CRIM, Machine Learning and Big Data: A Case Study on the Coimbra Manuscripts

Cory McKay
Marianopolis College, Canada

María Elena Cuenca
Universidad de Salamanca, Spain

Counterpoints: Renaissance Music and Scholarly Debate in the Digital Domain
Tours, France
November 15, 2019.

Topics

- Introduction to “features”
  - jSymbolic
- Musical style in the anonymous and doubtfully attributed mass movements of the Coimbra manuscripts
  - Qualitative analysis
  - Quantitative experiments
- Workshop
  - General discussion questions
  - CRIM data and the jSymbolic features

Big questions to think about

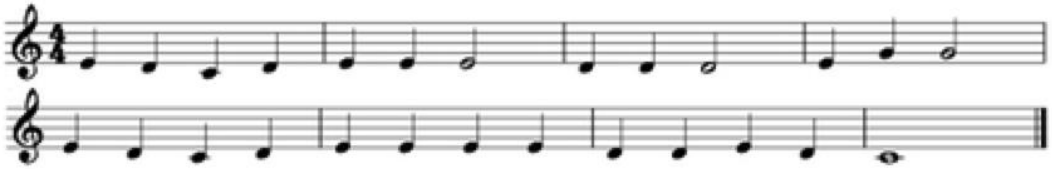
- What existing needs of music scholars can be addressed by computational approaches?
- What new, different opportunities for scholarship do computational approaches present?
- What challenges and pitfalls do computational approaches pose?
- How can we stimulate collaboration and discussion between domain experts (e.g. musicologists and data scientists)?

What is a “feature”?

- A piece of information that measures a characteristic of something (e.g. a piece of music) in a simple and consistent way
- Represented as a simple number
  - 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
- Can be extracted from pieces in their entirety, or from segments of pieces

Example: A basic feature

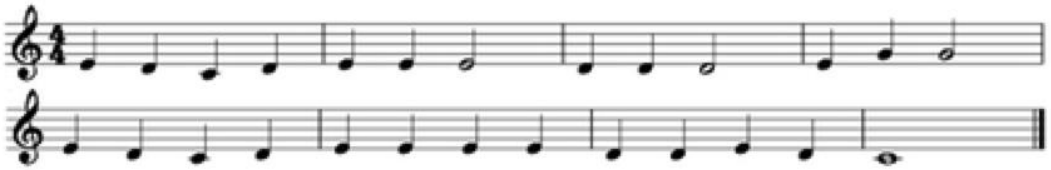
- Range (1-D): Difference in semitones between the highest and lowest pitches



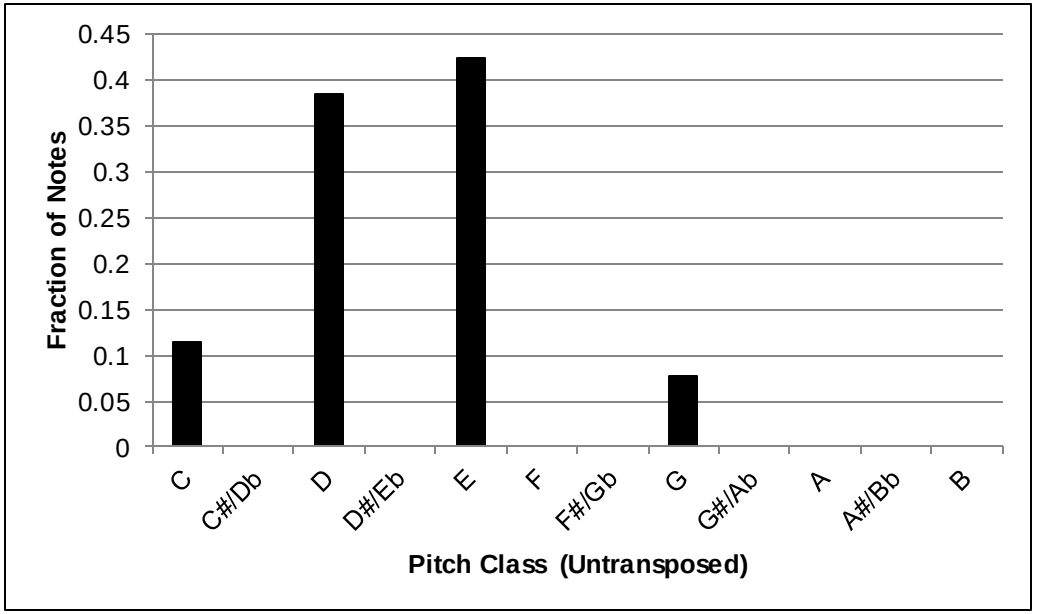
- Value of this feature: 7
  - G - C = 7 semitones

Example: A histogram feature

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

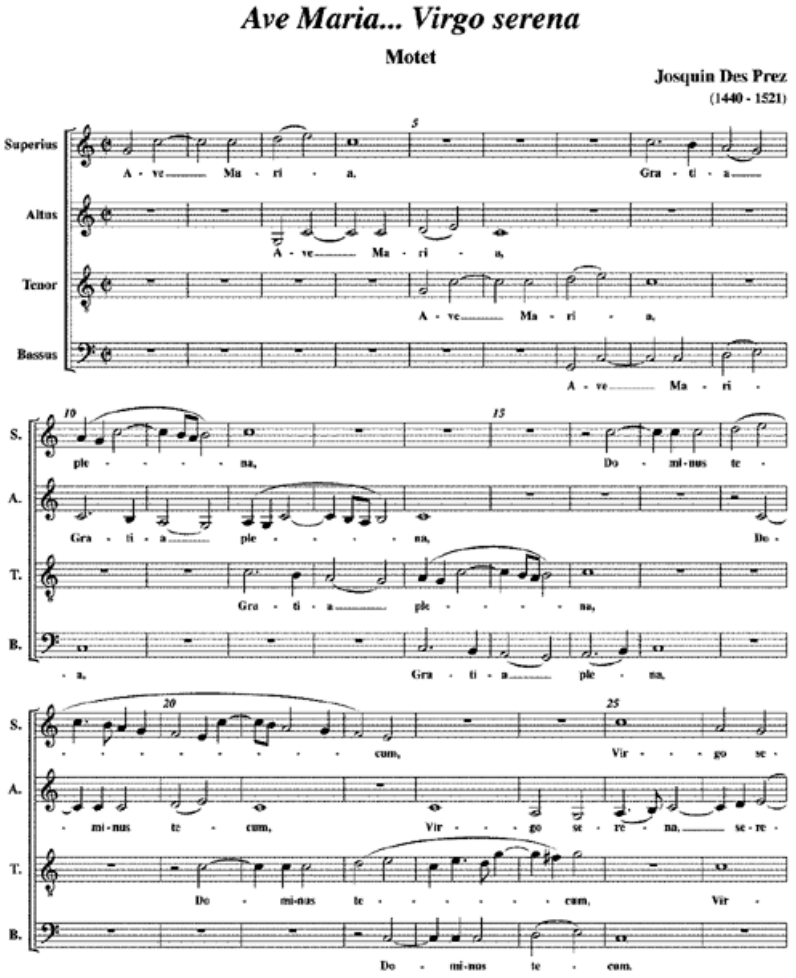


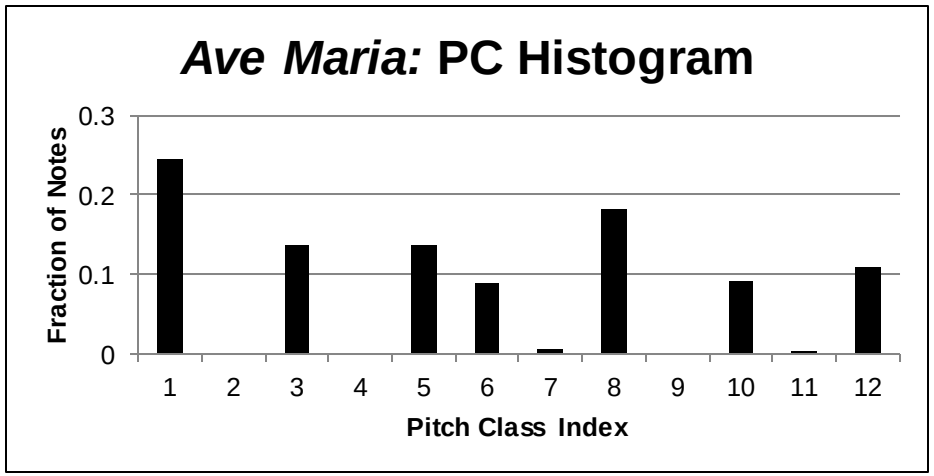
- Graph on right shows feature values
- Pitch class counts:
  - C: 3, D: 10, E: 11, G: 2
- Most common note is E:
  - 11/26 notes
  - Corresponds to a feature value of 0.423 for E



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

- Range: 34 (semitones)
- Repeated notes: 0.181 (18.1%)
- Vertical perfect 4^{ths}: 0.070 (7.0%)
- Rhythmic variability: 0.032
- Parallel motion: 0.039 (3.9%)



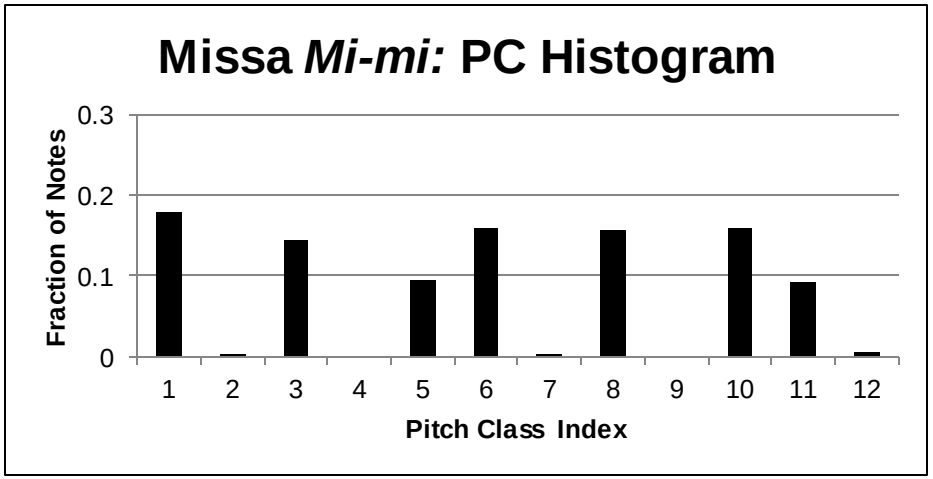


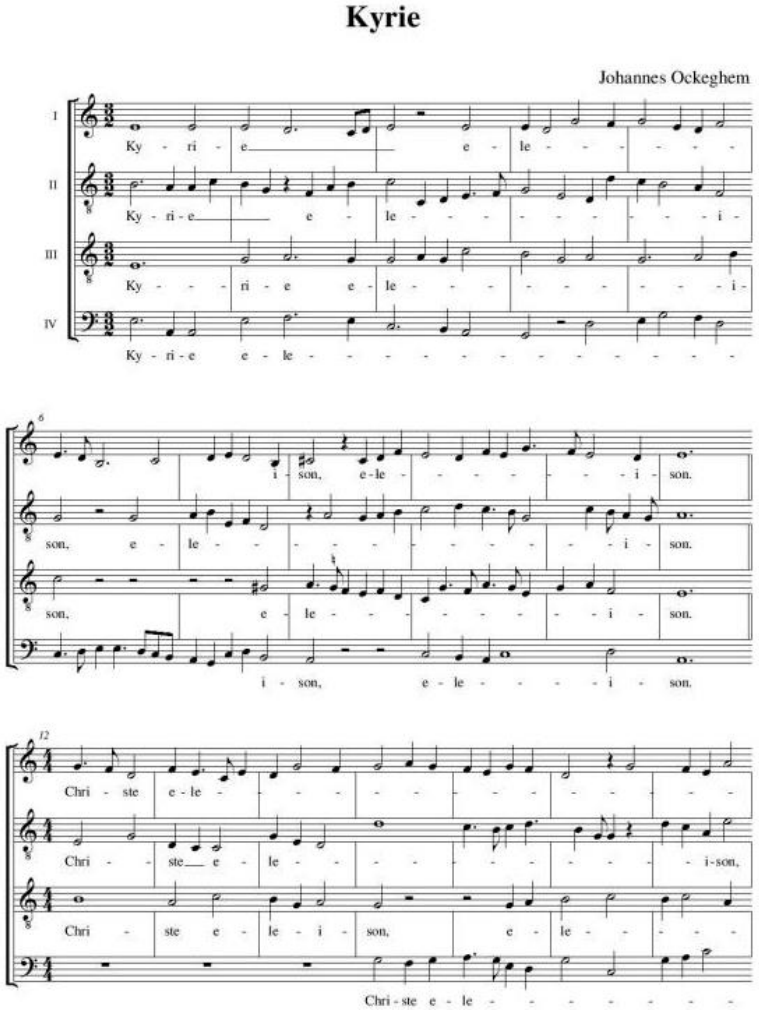
VNiVERSiDAD D SALAMANCA
CAMPUS DE EXCELENCIA INTERNACIONAL

