LinkedMusic, SIMSSA DB and Feature-Based Musicology

Cory McKay, Marianopolis College

(with some slides based on a deck by Ichiro Fujinaga)

Topics

- The LinkedMusic project
- SIMSSA DB
  - Extracting musical features
  - Musicological research with features

LinkedMusic: Scope

- Funded for 7 years (2022–2029): $3.2M
  - SSHRC Partnership Grant
  - FRQSC Research Team Support Grant
  - Based at McGill
- Broad international involvement
  - 7 co-investigators
  - 18 collaborators
  - 9 partners
  - 4 advisory board members

LinkedMusic: Co-Investigators

- *PI:* Ichiro Fujinaga (McGill University)

- Jennifer Bain (Dalhousie University)
- Houman Behzadi (McGill University)
- Julie Cumming (McGill University)
- Debra Lacoste (University of Waterloo)
- Audrey Laplante (Université de Montréal)
- Cory McKay (Marianopolis College)
- Laurent Pugin (RISM-Digital)

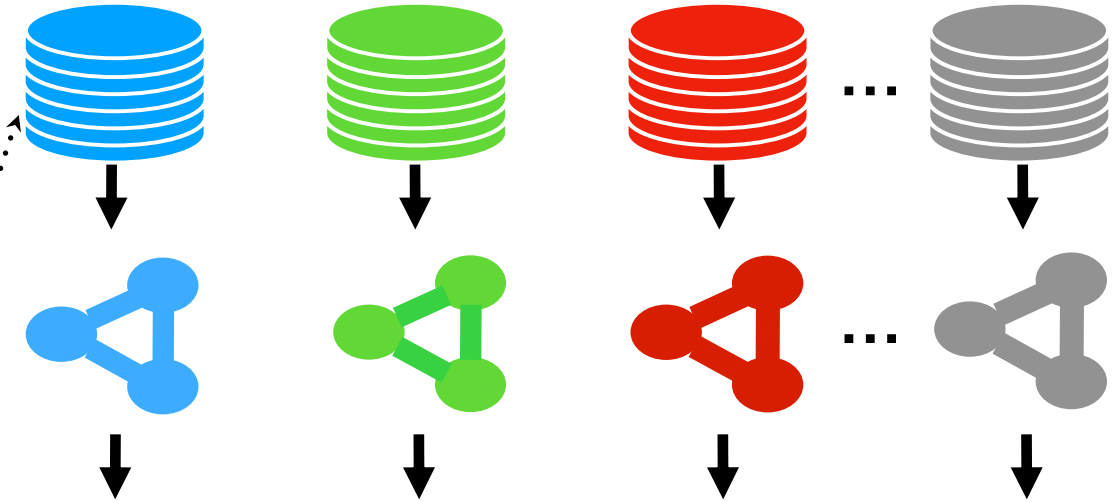
- Many CIRMMT students have also been involved

LinkedMusic: Goals

- Make more musical information accessible to more people in the world
  - With a particular focus on making queries available in languages other than English
- Use linked data and semantic web technologies to create a data lake infrastructure allowing one to search across multiple music databases from one website
  - Wikidata for authority control
  - OpenRefine to improve data hygiene
  - SPARQL and other search engines (e.g., Solr, ElasticSearch) for queries
- Create a Virtual Instrument Museum
  - A crowd-sourced website
  - Images and recordings of musical instruments
  - Name of each instrument in the local language, with translations

Initial 14 databases to import into data lake

1. SIMSSA DB
2. Cantus Ultimus
3. Cantus Database
4. DIAMM
5. RISM
6. Cantus Index
7. Canadian Chant Database
8. Global Jukebox
9. DTL1000 (Dig That Lick)
10. MusicBrainz
11. AcousticBrainz
12. CritiqueBrainz
13. ListenBrainz
14. MOTET Database (Jennifer Thomas)



