SUPPLEMENTS

Mode Classification and Natural Units in Plainchant

Bas Cornelissen, Willem Zuidema and John Ashley Burgoyne

Data and code. All data and code used in this study has been made available online at github.com/bacor/ISMIR2020. All randomness in the code has been fixed, so it should in theory be possible to reproduce our results exactly. The evaluations metrics of all experiments are already included in the repository, as is the data used in the first run of the experiment; this should be sufficient for reproducing most figures. We have included model predictions and tuning results only for the first run of the experiment. Detailed logs of everything from data generation to vizualization can also be in the repository, together with many more figures besides those included in the paper and the supplements. In particular, the repository contains heatmaps with multiple evaluation metrics (accuracy, precision, recall and F_1) for all models and all experimental conditions.

```
Filtering chants...
. Filter Chants Without Volpiano:
. Exclude all chants with an empty volpiano field
.  > 87.20% removed (433443 out of 497071; 63628 remain)
. Filter Chants Without Notes:
. Exclude all chants without notes
.  > 2.87% removed (1825 out of 63628; 61803 remain)
. Filter Chants Without Simple Mode:
. Include only chants with simple modes: 1-8, not transposed
.  * include_transposed=False
.  > 23.02% removed (14227 out of 61803; 47576 remain)
. Filter Chants Without Full Text:
. Filter chants without full text
.  > 20.65% removed (9823 out of 47576; 37753 remain)
. Filter Chants Where Incipit Is Full Text:
. Filter chants whose incipit is identical to the full text
.  > 14.59% removed (5507 out of 37753; 32246 remain)
. Filter Chants By Genre:
. Include only chants with a certain genre
.  * include=[’genre_a’]
.  > 52.06% removed (16787 out of 32246; 15459 remain)
. Filter Chants Not Starting With G Clef:
. Exclude chants that do not start with a G clef
.  > 0.05% removed (7 out of 15459; 15452 remain)
. Filter Chants With F Clef:
. Exclude chants that contain an F clef
.  > 0.00% removed (0 out of 15452; 15452 remain)
. Filter Chants With Missing Pitches:
. Filter chants with missing pitches: containing the substring 6------6
.  > 7.54% removed (1165 out of 15452; 14287 remain)
. Filter Chants With Nonvolpiano Chars:
. Exclude all chants with non-volpiano characters
.  > 0.03% removed (5 out of 14287; 14282 remain)
. Filter Chants Without Word Boundary:
. Only include chants with ’---’ in their volpiano
.  > 0.08% removed (11 out of 14282; 14271 remain)
. Filter Chants With Duplicated Notes:
. Filter duplicate chants: whose notes occur multiple times
.  > 2.84% removed (406 out of 14271; 13865 remain)
```

Figure S1. **Filtering.** As described in the main text, we filtered the total dataset of 497,071 chants to obtain a clean subset of responsories and antiphons. The effects of all of the filters are logged and will be made available online. As an example, this ‘figure’ shows the series of filters applied to obtain the full set of antiphons used in this study.

Genre	Subset	Split	# chants	# notes	Mean length (notes)
responsory	full	train	4 922	676 807	137.5
responsory	full	test	2 109	290 064	137.5
responsory	full	total	7 031	966 871	137.5
responsory	subset	train	1 234	169 642	137.5
responsory	subset	test	529	72 504	137.1
responsory	subset	total	1 763	242 146	137.3
antiphon	full	train	9 706	576 738	59.4
antiphon	full	test	4 159	248 405	59.7
antiphon	full	total	13 865	825 143	59.5
antiphon	subset	train	2 911	190 165	65.3
antiphon	subset	test	1 248	82 781	66.3
antiphon	subset	total	4 159	272 946	65.6

Figure S2. **Dataset statistics.** The number of chants, their average length and number of notes for each dataset. We sort datasets by genre, then by subset (include melody variants in the full set, or exclude them in the subset) and finally by train/test split (or total for the two combined). The train/test splits are different in each run of the experiment. These statistics are computed from the data used in the first run, and others are comparable.

genre	dataset	kind	top mode	frequency
responsory	full	train	8	20.85%
responsory	full	test	8	**21.13%**
responsory	subset	train	1	21.65%
responsory	subset	test	1	20.19%
antiphon	full	train	8	28.47%
antiphon	full	test	8	**28.13%**
antiphon	subset	train	1	23.50%
antiphon	subset	test	1	24.18%

Figure S3. **Majority baselines**. The frequency of the largest classes in each of the datasets. Boldfaced values correspond to the classification *accuracy* of the worst-performing conditions discussed in the main text. (The frequencies are marginally different in the five experimental runs; shown are the averages.)

