Performing Statistical Musicological Research Using jSymbolic and Machine Learning

Cory McKay

Marianopolis College and CIRMMT

2018 Anatomy of Polyphonic Music around 1500 International Conference
Cascais, Portugal

Topics

- Introduction to “features”
  - How they can be useful for musicologists and music theorists
- jSymbolic
- Overview of research performed with jSymbolic
  - Sidebar: Avoiding encoding bias

jMIR

Personal context

- I was originally trained as a physicist, an information scientist and a jazz guitarist before changing careers and focusing on music information retrieval
- As a former physicist, I am deeply attached to:
  - Abstract theoretical models
  - Empirical validation of those models
- I think we do a great job at the first of these in musicology and music theory
  - But there is still room for improvement with respect to the second
  - The time needed to study scores and manuscripts by hand has severely limited the breadth and scope of many results

Software and statistics

- Automated software tools and statistical analysis techniques allow us to:
  - Study huge quantities of music very quickly
    - More than any human could reasonably look at
  - Empirically validate (or repudiate) our theoretical predictions
  - Do purely exploratory studies of music
  - See music from fresh perspectives
    - Can inspire new ways of looking at music

Human involvement is crucial

- Of course, computers certainly cannot replace the expertise and insight of musicologists and theorists
  - Computers instead serve as powerful tools and assistants that allow us to greatly expand the scope and empirical supportability of our work
- Computers do not understand or experience music in ways at all similar to humans
  - We must pose the research questions for them to investigate
  - We must interpret the results they present us with
- Music is, after all, defined by human experience, not some “objective” externality

What are “features”?

- Pieces of information that can characterize something (e.g. a piece of music) in a simple way
- Usually numerical values
  - A feature can be a single value, or it can be a set of related values (e.g. a histogram)
- Can be extracted from pieces in their entirety, or from segments of pieces

Example: A basic feature

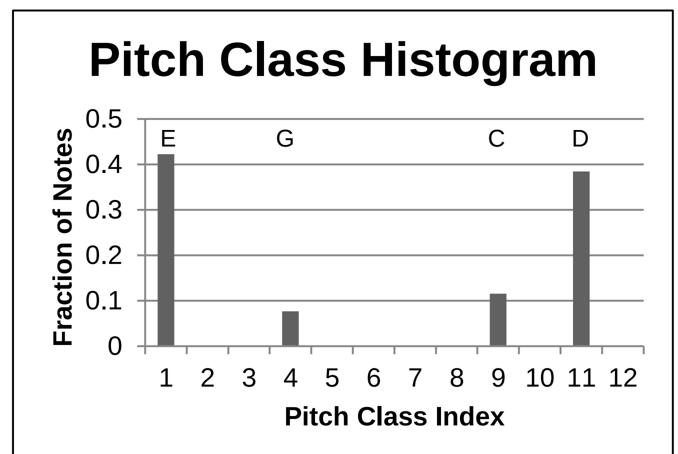
- Range (1-D): Difference in semitones between the highest and lowest pitches

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

Example: A histogram feature

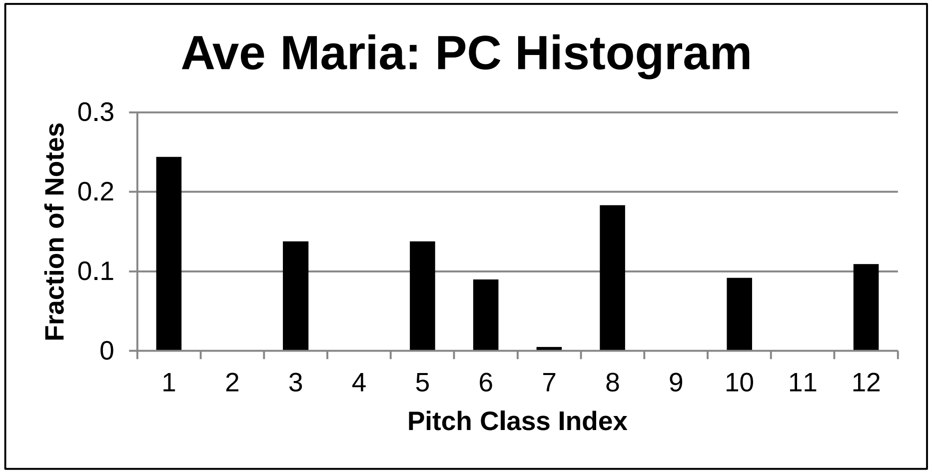
- Pitch Class Histogram: Each of its 12 values represents the fraction of all notes belonging to a particular pitch class
  - The first value corresponds to the most common pitch class
  - Each following value corresponds to a pitch class a semitone higher than the previous

- Values of this feature: see graph ->
  - Note counts: C: 3, D: 10, E: 11, G: 2
  - Most common note: E (11/26 notes)
    - Corresponding to 0.423 of the notes
  - E is thus pitch class 1, G is pitch class 4, C is pitch class 9, D is pitch class 11



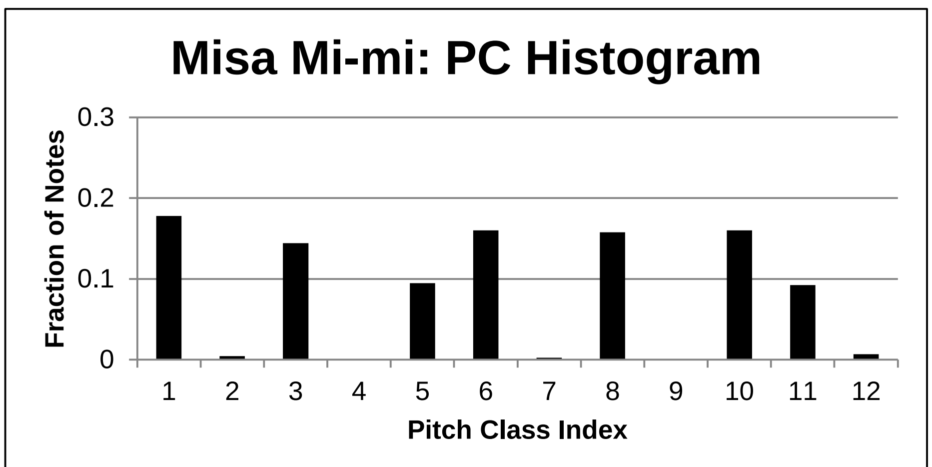
Josquin's *Ave Maria... Virgo serena*

- Range: 34
- Repeated notes: 0.181
- Vertical perfect 4ths: 0.070
- Rhythmic variability: 0.032
- Parallel motion: 0.039



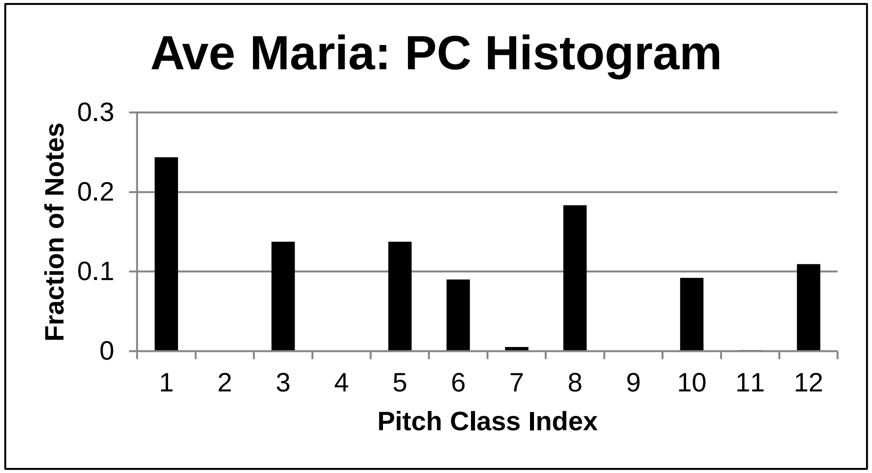
Ockeghem's *Missa Mi-mi* (Kyrie)

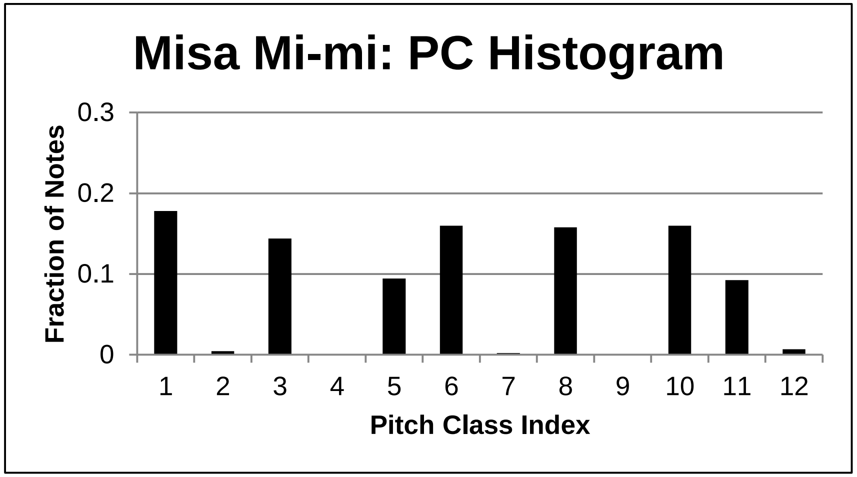
- Range: 26
- Repeated notes: 0.084
- Vertical perfect 4ths: 0.109
- Rhythmic variability: 0.042
- Parallel motion: 0.076



Feature value comparison

Feature	Ave Maria	Misa Mi-mi
Range	34	26
Repeated notes	0.181	0.084
Vertical perfect 4ths	0.070	0.109
Rhythmic variability	0.032	0.042
Parallel motion	0.039	0.076





Comparing features

- Comparing pairs of pieces like this in terms of features can be very revealing
  - Especially when that comparison involves hundreds or thousands of features, not just six
- Things get really interesting, however, when comparisons are made between hundreds or thousands of pieces, not just two
  - Especially when the music is aggregated into groups, which can then be contrasted collectively
  - e.g. comparing sets of composers, genres, regions, time periods, etc.