Different Music Databases (14) with incompatible metadata schemas

Export metadata to simple CSV based on the original metadata schemas

Data Lake

OpenRefine

WIKIDATA

Authority control and data hygiene

Search for All Types of Music Information from SESEMMI (Search Engine System to Enhance Music Metadata Interoperability) website

Virtual Instrument Museum (VIM) (provides vocabulary to Wikidata)

Once an item is found, the user is guided to the original database for detailed viewing

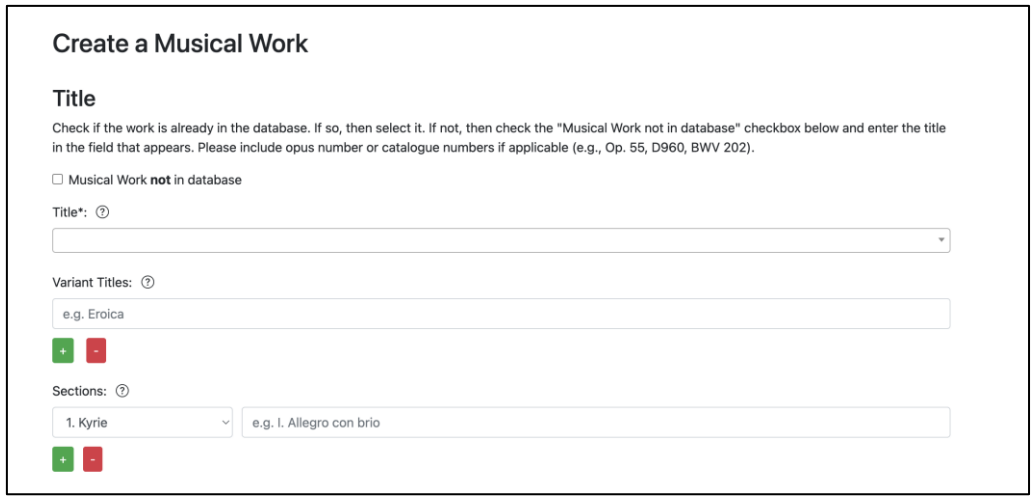
What is the SIMSSA DB?

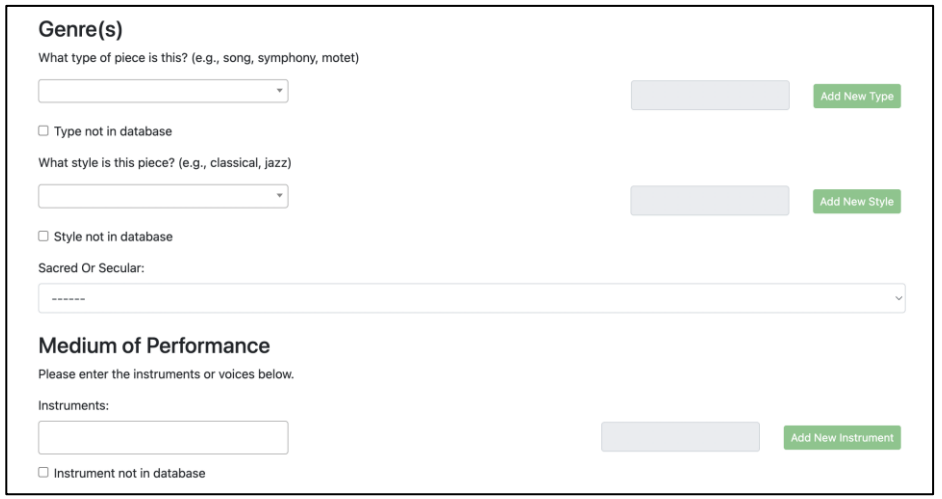
- Collaborative database prototype infrastructure for holding and accessing symbolic music files, associated auto-extracted content-based feature values, and musicologically-focused metadata
  - With a web Django-based browser interface
- Populated by:
  - **Now:** Samples from research datasets we have constructed
  - **Medium-term:** Import existing open symbolic datasets that musicologists, libraries and others have already constructed
    - We can import such datasets, or users can contribute them directly
  - **Long-term:** Auto-population via (verified) OMR
- Focused (for now) on early music

