedule and summarizes the contents of this course.

No.	Date	Time	Room	Topic	Corpus	Methods
1	Fr., 13.11.2020	16:00-17:20 Uhr	Neuer Seminarraum 1.315	Introduction / Background		
2		17:40-19:00 Uhr		Folk Songs, Melodies, Pitches and Intervals	Essen Folk Song Collection	frequencies, mean, variance
3	Sa., 14.11.2020	09:00-10:20 Uhr	Neuer Seminarraum 1.315	Jazz Solos, Melodies	Weimar Jazz Database	Regular Expressions
4		10:40-12:00 Uhr		Beethoven's string quartets Harmony	Annotated Beethoven Corpus	<i>n</i> -grams, Markov models
		12:00-13:00 Uhr		Lunch Break		
5		13:00-14:20 Uhr		Pop Charts Billboard 100, harmony,	McGill Billboard Dataset	Hidden Markov Models
6		14:40-16:00 Uhr		Free group work		
7	Fr., 11.12.2020	10:00-11:20 Uhr	Alter Seminarraum 1.408	Brazilian Choro, harmony, form,	Choro Songbook Corpus	Context-Free Grammars
8		11:40-13:00 Uhr		19th century piano music, harmony	DCML Piano Corpus	Probabilistic CFGs
9	Sa., 12.12.2020	09:00-10:20 Uhr	Neuer Seminarraum 1.315	Malian Percussion Music, rhythm, meter	Interpersonal Entrainment in Music Performance: Malian Jembe	
10		10:40-12:00 Uhr		Electronic Music 1950-1990	Curated Corpus of Historical Electronic Music	
		12:00-13:00 Uhr		Lunch Break		
11		13:00-14:20 Uhr		Free group work		
12		14:40-16:00 Uhr		Recapitulation and conclusion		

## **1.2 Credits**

Ich gehe in der Seminarplanung von 12 Semesterwochen à 2 SWS aus, für das gesamte Blockseminar also 24 SWS. Das Seminar wird mit 3 CP bewertet, was 90 Stunden aktiver Arbeit entspricht. Davon entfallen 24 SWS an die Präsenzzeit im Seminar plus 48 SWS an Vor- und Nachbereitung der Seminarsitzungen. Die verbleibenden 18 SWS sind für die Lektüre der Fachliteratur vorgesehen.



---

CHAPTER  
TWO

---

## INTRODUCTION / BACKGROUND



Fig. 1: Image by Victoria Alexander on Unsplash.

### 2.1 What are Musical Corpus Studies?

tbc... (text from diss?)

## 2.2 Epistemological goals

tbc...

## 2.3 Issues

tbc [[Coo06](#)][[Hon06](#)][[Hur13](#)][[Mar16](#)][[NR16](#)][[Pug15](#)][[Sch16](#)][[TV13](#)]

## 2.4 MCS and traditional musicology

tbc

---

CHAPTER  
**THREE**

---

## FOLK SONGS AND THE MELODIC ARC

Huron... / MusThe Tutorial



---

CHAPTER  
FOUR

---

## SOLOS IN THE *WEIMAR JAZZ DATABASE*



Fig. 1: Photo by Janine Robinson on Unsplash

The first project we will have a look at is the [Jazzomat](#) project. Transcriptions of Jazz solos. [Pfl17]



## HARMONY IN BEETHOVEN'S STRING QUARTETS

### 5.1 Access the data

The data lies on the GitHub repository [DCMLab/ABC](#). Either download the `.tsv` file directly and open it in pandas or load it from the URL as follows:

```
import pandas as pd  
  
df = pd.read_csv("https://github.com/DCMLab/ABC/corpus.tsv", sep="\t")
```

The corpus is now stored in the variable `df`.

### 5.2 Harmonic Annotations

- regular expressions

[NHMR18][MNHM19]

### 5.3 Chord Transitions

- n-grams



---

CHAPTER  
SIX

---

## BILLBOARD POP CHARTS



Fig. 1: Photo by [israel palacio](#) on [Unsplash](#).