How can we use features? (1/2)

- Manual analysis to look for patterns
- Apply statistical analysis and visualization tools to study features extracted from large collections of music
  - Highlight patterns
  - Measure how similar various types of music are
  - Study the relative musical importance of various features
  - Observe unexpected new things in the music
- Perform sophisticated content-based searches of large musical databases
  - e.g. find all pieces with less than X amount of chromaticism and more than Y amount of contrary motion

How can we use features? (2/2)

- Use supervised machine learning to classify music
  - Done by training on pre-labelled data
  - Can study music using whatever categories (“classes”) one likes
    - e.g. composer, genre, style, time period, culture, region, etc.
  - Many possible direct applications:
    - e.g. identify the composers of unattributed musical pieces
    - e.g. explore the stylistic origins of genres, such as madrigals
    - e.g. delineate regional styles, such as Iberian vs. Franco-Flemish
- Use unsupervised machine learning to cluster music
  - Done by training on unlabelled data
  - i.e. see how the system groups pieces based on statistical similarity
    - And then see if we can find meaning in these groups

Tools for examining features

- Manually:
  - Text editors
  - Spreadsheets
- With automatic assistance:
  - Statistical analysis software
    - e.g. SPSS, SAS, etc.
  - Machine learning and data mining software
    - e.g. Weka, Orange, etc.
- Many of these tools can produce helpful visualizations

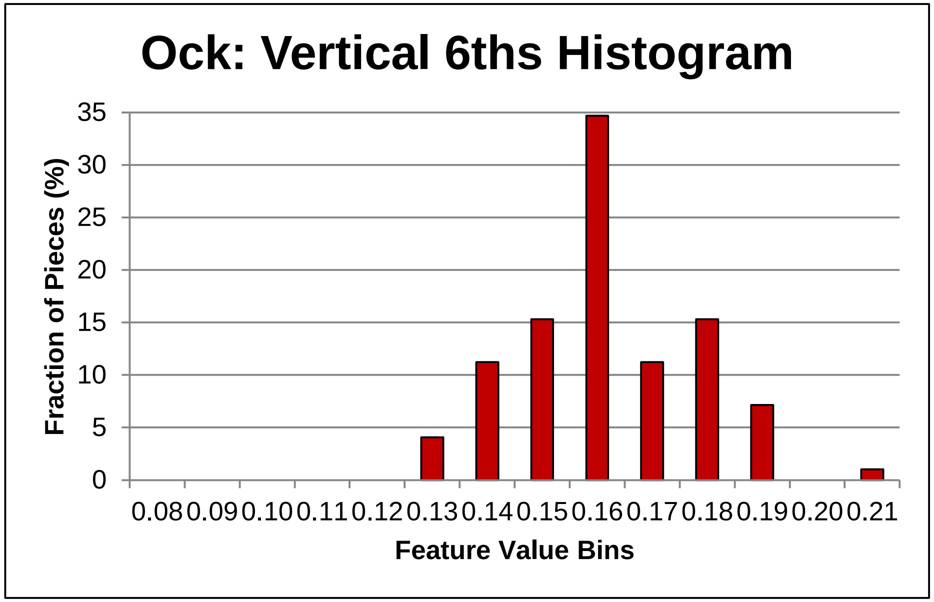
Feature visualization: Histograms (1/6)

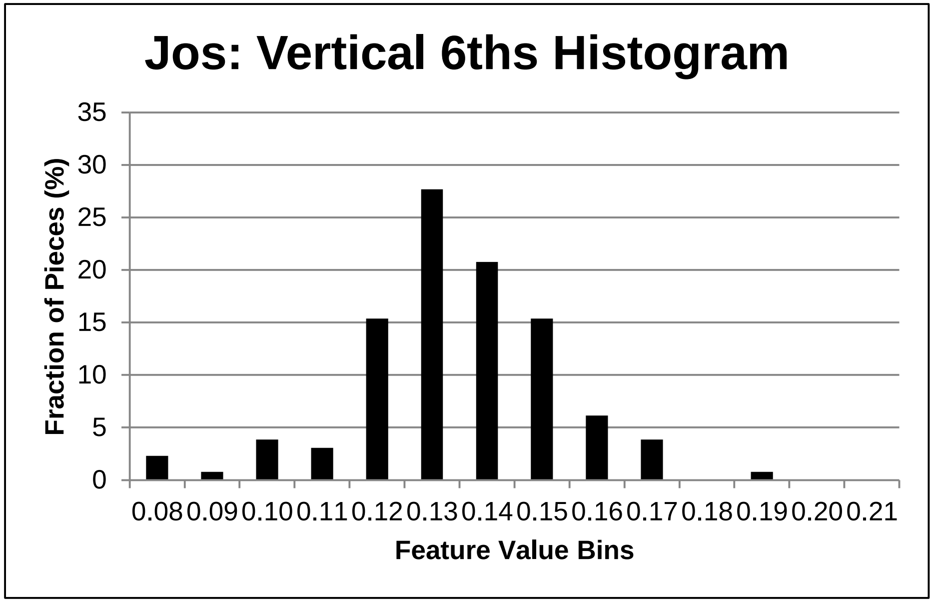
- Histograms offer a good way to visualize how the values of a feature are distributed across a corpus as a whole
  - As opposed to focusing on individual pieces
- The x-axis corresponds to a series of bins, with each corresponding to a range of values for a given feature
  - e.g. the first bin could correspond to Parallel Motion feature values between 0 and 0.1, the next bin to Parallel Motion values between 0.1 and 0.2, etc.
- The y-axis indicates the fraction of all pieces that have a feature value within the range of each given bin
  - e.g. if 30% of pieces in the corpus have Parallel Motion values between 0.1 and 0.2, then this bin (0.1 to 0.2) will have a y-coordinate of 30% (or, equivalently, 0.3)

Feature visualization: Histograms (2/6)

- In other words:
  - Each bar on a histogram represents the fraction of pieces in a corpus with a feature value falling in that bar's range of feature values
- Clarification: I am speaking here about a way to visualize a 1-dimensional feature as it is distributed across a corpus of interest
  - This is distinct from the multi-dimensional histogram features discussed earlier
    - e.g. Pitch Class Histograms
  - Although both are equally histograms, of course

Feature visualization: Histograms (3/6)

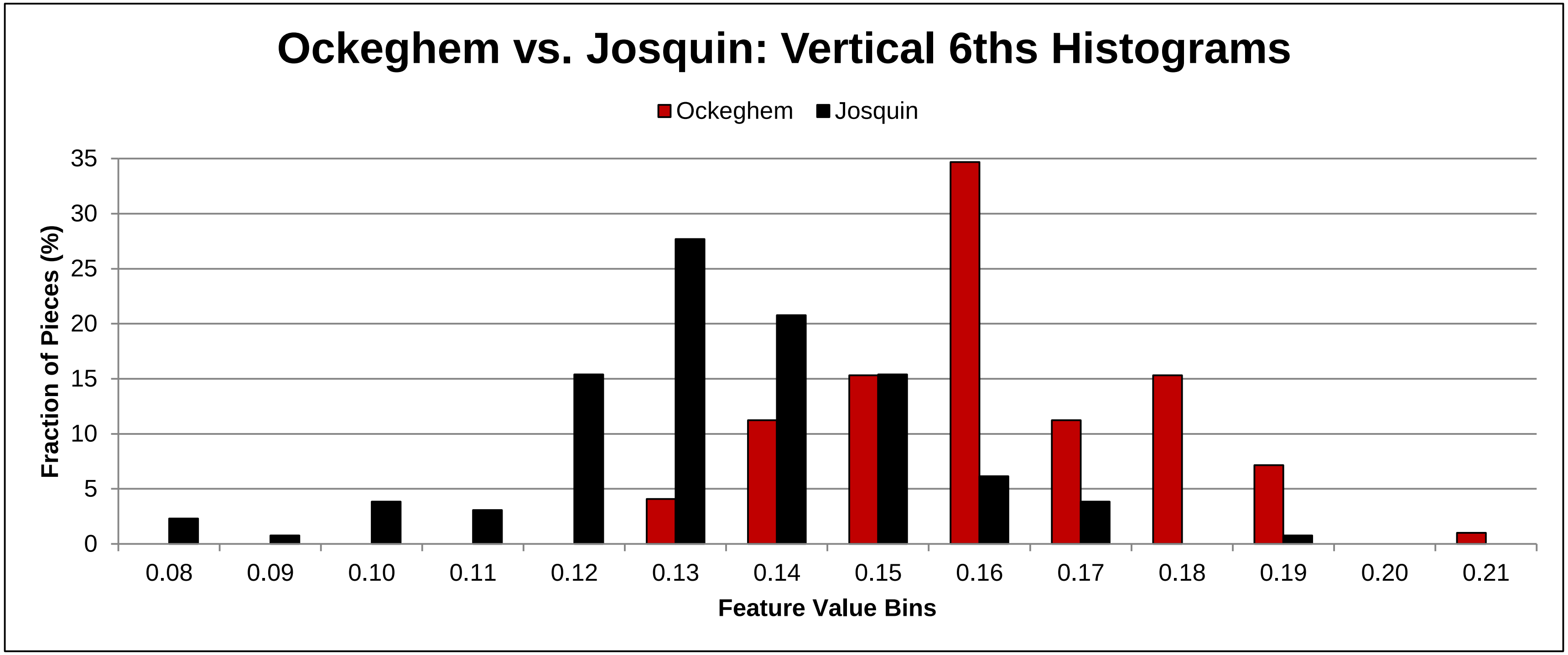




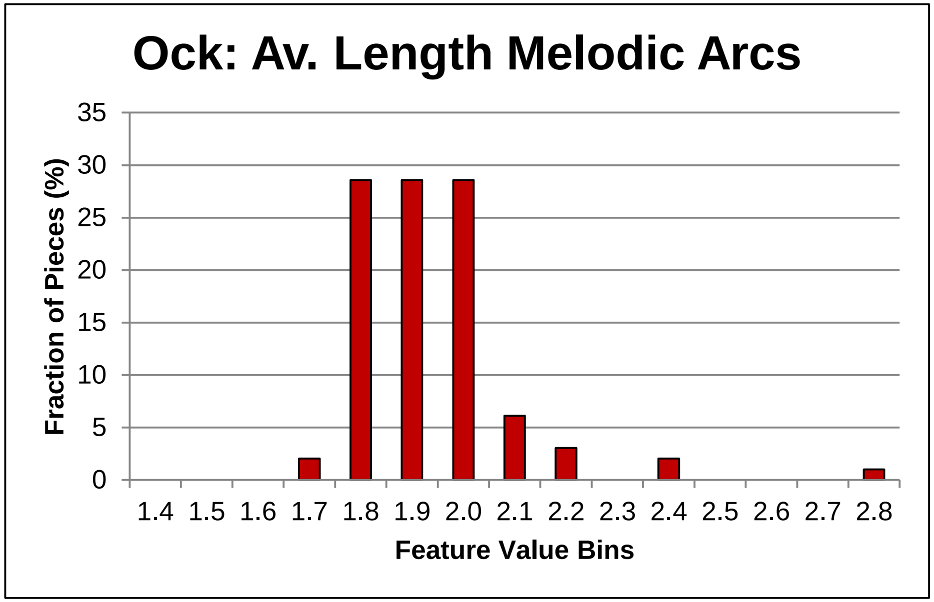
- These histograms show that Ockeghem tends to have more vertical 6ths (between all pairs of voices) than Josquin
  - Ockeghem peaks in the 0.16 to 0.17 bin, at nearly 35%
  - Josquin peaks in the 0.13 to 0.14 bin, at about 28%
- Of course, there are also clearly many exceptions
  - This feature is helpful, but is limited if only considered alone