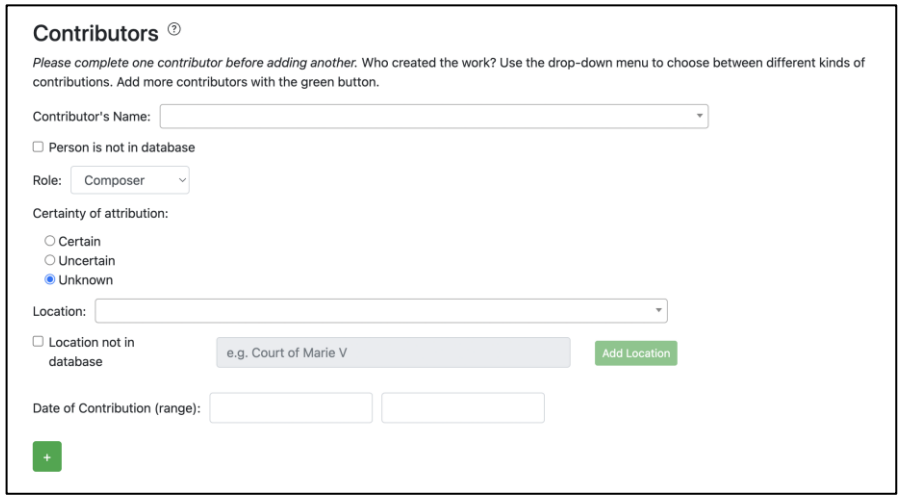
An infrastructure, not a corpus

- The SIMSSA DB is not intended just as a repository of music we have transcribed ourselves
  - Although are seeding it with datasets we have made, such as JLSDD (Cumming et al. 2018), Florence 164 (Cumming & McKay 2018), etc.
- Rather, it is a general unified infrastructure to which it is hoped other scholars can contribute and share symbolic music files (and more) that they have used in their own work

SIMSSA DB prototype contribution form







Core focus: Symbolic music files

- Research-grade symbolic music files are surprisingly difficult to access
  - e.g., MEI, MusicXML, MIDI, etc.

Metadata and feature searches

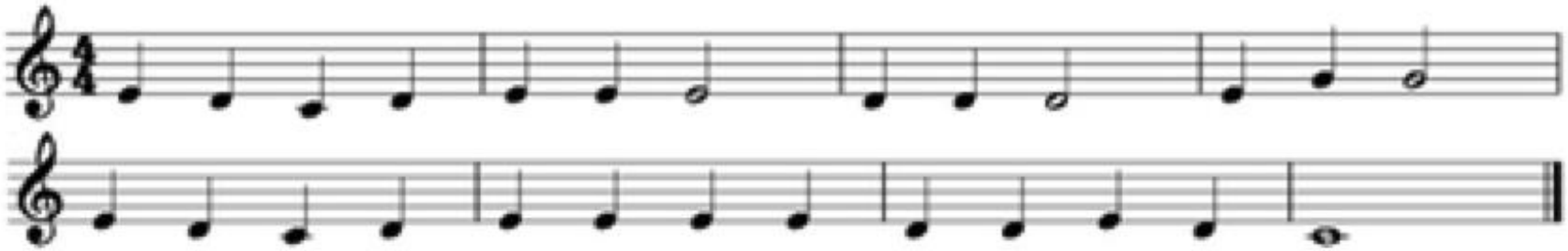
- SIMSSA DB may be searched using traditional metadata queries:
  - Free-text search
  - Faceted metadata filters, such as:
    - Contributor
      - Composer, arranger, author of text, transcriber, etc.
    - Instruments / voices
    - Sacred / secular
    - Genre / type of work
      - e.g. madrigal, motet, etc.
    - Etc.
- SIMSSA DB also permits content-based searches based on features

Wait, what is a "feature?"