MARIANOPOLIS COLLEGE

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

- Range: 26 (semitones)
- Repeated notes: 0.084 (8.4%)
- Vertical perfect 4^{ths}: 0.109 (10.9%)
- Rhythmic variability: 0.042
- Parallel motion: 0.076 (7.6%)



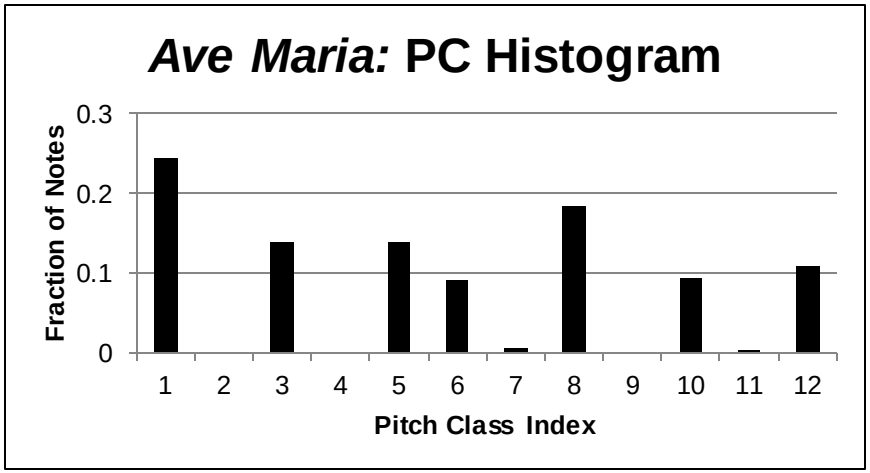


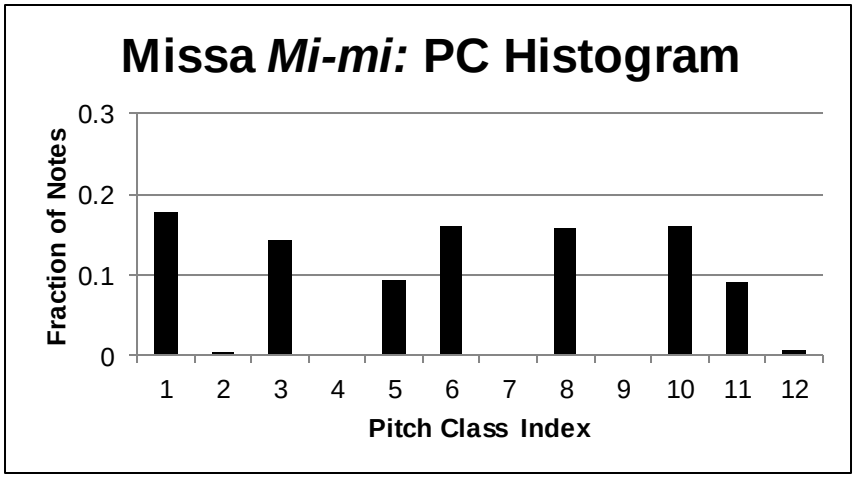
VNiVERSiDAD D SALAMANCA
CAMPUS DE EXCELENCIA INTERNACIONAL

MARIANOPOLIS COLLEGE

Feature value comparison

Feature	Ave Maria	Missa *Mi-mi*
Range	34	26
Repeated notes	0.181	0.084
Vertical perfect 4^{ths}	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 even more 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 composers, genres, regions, time periods, etc.

How can we use features? (1/3)

- Manual analysis to look for patterns
- Applying 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
  - e.g. the SIMSSA DB

How can we use features? (2/3)

- Use supervised machine learning to classify music
  - Done by training models on pre-labelled data
  - Can study music using whatever categories ("classes") one is interested in
    - e.g. composer, genre, style, time period, culture, region, etc.
  - Sample applications we have already explored:
    - Identify the composers of unattributed musical pieces
    - Explore the stylistic origins of genres (e.g. madrigals)
    - Delineate regional styles (e.g. Iberian vs. Franco-Flemish)

How can we use features? (3/3)

- Use unsupervised machine learning to cluster music
  - Done by training models on unlabelled data
  - Can study how the model groups pieces based on statistical similarity
    - And then see if we can find meaning in these groups

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
  - One can statistically condense many features into more interpretable 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

jSymbolic: Introduction

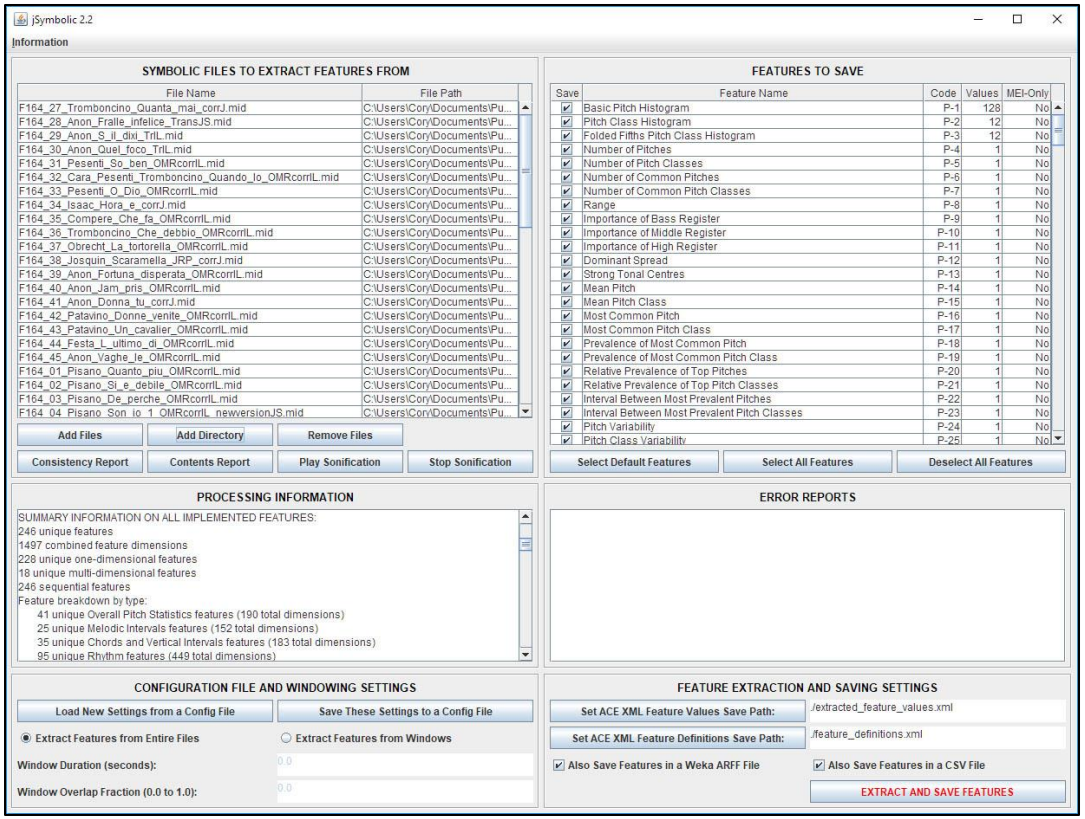
- jSymbolic is a software platform for extracting features from symbolic music
  - Part of the much larger (multimodal) jMIR package
- Compatible with Macs, PCs and Linux computers
- Free and open-source

jSymbolic: Features extracted

- The current release version (2.2) extracts 246 unique features
  - 1497 distinct values when multi-dimensional features (e.g. histograms) are expanded
- Characteristics examined include:
  - Pitch statistics
  - Melody / horizontal intervals
  - Chords / vertical intervals
  - Texture
  - Rhythm
  - Instrumentation
  - Dynamics

