Final Report

Research Assistant: Meiying Ding

Table of Contents

Part I: Introduction

Throughout the period of Summer 2019 for the Work Learn position in Dr. Ève Poudrier's Rhythm Computation and Cognition Lab, I worked on two main projects: "Modeling Polyrhythmic Experience" and the associated Suter (1980) Corpus and Polyrhythm website, and the Carter Beat Synchronization project. These projects used different experimental methods and computational tools to explore the historical development, structural features, and human perception of complex rhythms. In each of these projects, I performed data validation, management, and analysis using the command terminal as well as software (e.g., *Microsoft Office*) and open-source programs such as *Vega-Lite* and *R* (*R Commander* and *R Studio*). These tools were used to create various types of visualizations and templates, and to conduct statistical analysis, the results of which were incorporated into conference presentations, the Polyrhythm website, and will be used for future scholarly articles.

Part II: Projects & Tasks

I. MPE Suter (1980) Corpus

The MPE Project involves using geocoding software, specifically *ArcGIS* (*ArcMap* and *ArcGIS Online*) and *Carto*, to explore trends in the development of polyrhythm in Western music from 1900 to 1950. The Suter (1980) Corpus refers to 724 examples extracted by Louis-Marc Suter from 450 musical works and reproduced in his unpublished doctoral dissertation (Université de Berne), "Les polyrythmes dans la musique de vingt compositeurs de la première moitié du vingtième siècle, 1900-1950 [Polyrhythms in the music of twenty composers from the first half of the 20th century]." These examples constitute the Full Dataset (FDS), which was divided into two datasets—the Reserve Dataset (RDS) and the Testing Dataset (TDS), where the RDS constitute a stratified sample of 80 examples to be used for exploratory analysis.

I have formatted data for the new version of the FDS *Carto* Map and created timeline visualizations of Composers Birth Year using *Office Timeline Online* as well as World Map visualizations using *Excel*. Statistical analysis and visualizations were also performed using *R Commander*, to assist with the preparation of Dr. Poudrier's talks at the MusCan/CAML conference (June 6 and 7, 2019).

Work with ArcGIS

Previous work on *ArcGIS* was done by other undergraduate research assistants (Hailey, Jiayi, and Sarah) who prepared reports and tutorials on how specific maps were generated. In preparation for geocoding work, I reviewed and worked through the existing tutorials (linked document includes annotations). For the current work, however, we decided to focus on using *Carto* to generate the FDS map, thus there was no more work done with ArcGIS.

Some of the issues I experienced while working with the tutorials included not being able to create lines connecting Composer Death and Birth Place and locating the "Actions" tab under "layer properties" in order to filter data points. These are elements that could be further clarified in the tutorials.

Some useful information for further work with *ArcGIS* and *ArcMap*:

1. The UBC GIS Resources Research Guide and the *ArcGIS* Librarian Consultation (by appointments, evan.thornberry@ubc.ca)
2. Koerner Library GIS Lab: *ArcGIS Online* and ArcMap can be accessed through the account with user name: qianqianwu577_UBC, and password: Suter1980corpus
3. *ArcGIS Online Web App Builder* extends tools and templates that enable functionalities such as an interactive Time Slider
4. Online courses on *ArcGIS* offered by Esri Virtual Campus in order to learn *Web App Builder* add-ons
5. Other Possible Programs:

a. *d3.js*, online *JavaScript* code that can be used to create visualizations. Most widely used visualization tool
b. *leaflets.com*, also a *JavaScript* source that can be used to personalize map designs

