jSymbolic in 2019: Updates and Improvements

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

> SIMSSA Workshop XIX September 21, 2019 Montreal, 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

SYMBOLIC FILES TO EXTRACT FEATURES FROM					FEATURES TO SAVE					
	File Name		File Path	Sav	9	Feature Name		Code 1	Values	MEI-Only
164_27_Tromboncino_Q	uanta mai corrJ.mid	C:\User	s\CorvDocuments\Pu		Basic Pitch Histogram			P-1	128	No
164_28_Anon_Fralle_infe		C:\User	s\CorvDocuments\Pu		Pitch Class Histogram			P-2	12	No
164_29_Anon_S_il_dixi_		C:\User	s\CorvDocuments\Pu		Folded Fifths Pitch Class H	listogram		P-3	12	No
164 30 Anon Quel foco		C:\User	s\CorvDocuments\Pu	V	Number of Pitches			P-4	1	No
164 31 Pesenti So ber	n OMRcorrIL.mid	C:\User	s\Cory\Documents\Pu	V	Number of Pitch Classes			P-5	1	No
164 32 Cara Pesenti T	romboncino_Quando_lo_O	MRcorrlL.mid C:\User	s\Cory\Documents\Pu		Number of Common Pitche	s		P-6	1	No
164 33 Pesenti O Dio			s\Cory\Documents\Pu	~	Number of Common Pitch	Classes		P-7	1	No
164 34 Isaac Hora e d		C:\User	s\CorvDocuments\Pu	~	Range			P-8	1	No
164 35 Compere Che		C'\User	s\CorvDocuments\Pu	V	Importance of Bass Regist	er		P-9	1	No
164 36 Tromboncino C	he_debbio_OMRcorrlL.mid	C:\User	s\CorvDocuments\Pu		Importance of Middle Regis	ster		P-10	1	No
164 37 Obrecht La tort			s\CorvDocuments\Pu	~				P-11	1	No
164 38 Josquin Scaran	nella JRP corrJ.mid	C:\User	s\Cory\Documents\Pu	~	Dominant Spread			P-12	1	No
F164_39_Anon_Fortuna_disperata_OMRcorrlL.mid			s\Cory\Documents\Pu	V	Strong Tonal Centres			P-13	1	No
F164 40 Anon Jam pris OMRcorrlL.mid			s\CorvDocuments\Pu	1				P-14	1	No
F164 41 Anon Donna tu corrJ.mid			Users\Cor\Documents\Pu					P-15	1	No
F164_42_Patavino_Donne_venite_OMRcorrlL.mid		C'User	C:\Users\Cory\Documents\Pu		Most Common Pitch		P-16	1	No	
F164_43_Patavino_Un_cavalier_OMRcorrIL.mid			C:\Users\Cory\Documents\Pu		Most Common Pitch Class		P-17	1	No	
F164_44_Festa_L_ultimo_di_OMRcorrlL.mid			C:\Users\Cory\Documents\Pu		Prevalence of Most Common Pitch		P-18	1	No	
F164_45_Anon_Vaghe_le_OMRcorrIL.mid			C:\Users\Cory\Documents\Pu		Prevalence of Most Common Pitch Class		P-19	1	No	
F164_01_Pisano_Quanto_piu_OMRcorrIL.mid			Con/Documents\Pu Relative Prevalence of Top Pitches			P-20	1	No		
F164 02 Pisano Si e debile OMRcorrIL.mid			:\Users\Cory\Documents\Pu		Relative Prevalence of Top Pitch Classes		P-21	1	No	
164 03 Pisano De per			s\Cory\Documents\Pu	~	Interval Between Most Prev			P-22	1	No
	1 OMRcorrlL newversionJ		s\Cor/\Documents\Pu		Interval Between Most Prev	alent Pitch Classes		P-23	1	No
Add Files			1	~	Pitch Variability			P-24	1	No
	Add Directory	Remove Files	]		Pitch Variability Pitch Class Variability			P-24 P-25	1	
Consistency Report	Add Directory Contents Report	Remove Files Play Sonification	Stop Sonification			1	ll Features		1	No
	Contents Report PROCESSING	Play Sonification			Pitch Class Variability	Select A		P-25	1	No
SUMMARY INFORMATION 246 unique features 1497 combined feature di 228 unique one-dimension 246 sequential features 41 unique unti-dimension 41 unique Overall Pitch 25 unique Melodic Inte 35 unique Chords and	Contents Report PROCESSING ON ALL IMPLEMENTED FEA mensions nal features nal features	Play Sonification INFORMATION ATURES: al dimensions) tensions) 133 total dimensions)	Stop Sonification		Pitch Class Variability	Select A	II Features	P-25	1	No