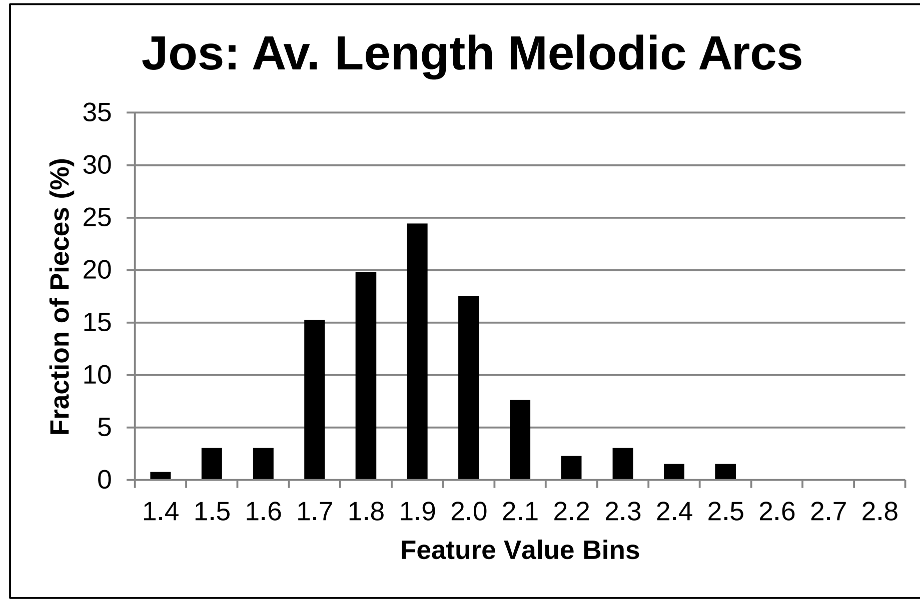
Feature visualization: Histograms (4/6)

- The histograms for both composers can be superimposed onto a single chart:



Feature visualization: Histograms (5/6)

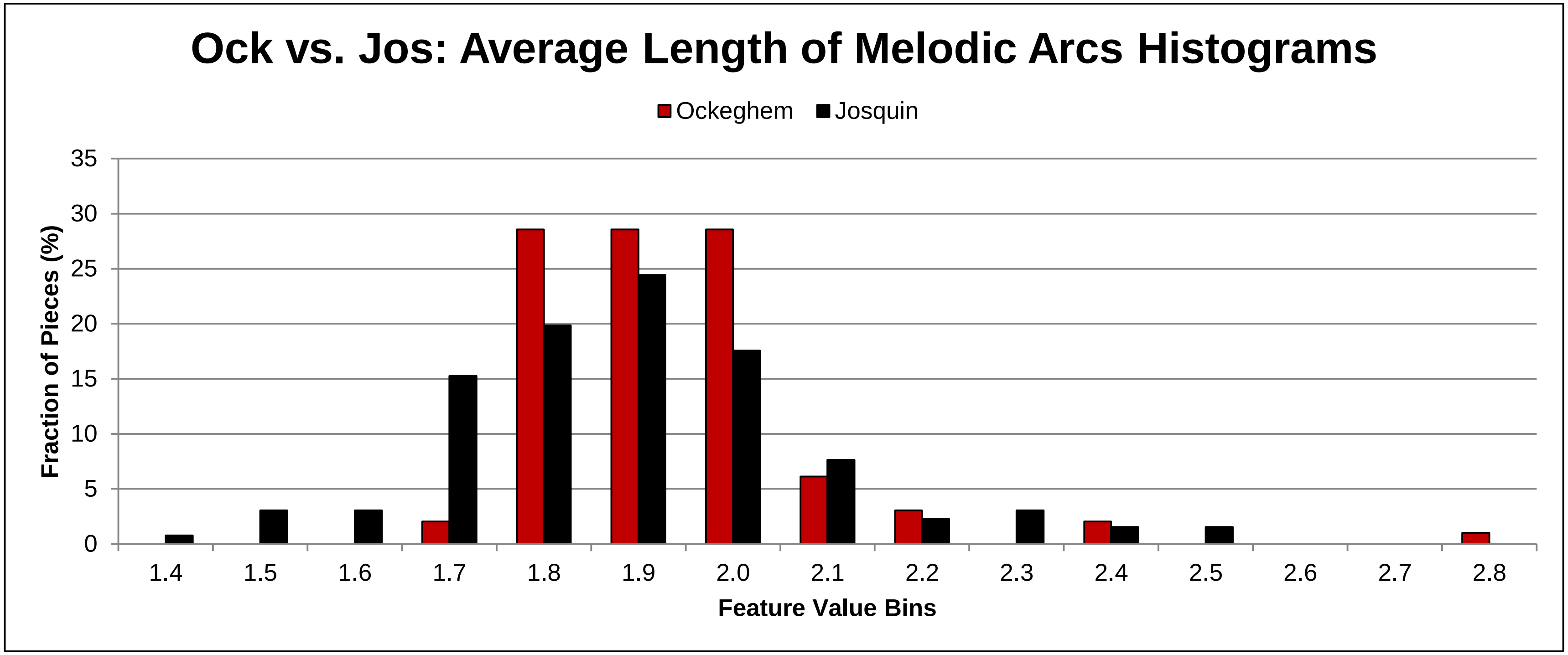




- These histograms show that Ockeghem tends to have longer melodic arcs (average number of notes separating peaks & troughs)
  - Both peak in the 1.9 to 2.0 bin
  - However, Josquin's histogram is (slightly) more skewed to the far left
- Of course, there are once again clearly many exceptions
  - This feature is also helpful, but also limited if considered alone

Feature visualization: Histograms (6/6)

- Once again, the histograms for both composers can be superimposed onto a single chart:

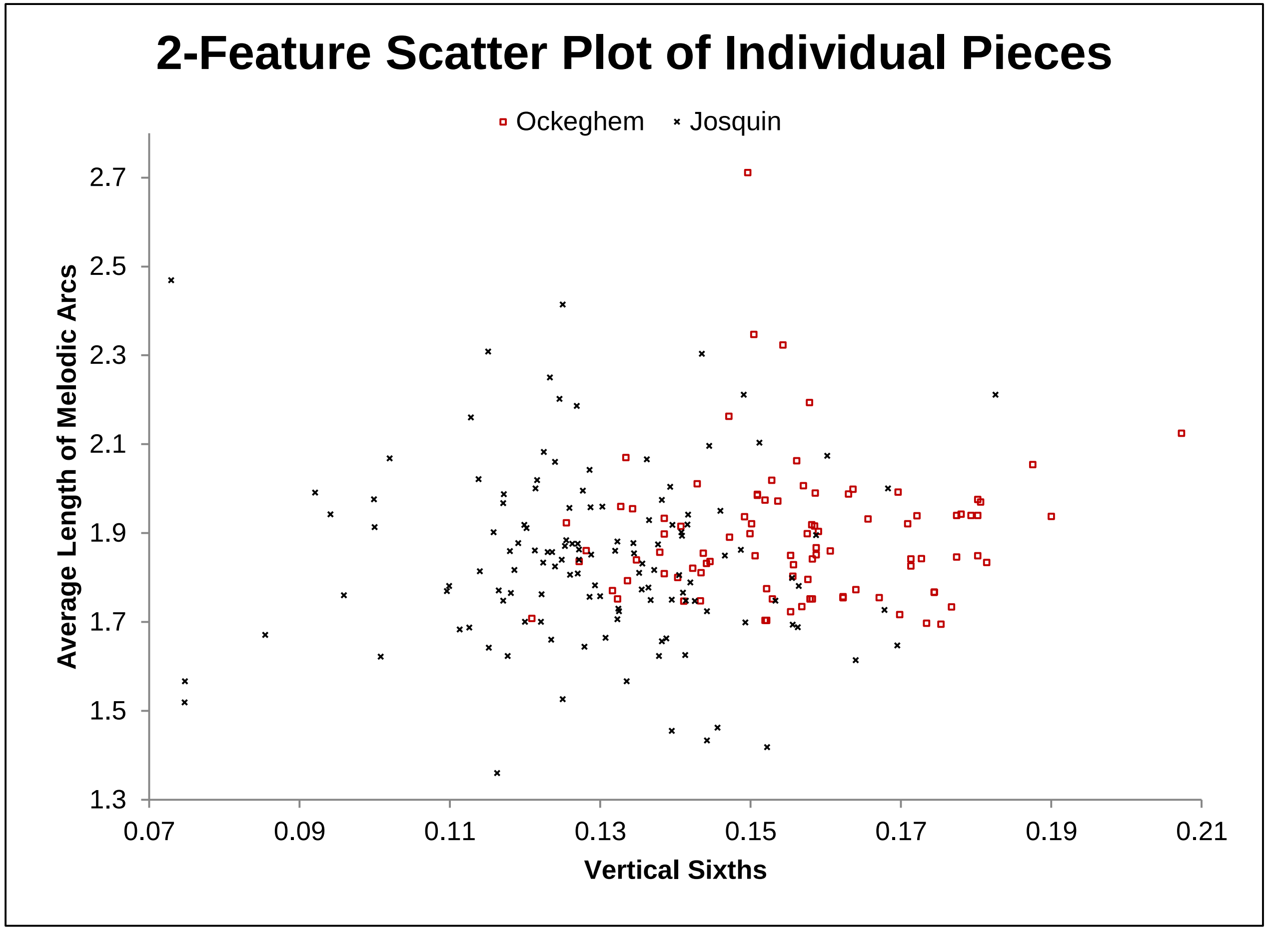


Feature visualization: Scatter plots (1/6)

- Scatter plots are another good way to visualize feature data
  - The x-axis represents one feature
  - The y-axis represents some other feature
  - Each point represents the values of these two features for a single piece
- Scatter plots let you see pieces individually, rather than aggregating them into bins like histograms
  - Scatter plots also let you see more clearly how the two features divide the different composers
- To make them easier to read, scatter plots typically have just 2 dimensions
  - Computer classifiers, in contrast, work with much larger n-dimensional scatterplots (one dimension per feature)