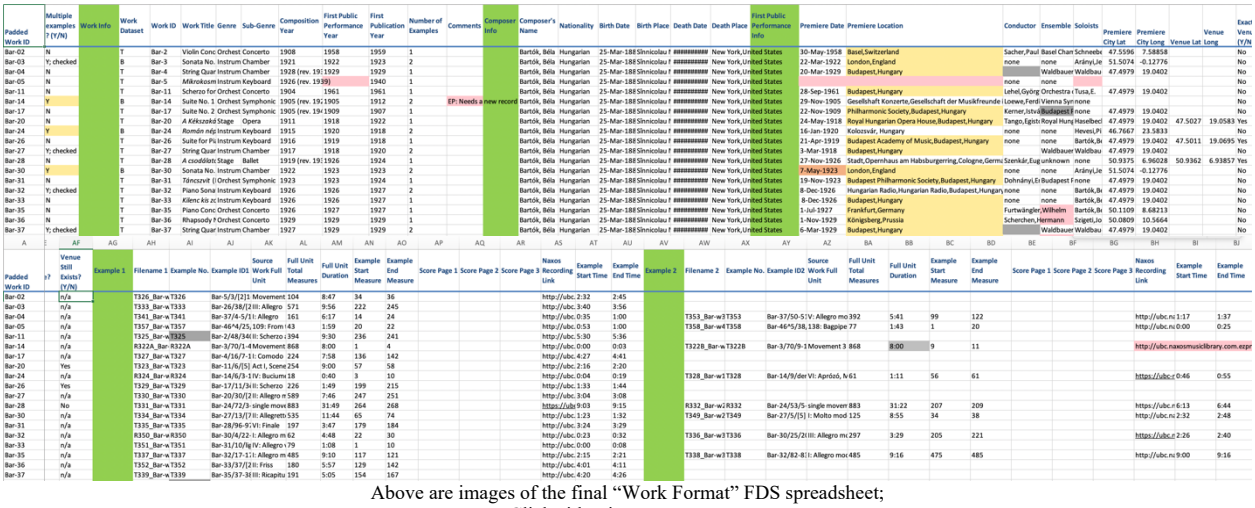
Data Formatting for the FDS Carto Map with Excel

In order to generate the FDS Carto map, the data needed to be formatted. The two main tasks I worked on were:

1. Inserting Filenames into FDS: New Filename system was created and need to be incorporated into the FDS. The new filenames added R/T in front of example numbers, and used only symbols that would be compatible with Carto.
2. Reformatting "Example Format" into "Work Format": FDS in "Example Format" (each example occupy one row in *Excel*) were wanted in "Work Format" (each work occupying one row, with all its examples tagged to the right). Examples can be from either RDS ("R") or TDS ("T") but not both, while works can contain examples from RDS, TDS or Both ("B"). Therefore, examples can be labelled as {R, T}, whereas works can be labelled as {R, T, B}. The reformatting was done one category at a time:

   [Example, Work] = {[R,R], [R,B], [T,B], [T,T]}

The re-formatted spreadsheet was transferred to Melissa Liang for further processing. The spreadsheet is shown below and can be accessed on Box by clicking on the image.

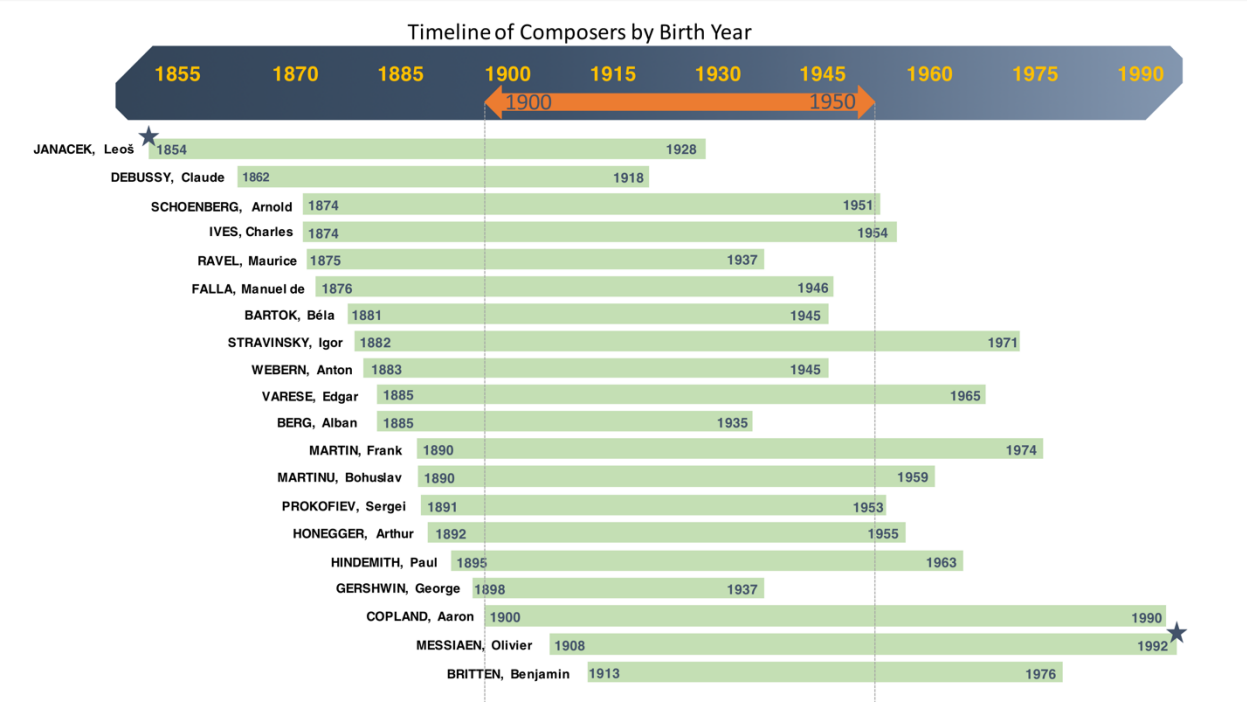


Above are images of the final "Work Format" FDS spreadsheet;
Click either images to access.

