jSymbolic in 2019: Updates and Improvements

Cory McKay (Marianopolis College, Canada)
Rían Adamian (McGill University, Canada)
Introduction to jSymbolic

• jSymbolic is software that extracts features from symbolic music files (MIDI or MEI)

• A feature is a piece of statistical information that characterizes some aspect of a piece of music using a simple, consistent measurement
  • Each feature is expressed as one or more simple numerical values
  • Features can reveal meaningful patterns in music at a macro scale
Uses of features

• Training classification models with machine learning
• Statistical feature analysis
• Content-based searches
  • e.g. SIMSSA DB
jSymbolic’s features (1/2)

• The current 2018 release version (2.2) extracts 246 unique features
  • 1497 distinct values when multi-dimensional features (e.g. histograms) are expanded
jSymbolic’s features (2/2)

- Feature types include:
  - 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
User interfaces

- Graphical user interface
- Command-line interface
- Java API
- Rodan workflow for distributed processing
New interface developments in 2019

• Expanded the already extensive tutorial and manual
• Expanded multilingual support
• Feature summary stat reports
• Many miscellaneous interface improvements
Extensibility

• jSymbolic is designed to encourage researchers to add their own bespoke features
  • Modular plug-in feature design
  • Easy to iteratively build new features of increasing sophistication by incorporating values of already-implemented features in new features

• jSymbolic’s feature catalogue has already expanded greatly
  • The original 2006 jSymbolic 1.0 had 160 features, compared to the 2018 jSymbolic 2.2’s 246 features
  • Tristano Tenaglia implemented a good share of these new features from 2015 to 2016
New features in 2019

• Rían Adamian has already implemented 190 additional new unique features this summer (comprising 422 new feature values):
  • 8 new pitch statistics features
  • 19 new rhythmic features
  • 112 new melody / horizontal interval features
  • 43 new chords / vertical interval features
  • 10 new instrumentation features

• There are now 436 unique features in total
Cory Monster want MORE FEATURES!

https://register.myrunti.me/sesamestreetrun/
Features areas remaining to be more fully explored by jSymbolic

• **Local melodic transitions** longer than one horizontal interval and strings of horizontal patterns

• **Local chord transitions** and strings of vertical patterns
  • Current vertical features aggregate vertical intervals independently of what directly precedes and follows them

• **Local rhythmic transitions** and strings of rhythmic patterns
  • Current rhythmic features aggregate attacks, rhythmic values and rests independently of what directly precedes and follows them

• **Complex textural behaviour**
  • e.g. measures of imitation
Infrastructure needed to do this

- Note onset slices

- N-grams

https://en.wikipedia.org/wiki/Salami
Note onset slices (1/2)

- A slice consists of vertical groups of notes sounding simultaneously.
- A new slice is started every time a new (pitched) note attack occurs.
- There are various (non-deli) flavours:
  - e.g. a slice may only contain notes starting at the beginning of the slice.
  - e.g. a slice may 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 grouped units of notes that permit the calculation of new features associated with:
  • Local harmonic transitions
  • Local melodic transitions
  • Local rhythmic transitions
  • Sophisticated textural behaviour

• Sets of such transitions can also be used to construct . . .
N-grams

- N-grams encode sequences of $n$ note onset slices
- Can be related to:
  - Harmonic sequences
  - Melodic sequences
  - Rhythmic sequences
- Examples:
  - 7-6-8 is a 3-gram showing the vertical intervals between outer voices
  - [7] (1 -2) [6] (-2 2) [8] is a 3-gram that also encodes melodic transitions in the outer voices
- There can be many varieties of n-grams
Current jSymbolic development status

• A variety of note onset slice and n-gram implementation are already implemented and undergoing code review and testing

• We are designing features we can extract from them
  • e.g. textural features
    • Such as density of imitation
  • e.g. features looking at general n-gram distributions
    • Such as histogram statistics
  • e.g. features looking at selected n-grams expected to be meaningful
    • Such as cadential patterns
to tell us about any features you think could be usefully added to jSymbolic!
Thanks for your attention!

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