Exploring Renaissance Music Using N-Gram Aggregates to Summarize Local Musical Content

Cory McKay, *Marianopolis College, Canada*
Rían Adamian, *McGill University, Canada*
Julie Cumming, *McGill University, Canada*
Ichiro Fujinaga, *McGill University, Canada*

Medieval and Renaissance Music Conference (MedRen)
July 4, 2020.
Edinburgh, Scotland, UK

What are n-grams?

- The notion of n-grams is drawn primarily from a substantial literature in computational linguistics
  - Typically used to represent sequences of *n* words
  - e.g. to-be-or-not-to-be is a 6-gram
    - i.e. a sequence of 6 words
  - e.g. to-be, be-or, or-not, not-to, to-be are the five 2-grams making up this 6-gram
- N-grams have also been used to many other disciplines
  - e.g. base pairs in DNA sequences, such as the A-G-C 3-gram

Outline of this talk

- This talk will focus on our work on using n-grams as the basis for learning about Renaissance music
- Will discuss experiments we performed using machine learning and statistical analysis to process n-gram data
  - Or, more specifically, n-gram features automatically extracted from music using our jSymbolic software

What is a "feature"?

- A piece of information that measures a characteristic of something (e.g. a piece of music) in a simple, consistent and precisely-defined way
- Represented using numbers
  - Can be a single value, or can be a set of related values (e.g. a histogram)
- Provides a summary description of the characteristic being measured
  - Usually macro, rather than local
- Usually extracted from pieces in their entirety
  - But can also be extracted from segments of pieces

Example: A simple feature

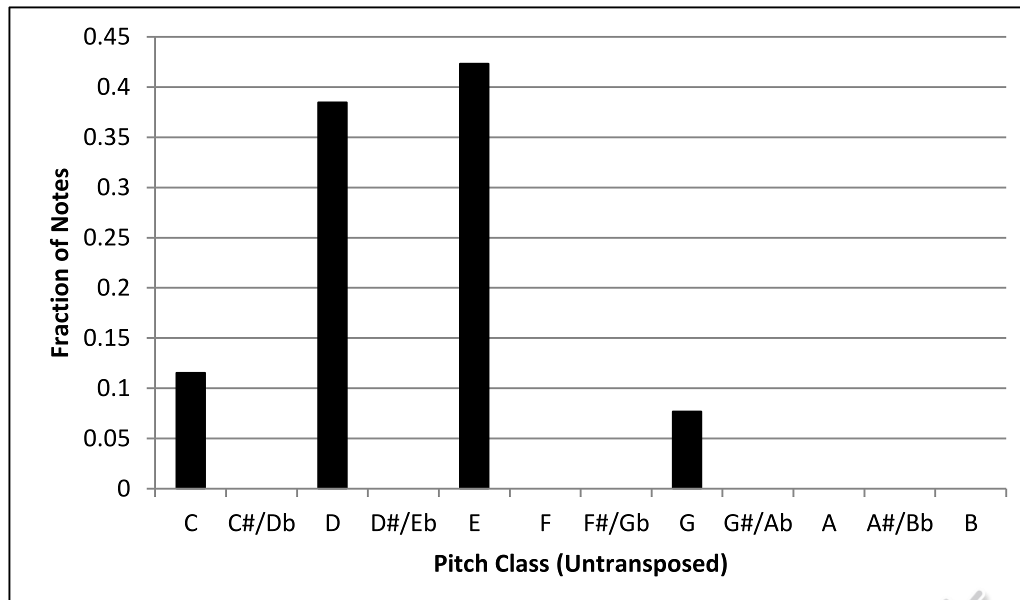
- Range (1-D): Difference in semitones between the lowest and highest pitches

- Value of this feature: 7
  - G - C = 7 semitones

Example: A histogram feature

- **Pitch Class Histogram:** Consists of 12 values, each representing the fraction of all notes belonging to an enharmonic pitch class

- Histogram graph on right shows feature values
- Pitch class counts:
  - C: 3, D: 10, E: 11, G: 2
- Most common note is E:
  - 11/26 notes
  - Corresponds to a feature value of 0.423 for E



Comparing features

- Comparing pairs of pieces in terms of features can be very revealing
  - Especially when that comparison involves hundreds or thousands of features
- Things get even more interesting, however, when comparisons are made between hundreds or thousands of pieces
  - Especially when the music is aggregated into groups, which can then be contrasted collectively
  - e.g. comparing composers, genres, regions, time periods, etc.