Visualizations with Microsoft Office

There were two visualizations performed using *Office*, a timeline and a map of the composers in the Suter (1980) Corpus.

1. Timeline of Composers by Birth Year: I created this tutorial for using *Office Timeline Online*. Note that *Office Timeline Online* only takes ten rows of data at a time. Further processing of timelines generated by this online tool can be downloaded as PPT files which can then be edited using *Microsoft PowerPoint*.



Above is an image of the final timeline visualization;
Click Image for alternate timeline formats.

2. Maps of Composer Nationality: *Excel* is also a very powerful tool to visualize data on world maps autogenerated by *Excel*'s built-in Bing extension, available for Version 2013 onward. The data can be formatted using the sample format below, and then the map can be inserted using the command “Insert” -> “Maps”. I created visualizations by Total Number of Works and Total Number of Examples per country on a World Map or a Europe Map by composer’s nationality/nationalities (All or Primary). (Underline is used to denote variables.)

Here is an extensive list of the maps that were generated (maps are included in the linked workbooks):

1. World: Number of Works by Composer Nationality (All; all nationalities of a composer receive a count if the composer has more than one nationality)
2. World: Number of Examples by Composer Nationality (All)
3. World: Number of Works by Composer Nationality (Primary; only the birth country of a composer receives a count if the composer has more than one nationality)
4. World: Number of Examples by Composer Nationality (Primary)
5. Europe: Number of Works by Composer Nationality (All)
6. Europe: Number of Examples by Composer Nationality (All)

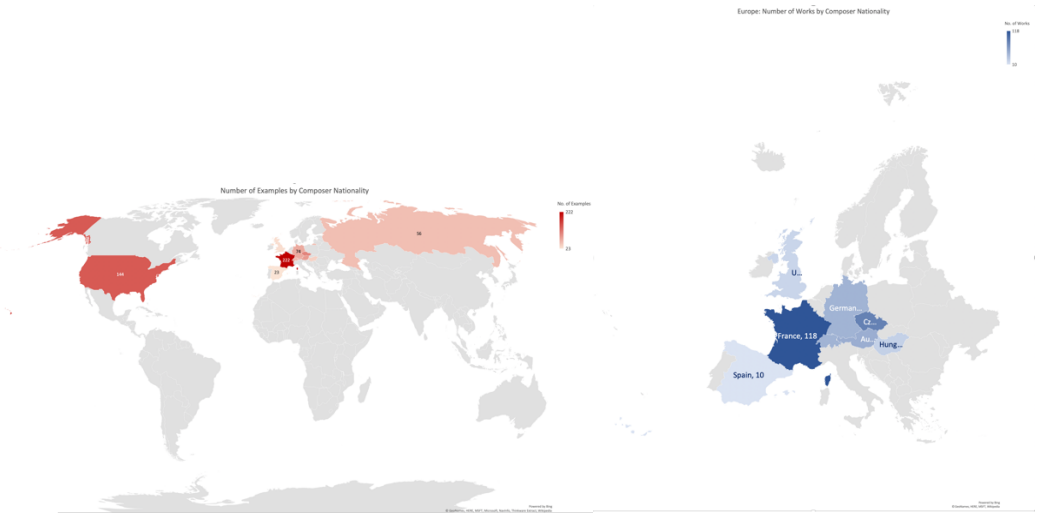
7. Europe: Number of Works by Composer Nationality (Primary)
8. Europe: Number of Examples by Composer Nationality (Primary)

This is a table with a sample format of the data:

Nationality	Number of Works/Examples by all Composers
France	10
Germany	15
England	12

Finally, *Excel* allows changing of some features displayed on the map, such as adding layers. Future possibilities would be to explore 3D charts and 3D world maps.

Two sample maps are shown below and additional maps and associated data can be accessed on Box by clicking on the image.



Above are images of the "Number of Works/Examples by Composer Nationality Maps";
World: Number of Examples by Composer Nationality (left); Europe: Number of Works by Composer Nationality (right);
Click either images for data and all other maps.