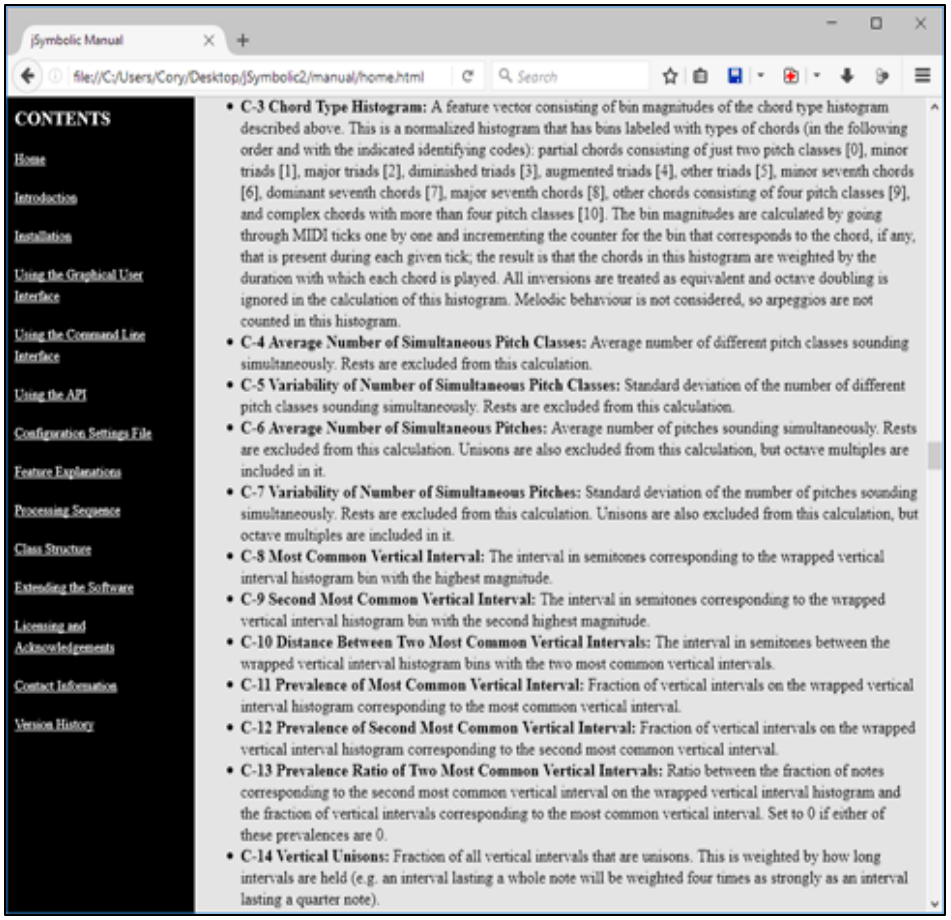
jSymbolic: User interfaces

- Graphical user interface
- Command line interface
- Java API



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



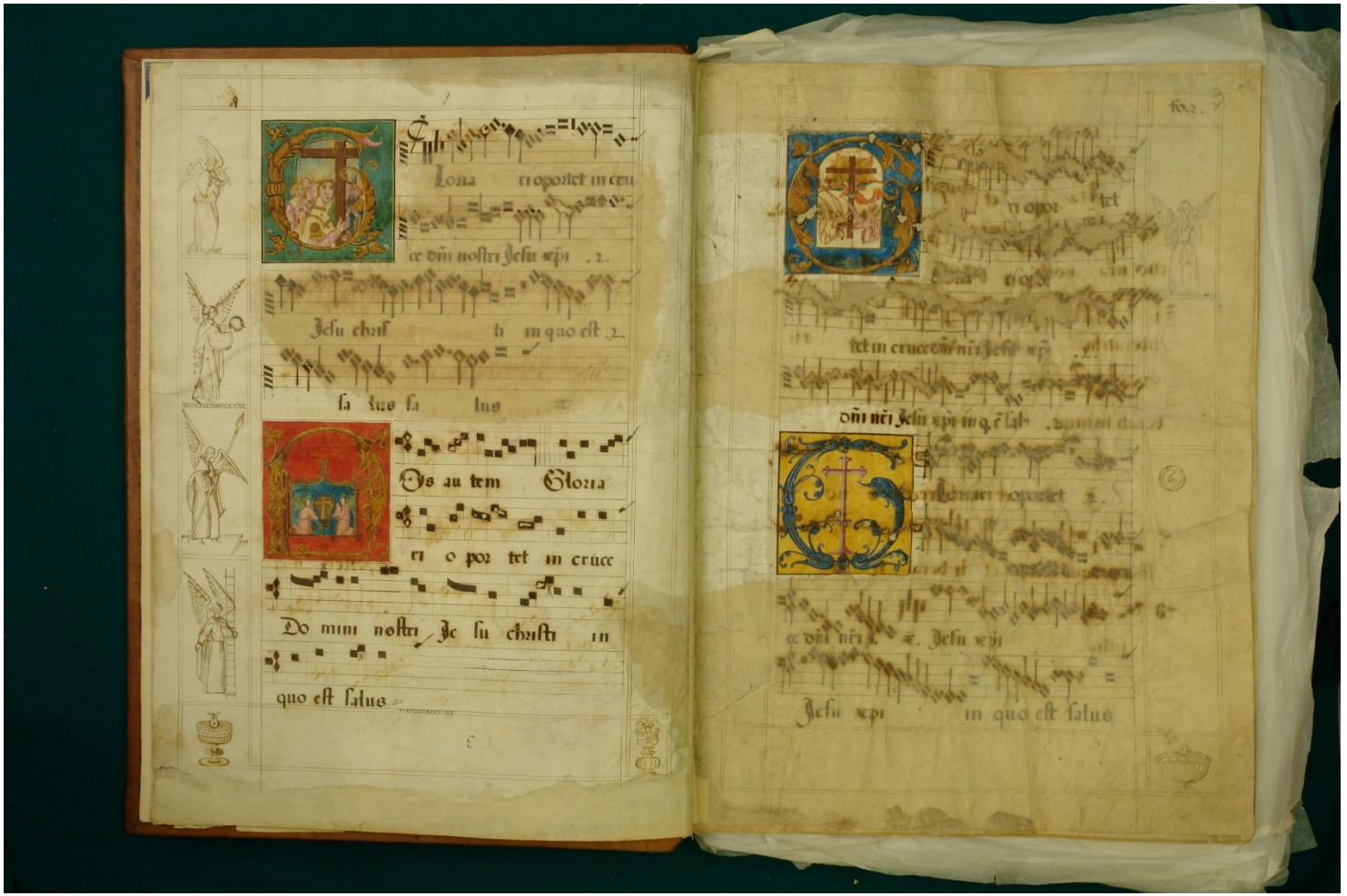
jSymbolic: Extensibility

- 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

The Coimbra research project

- Computational approaches, expert theoretical analyses and historical studies can complement one another extremely well
- There are many additional opportunities for joint future research of this kind in a wide range of musical domains

The Coimbra research project



P-Cug MM 9, ff. 1v-2r

Polyphonic repertoires in Portugal

- Rees, Owen (1994-5), Lisbon, Biblioteca Nacional, CIC Ms 60: the Repertories and their Context, *Revista Portuguesa de Musicologia* 4-5, 53-93.
- _____ (1995), *Polyphony in Portugal* c. 1530-c. 1620: *Sources from the Monastery of Santa Cruz, Coimbra.* New York & London: Garland.
- _____ (2004), Relaciones musicales entre España y Portugal, in John Griffiths & Javier Suárez-Pajares (eds), *Políticas y prácticas musicales en el mundo de Felipe II, Música Hispana, Textos, Estudios* 8. Madrid: Instituto Complutense de Ciencias Musicales, 455-487.
- d'Alvarenga, João Pedro (2010), 'Some Notes on the Reception of Josquin and of Northern Idioms in Portuguese Music and Culture', *Journal of the Alamire Foundation* 2 (1), 69-89.

Polyphonic repertoires in Portugal

- d'Alvarenga, João Pedro (2012), 'A Neglected Anonymous Requiem Mass of the Early (Sixteenth Century and its Possible Context', *Musica Disciplina* 57, 155-189.
- _____ (forthcoming), 'On the Transmission of Iberian Polyphonic Music in the Early Decades of the 16th Century: Some Philological Issues Revisited'.
- Nelson, Bernadette (2004-5), 'The Leiria Fragments: Vestiges of Fifteenth-Century Northern Polyphony in Portugal', *Revista Portuguesa de Musicologia* 14-15, 79-100.
- _____ (2015), 'Morales's Magnificats and Some Anonymous Settings in Portuguese Sources: Questions of Style and Authorship', *Revista Portuguesa de Musicologia*, NS 2/2, 193-214.

Polyphonic repertoires in Portugal

- Ferreira, Manuel Pedro (2005), '*L'homme armé* no Cancioneiro de Resende', *Revista da Facultade de Ciências Sociais e Humanas* 16, 259-268.
- Knighton, Tess (ed.) (2012), *Gonçalo de Baena, Arte para tanger*. Lisbon: Ediçoes Colibri, CESEM.
- ____ (ed.) (2017), *Companion to Music in the Age of the Catholic Monarchs*. Leiden / Boston: Brill, 205-241.
- *The Anatomy of Late 15th- and Early 16th-Century Iberian Polyphonic Music* project at the Lisbon Nova University and CESEM (FCT-funded project, PTDC/CPC-MMU/0314/2014, led by João Pedro d'Alvarenga).

VNiVERSiDAD D SALAMANCA
CAMPUS DE EXCELENCIA INTERNACIONAL

MARIANOPOLIS COLLEGE

Objective

- To provide insights on whether there was circulation of foreign repertoire or not, and on the possible prevalence of Franco-Dutch repertoires in the manuscripts copied in Coimbra.
  - To present an initial analysis of the anonymous and doubtfully attributed masses and loose movements.
  - To discuss a statistical analysis of these works using the jSymbolic software.

Number and percentages of masses and works of Franco-Flemish, Iberian or unknown origin

Sources	No. of masses	No. of Franco-Flemish works	No. of Iberian works	No. of anonymous works
P-Cug MM 2 [c.1530-1535]?	12/12 = 100%	11/12 = 91%	0/12 = 0%	1/12 = 1,5% (only one mass movement)
P-Cug MM 6 [c.1540-c.1555]	1/22 = 4,5%	0/22 = 0%	4/22 = 18,1%	18/22 = 81,8%
P-Cug MM 7 Mid-16th century	1/24 = 4,1%	0/24 = 0%	0/24 = 0%	24/24 = 100%
P-Cug MM 9 [c.1545-c.1550]	6/40 = 15%	5/40 = 12,5% ?	7/40 = 17,5%	28/40 = 70%
P-Cug MM 12 [c.1540-c.1550]	8/64 = 12,5%	3/64 = 4,6%	30/64 = 46,8%	31/64 = 48,4%
P-Cug MM 32 Mid- 16th century (c.1540-c.1555) and late 16th century	0/74 = 0%	7/74 = 9,4%	21/74 = 28,3%	46/74 = 62,1%

Anonymous and doubtfully attributed mass movements of the Coimbra Manuscripts selected as a case study

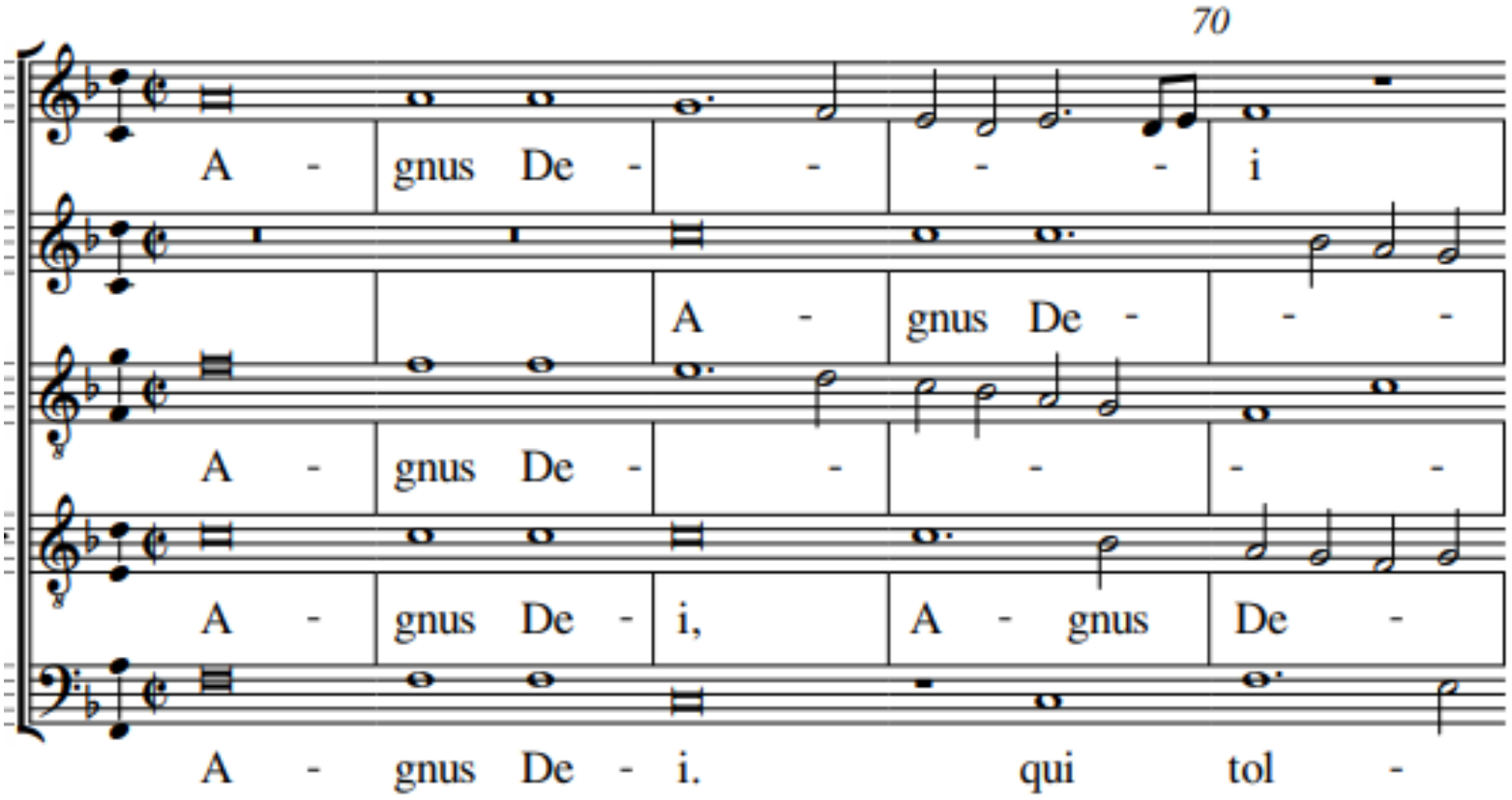
No.	ff.	Work	Vv.	Autorship /attributions	Concordant sources	Edited and observations
1	1v-7r	**[Missa *Salve regina*]** Kyrie, Gloria	4/5	(Do Pregador)	*P-Cug* MM 9, ff. 104v-105r	Cuenca's edition
2	8v-18r	Credo, Sanctus, Benedictus, Agnus Dei				Cuenca's edition
3	19v-22r	**[Missa]** Kyrie, Gloria, Sanctus, Agnus Dei	4			Cuenca's edition
4	73v-80r	**[Missa]** Kyrie, Gloria	4			Cuenca's edition