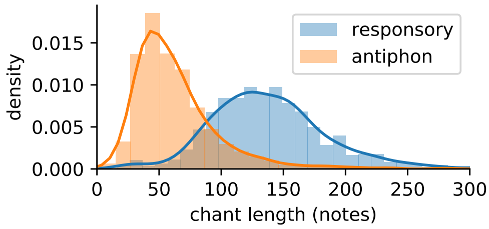


Figure S4. **Chant lengths in two genres.** Responsories are usually much longer than antiphons. The distribution is estimated from the training datasets without melody variants.

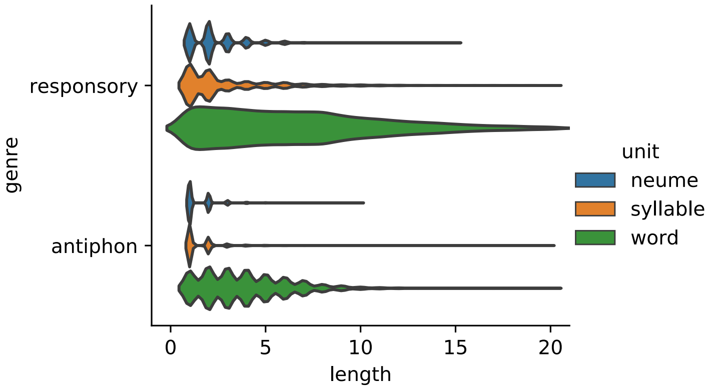


Figure S5. **Lengths of natural units.** Natural units have different lenghts in responsories and antiphons. Responsories are more *melismatic*: they use more notes per syllable. As a result, a typical word is also much longer. This is shown in the figure using violin plots, a visualization of the length distribution using a kernel density estimate. Note that the total area has no meaning in this plot; we normalized the widths of the violins for better readability. The distributions are estimated from the training datasets without melody variants).

	neume	syllable	word
antiphon	1.50	1.55	3.98
responsory	2.32	2.96	7.12

Figure S6. **Mean lengths of natural units.** Natural units have different lengths in responsories and antiphons, as the mean lenghts in number of notes shows. Figure S5 shows the full distribution. Means estimated from the training datasets without melody variants.

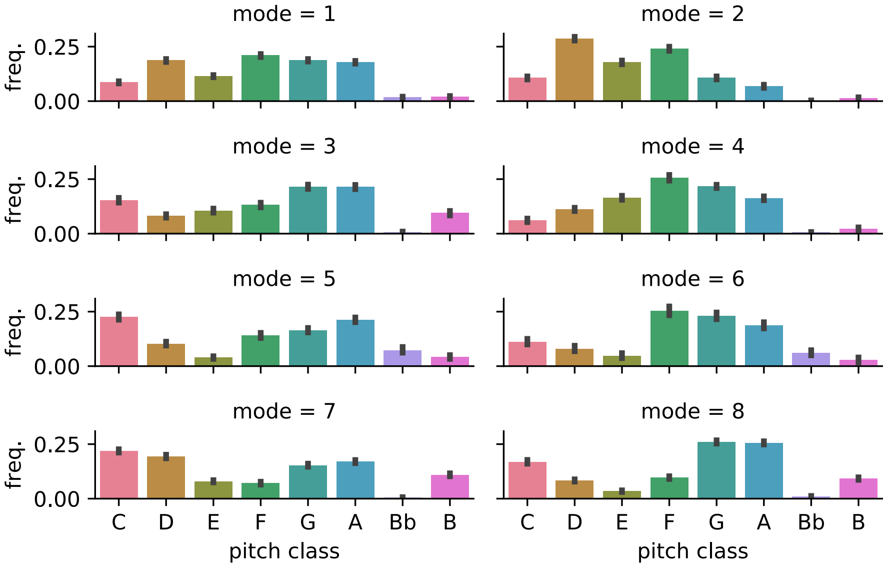


Figure S7. **Pitch class profiles** The pitch class profiles used in the profile approach (cf [12]). Shown are data for responsories, estimated from the training data without melody variants.