R Commander Plots and Statistics

R takes in data in CSV files (save *Excel* sheets into .csv format). *R Studio* (also used in the Carter Beat Synchronization Project) requires coding but allows more functionality, whereas *R Commander* generates plots and calculates statistical summaries with dropdown menus that are easier to manipulate. Eight scatterplots and three box plots were created to explore the relationship between Tempo (Example Start and Full Unit Start) and First Composition Year and Premiere Year as well as Example Start Location by measure and by time.

Here is an extensive list of all scatterplots and box plots that were generated:

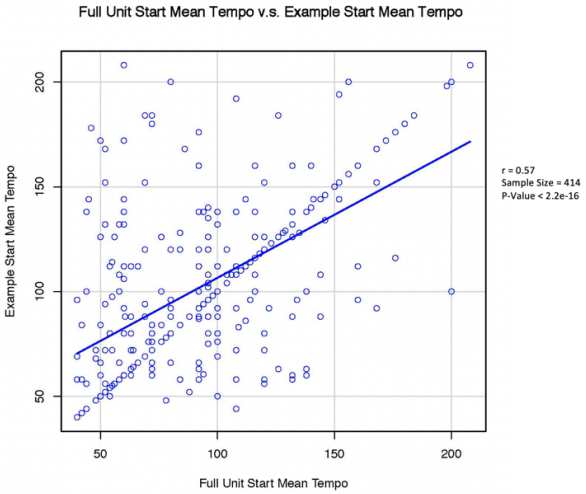
1. Distribution of Examples by Composition Year and Premiere Year boxplot
2. Full Unit Start Mean Tempo vs. Example Start Mean Tempo scatterplot
3. Composition Year vs. Full Unit Start Mean Tempo scatterplot
4. Composition Year vs. Example Start Mean Tempo scatterplot
5. Premiere Year vs. Full Unit Start Mean Tempo scatterplot
6. Premiere Year vs. Example Start Mean Tempo scatterplot
7. Conference Follow-up Plots: FUSMT_ESMT Excluding A Tempo (pages 3-4 of the linked report)

Below are the steps involved in creating scatterplots and boxplots:

Scatterplots:

For example, to plot First Composition Year vs. Example Start Mean Tempo:

- First calculate mean tempo
- Then, organize data into three columns "Filename", "First Composition Year", and "Example Start Mean Tempo" in *Excel*
- Load R Commander by typing *library(Rcmdr)* in *R*
- Import the CSV file into *R Commander*
- GRAPH->SCATTERPLOT (Choose "Linear Regression Line"; Name axes)
- To save graphs: GRAPHS-> SAVE GRAPH TO FILE



	Mean	SD	IQR	Slowest	25%	50%	75%	Fastest	Sample Size
Example Start Mean Tempo	104	37	53	40	73	100	126	208	414
Full Unit Start Mean Tempo	96	35	54	40	66	96	120	208	414

Above is an image of the "Full Unit Start Mean Tempo vs. Example Start Tempo" Scatterplot with a Summary Statistics Table.

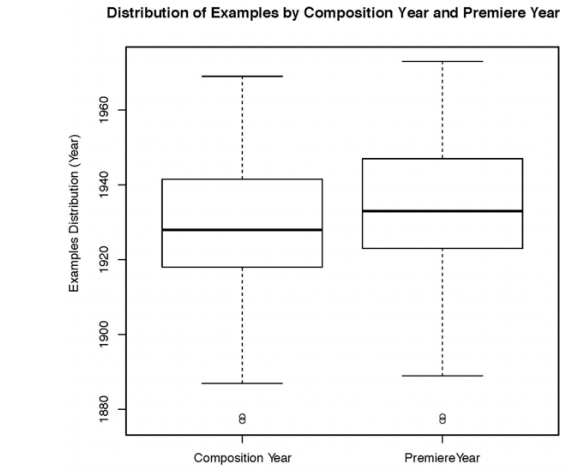
Boxplots:

For example, to plot First Composition Year vs. Premiere Year:

- Organize the data by the Categorical Variable "First Composition Year" or "Premiere Year" (all in one column)

Filename	First Composition Year	1908
...	...	...
...	Premiere Year	1921
...	...	...

- Load R Commander by typing *library(Rcmdr)* in *R*
- Once the CSV file is imported into *R Commander*, select GRAPH->BOXPLOT->"plot by category" (category is First Composition Year_Premiere Year, the second column above)



	Mean	SD	IQR	Earliest	25%	50%	75%	Lastest	Sample Size
Composition Year	1929	16	24	1877	1918	1928	1942	1969	703
Premiere Year	1934	17	24	1877	1923	1933	1947	1973	703

Above is an image of the "Distribution of Examples by Composition Year and Premiere Year" Boxplot with a Summary Statistics Table.