5	**80v**	**(1) Et incarnatus**	**4**			**Cuenca's edition**
6	81r	(2) Et incarnatus	4			Cuenca's edition
7	81v-88r	Credo, Sanctus	4/5	[Tordesillas]		Cuenca's edition
8	88v-89r	Agnus	4	Tordesillas?	Doubtful Agnus Dei. Identic beginning, but from bar 7 onwards it's different.	Cuenca's edition
9	90v-91r	**[Missa]** Kyrie	4	DE:RIBEIRA	*E-TZ* Ms. 2-3, ff. cxci^{v}-clxxi^{r}, 'Tordesillas'. This Kyrie is part of the mass attributed to Tordesillas	Cuenca's edition
10	91v-93r	Gloria	4	[Antonio de Ribera]	*E-TZ* Ms. 2-3, ff. cliii^{v}-clv^{r}, 'An. de ribera'	Cuenca's edition
11	93v-94r	Credo (opening of S and T parts), 94r blank		[Tordesillas]	= no. 7 f. 94r, different hand: '*Este Credo e os Sanctus ficam a tras a folhas 8i não tem Agnus*'. (= no. 8)	Kreitner's Tordesillas

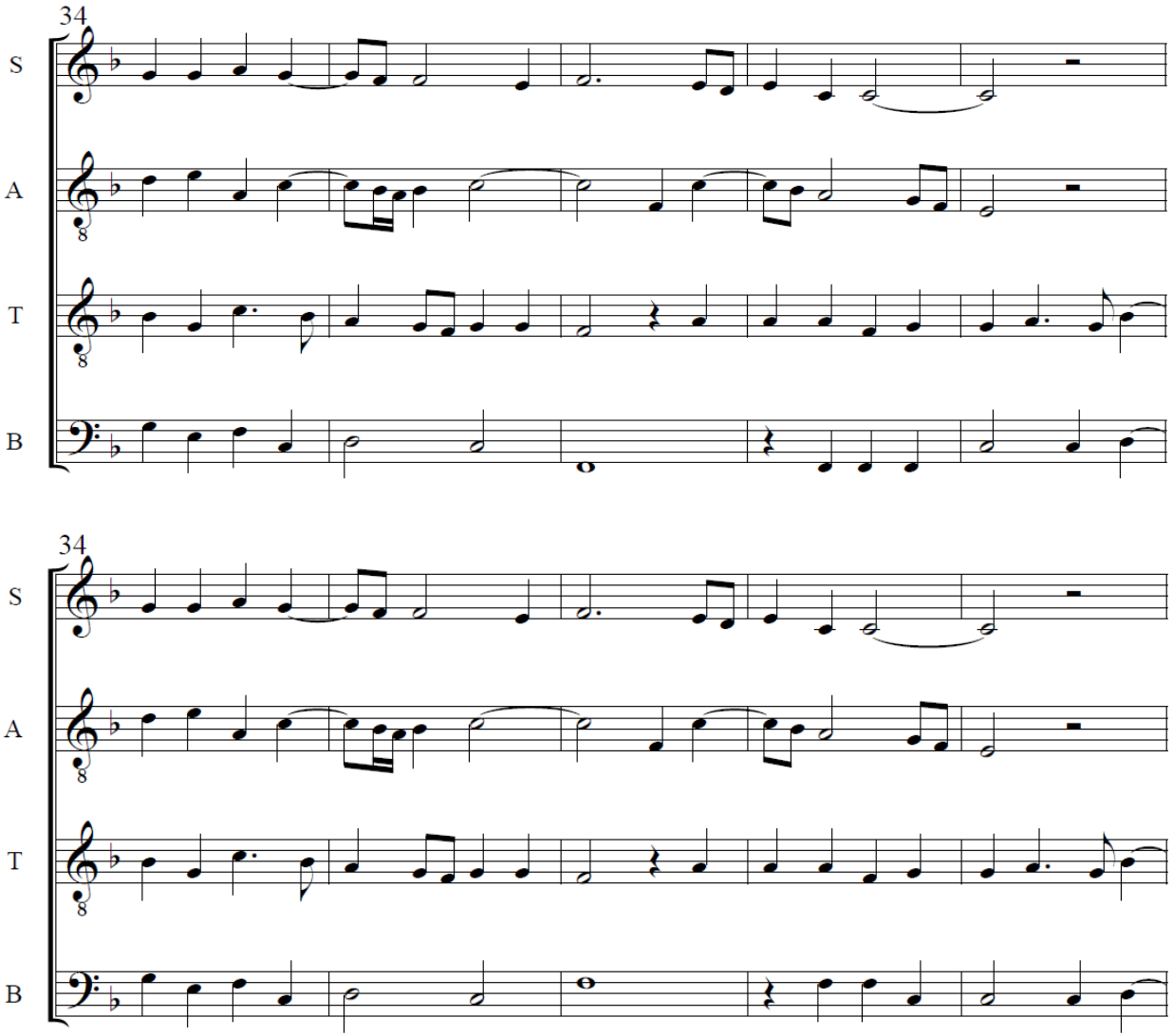
Masses and Mass movements in *P-Cug* MM 9 [mid or late 1540s]				
***Missa De leirea* [Leiria?]**	4		(*unicum*)	Cuenca's edition
***Missa A batalha* (K G S A)**	4	[Janequin]	*P-Cug* MM 6, ff. 28v-32r [Credo]; *I-CFm* Cod. LIII, ff. 167v–177r; *I-Bc* Q.25, ff. 1r-4v; *A-Wn* Mus.Hs. 15499 Mus, ff. 63v–88r; *I-CMac* P(E), ff. 54v–63r.	Cuenca's edition
***Missa, Bruxel* (K G S A)**	4	[Diego Bujel? or "from Brussels"?]	(*unicum*)	Cuenca's edition

Masses and Mass movements in *P-Cug* MM 9 [mid or late 1540s]						
18	88v-97r+[159]v-[165]r	*Missa Da morte et fortuna* (K G C)	4	[Jacquet de Berchem]	It's not the same as *E-Tc* Ms. 28, ff. 70v-98r.	Cuenca's edition (only Kyrie; the remaining movements can't be seen because of ink corrosion)
19	98v-103r	[15] *Missa* (K G S A)	4	Verdeloth [?]	(*unicum*)	Cuenca's edition (except tenor in S and the full Agnus Dei due to ink corrosion)
Masses and Mass movements in *P-Cug* MM 6 [mid or late 1540s]						
20	28v-32r	Credo	4			Cuenca's edition

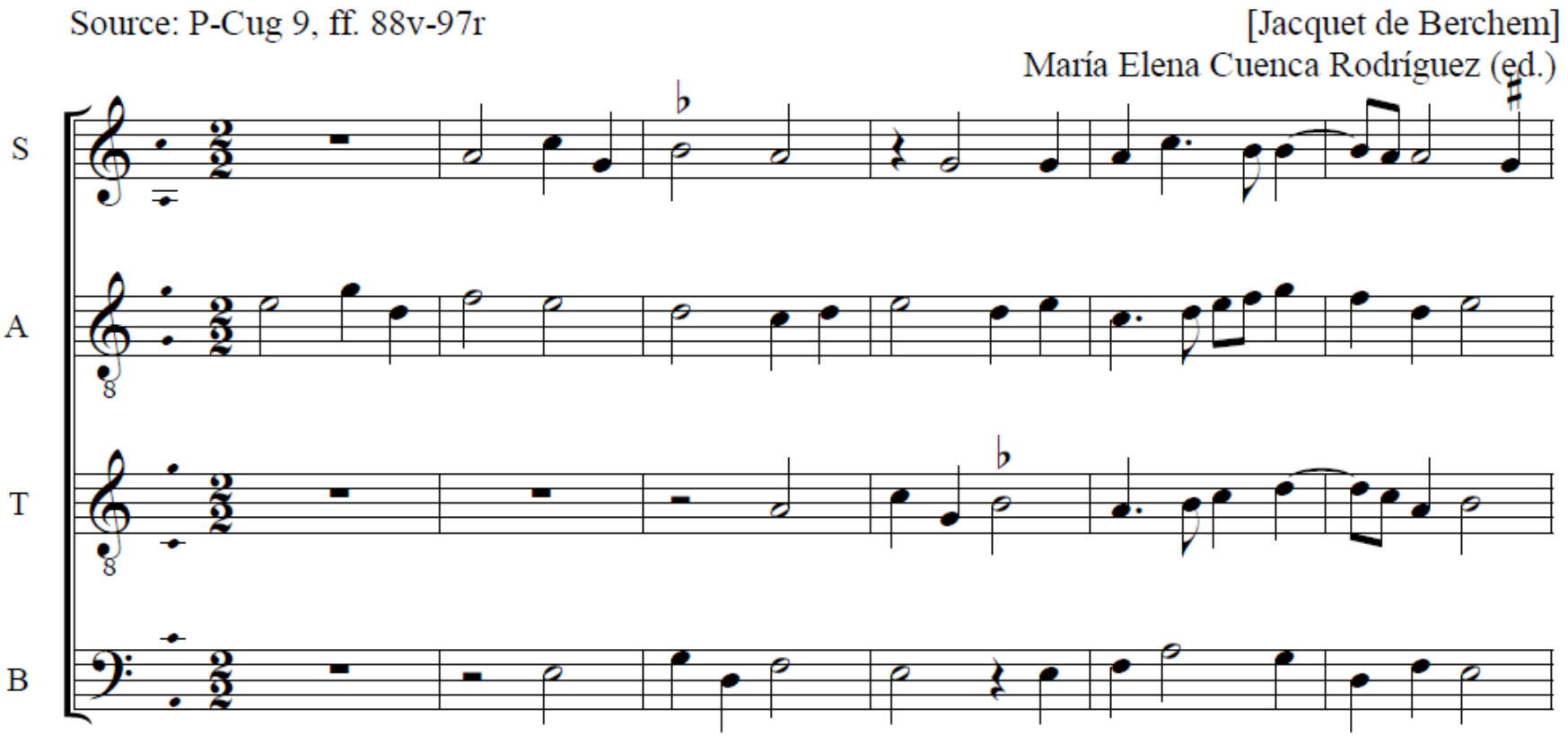
Janequin's *Missa La Bataille* (Agnus Dei III, bb. 66-71) in Moderne's *Liber decem missarum* (1532)



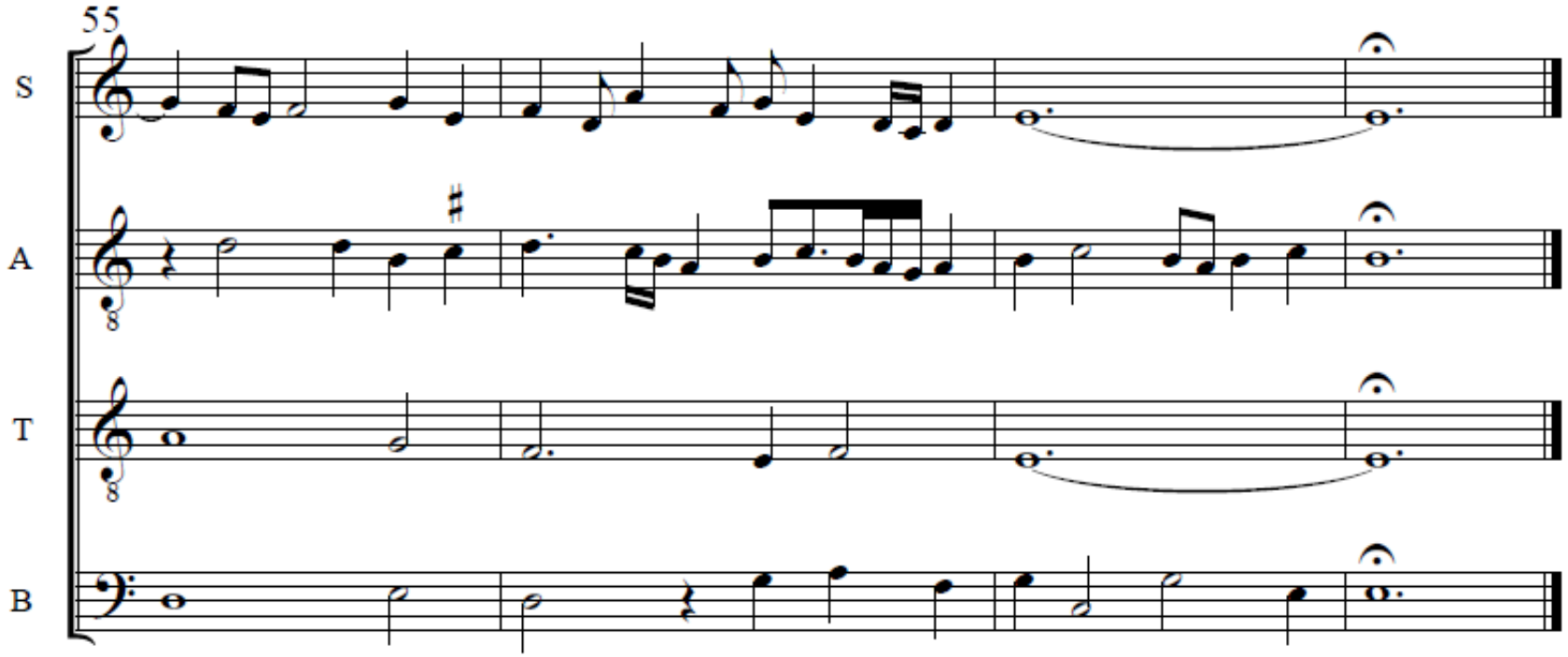
Janequin's *Missa La Bataille* (Gloria bb. 34-38) in Moderne's *Liber decem missarum* (1532) (above) and in *P-Cug* MM. 9, 68v-76r