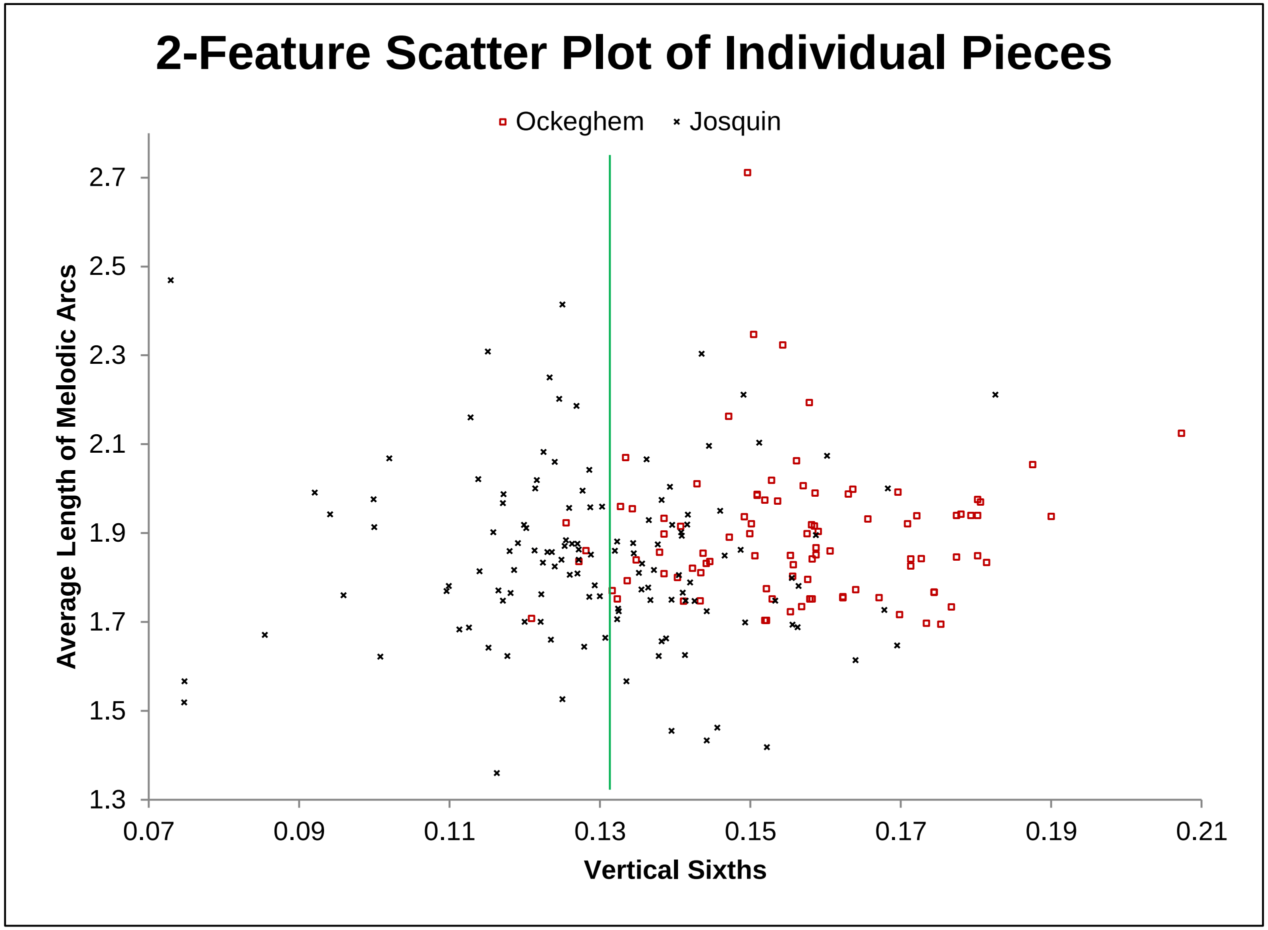
Feature visualization: Scatter plots (2/6)

- Josquin pieces tend to be left and low on this graph



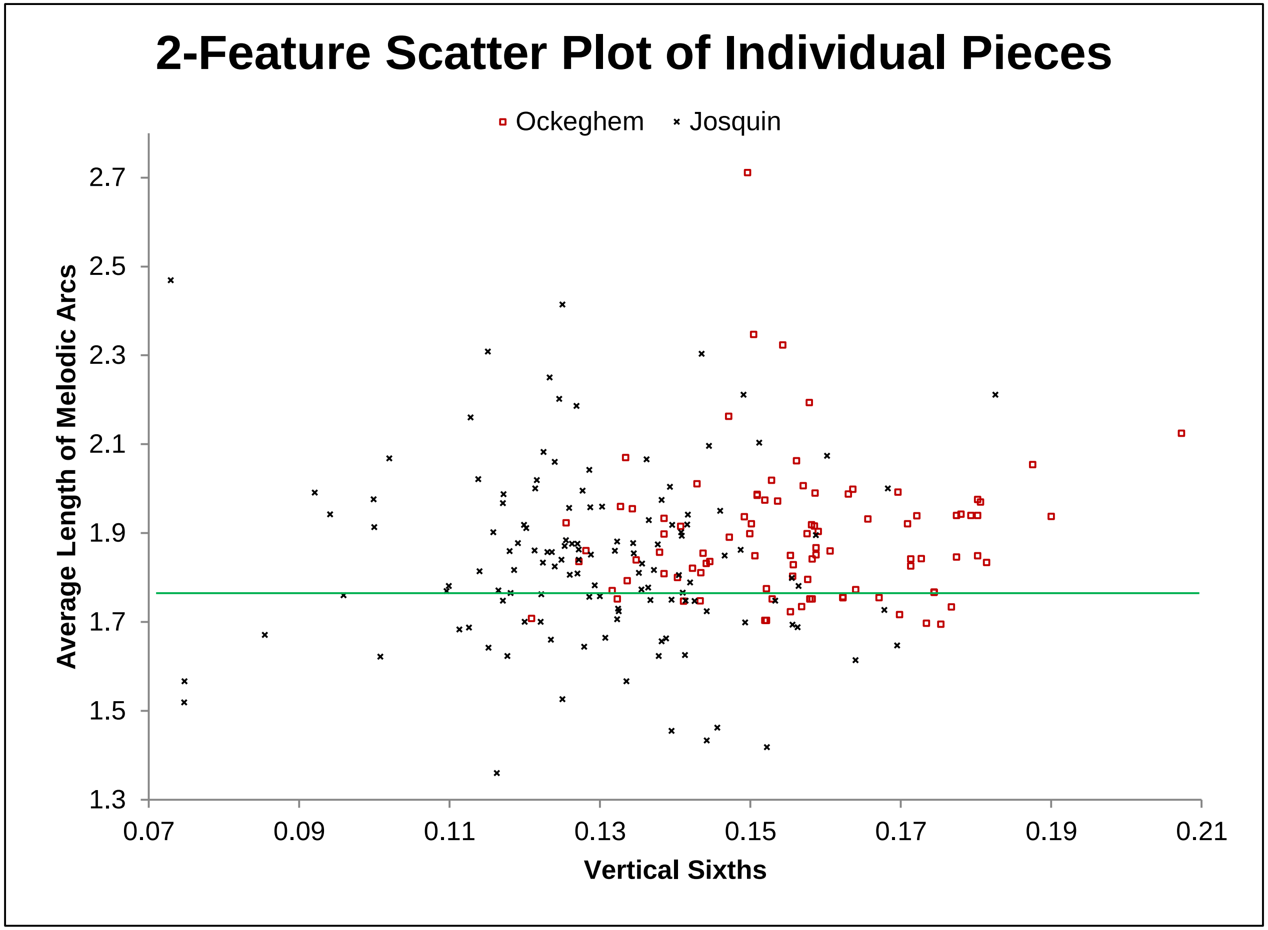
Feature visualization: Scatter plots (3/6)

- Simply drawing a single 1-D dividing line ("discriminant") results in a not entirely terrible classifier based only on Vertical Sixths
  - But many pieces would still be misclassified
  - Get 62% classification accuracy using an SVM and just this one feature



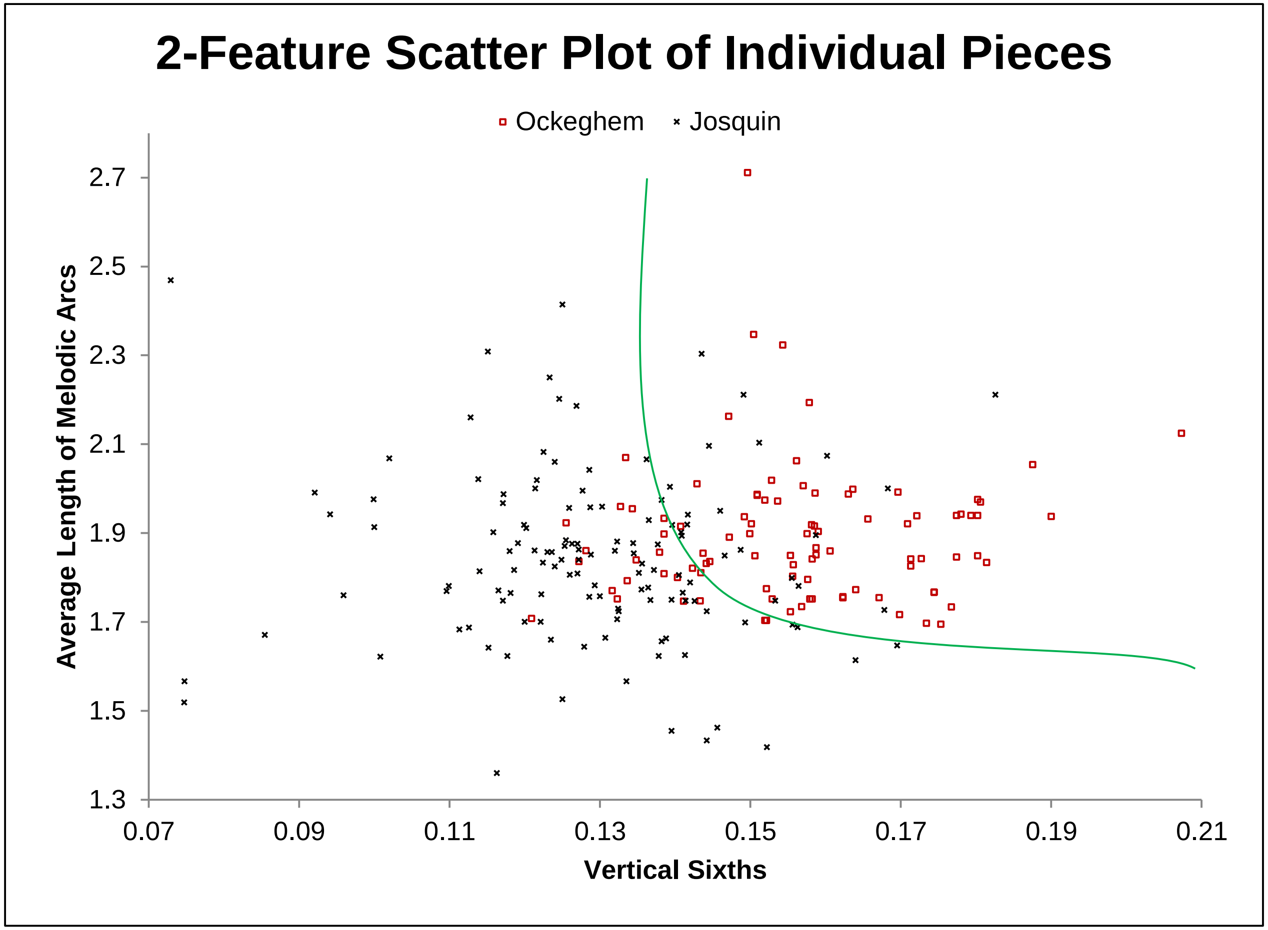
Feature visualization: Scatter plots (4/6)

- Could alternatively draw a 1-D discriminant dividing the pieces based only on the Average Length of Melodic Arcs
  - Get 57% classification accuracy using an SVM and just this one feature
  - Not as good as the Vertical Sixths discriminant (62%)