- Information that measures a characteristic of a segment 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 vector of histogram bin values)
- Provides a summary description of the characteristic being measured
  - Usually provides a macro rather than local view
- Usually extracted from pieces or distinct sections (e.g., mass movements) in their entirety
  - But can also be extracted from smaller segments of music

Example: A simple feature

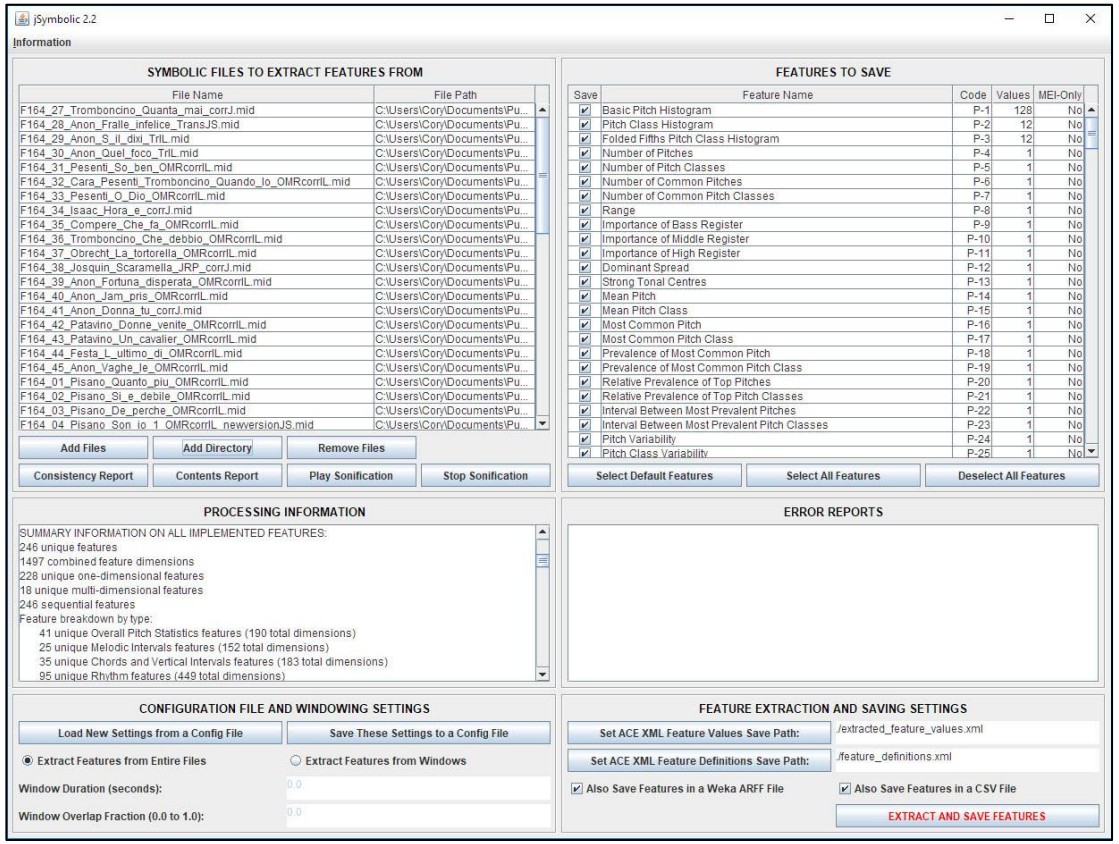
- Range: Difference in semitones between the lowest and highest pitches present
  - A 1-dimensional feature