Berchem's *Missa da morte et fortuna* (Kyrie, bb. 1-6)

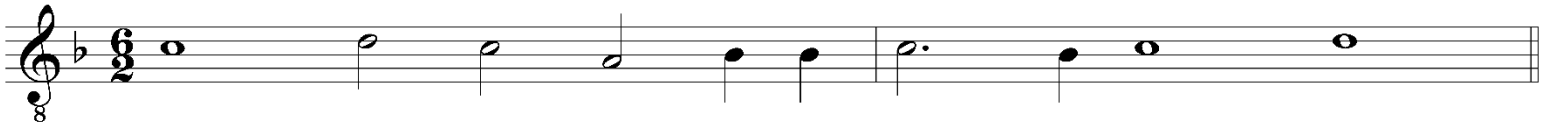


Elaborated cadence at the end of Kyrie in *Missa* no. 3 (bb. 55-58)

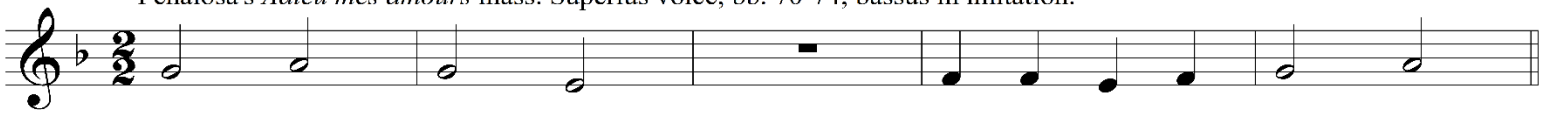


Peñalosa, Anchieta, Escobar, and Tordesillas' archetypical melody for "Crucifixus" (Credo)

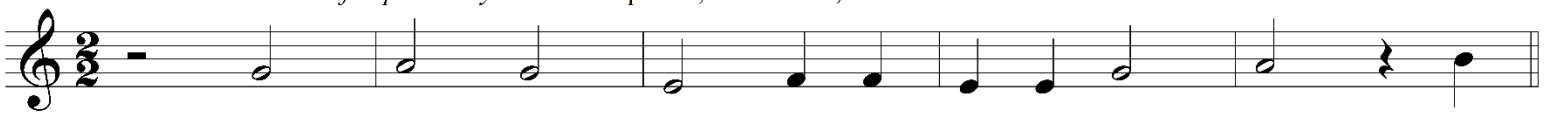
Peñalosa's *Por la mar* mass. Altus voice, bb. 97-100.



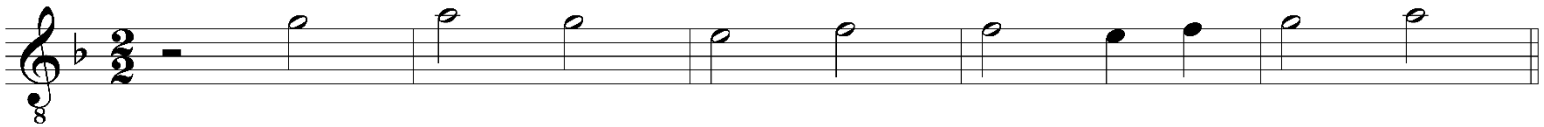
Peñalosa's *Adieu mes amours* mass. Superius voice, bb. 70-74; bassus in imitation.



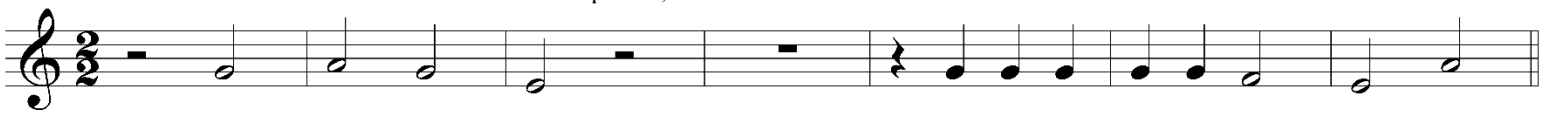
Peñalosa's *Nunca fue pena mayor* mass. Superius, bb. 96-100; altus and bassus in imitation.



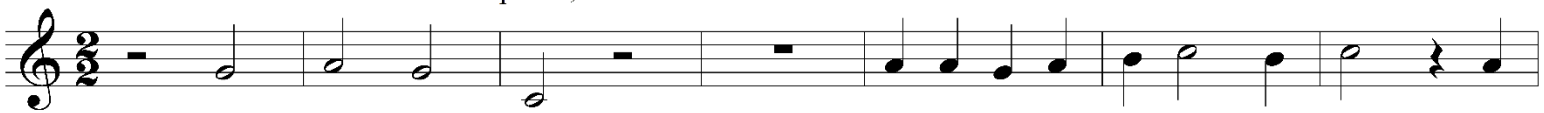
Peñalosa's *Ave María Peregrina* mass. Altus, bb. 118-122.



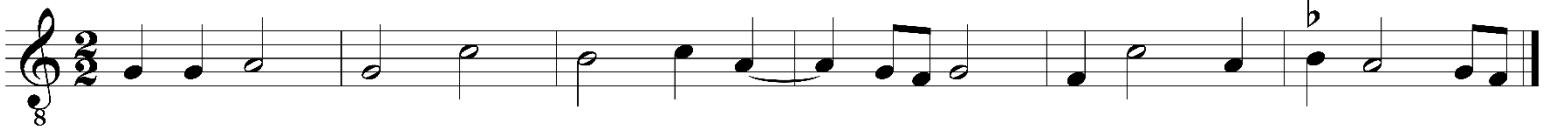
Anchieta's *De Nuestra Señora* mass. Superius, bb. 91-96.



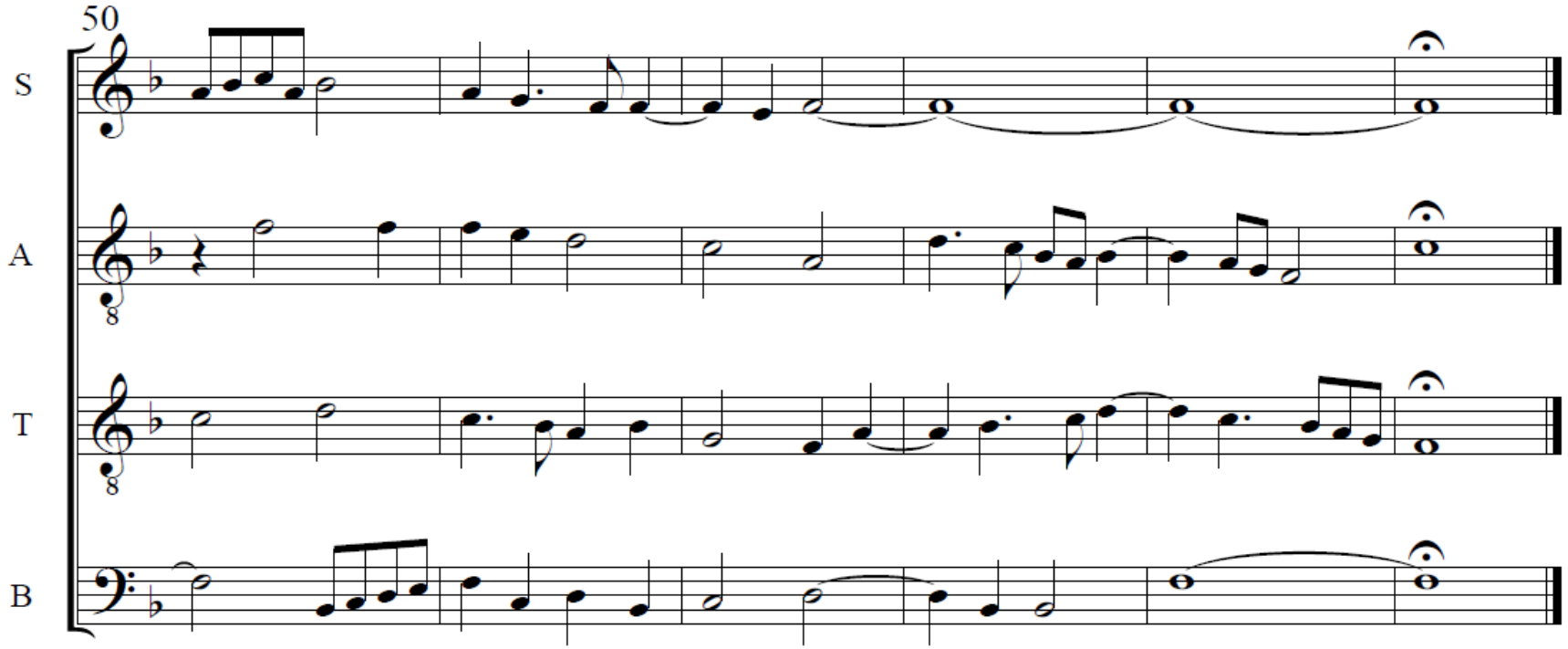
Escobar's *Sine nomine* mass. Superius, bb. 87-103.



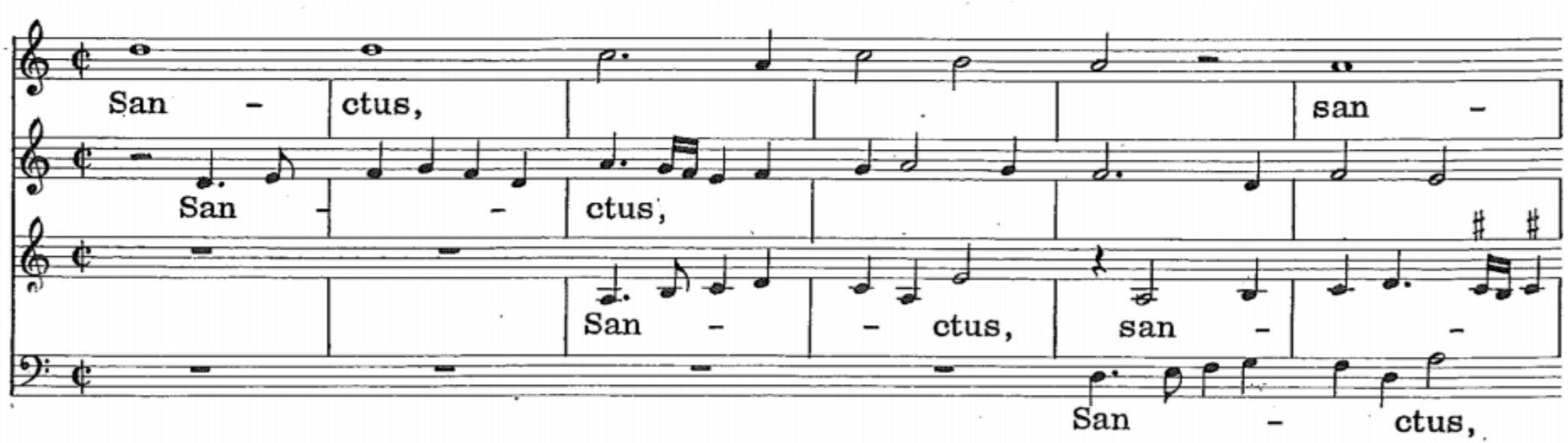
Tordesilla's *Sine nomine* mass. Tenor, bb. 98-103.