Benefits of features

- Provide an empirical basis for manual comparison by experts, machine learning or statistical analysis
- Permits studies involving huge quantities of music (thousands of pieces!)
- Can simultaneously explore a broad range of musical characteristics (thousands!) and their interrelationships
  - Including characteristics one may not have thought to consider
- No need to formally specify specific queries or heuristics before beginning analyses
  - But may do so if one wishes to, of course
  - Facilitates exploratory research
- Help to avoid potentially incorrect ingrained assumptions and biases
  - But only if treated properly

jSymbolic: Introduction

- jSymbolic is a software platform for extracting features from digital scores
- Compatible with Macs, PCs and Linux computers
- Free and open-source

jSymbolic (version 2.2)'s feature types

- Pitch statistics
  - e.g. Range
- Melody / horizontal intervals
  - e.g. Most Common Melodic Interval
- Chords / vertical intervals
  - e.g. Vertical Minor Third Prevalence
- Texture
  - e.g. Parallel Motion
- Rhythm
  - e.g. Note Density per Quarter Note
- Instrumentation
  - e.g. Note Prevalence of Unpitched Instruments
- Dynamics
  - e.g. Variation of Dynamics

jSymbolic 3.0 (alpha)

- Currently being tested and refined internally
- Many miscellaneous usability improvements
  - Including expanded multilingual support
- Many new features
  - 533 unique features and 2040 feature values, as of June 2020
    - Up from 246 and 1497, respectively
  - Including new features based on n-grams

Measuring local sequences (1/2)

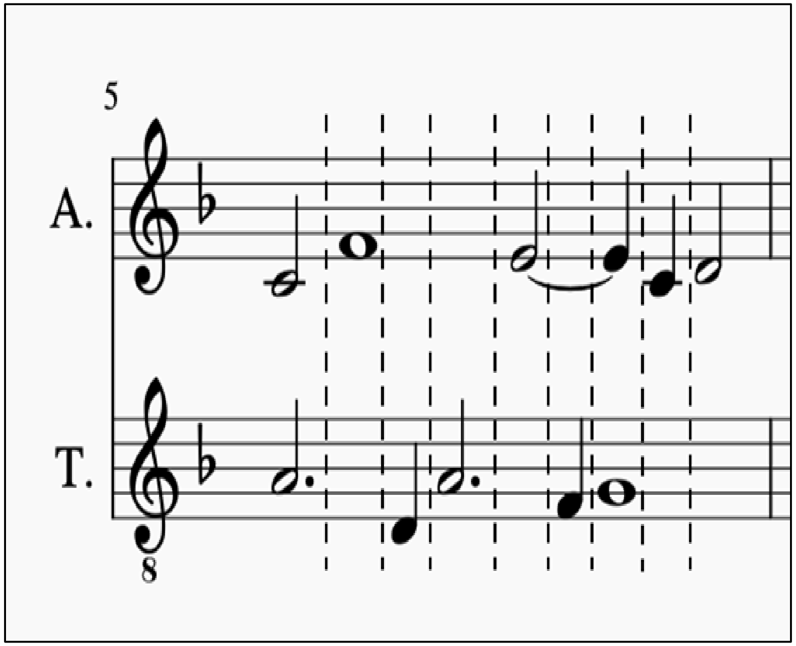
- jSymbolic (version 2.2) strongly emphasizes global summary statistics
  - i.e. describing an overall, aggregated characteristic of the music, like its range
- Features measuring local sequences are very limited in version 2.2
  - e.g. melodic transitions from just one note to the next or, at the most, single melodic arcs
  - Many features simply ignore the order in which events happen

Measuring local sequences (2/2)

- This is a limitation, since many musically interesting phenomena are associated with local patterns and sequences
- Challenge: How can one measure local behaviour while also maintaining the requirement that features be expressed as simple global numbers?
  - Through n-grams!
  - Let us begin by first defining “note onset slices” (sometimes called “salami slices” in the literature) . . .

