Computational Modelling in Musicology: The case of Medieval Chant

Peter van Kranenburg

Utrecht University, Meertens Institute

Geert Maessen

Independent Scholar



Universiteit Utrecht



Gregoriana Amsterdam



Modelling









Formalized data: musical scores neume sequences audio historical data

•••



Formalized data: musical scores neume sequences audio historical data

...

Classification Similarity Distance Clustering Model Patterns Rules

...









The Modelling Cycle





Tailored process of modelling

Fron a ·~ ~ ~ ~. N (\mathbf{H}) Offert. WY ev miv for truet ner-sur. Dómi- ne: di-xi: n te sperávi, برسر مر مر. برهم المر الم MAN V. - e-al.r Jerau N. T II .N - 11. IN in má- ni-bus Tu es Deus me- us, tu-7 1 1 1 1 P. 2 I. ...9 n:= 1" res 5 is témpo- ra me- a.

Medieval Christian Chant: Traditions

GRE (Gregorian) – Around 800 AD, Charlemange ordained the Roman liturgy (including chant) in all of Western Europe, replacing local traditions.

Local Traditions: GAL (Gallican) – Tradition from Gaul ROM (Old Roman) – Tradition from Rome itself BEN (Beneventan) – Tradition from Benevenum MIL (Milanese/Ambrosian) – Tradition from Milan HISP (Old Hispanic) – Tradition from Iberian Peninsula





Medieval Christian Chant: Traditions

Question: Which was first, Old Roman (ROM) or Gregorian (GRE)?



Lost pitches

Early music notation did not show exact pitches:



As do later sources:



If we do not have a later source, the melos is lost.

This applies to almost all of the Old Hispanic tradition.

Aim & Objective

Long Term Aim: Computational evaluation of the relations between the melodies from the different traditions to find clues on origin and development.

Current Objective: Classify chant melodies according to their origin.

Offertory

Longest genre (type) of chant.

Basic structure:

Response verse 1 Response' verse 2 ... verse n Response'

verses mostly come from the Book of Psalms.

This chant-type is present in all studied traditions

Dataset

Newly encoded dataset with Offertories from five traditions

tradition	century	offertories	parts	avg. notes/part	std. notes/part
GRE: Gregorian Chant	XI–XII	115	344	162.54	62.65
ROM: Old Roman Chant	XI–XIII	94	285	170.14	69.04
MIL: Milanese Chant	XII–XIII	104	147	177.63	98.50
BEN: Beneventan chant	XI–XII	39	41	152.98	64.00
MOZ: Mozarabic chant	XI–XVI	71	139	127.94	52.74

Encoding: Volpiano font (Univ. of Regensburg)

-	•	1	2		3	4	5	6	7
Ξ	Ē	Ś	2		Ē			Ī	+ - -
0045	0046	004	005	50 00	51 0	052	0053	0054	0055
i	Ι	w	W	x	X		Y	z	Z
	+44+		 		¥ 		 	Þ 	Þ
0105	0073	0119	0087	0120	0088	012	1 0089	0122	0090
8	9	a	b	с	d	e	f	g	h
•		•			 •	-	 Z		
0056	0057	0097	0098	0099	0100	010	1 0102	0103	0104
j	k	1	m	n	0	P	P	r	S
•	- - -	•	E 	•	e	•	e 	 ▲ ↓ ↓	



Encoding

Some Conventions:

- --- (\equiv) separates words
- -- (\equiv) separates syllables
- (⁼/₌) separates neumes

- 1 (💰) : start of chant
- 3 (1) : end of 'phrase'
- 4 ($\mathbbm{1}$) : end of part
 - 5 (\blacksquare) : end of chant





Which continuation is more likely?

P (green | The grass is) > P (happy | The grass is)

For a sentence *s* consisting of words *w* in language *L* with Vocabulary *V*

$$s = w_1, w_2, \ldots, w_l$$

model the probability of w_i given its preceding context of n-1 words

$$p(w_i | w_{i-(n-1)}, \dots, w_{i-1})$$
Learned from training data
+ smoothing
$$p(w_3) \approx P(w_3 | w_1 w_2)_4 | w_2 w_3) \approx P(w_6 | w_4 w_5)_7 | w_5 w_6)$$
for n=3:
$$w_1 w_2 w_3 w_4 w_5 w_6 w_7$$

Re-defining terms for chants:

Word w : one interval, measured in semi-tones
Vocabulary V : melodic intervals, measured in semi-tones
Sentence s : part of a offertory (i.e., a *response*, or a *verse*)

(Conklin & Witten, 1995; Pearce 2005)





'words': 3 2 2 - 2 3 - 3 2 1 2 - 2 0 0 0 - 1 - 4 2 2 1 - 1 - 2 0P(3|3 2 2 - 2)



'words': 3 2 2 -2 3
$$\overline{-3}$$
 2 1 2 -2 0 0 0 -1 -4 2 2 1 -1 -2 0
 $P(-3|2 2 -2 3)$



'words':
$$3 2 2 - 2 3 - 3 2 1 2 - 2 0 0 0 - 1 - 4 2 2 1 - 1 - 2 0$$

$$P(2|2 - 2 3 - 3)$$
etc.