I also provided summary statistics with the plots. To obtain mean, median, IQR and sample size, you can select "Statistics"-> "Summaries" -> "Numerical Summaries":

```
> numSummary(FUSMT_ESMT_Categorized[,"Mean.Tempo", drop=FALSE], groups=FUSMT_ESMT_Categorized$Category,
+   statistics=c("mean", "sd", "IQR", "quantiles"), quantiles=c(0,.25,.5,.75,1))
                            mean       sd   IQR 0%   25% 50% 75% 100% Mean.Tempo:n
Example Start Mean Tempo   104.30435 36.91349 53.25 40 72.75 100 126  208          414
Full Unit Start Mean Tempo  96.26329 35.09412 54.00 40 66.00  96 120  208          414
```

To obtain correlation coefficient, select "Statistics"-> "Summaries" -> "Correlation Test".

II. Polyrhythm Website (Vega-Lite and Terminal)

Vega-Lite (https://vega.github.io/vega-lite/) is a high-level language for creating expressive and concise interactive visualizations and mapping data fields to properties of graphical marks. Vega-Lite uses a portable JSON format that compiles to full specifications in the larger Vega language. Vega-Lite includes support for data transformations such as aggregation, binning, filtering, and sorting, as well as visual transformations such as stacking and faceting into small multiples. There are four components in a graph: layer, facet, concat, and repeat. Users can specify the type of selection desired (e.g., a point or interval selection) along with possible transformations. *Vega-Lite* uses defaults such as scale and axis, which don't need to be coded but can be overwritten to personalize. More information can be accessed in *Vega-Lite's* official introduction: OpenVis Conf Talk.

Project collaborator Craig Sapp (programmer and consultant for the MPE project) developed the *Accent-Features* program using *Humlib*. This program extracts desired features that allows analysis on rhythmic patterns. *Humdrum* is a software toolkit that use Kern language to encode music notation for computer aided analysis. *Humlib* is a set of C++ classes for parsing *Humdrum* data files. My role in this project (together with the senior research assistant, Melissa Liang) was to generate several templates to visualize accentuated features for each example in the Full Dataset. These templates were to be incorporated into the website polyrhythm.humdrum.org.

Exploratory Terminal Code and Plots

Melissa and I generated exploratory graphs to visualize features including trills, slur beginnings and endings, regular accents, marcatos, tenutos, number of notes, and position of notes in a stacked bar chart, a layered bar chart, or a stacked butterfly bar chart (click links for command terminal code instructions).

Before Craig developed the *accent-features* program, we could only extract one accent feature at a time when preparing the data for each of the three graphs below. Each extracted feature would be saved to a TSV file. Then, we combined the desired features from different TSV files into one.

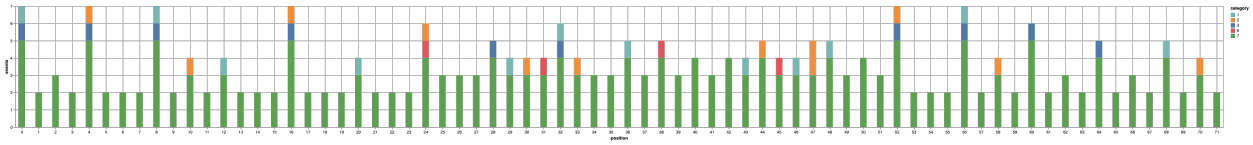
For example, to obtain the data for a stacked bar chart of number of slurs and accents for Example 161 from the Suter (1980) Corpus:

1. Extract number of slurs per tick position:
   grep -on '([\[]*[0-9]' KERN/161_Deb-13_34_25-26.krn | cut -d : -f1 | uniq -c | cut -d " " -f7 > 161_num_slr.tsv
   Extract tick position from composite beat where slurs occur:
   beat -tcp KERN/161_Deb-13_34_25-26.krn | grep '([\[]*[0-9]' | cut -f1 > 161_pos_slr.tsv
   Obtain one TSV file that combines tick position and number of slurs:
   paste 161_pos_slr.tsv 161_num_slr.tsv > 161_beatpos_numslr.tsv
2. Repeat Step 1 to obtain number of accents per tick position