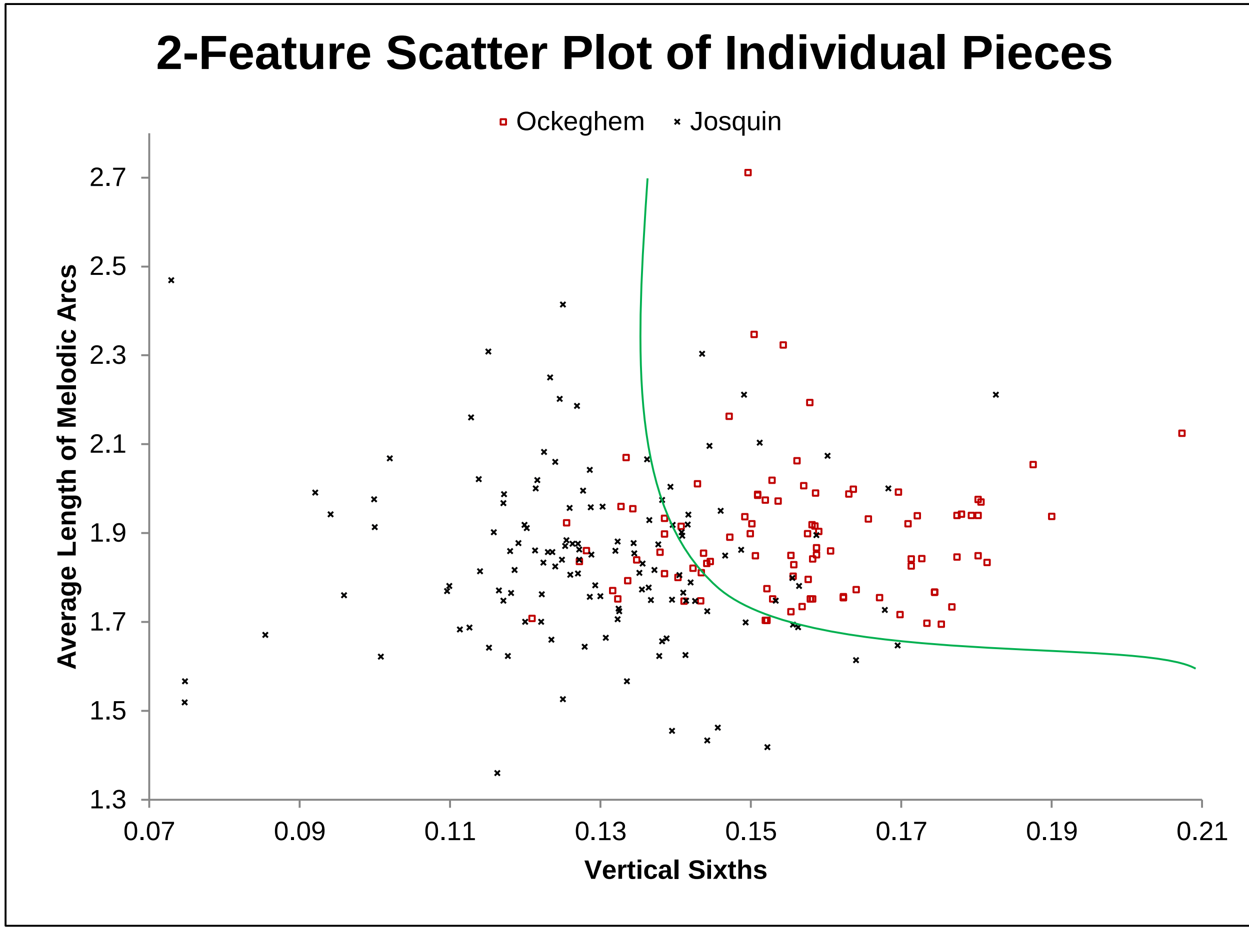
Feature visualization: Scatter plots (5/6)

- Drawing a curve (another kind of discriminant) divides the composers still better that either of the previous discriminants
  - Get 80% accuracy using an SVM and just these 2 features!
- More than 2 features are clearly needed to improve performance



Feature visualization: Scatter plots (6/6)

- In fact, many (but not all) types of machine learning in effect simply learn where to place these kinds of discriminants as they train
- But typically with many more then just two features, of course



Benefits of features

- Can quickly perform consistent empirical studies involving huge quantities of music
- Can be applied to diverse types of music in consistent ways
- Permit simultaneous consideration of thousands of features and their interrelationships
  - And can statistically condense many features into low-dimensional spaces when needed
- No need to formally specify any queries or heuristics before beginning analyses
  - Unless one wants to, of course
- Help to avoid potentially incorrect ingrained assumptions and biases

Salience

- A fundamental difference between traditional and feature-based approaches is linked to:
  - Perceived salience of particular pieces
  - Perceived salience of particular musical characteristics
- Human experts know (or assume they know?) what is important to look at
  - Due to time constraints, experts thus tend to focus primarily on the pieces (or excerpts) and the musical characteristics they expect to be important
  - This means that, in many research projects, the significant majority of a given repertory is left unstudied, and many musical characteristics are left unexplored
  - The selected pieces or characteristics may not be representative
- Computers, in contrast, have no expectations as to what is important, and time is much less of a constraint for them
  - So they can look at everything we let them look at

But . . .

- Certain essential areas of insight are left uninvestigated by content-based symbolic features (at least so far)
  - Qualities that are difficult to precisely define and measure consistently
    - e.g. amount and types of imitation
  - Text
  - Historical evidence

Computers need us!

- So, a feature-based approach is useless without:
  - Human experts to ask important questions
  - Human experts to interpret results
  - Human experts to place feature values in the larger context
- Automatically extracted features are thus a tool that expert musicologists and theorists can add to their already rich toolbox
  - Features are a great tool that opens up many new possibilities, but a tool that this is of very limited utility by itself

Features and potential bias

- But does a feature-based approach really avoid bias?
  - What if the composition of the research corpus computers are provided with is limited or biased?
  - What if the encoding of the music is biased?
    - A particular problem if files with inconsistent encodings (and editorial decisions) are compared
  - What if the particular features that are implemented are limited or biased?

Choosing features to implement

- Which features do we need?
  - The ones that are relevant to the kinds of music under consideration
  - The ones we already know or suspect are important
  - The ones that are important, but we do not know it yet
- So, we need a lot of diverse features!
  - So we can deal with many types of music
  - So we can address the interests of many different researchers
  - So we permit unexpected but important surprises
  - So we are less likely to miss out on important insights
- The same can be said for data
  - The more music and the more varied it is the better!
  - We'll return briefly to data in a bit, but let's focus on features for the moment . . .

jSymbolic: Introduction

- jSymbolic is a software platform for extracting features from symbolic music
  - Part of the much larger jMIR package
- Compatible with Macs, PCs and Linux computers
- Free and open-source

What does jSymbolic do?

- (Version 2.2) extracts 246 unique features
- Some of these are multi-dimensional histograms, including:
  - Pitch and pitch class histograms
  - Melodic interval histograms
  - Vertical interval histograms
  - Chord types histograms
  - Rhythmic value histograms
  - Beat histograms
  - Instrument histograms
- In all, (version 2.2) extracts a total of 1497 separate values

jSymbolic: Feature types (1/3)

- Pitch Statistics:
  - What are the occurrence rates of different pitches and pitch classes?
  - How tonal is the piece?
  - How much variety in pitch is there?
- Melody / horizontal intervals:
  - What kinds of melodic intervals are present?
  - How much melodic variation is there?
  - What kinds of melodic contours are used?
- Chords / vertical intervals:
  - What vertical intervals are present?
  - What types of chords do they represent?
  - How much harmonic movement is there?

jSymbolic: Feature types (2/3)

- Texture:
  - How many independent voices are there and how do they interact (e.g. moving in parallel, crossing voices, etc.)?
- Rhythm:
  - Rhythmic values of notes
  - Intervals between the attacks of different notes
  - Use of rests
  - What kinds of meter is used?
  - Rubato?
- Instrumentation:
  - What types of instruments are present and which are given particular importance relative to others?
- Dynamics:
  - How loud are notes and what kinds of dynamic variations occur?

jSymbolic: Feature types (3/3)

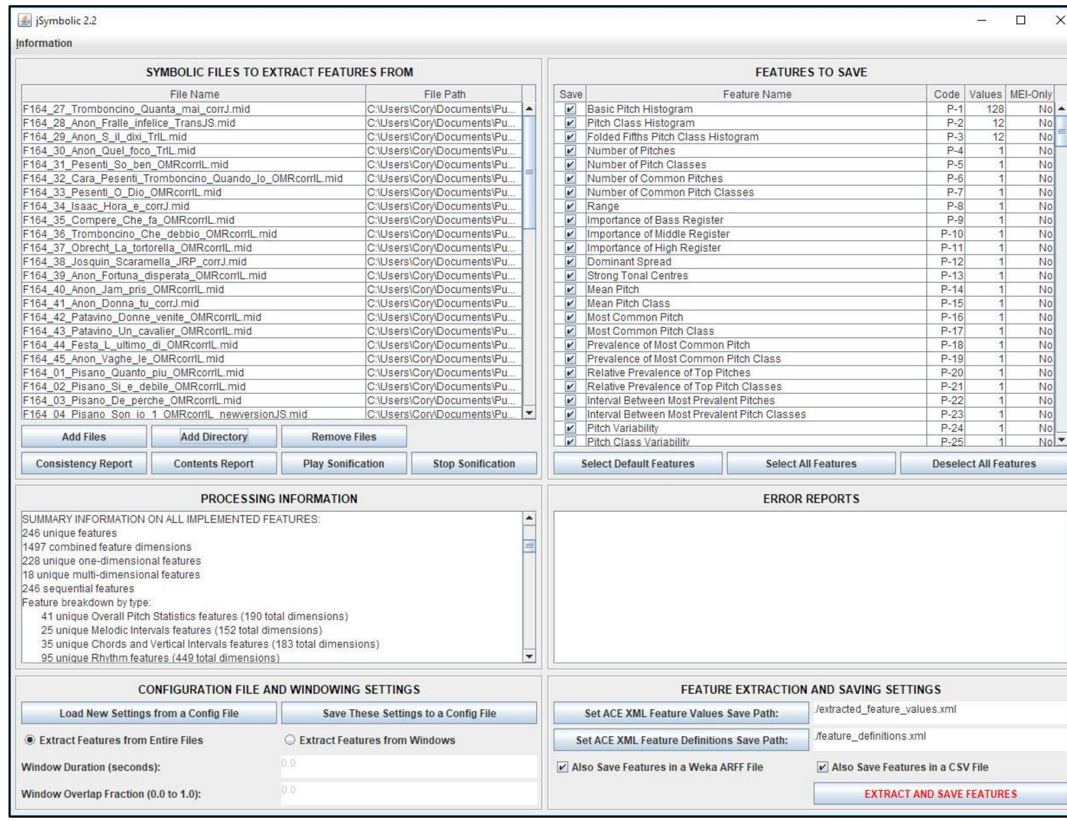
- jSymbolic only (for now) extracts features associated with musical content
- There are thus no features associated with:
  - Text
  - Historical evidence
- This is partly a disadvantage:
  - Obviously these kinds of information are essential
  - Researchers using jSymbolic features must of course use their expertise to consider extracted features in the larger context
- It is also an advantage, however:
  - It allows us to (temporarily) focus only on the music, so that we can find insights there that we might otherwise have missed

Other music research software

- jSymbolic is intrinsically different from other software used in empirical symbolic music research
  - e.g. music21 (includes a port of the original jSymbolic features)
  - e.g. Humdrum
  - e.g. VIS
- This other software is excellent for finding exactly where specific things one is searching for happen
  - Perfect for very targeted research based on specific searches
- jSymbolic, in contrast, allows one to acquire large amounts of summary information about music with or without knowing *a priori* what one is looking for
  - Good for general annotation of symbolic databases
  - Good for statistical analysis and machine learning
  - Good for free exploratory research
  - Good for large-scale validations of theoretical models