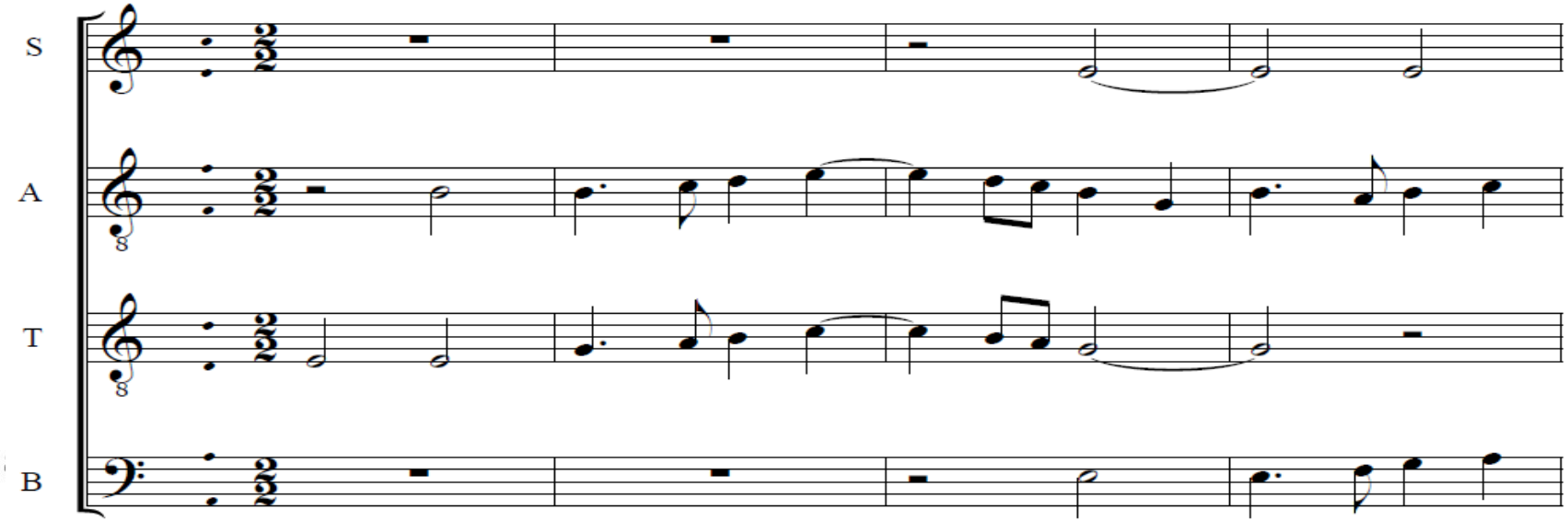


Prolonged cadence in Bruxel's Sanctus from his mass (no. 17 in Table 2) (bb. 50-55)

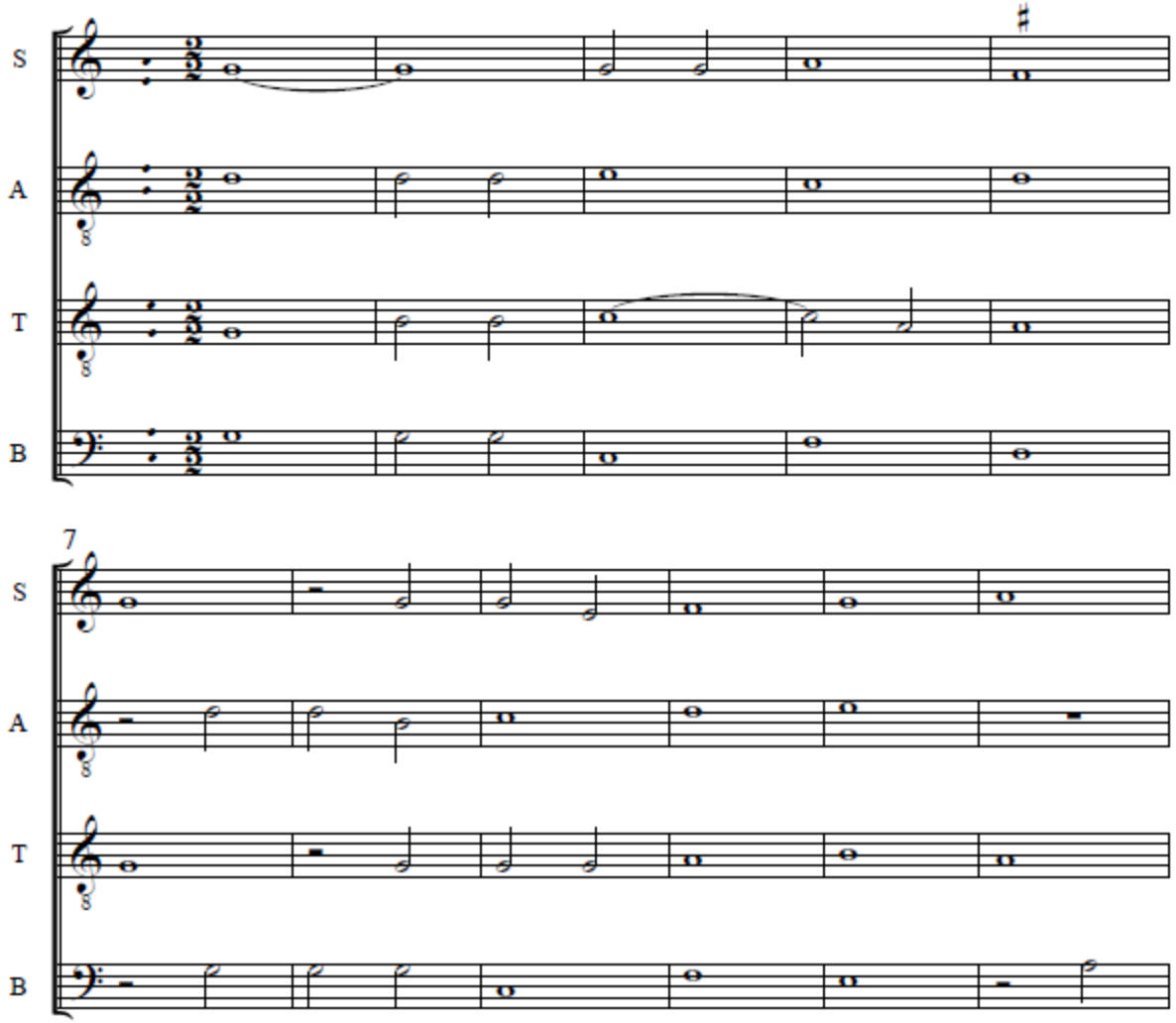


Beginning of Sanctus in Verdelot's *Missa Philomena* (above) and Mass no. 19 (below)