3. Paste two resulting TSV files of Step 1 and Step 2.

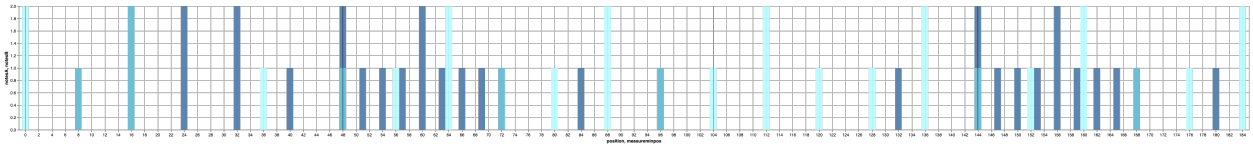
Below are the three types of exploratory graphs:

Example 389 Slur Beginnings, Endings, Accents, Trills and Attacks vs. Beat Position



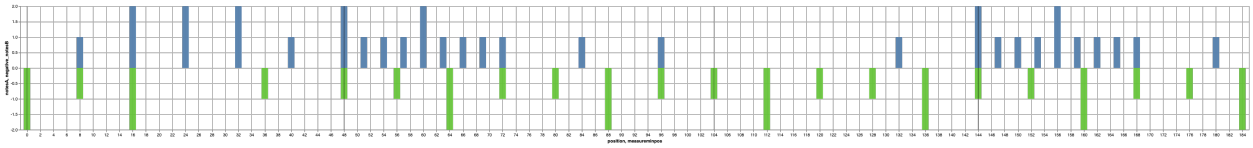
Above is an image of a stacked bar chart.

Jan_129 Regrouped Notes Position with Measure Separation Layered



Above is an image of a layered bar chart.

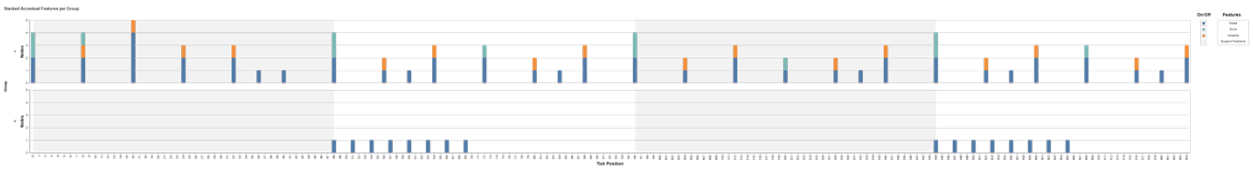
Jan_129 Regrouped Notes Position with Measure Separation Juxtaposed



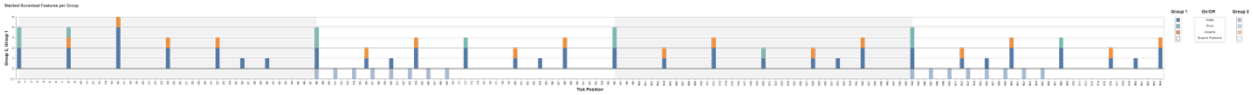
Above is an image of a butterfly bar chart.

Final Vega-Lite Templates and Plots

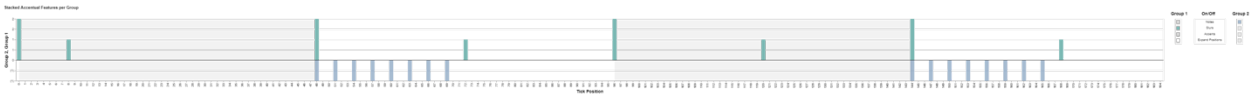
The three plots below are final templates that can be incorporated into the website. By clicking (SHIFT + Click) the boxes within the legend box, one can select only desired features to visualize.



Above is an image of the final template for stacked bar chart with separated groups;
Click image to view the plot;
Click HERE to view the *Vega-Lite* code;
Paste code into Vega-Lite Editor for interactive plot.



Above is an image of the final template for stacked butterfly bar chart;
Click image to view the Plot;
Click HERE to view the *Vega-Lite* code;
Paste code into Vega-Lite Editor for interactive plot.



Above is an image of the final template for stacked butterfly chart when selecting specific features;
Click image to view the plot.

Accent-Features (Humlib)

The *Accent-Features* program, which was created in collaboration with Craig Sapp and Melissa Liang enables us to extract all desired features with one function. I summarized how to use *Accent-Features* in this document:
https://app.box.com/file/498067674092

Here is an outline of the linked document:

1. What is *accent*-features?
2. A list of all spines that can be obtained
3. -k option: To print **kern, **pkern and **nkern
4. -p option: To print **code, **abbr and **name
5. -m option: To print *num, **start, **stop (in quarter notes)
6. To Install and Update *Accent-Features*
7. Extracting *Accent-Features* Data from Kern Files
8. Extracting Specific Columns of *Accent-Features*
9. Extracting Parts Name and Measure Data from Kern Files

Basic Terminal Tutorial

I also created this tutorial on how to get started with using Bash command terminal.