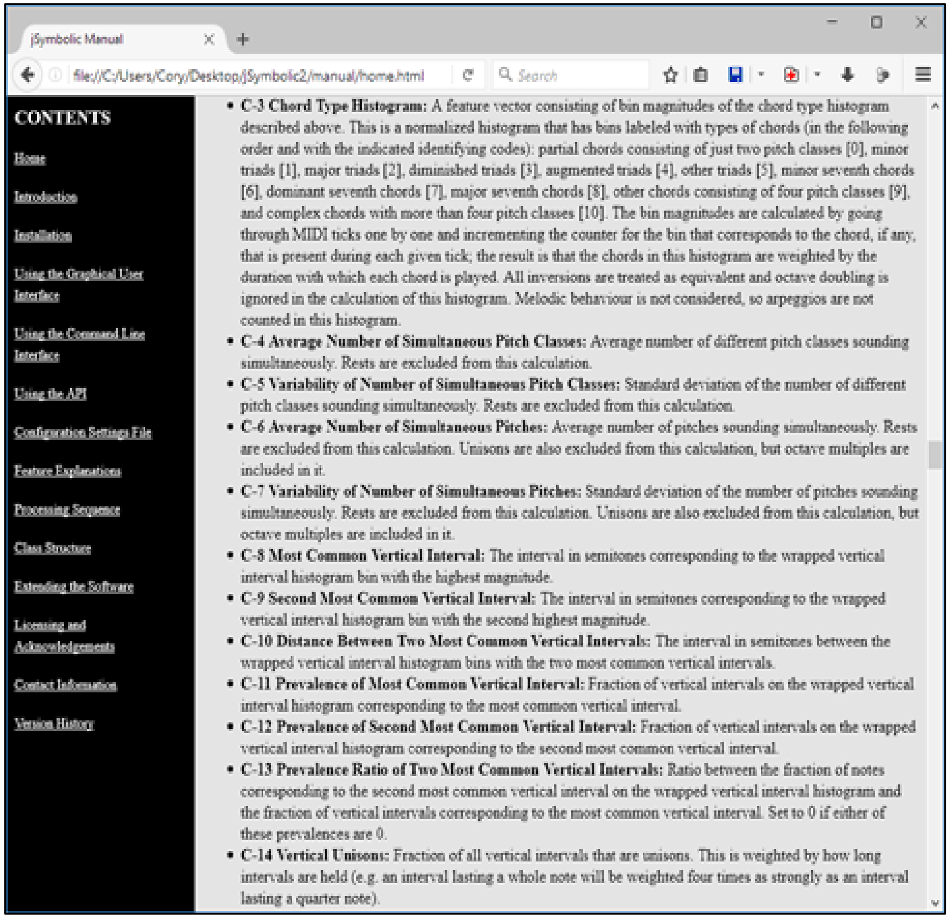
jSymbolic: User interfaces

- Graphical user interface
- Command line interface
- Java API
- Rodan workflow



jSymbolic: Manual

- Extensive manual includes:
  - Detailed feature descriptions
  - Detailed instructions on installation and use
- There is also a step-by-step tutorial with worked examples


jSymbolic Manual

file:///C:/Users/Cory/Desktop/jSymbolic2/manual/home.html

CONTENTS

Home

Introduction

Installation

Using the Graphical User Interface

Using the Command Line Interface

Using the API

Configuration Settings File

Feature Explanations

Processing Sequence

Class Structure

Extending the Software

Licensing and Acknowledgements

Contact Information

Version History

- **C-3 Chord Type Histogram:** A feature vector consisting of bin magnitudes of the chord type histogram described above. This is a normalized histogram that has bins labeled with types of chords (in the following order and with the indicated identifying codes): partial chords consisting of just two pitch classes [0], minor triads [1], major triads [2], diminished triads [3], augmented triads [4], other triads [5], minor seventh chords [6], dominant seventh chords [7], major seventh chords [8], other chords consisting of four pitch classes [9], and complex chords with more than four pitch classes [10]. The bin magnitudes are calculated by going through MIDI ticks one by one and incrementing the counter for the bin that corresponds to the chord, if any, that is present during each given tick; the result is that the chords in this histogram are weighted by the duration with which each chord is played. All inversions are treated as equivalent and octave doubling is ignored in the calculation of this histogram. Melodic behaviour is not considered, so arpeggios are not counted in this histogram.
- **C-4 Average Number of Simultaneous Pitch Classes:** Average number of different pitch classes sounding simultaneously. Rests are excluded from this calculation.
- **C-5 Variability of Number of Simultaneous Pitch Classes:** Standard deviation of the number of different pitch classes sounding simultaneously. Rests are excluded from this calculation.
- **C-6 Average Number of Simultaneous Pitches:** Average number of pitches sounding simultaneously. Rests are excluded from this calculation. Unisons are also excluded from this calculation, but octave multiples are included in it.
- **C-7 Variability of Number of Simultaneous Pitches:** Standard deviation of the number of pitches sounding simultaneously. Rests are excluded from this calculation. Unisons are also excluded from this calculation, but octave multiples are included in it.
- **C-8 Most Common Vertical Interval:** The interval in semitones corresponding to the wrapped vertical interval histogram bin with the highest magnitude.
- **C-9 Second Most Common Vertical Interval:** The interval in semitones corresponding to the wrapped vertical interval histogram bin with the second highest magnitude.
- **C-10 Distance Between Two Most Common Vertical Intervals:** The interval in semitones between the wrapped vertical interval histogram bins with the two most common vertical intervals.
- **C-11 Prevalence of Most Common Vertical Interval:** Fraction of vertical intervals on the wrapped vertical interval histogram corresponding to the most common vertical interval.
- **C-12 Prevalence of Second Most Common Vertical Interval:** Fraction of vertical intervals on the wrapped vertical interval histogram corresponding to the second most common vertical interval.
- **C-13 Prevalence Ratio of Two Most Common Vertical Intervals:** Ratio between the fraction of notes corresponding to the second most common vertical interval on the wrapped vertical interval histogram and the fraction of vertical intervals corresponding to the most common vertical interval. Set to 0 if either of these prevalences are 0.
- **C-14 Vertical Unisons:** Fraction of all vertical intervals that are unisons. This is weighted by how long intervals are held (e.g. an interval lasting a whole note will be weighted four times as strongly as an interval lasting a quarter note).

jSymbolic: File formats

- Input:
  - MIDI
  - MEI
- Output:
  - CSV
  - ACE XML
  - Weka ARFF

jSymbolic: Miscellany

- Windowed feature extraction
  - Including overlapping windows
- Configuration files
  - Pre-set feature choices
  - Pre-set input and output choices
  - More
- Can combine jSymbolic with other jMIR components to perform multimodal research
  - i.e. combine symbolic features with other features extracted from audio, lyrics and cultural data
  - This improves results substantially! (McKay et al. 2010)

jSymbolic: Extensibility

- jSymbolic is specifically designed such that music scholars can design their own features and work with programmers to then very easily add these features to the jSymbolic infrastructure
  - Fully open source
  - Modular plug-in feature design
  - Automatically handles feature dependencies and scheduling
  - Very well-documented code

Important software principles

- As Frans Wiering wisely pointed out at IMS 2017, those of us who produce research software must be careful to give musicologists what they want and need
  - Rather than trying to impose choices on them
- This emphasizes the importance of establishing on on-going dialog
  - Software designers should find out from musicologists what will be valuable to them
  - Software designers can also present musicologists with the possibility of options that they would not necessarily have though of, or thought possible
- So, please let me know what you need or want!

Research involving jSymbolic

- I will now briefly highlight several research projects that have been carried out based on jSymbolic features
  - To give you an idea of what is possible
- I will place a special focus on a study comparing Renaissance composers
  - It is particularly illustrative
- I will also discuss new research on distinguishing Iberian Renaissance music from Franco-Flemish Renaissance music
- Several other studies will also be discussed
  - In less detail

Composer identification study

- Related publication: MedRen 2017
- Used jSymbolic features to automatically classify pieces of Renaissance music by composer
  - As an example of the kinds of things that can be done with jSymbolic
  - As a meaningful research project in its own right

RenComp7 dataset

- Began by constructing the "RenComp7" dataset:
  - 1584 MIDI files
  - By 7 Renaissance composers
- Combines:
  - Top right: Music drawn from the Josquin Research Project (Rodin, Sapp and Bokulich)
  - Bottom right: Music by Palestrina (Miller 2004) and Victoria (Sigler, Wild and Handelman 2015)

Composer	Files
Busnoys	69
Josquin *(only includes the 2 most secure Jesse Rodin groups)*	131
La Rue	197
Martini	123
Ockeghem	98

Composer	Files
Palestrina	705
Victoria	261

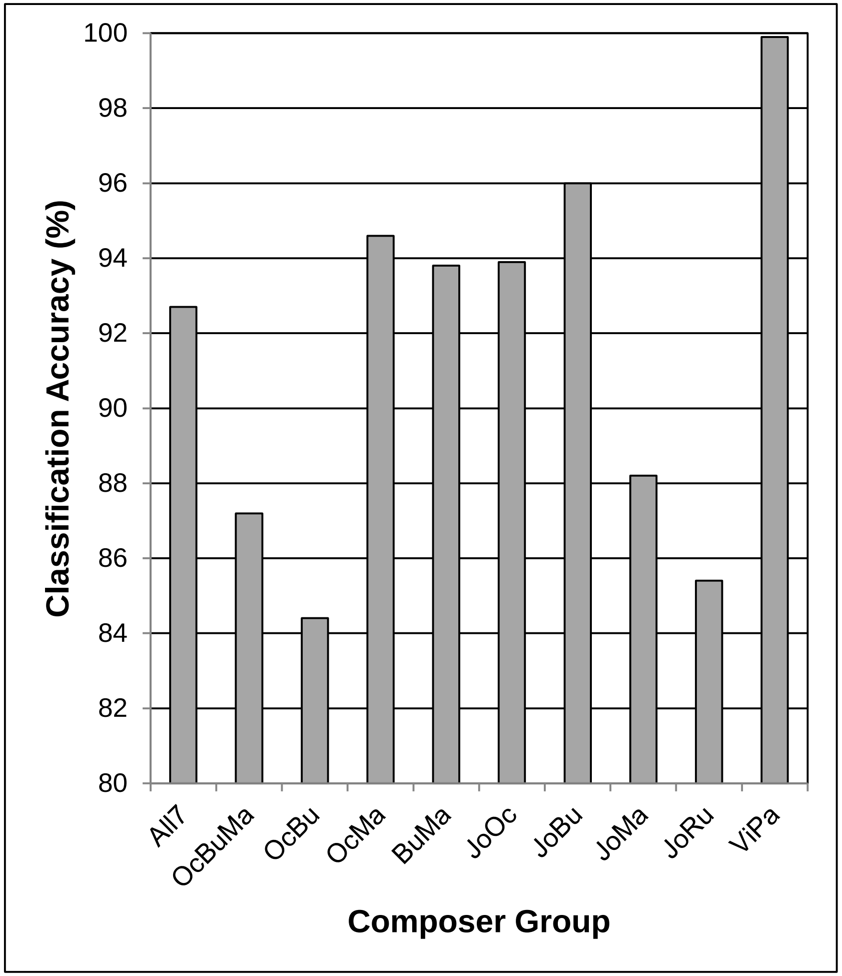
Methodology

- Extracted 721 feature values from each of the 1584 RenComp7 files using jSymbolic 2.0
- Used machine learning to teach a classifier to automatically distinguish the music of the composers
  - Based on the jSymbolic features
- Used statistical analysis to gain insight into relative compositional styles
- Performed several versions of this study
  - Classifying amongst all 7 composers
  - Focusing only on smaller subsets of composers
    - Some more similar, some less similar

jMIR

Classification results

Composer Group	Classification Accuracy
All 7	92.7%
Ockeghem / Busnoys / Martini	87.2%
Ockeghem / Busnoys	84.4%
Ockeghem / Martini	94.6%
Busnoys / Martini	93.8%
Josquin / Ockeghem	93.9%
Josquin / Busnoys	96.0%
Josquin / Martini	88.2%
Josquin / La Rue	85.4%
Victoria / Palestrina	99.9%