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

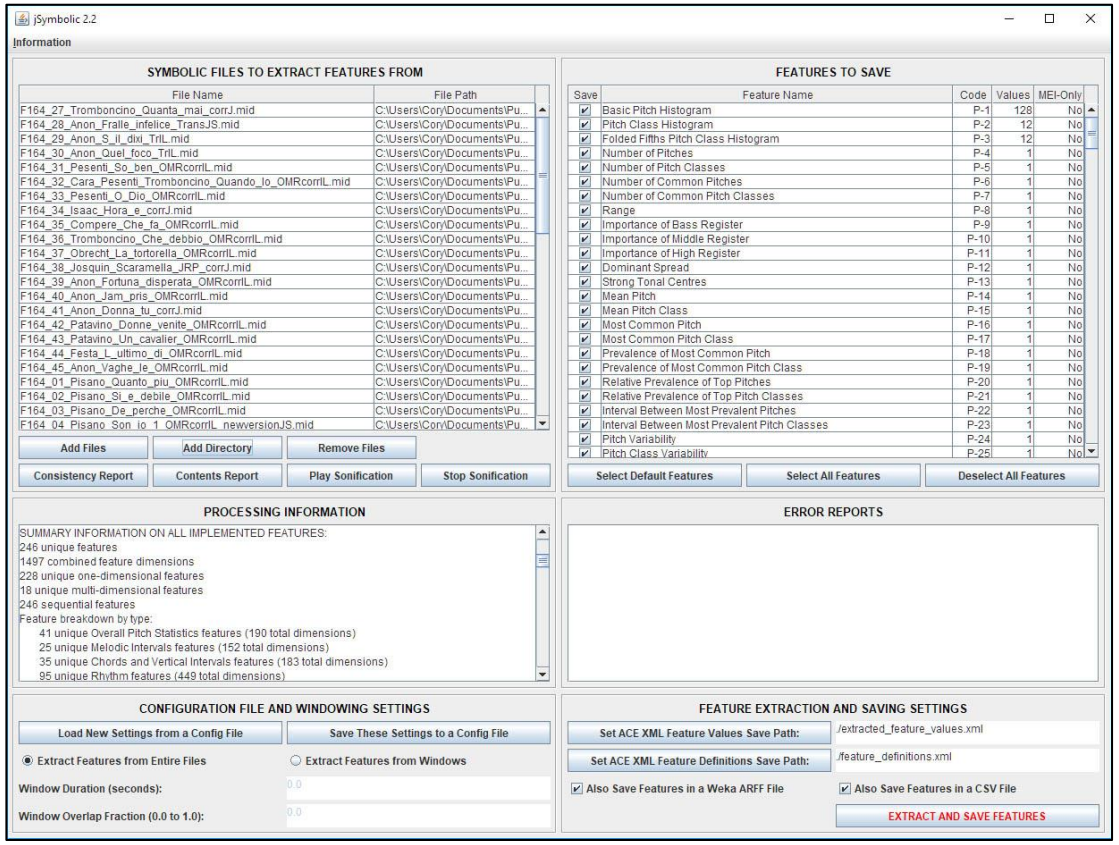
How might one calculate features?

- The jSymbolic research software (McKay et al. 2018) can be used to automatically extract features from symbolic digital scores
  - Open source
  - Applicable to diverse musics
- Version 2.2 extracts 246 unique features
  - 1497 separate feature values, since many features a multi-dimensional (e.g. histogram vectors)
- The upcoming Version 3 extracts 533 unique features
  - 2040 feature values, including n-gram features