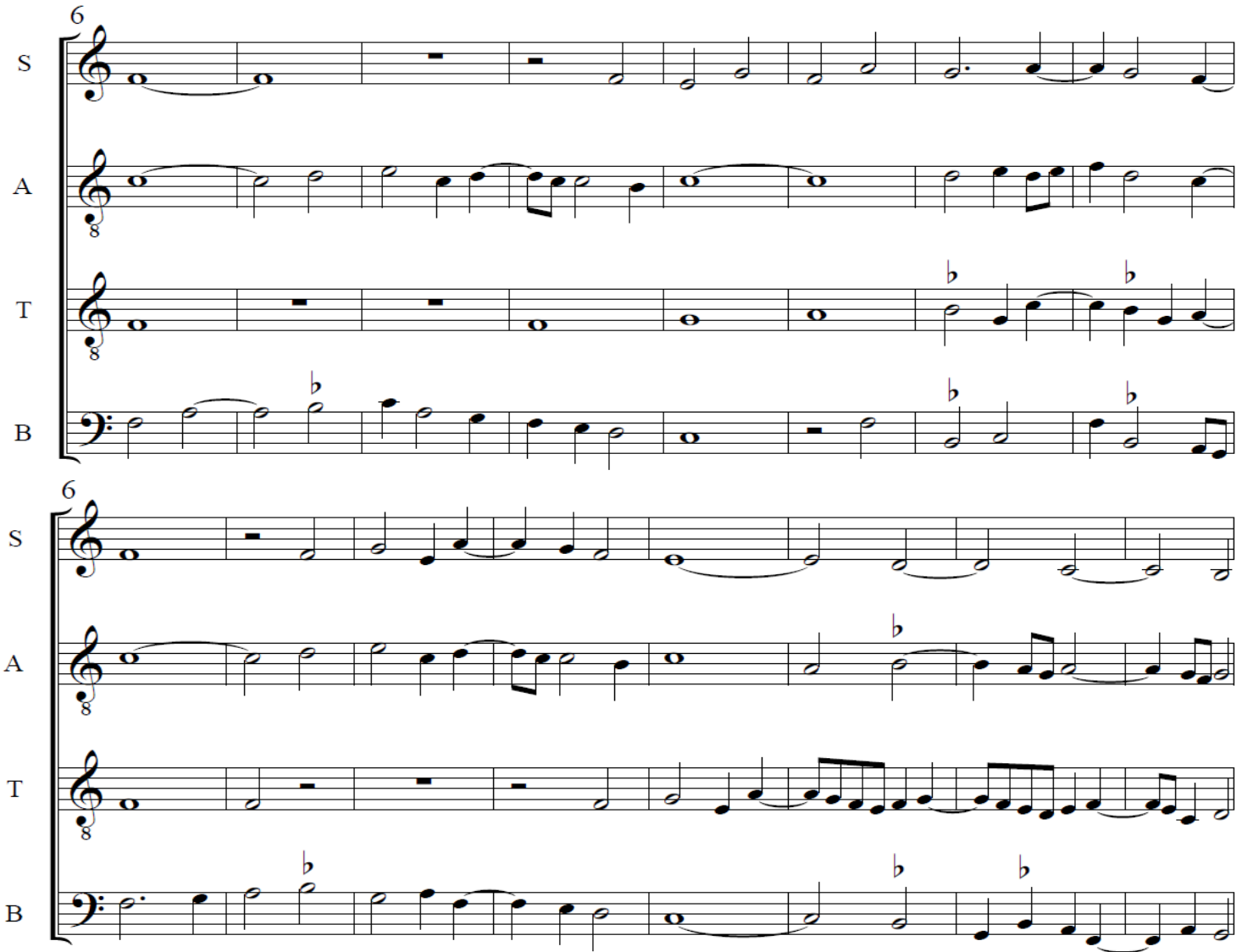




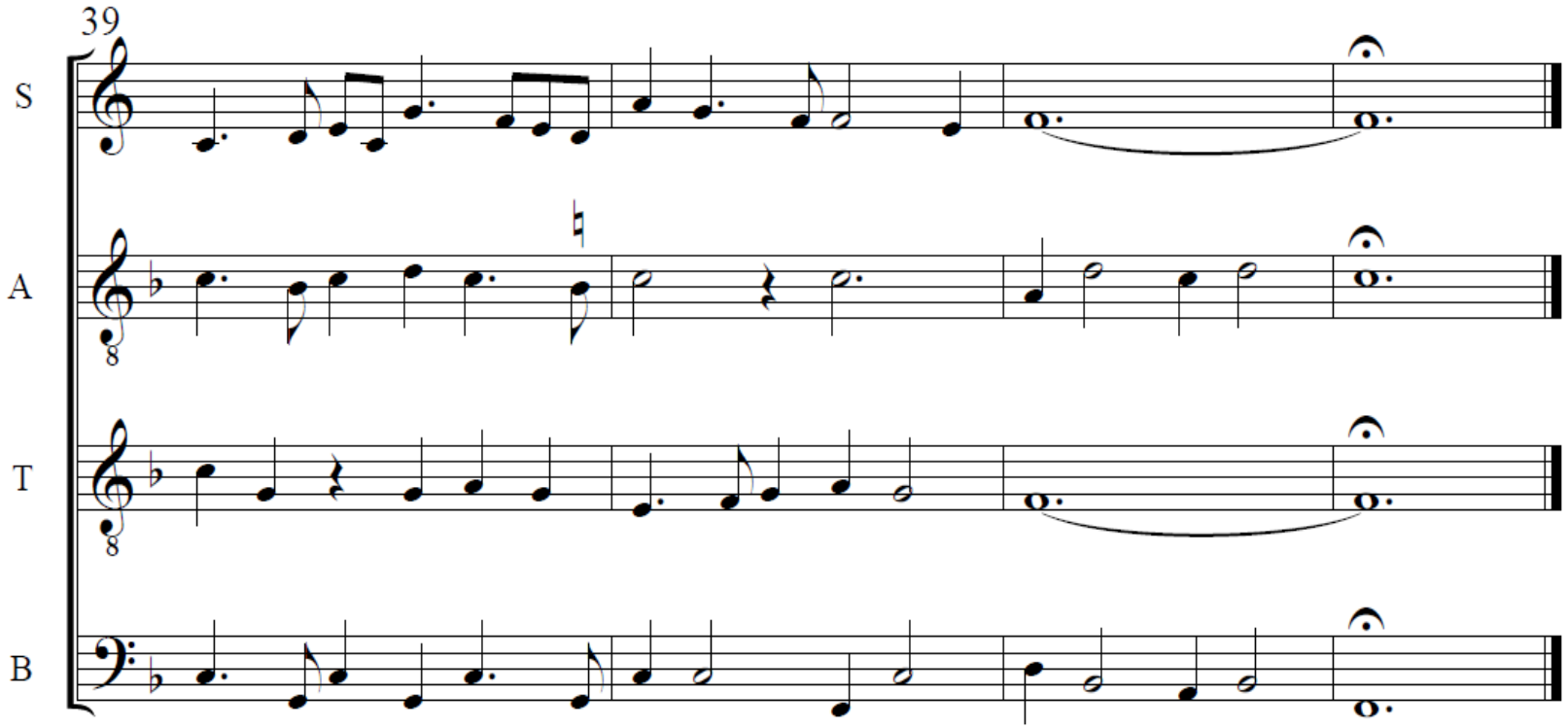
‘Et incarnatus’ from Credo no. 5 in *P-Cug* MM 12, ff. 80v



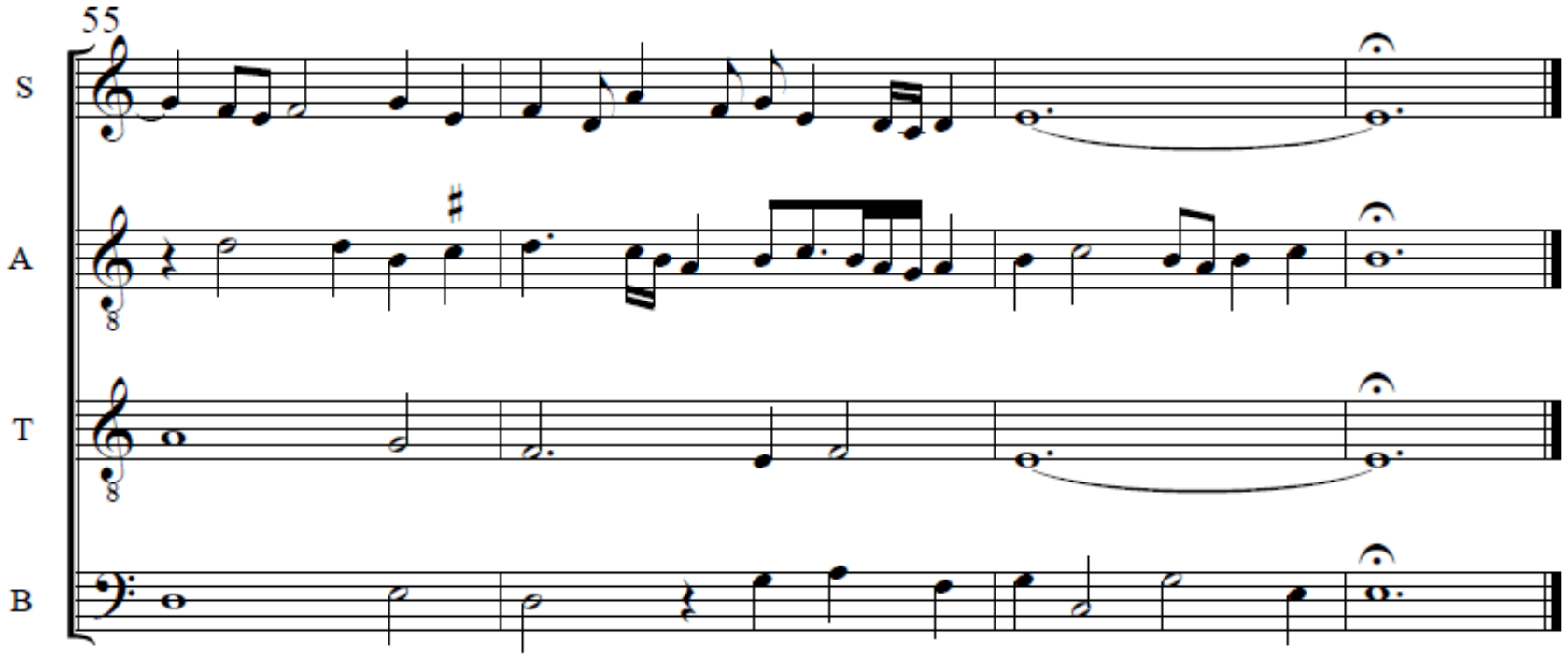
Agnus Dei no. 8 (above) and Tordesillas's Agnus Dei from *Missa Sine nomine* in *E-Tz* 2-3 (below)



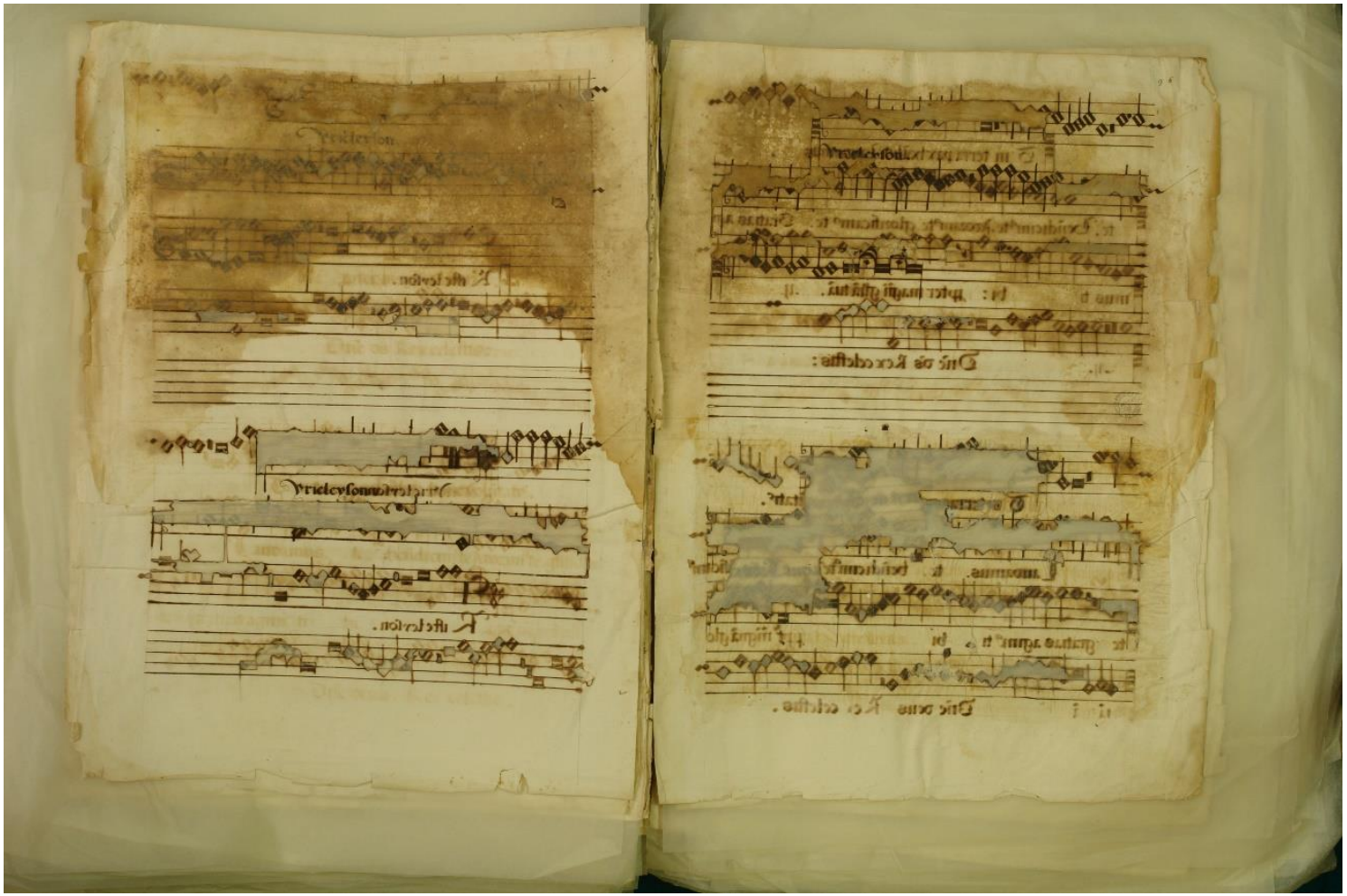
Elaborated cadence at the end of Kyrie no. 9



Elaborated cadence at the end of Kyrie in *Missa* no. 3 (bb. 55-58)



Ink corrosion in P-Cug MM 12, ff. 95v-96r (mass no. 12 in handout)



Quantitative Coimbra experiments

- We also performed a series of quantitative experiments using features, statistical analysis and machine learning

Our dataset: 603 MIDI files

Dataset	Mass Movements	Motets
Coimbra	38	0
Franco-Flemish	245	151
Iberian	78	91

- Secure Franco-Flemish composers (*from the Josquin Research Project*):
  - Alexander Agricola, Antoine Busnois, Loyset Compère, Josquin des Prez, Jacob Obrecht, Johannes Ockeghem, Marbrianus de Orto, Pierre de la Rue
- Secure Iberian composers (*from the Anatomy of Late 15th- and Early 16th-Century Iberian Polyphonic Music* project):
  - Alonso de Alba, Juan de Anchieta, Pedro de Escobar, Alonso Mondejar, Francisco de Peñalosa, Antonio de Ribera, Rivafrecha, Sanabria, Tordesillas, Juan de Urrede, Vasco Pires, Juan Illario, a few anonymous works
- All are 15th or early 16th century works

VNiVERSiDAD D SALAMANCA
CAMPUS DE EXCELENCIA INTERNACIONAL

MARIANOPOLIS COLLEGE

Feature extraction

- The same 553 features were extracted from all these MIDI files
  - These served as the basis of all experiments described on the following slides
- The remaining 944 jSymbolic features were excluded because of inconsistencies in data transcription and encoding
  - e.g. varying note durations
  - e.g. notes encoded as piano rather than voice
- These inconsistencies exist because the data was drawn from different sources
  - Each corpus used different preparation workflows

Experiment 1: Cross-validation

- Research questions:
  - How well are the secure Franco-Flemish and Iberian groups separated from one another stylistically?
  - Are these stylistic differences evident in both masses and motets?
  - Are the Coimbra mass movements statistically distinguishable from the Franco-Flemish and Iberian groups?
- Methodology:
  - Used machine learning to train classifiers to automatically distinguish between the different groups, based on the extracted jSymbolic features
  - Tested masses and motets separately, as well as together

Experiment 1: Classification accuracies

Music Being Compared	Average Classification Accuracy
FF and Ib, masses and motets	93.6%
FF and Ib, only motets	91.7%
FF and Ib, only masses	95.4%
FF, Ib and Coimbra, only masses	89.5%
FF, Ib and Coimbra, masses and motets	90.4%

- Rows 1 to 3 indicate that:
  - The Franco-Flemish and Iberian works are well-separated stylistically
  - This is true for both motets and masses, but mass movements are especially easily distinguishable (95.4%)