Direct applications of such work

- Validating existing suspected but uncertain attributions
- Helping to resolve conflicting attributions
- Suggesting possible attributions of currently entirely unattributed scores

Comparison with other work

- Brinkman, Shanahan and Sapp (2016) used 53 features to classify amongst 6 composers (J. S. Bach and five Renaissance composers)
  - Obtained success rates of roughly 63% on average
  - Did very well in distinguishing Bach from the Renaissance composers (97% on average)
  - This highlights both the high quality of their approach and the particular difficulty of differentiating the music of Renaissance composers
    - Which, in turn, makes the success of the jSymbolic 2.0 features on exclusively Renaissance (92.7% amongst 7 composers) music all the more encouraging
    - Of course, non-identical datasets make direct comparisons problematic

How do the composers differ?

- Some very interesting questions:
  - What musical insights can we learn from the jSymbolic feature data itself?
  - In particular, what can we learn about how the music of different composers differs?
- Chose to focus on two particular cases:
  - Josquin vs. Ockeghem: Relatively different
  - Josquin vs. La Rue: Relatively similar

A priori expectations (1/3)

- What might an expert musicologist expect to differentiate the composers?
  - Before actually examining the feature values
- Once formulating these expectations, we can then see if the feature data confirms or repudiates these expectations
  - Both are useful!
- We can also see if the feature data reveals unexpected insights

A priori expectations (2/3)

- What do you think might distinguish the composers?
  - Josquin vs. Ockeghem?
  - Josquin vs. La Rue?
- I consulted one musicologist (Julie E. Cumming) and one theorist (Peter Schubert), both experts in the period . . .

A priori expectations (3/3)

- Josquin vs. Ockeghem: Ockeghem may have . . .
  - Slightly more large leaps (larger than a 5th)
  - Less stepwise motion in some voices
  - More notes at the bottom of the range
  - Slightly more chords (or simultaneities) without a third
  - Slightly more dissonance
  - A lot more triple meter
  - More varied rhythmic note values
  - More 3-voice music
  - Less music for more than 4 voices
- Josquin vs. La Rue: La Rue may have . . . Hard to say!
  - Maybe more compressed ranges?

jMIR

Were our expectations correct?

- Josquin vs. Ockeghem: Ockeghem may have . . .
  - OPPOSITE: Slightly more large leaps (larger than a 5^{th})
  - SAME: Less stepwise motion in some voices
  - SAME: More notes at the bottom of the range
  - SAME: Slightly more chords (or simultaneities) without a third
  - OPPOSITE: Slightly more dissonance
  - YES: A lot more triple meter
  - SAME: More varied rhythmic note values
  - YES: More 3-voice music
  - YES: Less music for more than 4 voices
- Josquin vs. La Rue: La Rue may have . . .
  - SAME: Maybe more compressed ranges?

Importance of empiricism

- These results show that even some of the most highly informed experts in the field can have a number of inaccurate assumptions
  - And so, it is certain, do we all
- These results highlight the important need for empirical validation in general in musicology and music theory
  - There are very likely a range of widely held beliefs and theoretical models that will in fact turn out to be incorrect when they are subjected to exhaustive and rigorous empirical examination

(Free) diving into the feature values

- There are a variety of statistical techniques for attempting to evaluate which features are likely to be effective in distinguishing between types of music
- We used seven of these statistical techniques to find:
  - The features and feature subsets most consistently statistically predicted to be effective at distinguishing composers
- We then manually examined these feature subsets to find the features likely to be the most musicologically meaningful

Novel insights revealed (1/2)

- Josquin vs. Ockeghem (93.9%):
  - Rhythm-related features are particularly important
    - Josquin tends to have greater rhythmic variety
      - Especially in terms of both especially short and long notes
    - Ockeghem tends to have more triple meter
      - As expected
    - Features derived from beat histograms also have good discriminatory power
  - Ockeghem tends to have more vertical sixths
  - Ockeghem tends to have more diminished triads
  - Ockeghems tends to have longer melodic arcs

Novel insights revealed (2/2)

- Josquin vs. La Rue (85.4%):
  - Pitch-related features are particularly important
    - Josquin tends to have more vertical unisons and thirds
    - La Rue tends to have more vertical fourths and octaves
    - Josquin tends to have more melodic octaves

jMIR

Excluded features

- Only 721 of the available 1230 jSymbolic 2.0 features were used in order to avoid bias
  - Some excluded features were irrelevant to the data under consideration
  - Some excluded features were correlated with the source of the data
- This primarily meant removing features linked to instrumentation, dynamics and tempo

jMIR

Sidebar: Avoiding encoding bias (1/2)

- If music from multiple different sources is included in a study, then one must be careful to avoid making conclusions based on the source of the music rather than the underlying music itself
  - As this could corrupt the results
- Problems can occur when inconsistent editorial decisions are present. To be careful of in early music:
  - Inconsistent additions of accidentals (*musica ficta*)
  - Transposition to different keys
  - Choice of different note values to denote the beat
  - Differing metrical interpretations of mensuration signs
- Inconsistent encoding practices can also have an effect
  - e.g. if one set of files has precise tempo markings but another is arbitrarily annotated at 120 BPM

jMIR

Sidebar: Avoiding encoding bias (2/2)

- How to avoid corrupted feature-based results associated with the kinds of corpus inconsistencies and biases described above:
  - Ideally, use music files that were all consistently generated using the same methodology
    - All editorial decisions (e.g. *musica ficta*) should be applied consistently and should be documented
  - If this is not possible, then exclude all features that are sensitive to the particular bias present
- jSymbolic includes functionality that can help detect and identify these kinds of problems

Building valid digital symbolic music research corpora

- Related publication: ISMIR 2018
- Presents techniques and workflows for building large collections of symbolic digital music that avoid bias and facilitate statistically valid large-scale empirical studies
- Presents a corpus of Renaissance duos as a sample of how this can be done
    - Includes experiments with jSymbolic 2.2 features empirically demonstrating the negative effects that improper methodologies can produce

Josquin attribution study (1/3)

- Related publication: ISMIR 2017
- We also did a second composer-related study using the JRP data
  - This one investigated the attribution of pieces suspected to be by Josquin

Josquin attribution study (2/3)

- Jesse Rodin has broken Josquin's music into 6 levels of attribution certainty
  - Based on historical sources, not musical content
- We used the jSymbolic 2.0 features to train a 2-class SVM classifier
  - First class: Josquin
    - The Josquin music in the 2 most secure Rodin levels
  - Second class: NotJosquin
    - All the JRP music available from 21 other Renaissance composers similar to Josquin
- This model was then used to classify the Josquin music in the remaining 4 Jesse Rodin levels

Josquin attribution study (3/3)

- It turns out that, the more insecure a piece is according to Rodin's classification, the less likely it was to be classified as being by Josquin by our classifier
- This demonstrates some good empirical support for Rodin's categorizations
  - This is a great example of how features extracted by a computer and human expert knowledge can complement each other

Rodin Certainty Level	% Classified as Josquin
Level 3 "Tricky"	48.6%
Level 4 "Questionable"	17.2%
Level 5 "Doubtful"	14.0%
Level 6 "Very doubtful"	5.5%

Origins of the Italian madrigal (1/2)

- Related publication: MedRen 2018
- Where did the madrigal come from?
  - The frottola (Einstein 1949)?
  - The chanson and motet in Florence (Fenlon and Haar 1988)?
  - The Florentine carnival song, villotta, and improvised solo song (A. Cummings 2004)?
- How can we decide, based on the music?
  - Extract jSymbolic 2.2 features
  - Apply machine learning and feature analysis techniques
    - As we did with composers in the MedRen 2017 study

Origins of the Italian madrigal (2/2)

- Julie Cumming and I will present our results in Dublin next week

jMIR

Distinguishing Iberian Renaissance music from Franco-Flemish music

- New research presented for the first time
- Research question:
  - Is Iberian Renaissance music demonstrably stylistically distinct from Franco-Flemish music of the time?
- Investigated empirically:
  - Extracted jSymbolic 2.2 features from a dataset of Iberian and Franco-Flemish masses and motets
  - Trained machine learning models that could distinguish between Iberian and Franco-Flemish music
    - Based on these features
  - Tested expert predictions to see if they match the actual musical data
  - Used statistical analysis techniques to find those features that very strongly (statistically) distinguish Iberian and Franco-Flemish music

jMIR

Dataset used

- Used the "FraFle/Iber" dataset provided by the Anatomy project's team
- Consists of masses and motets
- 467 MIDI files total

- IMPORTANT CAVEAT:
  - This dataset was prepared for initial rough exploration
  - It was no yet fully cleaned, so it (and the results about to be presented) may be subject to a certain amount of encoding bias

Region	Composers	Files
Franco-Flemish Mass movements	3	286
Franco-Flemish Motets	3	59
Iberian Mass movements	7	79
Iberian Motets	10	43

Composer	FrancFlem Mass Mvts	FrancFlem Motets	Iberian Mass Mvts	Iberian Motets
La Rue	156	19		
Josquin	62	36		
Ockeghem	68	4		
Alba			6	5
Anchieta			8	9
Escobar			11	4
Fernandez de Castilleja			1	1
Illario				2
Mondejar				2
Peñalosa			42	16
Ribera			5	2
Rivafrecha				1
Sanabria				1
Tordesillas			6	