Note onset slices (1/2)

- A slice consists of vertical groups of notes sounding simultaneously
  - e.g. the first slice on the right contains the pitches A and C
- A new slice is started every time a new (pitched) note onset occurs
- Slices are separated by dotted lines on the right
- There can be different kinds of onset slices:
  - e.g. a slice may only contain notes starting at the beginning of the slice
  - e.g. a slice may alternatively also contain notes held from previous slices
  - e.g. a slice may omit notes that are only held for less than some fraction of the slice

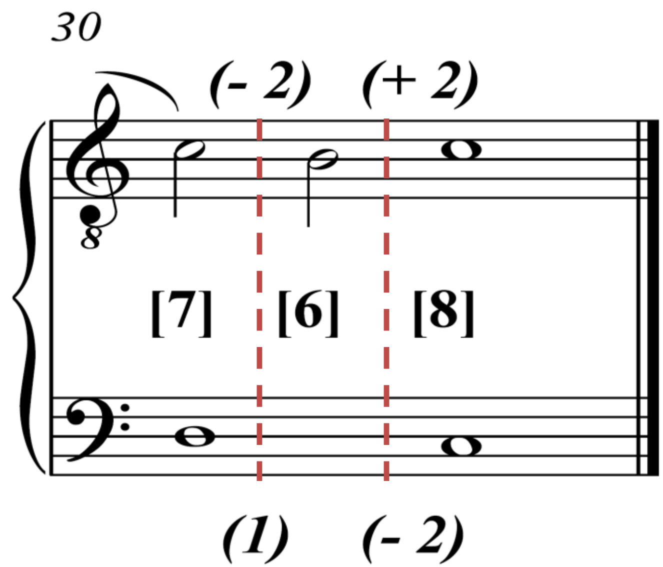


Note onset slices (2/2)

- Note onset slices provide units of grouped notes that permit the formation of structures based on local transitions and sequences associated with:
  - Vertical intervals
  - Horizontal intervals
  - Contrapuntal (vertical + horizontal) movement
  - Rhythm
  - More, such as textural patterns
- Sequences of such slices can be used to construct “n-grams” . . .

N-grams

- Our musical n-grams encode sequences of *n* (or *n+1*) note onset slices
  - Enumerated by exhaustively breaking the music into overlapping sliding windows
- Examples of 4 different kinds of diatonic n-grams (right):
  - 7 6 8 is a 3-gram indicating vertical intervals between the two voices
  - -2 2 is a 2-gram indicating horizontal intervals in the upper voice
  - [7] (1 -2) [6] (-2 2) [8] is a contrapuntal 3-gram that encodes both vertical and horizontal transitions
  - 2 2 4 (half half whole) is a 3-gram indicating the upper voice's rhythmic sequence



N-gram variants (1/2)

- What value(s) of n are best?
  - n=2 is typically too local (not enough musical content)
  - n>3 tends to lead to an explosion of rare n-grams
  - Is n=3 the Goldilocks n-gram?
    - It depends what one is interested in
- Which voices should n-grams consider?
  - All voices together?
  - Outer voices only?
  - Individual voices?

N-gram variants (2/2)

- What kinds of note onset slices are best?
  - Should brief “decorative” notes be included?
  - Should notes held from earlier slices be included?
  - Should duplicate notes (unisons) be included?
  - Should any new note trigger a new slice, or only a new pitch class?
- Other details?
  - Should melodic direction be encoded?
  - How should intervals be represented?
    - e.g. M3 vs. 3 vs. 4 (semitones)
  - Should intervals be octave-wrapped?
    - e.g. a 10^{th} to a 3^{rd}

Earlier work on musical n-grams

- Unlike our work with jSymbolic, earlier musicological research involving n-grams has focused mainly on contrapuntal n-grams
  - Often using a “modules” terminology to refer to repeated contrapuntal combinations
- Jessie Ann Owens (1998) did ground-breaking work relating to musical n-grams
- Peter Schubert (2007) further developed this into an analytical approach
- Our group has also since experimented with n-grams in previous work, but not in a feature-based way
  - e.g. Antila and Cumming (2014); Schubert and Cumming (2015); Arthur (2017); Cumming and McKay (2018, with Schubert, Condit-Shultz and Stuchbery)
- Others are also doing fascinating work involving concepts linked to n-grams
  - e.g. Richard Freedman et al.’s CRIM Project

Representing n-grams as features

- Features must be represented as standardized simple numbers that can be consistently compared
- Extracting sequences of n-grams from a piece can result in differently sized lists of n-grams
  - Does not fit the above requirements for a feature
- We thus need to extract features from the n-gram sequences
  - Rather than use n-gram sequences directly as features

Sample possible n-gram feature types (1/2)