jSymbolic 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

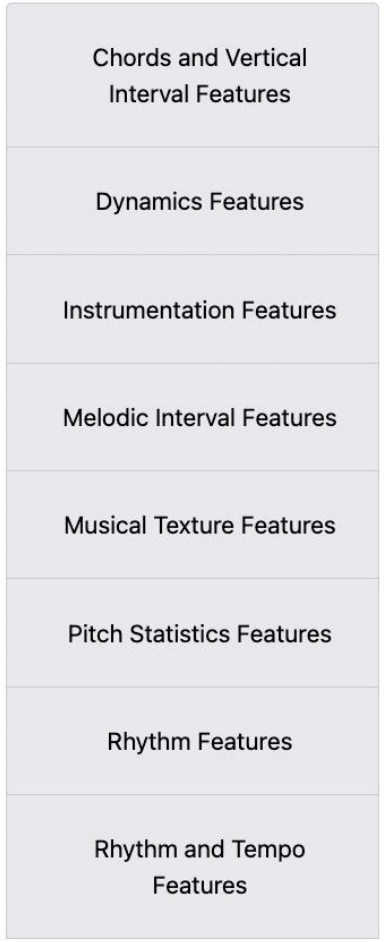


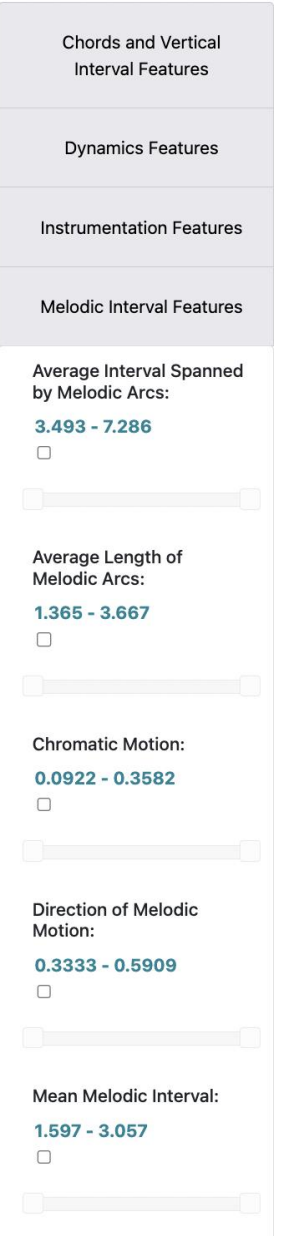
Sample musicological feature-based research

- Musical genre
  - Origins of the madrigal *(with Julie Cumming and others)*
  - Delineating popular music genres *(with Ichiro Fujinaga and others)*
- Compositional style *(with Julie Cumming and others)*
  - Empirically differentiating the styles of similar composers
  - Confirming historical evidence for Josquin attribution certainty
- Attribution of anonymous and doubtfully attributed works *(with Esperanza Rodríguez-García and Maria Elena Cuenca)*:
  - Masses transcribed by Siro Cisilino
  - Coimbra manuscripts
  - *Ave verum corpus* and *O decus virgineum*
  - *Ave festiva ferculis*
  - Gaffurius Codices
- Regional style in Iberian Renaissance music *(with Maria Elena Cuenca)*:
  - Musical influences of Pedro Fernández Buch
  - Musical Influences of Cristóbal de Morales and Francisco Guerrero

SIMSSA DB and features (1/2)

- jSymbolic 2.2 has been integrated into the SIMSSA DB
  - Whenever an extractable file is uploaded to the SIMSSA DB, features are automatically pre-extracted, stored and indexed
- Users can specify feature-range queries via a slider for each feature they are interested in