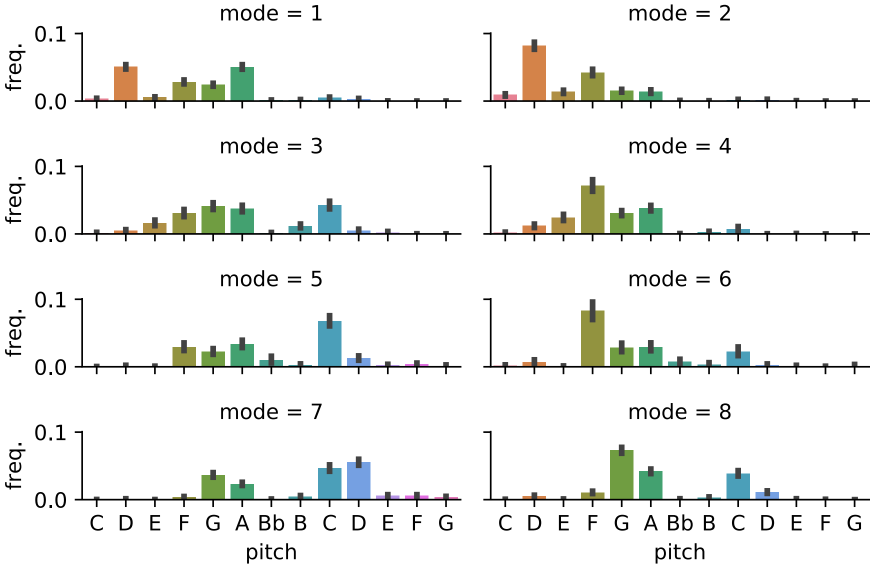


Figure S8. **Repetition profiles** The repetition profiles used in the profile approach. Every bar shows the average number of repetitions of that note in a chant (see main text for details). Shown are data for responsories, estimated from the training data without melody variants.