III. Carter Beat Synchronization (R ggplot2)

The Carter Beat Synchronization Project is based on the Carter Tapping Experiment. It is a behavioral experiment involving participants' listening and tapping to a passage from Elliot Carter's *90+* for piano (1994). Five musical parameters, including texture (segregated or combined layers), timbre and frequency (pitch on or off), attack velocity (accent on or off), event duration (staccato or tenuto accent), and register, were analyzed for their effect on perceived temporal structure.

Stimuli's Hits Total and Percentage (Excel)

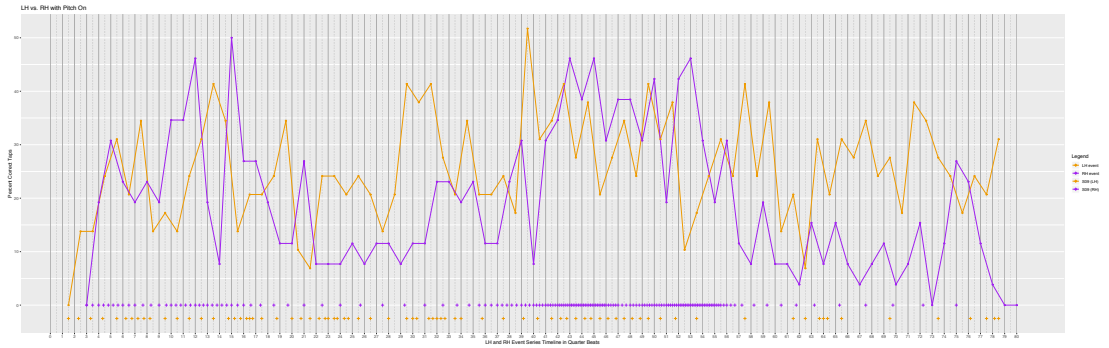
The hits total and hits percentage of some stimuli had been calculated (e.g., S01, 2, 4 and 5). As part of the Carter Beat Synchronization Visualization work, other stimuli's hits total and percentage were calculated (e.g., S03a, b, c, S06a, b, c, S09, and S07, 8, 10a, b). All Stimuli's

Hits Total and Percentage and the intermediate spreadsheets used to calculate them can be found on *Box*: https://app.box.com/folder/83279071694.

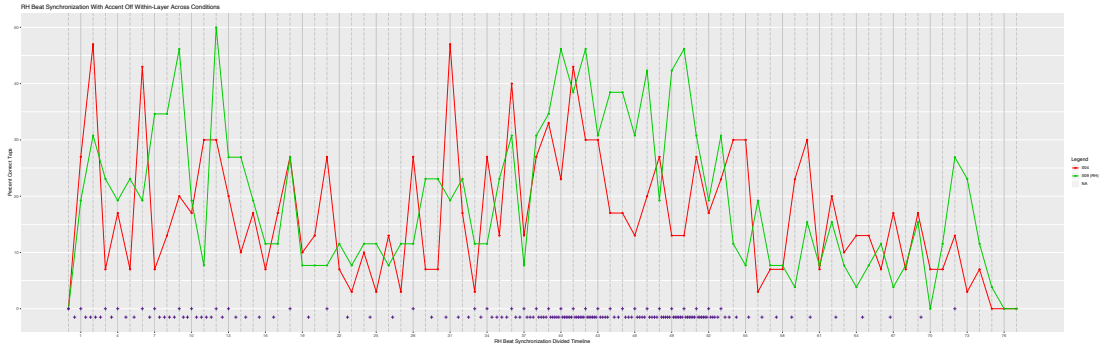
Carter Beat Synchronization Visualization Tutorial

I have created a detailed Carter Beat Synchronization Visualization tutorial on how to generate plots with the three templates that we have worked with; the inserted comments provide additional details on the process. All codes can be found in this .R file. All datasets that were imported into *R studio* can be found HERE.