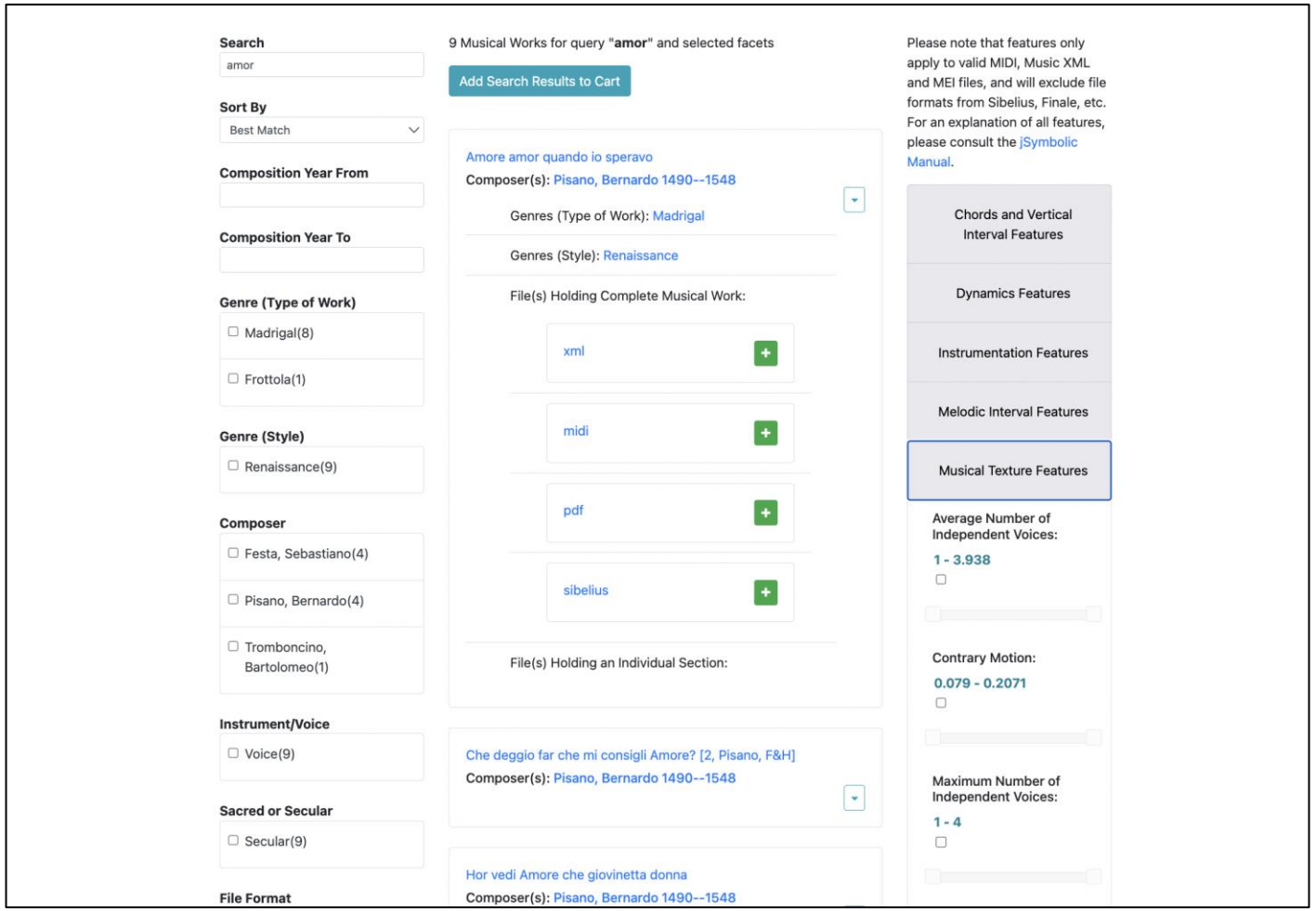


SIMSSA DB and features (2/2)

- Can download complete feature sets directly and use them as input to statistical analysis and machine learning tools (or analyze them manually)

- Feature searches can also be combined with metadata searches
  - e.g. retrieve all sacred pieces attributed to Josquin that contain parallel fifths

Sample query combining metadata and features



Other aspects of the SIMSSA DB

- Chains of provenance
- Conceptual separation between abstract musical works, sections and parts and particular instantiations of them
- Authority control
- Grouping into corpora
- Associations with specific experimental studies
- Links to other types of data (text, audio, images, etc.)

SIMSSA DB: Credit to the deserving

- I designed the original data model and provided high-level guidance to the project, along with Julie Cumming and Emily Hopkins
- Gustavo Polins Pedro and Yaolong Ju implemented the first version
- Rebecca Mizrahi recently resurrected the DB implemented substantial improvements
- Hong Van Pham has worked on deployment and towards LinkedMusic integration
- Ichiro Fujinaga generously hosted SIMSSA DB development in his lab

Please try the prototype yourself

- https://db.simssa.ca

LinkedMusic

SIMSSA | Single Interface for Music Score Searching and Analysis

Thanks for your attention!

cory.mckay@mail.mcgill.ca