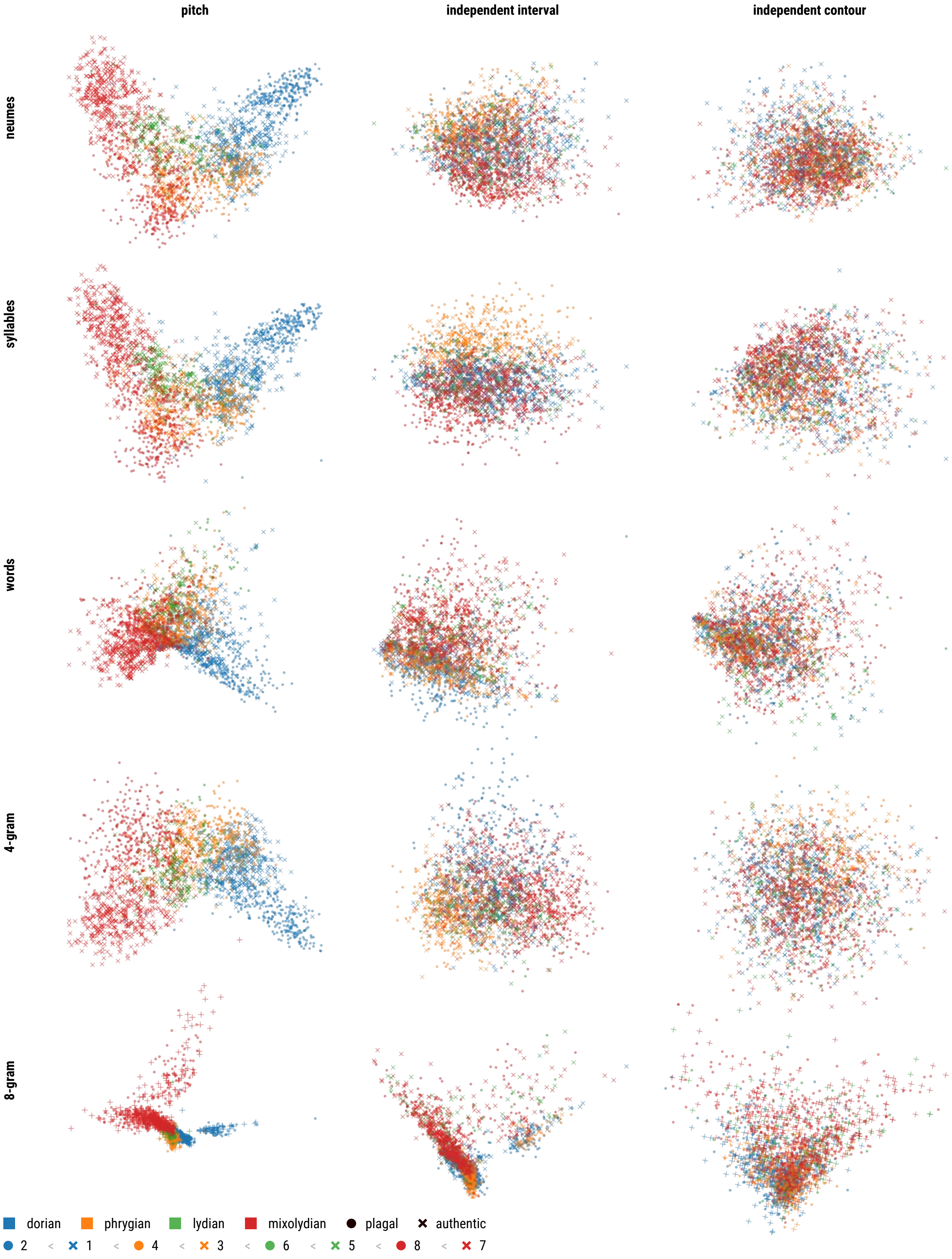


Figure S9. **PCA plots.** Two-dimensional representation of the high-dimensional feature space in several different conditions. A sample of TF–IDF vectors is shown, after reducing their dimensionality using a principal component projection.