Clustering analysis in [SVJ+20].



## BRAZILIAN CHORO



Fig. 1: A 2019 musical tribute to renowned choro composer Jacob do Bandolim, on the 50th anniversary of his death in São Paulo, Brazil. (c) Governo do Estado de São Paulo / CC BY 2.0

### 7.1 Data

The *Choro Songbook Corpus* contains transcriptions of the chord symbols of 295 Choro pieces [MSFR20]. It is available on [Zenodo](#) or [Github](#). See this link for a performance of the piece “Lamentos”

As before, we can import our data from a public resource.

```
df = pd.read_csv("https://raw.githubusercontent.com/DCMLab/choro/1.1/data/choro.tsv",  
sep='\t')
```

And display the first rows of the data frame:

```
df.head()
```

---

CHAPTER  
**EIGHT**

---

## HISTORY OF TONALITY

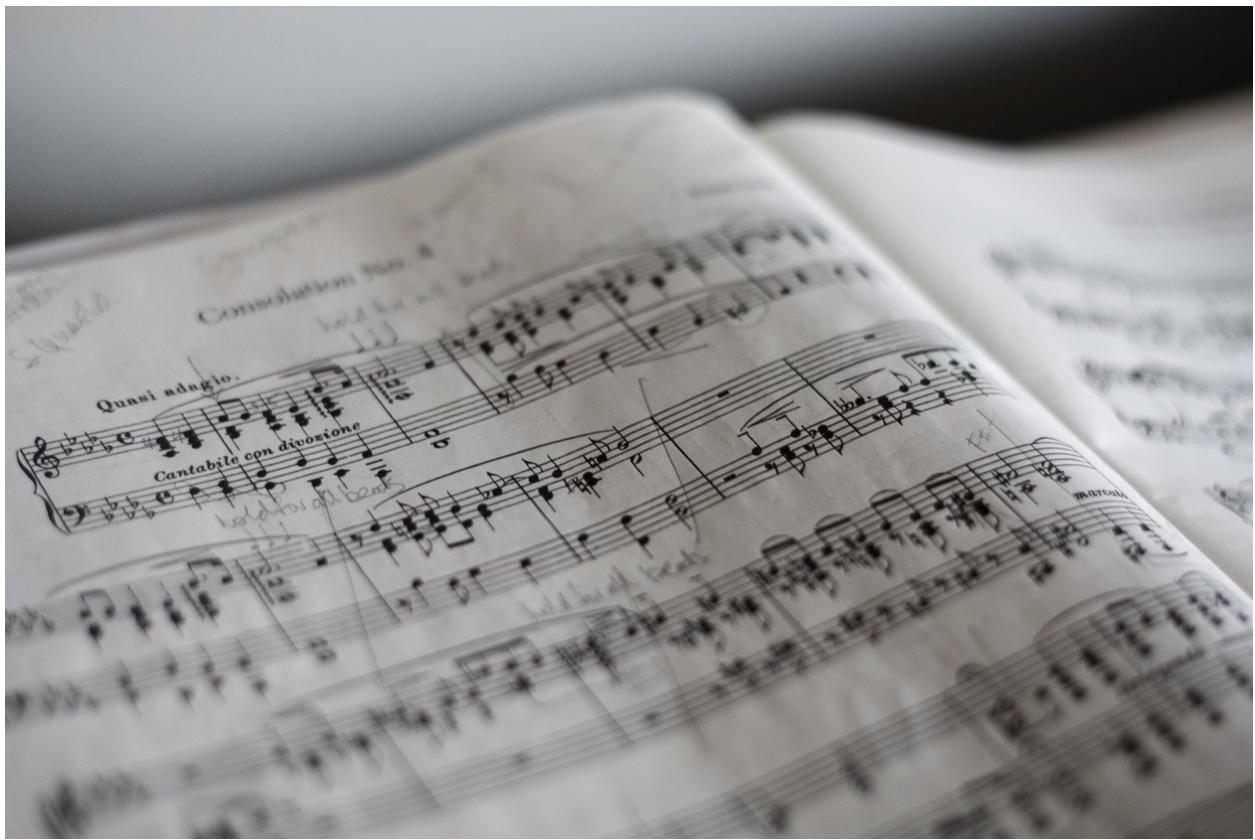


Fig. 1: Photo by Marius Masalar on Unsplash.