Experiment 1: Classification accuracies

Music Being Compared	Average Classification Accuracy
FF and Ib masses and motets	93.6%
FF and Ib only motets	91.7%
FF and Ib only masses	95.4%
FF, Ib and Coimbra only masses	89.5%
FF, Ib and Coimbra masses and motets	90.4%

- Rows 4 and 5 suggest that:
  - The Coimbra mass movements are also well-separated from the Franco-Flemish and Iberian music . . . or are they?
  - Actually, we need to look at the confusion matrices to verify
    - The Coimbra mass movements only represent 6.3% of the dataset
    - Their particular performance can thus be obscured

Experiment 1: Confusion matrices

True Label	Classified as Coimbra	Classified as FranFlem	Classified as Iberian
Coimbra masses	27	0	11
FF masses	2	236	7
Iberian masses	8	10	60

True Label	Classified as Coimbra	Classified as FranFlem	Classified as Iberian
Coimbra masses & motets	27	1	10
FranFlem masses & motets	7	377	12
Iberian masses & motets	7	21	141

- So, only some (a little under ¾) Coimbra mass movements are easily separable from the Franco-Flemish and Iberian music
  - So there is something distinctive about them, but there is also overlap

VNiVERSiDAD D SALAMANCA
CAMPUS DE EXCELENCIA INTERNACIONAL

MARIANOPOLIS COLLEGE

Experiment 1: Confusion matrices

True Label	Classified as Coimbra	Classified as FranFlem	Classified as Iberian
Coimbra masses	27	0	11
FF masses	2	236	7
Iberian masses	8	10	60

True Label	Classified as Coimbra	Classified as FranFlem	Classified as Iberian
Coimbra masses & motets	27	1	10
FranFlem masses & motets	7	377	12
Iberian masses & motets	7	21	141

- For those that are “misclassified”, they are almost always (95.4% of the time) classified as Iberian!
  - This suggests that at least some of them are closer in style to Iberian than Franco-Flemish music

Experiment 2: Classifying individual Coimbra mass movements

- Research question:
  - Are the individual Coimbra mass movements more Iberian or Franco-Flemish in style?
- Methodology:
  - Trained two classification models on the secure Franco-Flemish and Iberian music (only)
    - One classifier was trained on both motets and mass movements
    - One classifier was trained on only mass movements
  - Used these trained models to classify each Coimbra mass movement separately
    - Each could only be classified as Franco-Flemish or Iberian (i.e. there was no longer a Coimbra class)

Experiment 2: Results

Coimbra Mass Movement	Trained on Masses and Motets	Trained on Masses Only
2. Missa Salve Regina, Benedictus, P-Cug 12	Franco-Flemish	Iberian
3. Missa Kyrie P-Cug 12	Franco-Flemish	Iberian
3. Missa Sanctus P-Cug 12, f. 26v	Franco-Flemish	Iberian
4. Missa Sine nomine anónima, Kyrie P-Cug 12	Franco-Flemish	Franco-Flemish
16. Missa A Batalha. Janequin. Sanctus. P-Cug 9	Franco-Flemish	Iberian
18. Missa Da Morte e fortuna, Berchem, Kyrie P-Cug 9	Franco-Flemish	Iberian

- 84.2% (all but 6 / 38) of the Coimbra mass movements were classified as Iberian by both of the 2 classifiers
  - The 6 exceptions are shown on the table above
- 97.3% (all but one) were classified as Iberian by the model specialized in mass movements
  - Recall that all the Coimbra pieces are mass movements

VNiVERSiDAD D SALAMANCA
CAMPUS DE EXCELENCIA INTERNACIONAL

MARIANOPOLIS COLLEGE

Experiment 2: Results

Coimbra Mass Movement	Trained on Masses and Motets	Trained on Masses Only
2. Missa Salve Regina, Benedictus, P-Cug 12	Franco-Flemish	Iberian
3. Missa Kyrie P-Cug 12	Franco-Flemish	Iberian
3. Missa Sanctus P-Cug 12, f. 26v	Franco-Flemish	Iberian
4. Missa Sine nomine anónima, Kyrie P-Cug 12	Franco-Flemish	Franco-Flemish
16. Missa A Batalha. Janequin. Sanctus. P-Cug 9	Franco-Flemish	Iberian
18. Missa Da Morte e fortuna, Berchem, Kyrie P-Cug 9	Franco-Flemish	Iberian

- These results suggests that the Coimbra mass movements are, as a whole, more Iberian than Franco-Flemish in character
- Although results for individual mass movements should not be interpreted as perfectly authoritative, the overall pattern is clear and convincing

Experiment 3: Feature analysis

- Research question (putting the Coimbra mass movements aside for the moment):
  - Which particular musical characteristics best separate the Franco-Flemish and Iberian masses and motets?
- Methodology:
  - Used statistical analysis (absolute Pearson correlation co-efficient) to see which features are most strongly correlated with each type of music

Experiment 3: Overall results

- Motets:
  - Differences between Iberian and Franco-Flemish music are primarily melodic
  - Vertical elements play a secondary role
- Mass movements:
  - Features differentiating Iberian music from Franco-Flemish music are more varied than in motets
  - Vertical aspects now play a greater role than melodic aspects
  - Other individual features like range and diversity in the number of distinct pitches are more important still
- Mass movements and motets combined:
  - Melodic features once again emerge as the most important

Experiment 3: MotetsOnlyNoCoimbra most discriminating features

- 0.640719 Prevalence_of_Most_Common_Melodic_Interval
- 0.545375 Mean_Melodic_Interval
- 0.535871 Direction_of_Melodic_Motion
- 0.519188 Melodic_Sevenths
- 0.509731 Voice_Separation
- 0.506414 Melodic_Interval_Histogram_10
- 0.494385 Melodic_Large_Intervals
- 0.489793 Melodic_Octaves
- 0.489793 Melodic_Interval_Histogram_12
- 0.488229 Melodic_Sixths
- 0.476709 Melodic_Interval_Histogram_9
- 0.473214 Wrapped_Vertical_Interval_Histogram_4
- 0.473127 Melodic_Interval_Histogram_16
- 0.468809 Vertical_Interval_Histogram_7
- 0.467444 Melodic_Interval_Histogram_17
- 0.462648 Melodic_Interval_Histogram_14
- 0.445186 Melodic_Interval_Histogram_8
- 0.442547 Melodic_Interval_Histogram_5
- 0.442547 Melodic_Perfect_Fourths
- 0.441235 Average_Interval_Spanned_by_Melodic_Arcs

Experiment 3: MassesOnlyNoCoimbra most discriminating features

- 0.695843 Range
- 0.632123 Number_of_Pitches
- 0.458177 Chord_Duration
- 0.456041 Relative_Size_of_Melodic_Intervals_in_Lowest_Line
- 0.450811 Pitch_Variability
- 0.416166 Vertical_Interval_Histogram_15
- 0.408743 Complete_Rests_Fraction
- 0.384314 Prevalence_of_Most_Common_Pitch
- 0.384138 Vertical_Interval_Histogram_27
- 0.377422 Number_of_Pitch_Classes
- 0.372288 Vertical_Interval_Histogram_19
- 0.360324 Partial_Rests_Fraction
- 0.343436 Vertical_Interval_Histogram_5
- 0.341496 Number_of_Common_Pitches
- 0.339544 Prevalence_of_Dotted_Notes
- 0.32967 Wrapped_Vertical_Interval_Histogram_3
- 0.324811 Wrapped_Vertical_Interval_Histogram_5
- 0.324811 Vertical_Perfect_Fourths
- 0.323448 Total_Number_of_Notes
- 0.320974 Relative_Note_Density_of_Highest_Line

Experiment 3: MassesAndMotetsNoCoimbra most discriminating features

- 0.44709 Prevalence_of_Most_Common_Melodic_Interval
- 0.41496 Direction_of_Melodic_Motion
- 0.41302 Mean_Melodic_Interval
- 0.41007 Melodic_Sevenths
- 0.40328 Melodic_Interval_Histogram_10
- 0.40238 Vertical_Interval_Histogram_27
- 0.39769 Melodic_Large_Intervals
- 0.38499 Melodic_Interval_Histogram_12
- 0.38499 Melodic_Octaves
- 0.38236 Melodic_Interval_Histogram_16
- 0.37937 Melodic_Interval_Histogram_17
- 0.37482 Melodic_Interval_Histogram_14
- 0.3745 Voice_Separation
- 0.37384 Melodic_Sixths
- 0.37093 Melodic_Interval_Histogram_9
- 0.35815 Total_Number_of_Notes
- 0.35737 Partial_Rests_Fraction
- 0.35694 Average_Interval_Spanned_by_Melodic_Arcs
- 0.35501 Melodic_Interval_Histogram_13
- 0.35398 Number_of_Pitches

Experiment 3: Caveats

- The Pearson correlation co-efficient only considers features individually
  - In practice, how features vary in groups can be more important
  - We will leave a more sophisticated analysis to future research
- Many Rhythmic features were excluded from this study
  - Due to transcription and encoding inconsistencies in the data
  - Informal initial studies we performed suggest that rhythm may indeed play an important role
  - Studies are needed with more consistently encoded data
- jSymbolic does not (yet!) measure features based on:
  - Cadences
  - Imitation
  - Text

Final comments (1/3)

- The Coimbra manuscripts, and Portugal, were strongly influenced by international styles (and vice versa?):
  - Spanish-style repertoires
  - Northern-style masses

Final comments (2/3)

- Influences include:
  - Foreign masses circulated in Portugal
  - Iberian composers influenced by Northern styles
- The Coimbra works were adapted to the performative context of the Santa Cruz chapel
  - As happened in the case of Janequin's mass

Final comments (3/3)

- Franco-Flemish elements in the polyphony seem to have been received mostly through Spanish masses
  - Which were influenced by Burgundian and French composers
- This Spanish-influenced merged style was transmitted to Portugal through numerous anonymous and doubtfully attributed masses

Future related research

- Learn more about the reception of French works by Janequin, Verdelot, or Richafort
  - And their influence on Iberian polyphony
- Extend this case study to other genres
  - e.g. motets, hymns, or anonymous lamentations in Portuguese manuscripts
  - Others are doing promising work already

Thanks for your attention!

- **E-mail:** cory.mckay@mail.mcgill.ca
- **E-mail:** elenacrod@usal.es

VNiVERSiDAD D SALAMANCA

CAMPUS DE EXCELENCIA INTERNACIONAL

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

SIMSSA | Single Interface for Music Score Searching and Analysis

jMIR

General discussion questions

- What musical questions and problems can be most interestingly addressed by machine learning?
  - Both in general and with respect to the CRIM project?
- How can music scholars work interactively with machine learning algorithms in ways that the algorithms benefit from their expertise and they benefit from computational processing?
  - Can scholars use “opaque” models trained by machine learning (e.g. from deep learning), or are only less powerful but more transparent machine learning algorithms (e.g. decision trees) useful?
- What additional statistical analysis tools outside of machine learning can scholars make use of?
- What kinds of musical features would scholars most like to see computationally extracted from music?
- How can existing frameworks like CRIM, jSymbolic, music21 and Humdrum be improved to meet the needs of music scholars, both in general and with respect to machine learning?
  - Is there a need for new kinds of frameworks?

Working with the CRIM corpus (1/2)

- How can we use these methods to explore the CRIM corpus?
- What patterns of similarity might be revealed?
  - Could these methods reveal clusters of music that we might expect?
    - e.g. associate the component movements of Masses with one another
  - Could these methods reveal meaningful patterns that we might not anticipate?
    - And how can we evaluate what these patterns mean, and if they are useful?
- How can we use the CRIM observational metadata and relationships to inform machine learning?

Working with the CRIM corpus (2/2)

- The jSymbolic features have been pre-extracted from the CRIM data:
  - Posted as CSV files, for ease of access
  - https://drive.google.com/open?id=1MoyAyM01-gZNZTDlbZc8bp4rYfS2vInN
- Exists in two versions:
  - Full version: all features
  - Safe(r) version: features resilient to inconsistent data encoding practices
- Please feel free to download jSymbolic and use it to explore and experiment with this data
  - http://jmir.sourceforge.net/index_jSymbolic.html