Methodological details

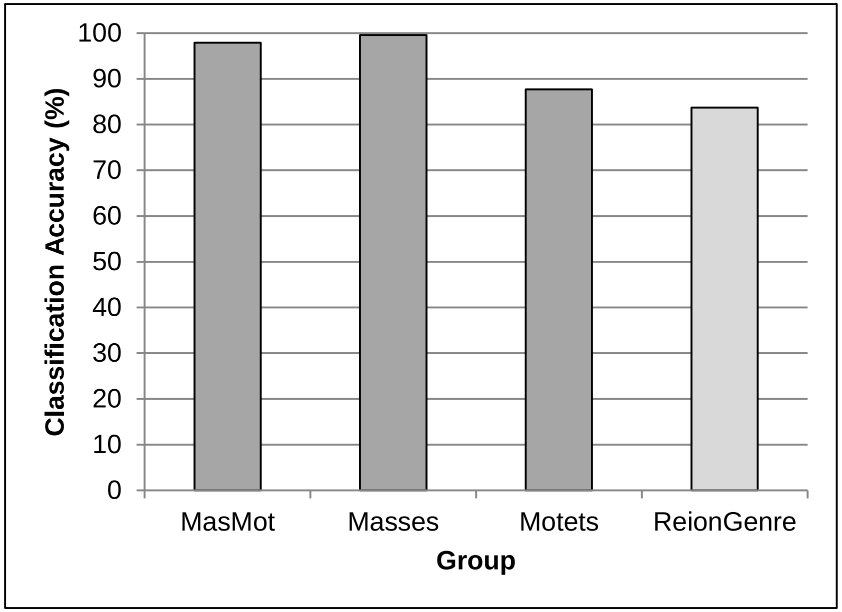
- Extracted 801 feature values from each of the 467 FraFle/Iber MIDI files using jSymbolic 2.2
  - jSymbolic 2.2 can extract 1497 features in total, but only 801 are relevant and “safe” for this particular corpus
  - i.e. excluded features associated with tempo, dynamics, instrumentation, etc. that are not relevant to this corpus
- Used machine learning to teach a classifier to automatically distinguish the music belonging to each of the regions
  - Based on the jSymbolic 2.2 features
  - Using Weka’s SMO SVM implementation
- Combined results from 10 different statistical analysis algorithms to gain insight into specific musical differences between the two regions

Classification experiments

- Performed three versions of the classification experiment, where a classifier was trained to distinguish the music by region:
  - Iberian masses and motets vs. Franco-Flemish masses and motets
  - Iberian masses vs. Franco-Flemish masses
  - Iberian motets vs. Franco-Flemish motets
- Also did a fourth study where I classified by both region and genre at once, just for fun
  - i.e. Iberian masses vs. Iberian motets vs. Franco-Flemish masses vs. Franco-Flemish motets

Classification results

Group	Classification Accuracy
Masses and motets combined	97.9%
Masses only	99.6%
Motets only	87.7%
Region AND Genre	*83.7%*



- The classifiers were very effective in distinguishing the music of the two regions!
  - This suggests that the Iberian music is observably distinct stylistically from the Franco-Flemish music
- More interpretation on these results in a bit …

How do the regions differ?

- What can we learn from the feature data about how the music belonging to each of the two regions differs?

A priori expectations

- What musical characteristics do you, as an expert musicologist, think might distinguish Iberian Renaissance music from Franco-Flemish Renaissance music?
  - Masses and motets combined?
  - Masses only?
  - Motets only?
- These predictions can then be tested against the feature data extracted across the corpus
- A number of features were tested, based on the responses many of you sent in to the comparison game call sent out before this conference started
  - Unfortunately, only characteristics for which jSymbolic features currently exist were tested
  - Some of you submitted ideas that will serve as excellent inspiration for future features!

Predicted results: Masses and motets combined

- **Prediction:** Iberian has more parallel motion
  -
- **Prediction:** Iberian has more long rhythmic values
  -
- **Prediction:** Franco-Flemish has more triple meter
  -
- **Prediction:** Iberian has more melodic leaps wider than a third
  -
- **Prediction:** Franco-Flemish mas more music with 5 or more voices
  -
- **Prediction:** Franco-Flemish has a wider pitch range
  -
- **Prediction:** Franco-Flemish pieces are longer
  -

jMIR

Predicted vs. empirical results: Masses and motets combined

- **Prediction:** Iberian has more parallel motion
  - Iberian tends to have more parallel motion
- **Prediction:** Iberian has longer rhythmic values
  - Franco-Flemish tends to have longer rhythmic values **(substantially)**
- **Prediction:** Franco-Flemish has more triple meter
  - Franco-Flemish tends to use more triple meter **(substantially)**
- **Prediction:** Iberian has more melodic leaps wider than a third
  - Iberian tends to have more leaps wider than a third **(substantially)**
- **Prediction:** Franco-Flemish mas more music with 5 or more voices
  - Franco-Flemish is more likely to have 5 or more voices **(substantially)**
- **Prediction:** Franco-Flemish has a wider pitch range
  - Franco-Flemish tends to have a wider range
- **Prediction:** Franco-Flemish pieces are longer
  - Franco-Flemish tends to be longer **(substantially)**

jMIR

Predicted results: Masses only

- **Prediction:** Franco-Flemish has more parallel motion
  -
- **Prediction:** Franco-Flemish has more unequal rhythmic activity in the voices
  -
- **Prediction:** No meaningful difference in the amount of variation in the number of voices sounding at once
  -
- **Prediction:** Franco-Flemish has more sixth chords
  -
- **Prediction:** No meaningful difference in the amount of vertical dissonance:
  -

Predicted vs. empirical results: Masses only

- **Prediction:** Franco-Flemish has more parallel motion
  - Iberian tends to have more parallel motion **(slightly)**
- **Prediction:** Franco-Flemish has more unequal rhythmic activity in the voices
  - Franco-Flemish rhythmic values tend to vary more between voices **(slightly)**
- **Prediction:** No meaningful difference in the amount of variation in the number of voices sounding at once
  - Franco-Flemish tends to have more variation in the number of active voices **(substantially)**
- **Prediction:** Franco-Flemish has more sixth chords
  - Franco-Flemish tends to have more vertical sixths **(slightly)**
- **Prediction:** No meaningful difference in the amount of vertical dissonance:
  - Iberian tends to have more vertical dissonance

Predicted results: Motets only

- **Prediction:** Iberian has more parallel motion
  -
- **Prediction:** Franco-Flemish has more variation in the number of voices sounding at once
  -
- **Prediction:** Franco-Flemish has more vertical dissonance
  -
- **Prediction:** Iberian has more rests in all voices
  -
- **Prediction:** Franco-Flemish pieces are longer
  -

Predicted vs. empirical results: Motets only

- **Prediction:** Iberian has more parallel motion
  - Iberian tends to have more parallel motion
- **Prediction:** Franco-Flemish has more variation in the number of voices sounding at once
  - Franco-Flemish tends to have more variation in the number of active voices **(but less markedly than masses)**
- **Prediction:** Franco-Flemish has more vertical dissonance
  - Franco-Flemish tends to have more vertical dissonance **(reverse of masses!)**
- **Prediction:** Iberian has more rests in all voices
  - Iberian tends to have more rests in all voices **(substantially)**
- **Prediction:** Franco-Flemish pieces are longer
  - Franco-Flemish tends to be longer **(substantially)**

jMIR

Comparing theoretical predictions and empirical evidence

- So, how do the empirical measurements match up overall with the *a priori* predictions?
  - Masses and motets combined?
    - Mostly the same, but not entirely
  - Masses only?
    - Some the same, some not (40% correct)
  - Motets only?
    - Identical
- Interesting observation:
  - Expert predictions were more accurate for motets than masses, but motets were harder to classify with machine learning

Diving into the feature values

- The next step was to determine which features are most statistically predictive
  - And whether they match or differ from the expert predictions
- Done only for both masses and motets combined
  - Not done for each individually

Most statistically significant differences: Characteristics highlighted by experts

- Triple meter is much more common in Franco-Flemish music
- Iberian music tends to have more complete rests
  - But Franco-Flemish masses tend to have longer complete rests than Iberian masses (although they are still fewer in number than in Iberian masses)
- Franco-Flemish tends to have more long rhythmic values (especially masses)
- Franco-Flemish tends to have a wider range

Most statistically significant differences: Characteristics not highlighted by experts

- Iberian music tends to have a much higher note density (especially masses)
- Iberian music tends to have more (but still relatively rare) shorter notes
- Chords tend to last longer in Franco-Flemish masses
  - But the reverse is true for motets

Important caveat

- These results (and the conclusions drawn from themin the following slides) should all be seen as more of a rough initial exploration than a rigorous study
  - The dataset, kindly provided by the Anatomy team, is still in its preliminary stages, and is still not complete nor fully controlled for encoding
  - The short time between the submission of expert predictions and the start of the conference made it impossible to thoroughly vet the data (although efforts to do so were certainly made)
- It is therefore possible that there are certain biases in the data
  - e.g. the rhythmic duration used to indicate the beat may vary across pieces, due to different editorial decisions in annotation, which could have affected results

Experimental conclusions (1/4)

- Using content-based features and machine learning:
  - It is very easy (99.6%) to distinguish the Iberian masses from the Franco-Flemish masses
  - The Iberian motets are harder to distinguish from the Franco-Flemish motets, although this can still be done quite well (87.7%)
    - Perhaps this is due to the fact that the data set had fewer motets than masses?
    - Perhaps the motets are in fact more similar in style, regardless of region, than the masses?
    - Need more (and more varied) data to be sure

Experimental conclusions (2/4)

- So, the Iberian music is stylistically distinct from the Franco-Flemish music!
  - Especially masses, but motets as well
- Since the classifier could distinguish the Iberian music from the Franco-Flemish music 97.9% of the time based only on musical content, there must be significant differences in content

Experimental conclusions (3/4)

- Comparing *a priori* predictions with empirical data:
  - Expert predictions matched the data very well for motets, but less well for masses
- Analysis of statistically very predictive features:
  - Matched four of the features highlighted by experts
  - Revealed three features not highlighted by experts

Experimental conclusions (4/4)

- Experts already have a very good overall understanding of the differences between the regions
  - And now what were certain previously unproven suspicions are now empirically validated truths
    - Based on a systematic analysis a lot of music (467 files)
  - The fact that experts are able to express the details of the stylistic differences between the regions also provides further evidence of clear differences in style between the two regions
- Computational feature-based approaches can also reveal still more useful insights that experts are not aware of

Research potential

- The statistical results reported here are the product of an initial relatively simple analysis
  - And the dataset needs to be further vetted, as noted earlier
- There is substantial potential to expand this study
  - Implement new specialized features
  - Use more (and cleaner) data
    - More composers, more genres
  - Apply more sophisticated and detailed statistical analysis techniques
  - Perform a detailed manual exploration of the feature data
    - So that experts can apply their expertise to reveal insights hidden in the data

Research collaborations (1/2)

- I enthusiastically welcome research collaborations with musicologists and theorists
- I am always looking for new and interesting musicological problems to apply features to
- I am always looking for ideas for interesting new features to implement
  - jSymbolic makes it relatively easy to add bespoke features
  - Can iteratively build increasingly complex features based on existing features
- I am always looking for new datasets to study with jSymbolic

jMIR

Research collaborations (2/2)

- Please do not hesitate to speak to me if you would like more information on:
  - Using jSymbolic
  - How one can apply statistical analysis or machine learning to extracted features
  - How feature values can be visualized and explored manually
- I would be more than happy to help you with jSymbolic if you need help

Acknowledgements

- Thanks to my colleagues and the students in the SIMSSA and MIRAI research groups, especially:
  - Julie E. Cumming
  - Ichiro Fujinaga
  - Tristano Tenaglia
- Thanks to the Fonds de recherche du Québec - Société et culture (FRQSC) and the Social Sciences and Humanities Research Council of Canada (SSHRC) the for their generous funding
- Thanks to the organizers of this conference and everyone involved in the Anatomy project for giving me the opportunity to present this work
  - And for providing the data!

Thanks for your attention!

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