MUSI 4843/ 8803: Special Topics - Computational Musicology

Fall 2022

Midterm Project

**Description:** You and your partner (teams of 2-3) will present, in written and oral form, your complete plan/outline for your final class project. Your final project will work towards computationally answering a hypothesis-driven question or problem using computational musicology data and tools.

In order to successfully complete the final, you will need to settle on a clear hypothesis *a priori* and assemble a corpus of symbolic data that appropriately reflects the question or problem you are trying to address. It is strongly recommended that you use already-available corpora to perform your analysis; however, if you prefer to build your own corpus you must have the data ready by the midterm (i.e., assembled and looked-through and committed to using, but not necessarily analyzed).

**An important aside:** Your *final projects* will ultimately be a musical analysis or inquiry. As such, any data that is considered symbolic that allows you to investigate your musical question is valid. However, at the heart of computational musicology is not computers, but musicology. It is a humanistic discipline. As such, it is expected that the corpus you create is curated, pre-processed, and cleaned (i.e., error-checked) with a *human* behind each stage (not merely an algorithm). That is, I expect a basic level of quality that accompanies human-centered data curation. *Note that the scale of the corpus will be necessarily be small compared with a typical published/distributed musical corpus. Students will not be penalized for insufficient statistical power. It is the process and logic that counts.*

**Deliverables:**

* A peer-review of another group’s midterm – randomly assigned (1 page document that you will fill-in according to a template). Each group will review one other group. (This will be due *following* the official deadline approximately one week later.)
* A zipped file or Gatech Github repository with all your necessary corpus materials (i.e., data, parsers, etc.)
* A 2-3 page document outlining your analysis plan
  + Introduction: Motivation, problem/question, and your specific hypothesis (or hypotheses)
  + Corpus description (not necessarily in this order):
    - sampling procedure, including operationalizations, as necessary (e.g., “Romantic music,” “pop songs,” “traditional song,” etc.)
    - Basic descriptive statistics for your corpus (i.e., what’s in it? How is it divided? What are the file formats?)
      * Format**:** Your corpus can be in any “official” musical format discussed (Humdrum, MusicXML, MEI, MIDI), or it may be in some form of alternative format (such as a CSV file or other plain text file).
    - Description of ‘quality’ and ‘quantity’ (where did it come from? Expert or crowd-sourced? Corrections made or not? How many files per category, etc.)
    - Detailed descriptions of the procedures or operationalizations associated with the *collection or application* of tags/labels/metadata/scoring, etc. (Include images, audio clips, or music notation to explain where necessary.)
    - Caveats: Min. 1 paragraph detailing the presumed “problems” or issues that may arise as a result of using the corpus as chosen (e.g., due to sampling method(s), operationalization, data format, etc., etc.)
  + Analysis plan
    - What is/are the musical parameter(s) that need(s) to be analyzed/searched/modeled? (E.g., rhythm patterns, chord changes, melody). Given that, how might you need to pre-format your data? Will you analyze your feature in counts? Proportions? Or as a function of time/duration? (E.g., counting pitch classes largely ignores duration unless you “timebase” or “slice” them first).
    - If your hypothesis requires operationalization that impacts how you do the analysis (e.g., “…more repetitive than…”) you must include your operationalization in this analysis plan section (e.g., “we will examine how frequently 3-, 4-, and 5-note patterns recur in each song using a window that slides every 1 note.”)
    - Given the computational musicology tools discussed so far, which (if any) is likely best suited to your analytical problem? Thinking about the “shape” and complexity of your expected data may help here.
    - Though at this time you will not likely have sufficient statistical understanding to know which statistical method you will chose to test your hypothesis, you *can* answer the following questions:
      * Do you have a linear/continuous question (e.g., is something changing over time? Changing over frequency/pitch height?), or do you have things that fall into categories (e.g., *composer X* uses melodic riff ‘*a’* more than *composer Y*) and you wish to examine differences between the categories? If the latter, how *many* categories?
      * Is the data you are collecting continuous or categorical? (E.g., pitch height in semitones would be considered continuous, whereas chord labels would be categorical).
  + Timeline. Write up a timeline (with deadline dates!) for your full analysis plan from midterm through final presentation date. Preference is for a table-style with overlapping boxes to help illustrate parallel tasks. It should include (these items may overlap):
    - Corpus pre-processing and cleaning (*max 2 weeks*) – what data do you need? What format do you need it in? What data needs to be extracted or added?
    - Reading in the data and unifying the corpus if necessary
    - Coding the analysis (include as many events/steps as possible) (*min 2 weeks*)
    - Some initial/basic graphing of analytical outcomes
    - Setting up data for testing
    - Performing statistical tests
    - Final graphs / tweaking / cleaning-up
    - Writing final paper
    - Final presentation (final exam slot).

**Grading:**

Corpus 10

Plan writeup 70

Peer review 10

Presentation 10

Total 100%

(15 points to total grade)