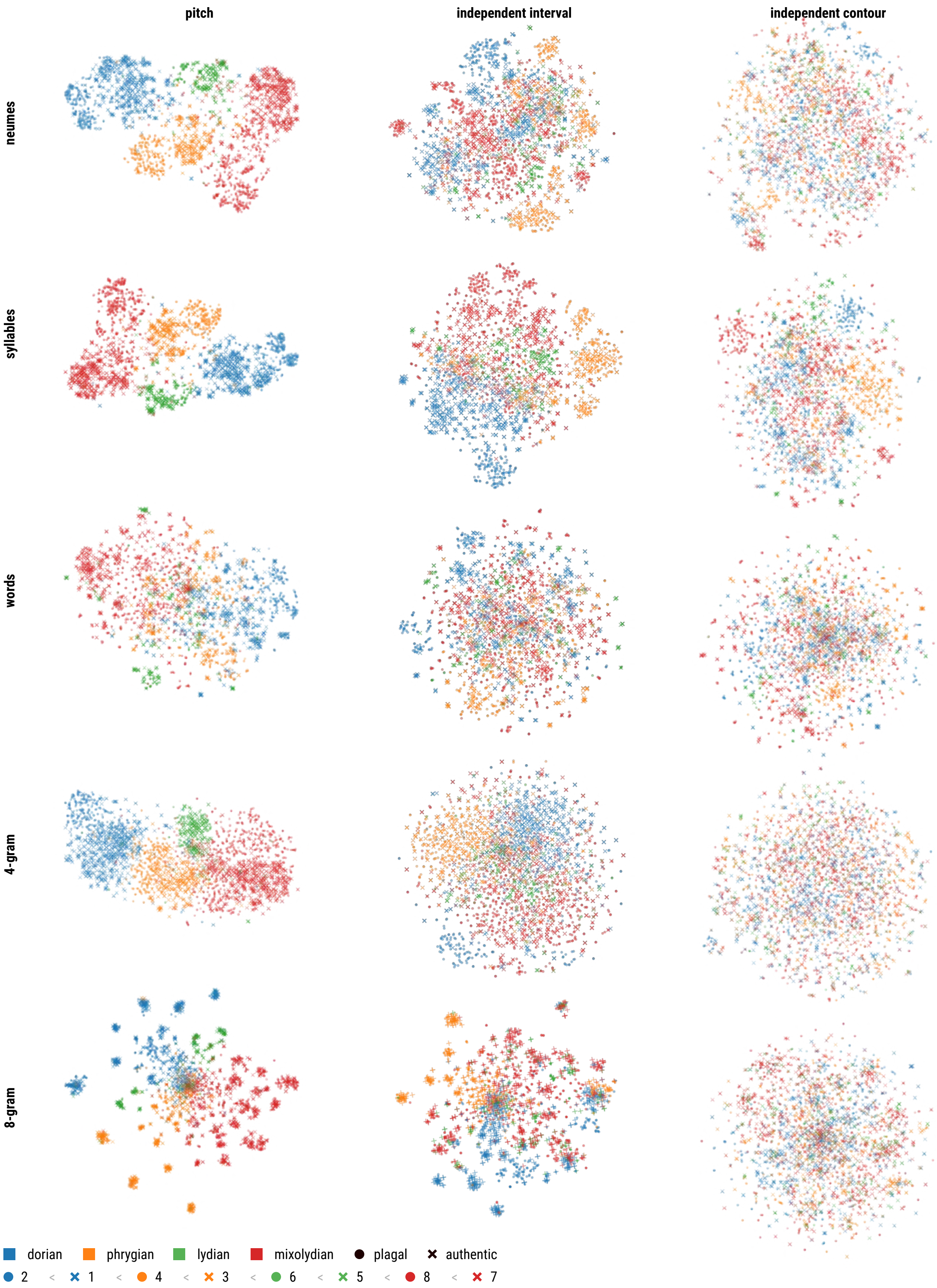
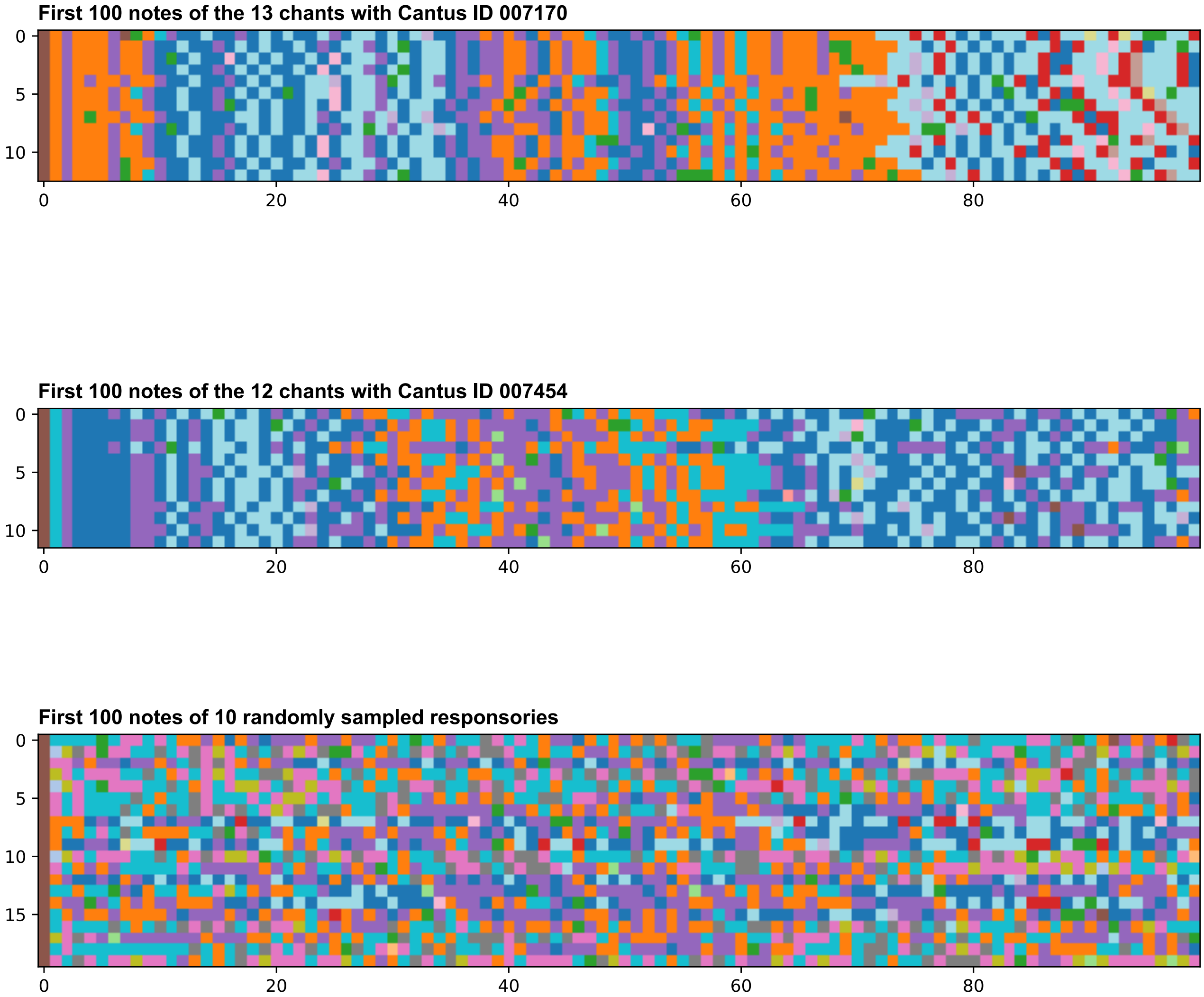