Perplexity

Given a sentence s

$$s = w_1, w_2, \ldots, w_l$$

and a (*n*-gram-)model for *p*(*s*), the perplexity *PP* is defined as:

$$PP(s) = \sqrt[l]{\frac{1}{p(w_1, w_2, \dots, w_l)}}$$

We use perplexity as a measure of fitness of a chant part in a tradition. The higher the perplexity, the more 'alien' the chant is to the tradition.

Perplexity: Example



Perplexity given the 5-gram model trained on GRE\GRE-21: 3.99

Perplexity given the 5-gram model trained on ROM: 6.47

Internal 'diversity'



Outliers (residue of misfits)







 $P(w_i) < 0.05$

Gradual, Beneventum, c. 1100

Discriminative model

Use perplexities for classification:

	BEN	GRE	MIL	MOZ	ROM
GRE-01-0	7,646661812	4.549781277	5.883542923	7.168091561	6.21939177
GRE-01-1	8.696976996	4.960571324	7.958359689	10.28244106	8.118070524
GRE-01-2	9.174275588	5.855445154	7.339194913	8.908588143	8.662131751
GRE-02-0	9.643594395	4.941502746	7.346635142	11.1504086	9.031562113
GRE-02-1	7.859347922	4.15448816	5.551899549	7.201198482	6.96581904
GRE-02-2	10.4675142	4.018135123	7.309646705	10.01487155	10.20867903
GRE-03-0	9.623284102	4.850433351	6.72254506	7.261196802	8.180618277
GRE-03-1	8.331485178	4.760134213	7.53795515	8.564636611	8.094603166
				•••	
GRE-21-0	7.83014702	3.98680402	6.45858974	7.83937877	6.460733899

Minimal Perplexity

tradition	precision	recall	F1-score	support
BEN	0.98	0.98	0.98	41
GRE	0.98	0.98	0.98	344
MIL	0.94	0.99	0.96	147
MOZ	0.99	0.94	0.96	139
ROM	1.00	0.99	0.99	285
avg/total	0.98	0.98	0.98	956

trad	BEN	GRE	MIL	MOZ	ROM
BEN	40	0	1	0	0
GRE	0	338	5	1	0
MIL	0	1	146	0	0
MOZ	1	4	3	130	1
ROM	0	1	1	0	283

Misclassified chants correspond to earlier observed outliers

Old Hispanic Tradition (HISP)

• c. 5,000 chants preserved in neumatic notation only.



Number of notes per syllable: 17 - 1 - 7 - 2 - 2 - 1 - 2 - 1 - 3 - 2 - 9 - ...

New vocabulary: length-classes:

1 2 3 4 5 6 7 8 9 10 11-15 16-20 21-25 26-50 >50 a b c d e f g h i j k l m n o

Number of notes per syllable: 1 - a - g - b - b - a - b - a - c - b - i - ...



Pitched Old Hispanic Chants

- We DO have pitch information for 26 Old Hispanic Chants (HISA)
- Furthermore, there are 15 expert reconstructions (HISB)
- There are 29 melodies from later sources that are believed to be Old Hispanic melodies (HISC)
- There are melodies in other traditions that are believed to be related to Old Hispanic counter-parts (HISD)
- We computationally generated 28 melodies based on the neumes (HISE)

	GRE	ROM	MIL	BEN	MOZ	Α	В	С	D	E	parts	GRE	ROM	MIL	BEN	MOZ	Α	В	С	D	Е
GRE	369	0	5	0	1	0	0	1	14	4	394	0	3	56	2	4	3	5	4	111	206
ROM	2	279	1	0	0	0	0	0	3	0	285	46	0	36	18	6	0	0	7	160	12
MIL	1	0	139	0	0	1	0	0	6	0	147	58	7	0	1	0	2	4	9	51	15
BEN	1	0	1	39	0	0	0	0	0	0	41	6	1	21	0	0	2	0	3	5	3
MOZ	5	1	2	0	126	0	0	5	0	0	139	12	8	56	0	0	0	2	43	5	13
Α	0	1	2	0	1	14	3	2	1	7	31	2	1	3	0	1	0	4	5	1	14
В	1	0	2	0	0	0	4	3	0	5	15	1	0	2	0	0	3	0	3	0	6
С	5	0	8	1	7	1	2	5	0	3	32	5	0	10	1	7	4	2	0	0	3
D	29	21	8	0	4	0	0	1	36	1	100	47	30	13	0	5	0	0	3	0	2
E	24	0	6	0	0	2	1	0	1	50	84	59	1	10	0	0	3	4	4	3	0

True class included as target

True class not included as target

- A: Original
- B: Expert reconstructions
- C: Later sources
- D: Other traditions
- E: Generated









Historical Musicology Computational Musicology

Computing and Information Science

Thank You



Universiteit Utrecht



Gregoriana Amsterdam