SUMMARY INFORMATION 246 unique features 1497 combined feature di 228 unique ont-dimension 246 sequential features 246 sequential features 246 unique Meiodic Inte 25 unique Meiodic Inte 35 unique Meiodic Inte 35 unique Meiodic Inte 35 unique Rivolta features 35 uni	Contents Report PROCESSING PROCESSING Reasions ana factures and factures and factures c Statistics features (190 tot vals features (190 tot vals features (192 total dim vals features (192 total dim vals features (192 total dim vals features (190 tot)	Play Sonification INFORMATION ATURES: al dimensions) ensions) 38 total dimensions)			IPitch Class Vanability	Select A	II Features	P-25 Deselec	1	No
SUMMARY INFORMATION. 246 unique features 149 combined feature din 280 unique one-dimension 280 unique muiti-dimension 240 sequential features 240 tra toreat/040 verall Pikch 250 unique Melodic Inter 350 unique Rhoths feat 950 unique Rhoths feat	Contents Report PROCESSING ON ALL IMPLEMENTED FE/ mensions tal features tal features Stalistics features (190 lot valis features (192 lotal dim valis features (192 lotal dim valis features (192 lotal dimensions)	Play Sonification INFORMATION ATURES: al dimensions) ensions) IS3 total dimensions)			IPitch Class Vanability	Select A ERROR	II Features	Deselec	1	No
SUMMARY INFORMATION. 449 unique features 449 combined feature din 230 unique onu-dimension 230 unique onu-dimension 240 sequenhal features 441 unique Overall Pitch 250 unique Meiodic Inte 350 unique Meiodic Inte 50 unique Rhythm feat CC Load New Settings	Contents Report PROCE SSING ON ALL IMPLEMENTED FE/ mensions hal features hal features Statistics features (190 tot vals features (190 tot vals features (190 tot vals features (190 tot vals (190 tot) features (190 tot) outers (149 total dimensions) ONFIGURATION FILE AN s from a Config File	Play Sonification INFORMATION ATURES: al dimensions) ensions) IS3 total dimensions)	GS ngs to a Config File		IPitch Class Variability Select Default Features FEATU	Select A ERROR JRE EXTRACTION os Save Path:	II Features REPORTS AND SAVING SE	P-25 Deselec TTINGS values.xml	1	No
SUMMARY INFORMATION. 246 unique features 1497 combined feature din 280 unique en-dimension 180 unique multi-dimension 180 unique multi-dimension 240 sequential features 240 sequential 250 unique Rhythm features 240 sequential 250 unique Rhythm features 250 sequential 250 sequential 25	Contents Report PROCE SSING PROCE SSING ON ALL IMPLEMENTED FE/ mensions hal features ial features is features (149 total dimensions) ONFIGURATION FILE AN from a Config File Entire Files	Play Sonification INFORMATION ATURES: al dimensions) ensions) 132 total dimensions) D WINDOWING SETTIN Save These Setti	GS ngs to a Config File		Fitch Class Variability Select Default Features FEATU Set ACE XML Feature Value	Select A ERROR JRE EXTRACTION is Save Path: ons Save Path:	Il Features REPORTS AND SAVING SE	P-25 Deselec TTINGS _values.xml is.xml	1 t All Fea	No

### 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 alreadyimplemented 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



https://en.wikipedia.org/wiki/Salami

• N-grams

## 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:



(1) (-2)

- 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

https://www.pinterest.ca/pin/584693964102162828/



# 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







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









DISTRIBUTED DIGITAL MUSIC ARCHIVES مج LIBRARIES LAB