Figure S10. *t*-SNE **plots.** Two-dimensional representation of the high-dimensional feature space in several different conditions. A sample of TF–IDF vectors is shown, after reducing their dimensionality using *t*-SNE, a nonlinear dimensionality-reduction technique that maximizes the probability of mapping nearby points in high-dimensional space to nearby points in a lower dimensional space. The axes have no natural interpretation, but the graphs suggest that clusters are most clearly separated at the left, and mostly overlapping at the right.



First 100 notes of the 13 chants with Cantus ID 007170

First 100 notes of the 12 chants with Cantus ID 007454

First 100 notes of 10 randomly sampled responsories

Figure S11. **Melody variants in Cantus.** The top three panels show examples of sets of melody variants: the first 100 notes of melodies sharing a Cantus ID. Different colors correspond to different pitches, or more precisely, different Volpiano characters after discarding dashes. As a comparison, the bottom panel shows 100 notes of 20 random melodies.

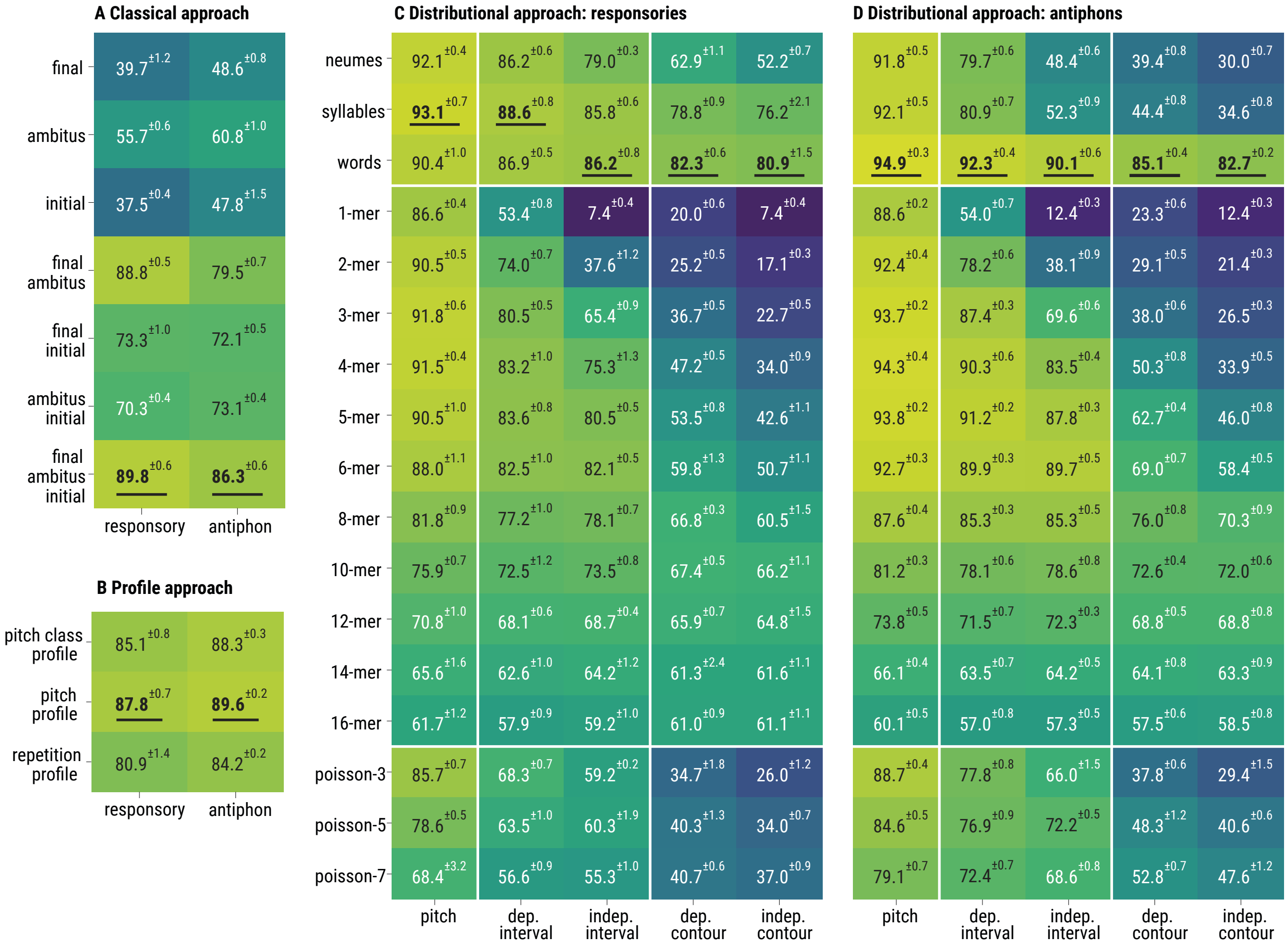
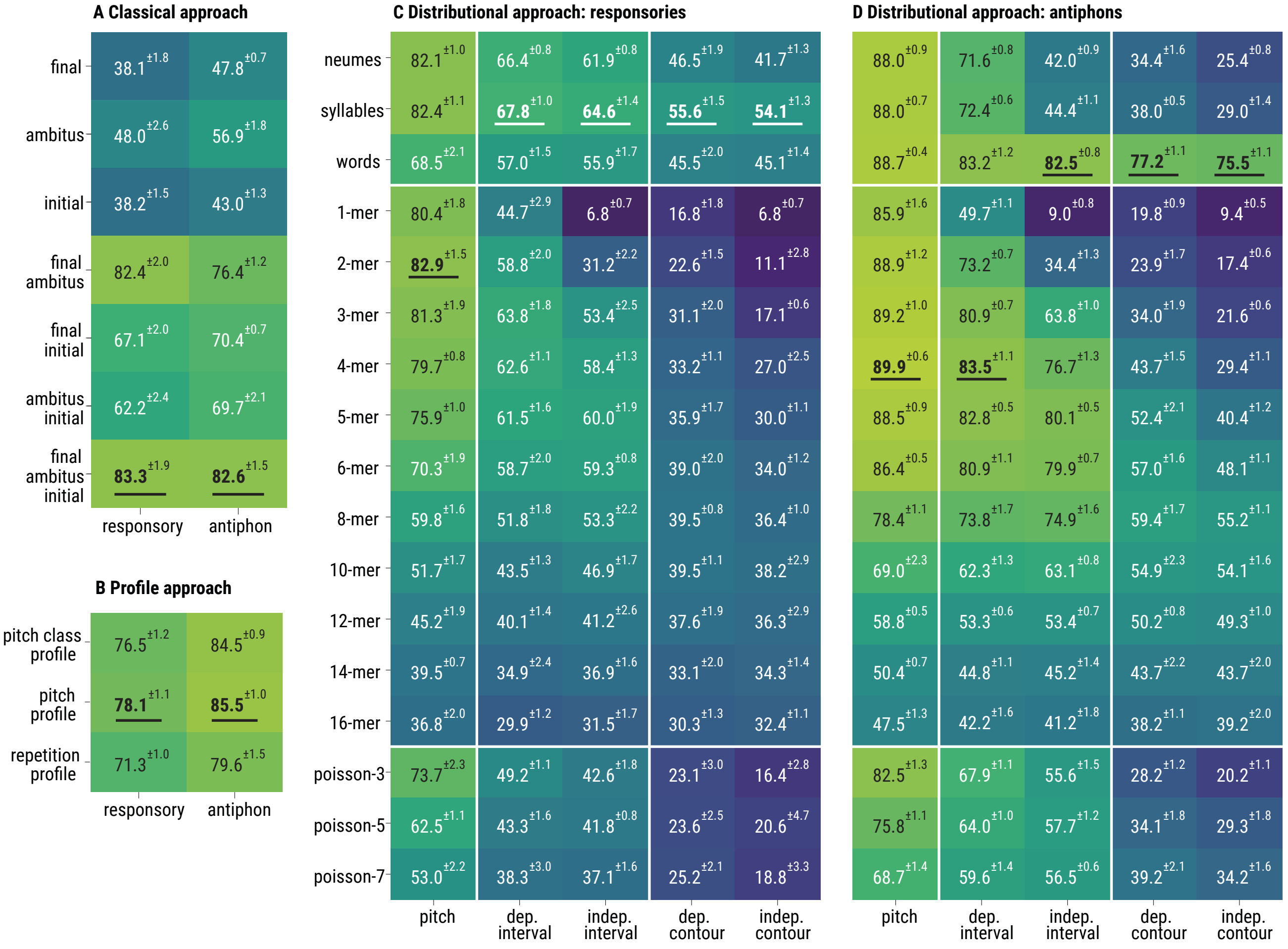