---

**CHAPTER  
NINE**

---

**MALIAN PERCUSSION MUSIC**



---

CHAPTER  
TEN

---

**ELECTRONIC MUSIC 1950-1990**



Fig. 1: Photo by Antoine Julien on Unsplash.



---

**CHAPTER  
ELEVEN**

---

**CONCLUSION**

Final thoughts, critical discussion...

[Some image]

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.



## BIBLIOGRAPHY

- [Coo06] Nicholas Cook. Border Crossings: A Commentary on Henkjan Honing’s “On the Growing Role of Observation, Formalization and Experimental Method in Musicology”. *Empirical Musicology Review*, 1(1):7–11, 2006.
- [Hon06] Henkjan Honing. On the Growing Role of Observation, Formalization and Experimental Method in Musicology. *Empirical Musicology Review*, 1(1):2–6, 2006.
- [Hur13] David Huron. On the Virtuous and the Vexatious in an Age of Big Data. *Music Perception: An Interdisciplinary Journal*, 31(1):4–9, 2013.
- [Mar16] Alan Marsden. *Music Analysis by Computer: Ontology and Epistemology*. Springer, 2016.
- [MNHM19] Fabian C. Moss, Markus Neuwirth, Daniel Harasim, and Rohrmeier Martin. Statistical characteristics of tonal harmony: a corpus study of beethoven’s string quartets. *PLoS ONE*, 14(6):e0217242, 2019. doi:10.1371/journal.pone.0217242.
- [MSFR20] Fabian C. Moss, Willian Souza Fernandes, and Martin Rohrmeier. Harmony and form in Brazilian Choro: A corpus-driven approach to musical style analysis. *Journal of New Music Research*, 2020. doi:10.1080/09298215.2020.1797109.
- [NHMR18] Markus Neuwirth, Daniel Harasim, Fabian C. Moss, and Martin Rohrmeier. The Annotated Beethoven Corpus (ABC): A Dataset of Harmonic Analyses of All Beethoven String Quartets. *Frontiers in Digital Humanities*, 5(July):1–5, 2018. doi:10.3389/fdigh.2018.00016.
- [NR16] Markus Neuwirth and Martin Rohrmeier. Wie wissenschaftlich muss Musiktheorie sein? Chancen und Herausforderungen musikalischer Korpusforschung. *Zeitschrift der Gesellschaft für Musiktheorie*, 13(2):171–193, 2016. doi:10.31751/915.
- [Pfl17] Martin Pfleiderer. *Inside the Jazzomat: New Perspectives for Jazz Research*. Schott, 2017.
- [Pug15] Laurent Pugin. The challenge of data in digital musicology. *Frontiers in Digital Humanities*, 2(4):1–3, 2015. doi:10.3389/fdigh.2015.00004.
- [Sch16] Kris Schaffer. What is computational musicology? online, 2016. URL: <https://medium.com/@krisschaffer/what-is-computational-musicology-f25ee0a65102>.
- [SVJ+20] Kris Schaffer, Esther Vasiete, Brandon Jacquez, Aaron Davis, Diego Escalante, Calvin Hicks, Joshua McCann, Camille Noufi, and Paul Salminen. A cluster analysis of harmony in the McGill Billboard dataset. *Empirical Musicology Review*, 14(3–4):146–162, 2020. doi:10.18061/emr.v14i3-4.5576.
- [TV13] David Temperley and Leigh VanHandel. Introduction to the Special Issue on Corpus Methods. *Music Perception: An Interdisciplinary Journal*, 31(1):1–3, 2013.