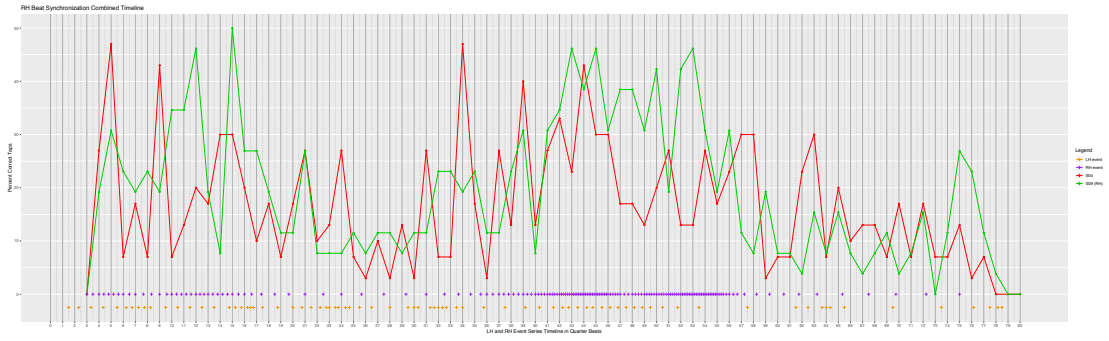
Future plots should have LH=orange and RH=purple. All current plots related to S09 should have LH=orange and RH=purple.



Above is an image of Template I: Across-Layer Comparison Against Combined Event Series (transformed data points)



Above is an image of Template II: Within-Layer Comparison Against Divided Event Series (offset events)



Above is an image of Template III: Within-Layer Comparison Against Combined Event Series (transformed data points)

Part III: Appendix

This appendix contains the in-text links as well as progress work reports not included in the main text.

I. MPE Suter (1980) Corpus Progress Work Report
   i. *ArcGIS* Work Summary
      a. Jiayi's Tutorial
      b. UBC GIS Resources
   ii. FDS *Excel* Formatting for Carto Map
      a. Final FDS Work Format Workbook
   iii. *Microsoft Office* Visualization
      a. *Office Online* Timeline of Composers Birth and Death Year
         i. Timeline Tutorial
         ii. Final Timelines
      b. *Excel* Maps of Composer Nationality
         i. Final Workbook Including All Maps
   iv. *R Commander* Plots and Statistics
      a. All Scatterplots and Box Plots
      b. Conference Follow Up Scatterplots (in 2nd Section of this linked report)

II. The Polyrhythm Humdrum Website—*Vega-Lite* and Terminal
   i. Exploratory Terminal Code and Plots
      a. Progress Work Report 1
      b. Progress Work Report 2
      c. Progress Work Report 3
   ii. Final Vega-Lite Templates and Plots
      a. stacked bar chart with separated groups CODE HERE
      b. stacked butterfly bar chart CODE HERE
      c. stacked butterfly chart when selecting specific features
   iii. Bash Tutorial

III. Carter Beat Synchronization Visualization tutorial
   i. .R file
   ii. All Spreadsheets
      a. Stimuli Hits Total & Percentage
      b. Structured Workbooks (convertible to .csv files)
      c. Original Data Format
      d. All Plots with Individual Code and .CSV Files

Part IV: Reflections

The Work Learn position in Dr. Ève Poudrier's lab has been extremely valuable for me. From close-reading of Carter's score with complex rhythm to sophisticated experimental designs, the content of the research itself has been inspiring. Beyond the content, I not only became more experienced with the software and tools used while working on the two projects, but also gained familiarity with how research is conducted, specifically in what ways researchers can collaborate.

I completed data management and visualization tasks using tools that were either familiar or unfamiliar to me. Firstly, I learned more functionalities with software and programs that I knew such as *Microsoft Office* and *R*. For example, to prepare materials for Dr. Poudrier's MusCan/CAML Conference, I used *R Commander* to perform statistical analysis and data visualization. I also tried many possible functions to generate plots using *ggplot2*. Secondly, I encountered completely new software and programs such as *ArcGIS*, *Carto*, *Vega-Lite* and the Bash command terminal. For example, I was introduced to the idea of geocoding through *ArcGIS*, *Carto,* and even *Excel Maps* while I performed data management and visualization for the MPE project. Melissa and I also explored *Vega-Lite* and were able to build a template in a new computer language.

Being a part of this research team also taught me how to collaborate with other researchers. Tasks are assigned based on the specific skillsets that each one has. Team meetings and video meetings are essential to share and discuss everyone's progress and identify goals for the next stage of work. Dr. Poudrier has been very instructive and responsive in guiding me with my day-to-day tasks.

I am amazed with the ability of how data can inform our knowledge about music, particularly how visualizations can assist music theory analysis. From our programming and computational analysis collaborators, I also envisioned a future possibility of combining my two fields of study, statistics and music, into a career. Overall, it has been a pleasure to work with such an intelligent and welcoming team.