- N-gram frequency histograms
  - How often each n-gram occurs relative to all n-grams of the same type in a piece
  - Many possible variants, including:
    - All possible n-gram permutations
    - Specific n-grams of interest (e.g. cadential n-grams)
    - Sorted from most common to least on a piece-by-piece basis
- Many features can be derived from such histograms
  - Or they can be used directly as feature vectors themselves

Sample possible n-gram feature types (2/2)

- Particular n-grams
  - Most common, second most common, etc.
  - e.g. if the most common rhythmic 3-gram is half-half-half, then this feature value would be the 3-dimensional vector [2,2,2]
- Prevalence of such n-grams
  - e.g. the most common rhythmic 3-gram might represent 0.204 (20.4%) of all rhythmic 3-grams
- Diversity of n-grams
  - e.g. number of unique n-grams
  - e.g. number of very common n-grams (those comprising, say, 15% or more of all rhythmic 3-grams)
  - e.g. number of rare n-grams (those comprising, say, 3% or less of all rhythmic 3-grams)
- etc.

jSymbolic 3.0 alpha's current n-grams and features derived from them

- Calculates three main types of n-grams:
  - Vertical, horizontal and rhythmic
  - Also calculates variants of these
    - e.g. vertical all voices vs. vertical outer voices only
- Uses n=3
  - At least for now
- Extracts 76 unique features from these n-grams
- Both n-grams and their note onset slices can be set to be calculated in a variety of ways
  - Permits experimentation with n-gram approaches

Experiments (1/2)

- We reran 3 feature-based experiments we have presented at previous MedRens, each with these variants:
  - Original features (no n-gram features)
  - New n-gram features only
  - Original features + new n-gram features

Experiments (2/2)

- All experiments involved using machine learning to train statistical models that could automatically classify music
  - The models performed classifications based only on the jSymbolic features they were given
  - Each experiment involved the same corpus described in the original conference presentation
  - Support vector machines with linear kernels were used in all cases
  - Results on the next slide are all classification accuracies based on 10-fold cross-validation

Cross-validation classification accuracies

Experiment	Original Features (No N-Grams)	N-Gram Features Only	Combined Features
Composers (2017) Josquin vs. La Rue	85.4%	69.5%	86.0% (statistically indistinguishable from performance of original features)
Genre (2018) Madrigals vs. Motets vs. Frottole/Villotte	68.4%	82.8%	74.1%
Region (2019) Franco-Flemish masses & motets vs. Iberian masses & motets	93.6%	81.6%	98.6%

Experimental discussion (1/2)

- For the composer and region experiments, when looked at individually, none of the n-gram features particularly stood out with respect to "information gain" (a rough measure of discriminative power)
  - i.e. the n-gram features performed well when grouped collectively, but none were particularly discriminative individually
  - When the original and n-gram features were combined, the features with the highest (individual) information gain were overwhelming not n-gram features

Experimental discussion (2/2)

- The genre experiment, in contrast, showed the following n-gram features had the highest (individual) information gain (they were the top 3 features when both groups were combined, and were 5 of the top 7):
  - Number of Distinct Melodic Interval 3-gram Types in Highest Line
  - Prevalence of Median Melodic Interval 3-gram Type in Highest Line
  - Number of Distinct Vertical Interval 3-gram Types
  - Number of Rare Melodic Interval 3-gram Types
  - Number of Distinct Melodic Interval 3-gram Types

Conclusions and future research

- Our experiments show that the new jSymbolic n-gram features clearly encapsulate useful information about Renaissance music
- We have only begun to scratch the surface of what can be done with n-grams
  - Can we interpret musical meaning from n-grams the way we have from other features in the past?
    - i.e. what insights can n-grams teach us about the particular style of a composer, region, etc.?
  - Can we come up with better, more useful n-grams?
    - And better features to extract from them?

Thanks for your attention

- **E-mail:** cory.mckay@mail.mcgill.ca
- **jSymbolic:** http://jmir.sourceforge.net
- **SIMSSA:** https://simssa.ca

McGill

Schulich School of Music
École de musique Schulich

CIRMMT Centre for Interdisciplinary Research in Music Media and Technology

Social Sciences and Humanities Research Council of Canada

Conseil de recherches en sciences humaines du Canada

Canada

Marianopolis College

Fonds de recherche sur la société et la culture Québec

DDMAL Distributed Digital Music Archives & Libraries Lab

SIMSSA Single Interface for Music Score Searching and Analysis

jMIR