Figure S12. **Classification results with standard deviation.** This is essentially the same figure as figure 5 but now with the mean F_1-score μ and its standard deviation σ shown as $\mu^{\pm\sigma}$, computed from five independent runs of the experiment.



A Classical approach

	responsory	antiphon
final	38.1$^{\pm 1.8}$	47.8$^{\pm 0.7}$
ambitus	48.0$^{\pm 2.6}$	56.9$^{\pm 1.8}$
initial	38.2$^{\pm 1.5}$	43.0$^{\pm 1.3}$
final ambitus	82.4$^{\pm 2.0}$	76.4$^{\pm 1.2}$
final initial	67.1$^{\pm 2.0}$	70.4$^{\pm 0.7}$
ambitus initial	62.2$^{\pm 2.4}$	69.7$^{\pm 2.1}$
final ambitus initial	**83.3**$^{\pm 1.9}$	**82.6**$^{\pm 1.5}$

B Profile approach

	responsory	antiphon
pitch class profile	76.5$^{\pm 1.2}$	84.5$^{\pm 0.9}$
pitch profile	**78.1**$^{\pm 1.1}$	**85.5**$^{\pm 1.0}$
repetition profile	71.3$^{\pm 1.0}$	79.6$^{\pm 1.5}$

C Distributional approach: responsories

	pitch	dep. interval	indep. interval	dep. contour	indep. contour
neumes	82.1$^{\pm 1.0}$	66.4$^{\pm 0.8}$	61.9$^{\pm 0.8}$	46.5$^{\pm 1.9}$	41.7$^{\pm 1.3}$
syllables	82.4$^{\pm 1.1}$	**67.8**$^{\pm 1.0}$	**64.6**$^{\pm 1.4}$	**55.6**$^{\pm 1.5}$	**54.1**$^{\pm 1.3}$
words	68.5$^{\pm 2.1}$	57.0$^{\pm 1.5}$	55.9$^{\pm 1.7}$	45.5$^{\pm 2.0}$	45.1$^{\pm 1.4}$
1-mer	80.4$^{\pm 1.8}$	44.7$^{\pm 2.9}$	6.8$^{\pm 0.7}$	16.8$^{\pm 1.8}$	6.8$^{\pm 0.7}$
2-mer	**82.9**$^{\pm 1.5}$	58.8$^{\pm 2.0}$	31.2$^{\pm 2.2}$	22.6$^{\pm 1.5}$	11.1$^{\pm 2.8}$
3-mer	81.3$^{\pm 1.9}$	63.8$^{\pm 1.8}$	53.4$^{\pm 2.5}$	31.1$^{\pm 2.0}$	17.1$^{\pm 0.6}$
4-mer	79.7$^{\pm 0.8}$	62.6$^{\pm 1.1}$	58.4$^{\pm 1.3}$	33.2$^{\pm 1.1}$	27.0$^{\pm 2.5}$
5-mer	75.9$^{\pm 1.0}$	61.5$^{\pm 1.6}$	60.0$^{\pm 1.9}$	35.9$^{\pm 1.7}$	30.0$^{\pm 1.1}$
6-mer	70.3$^{\pm 1.9}$	58.7$^{\pm 2.0}$	59.3$^{\pm 0.8}$	39.0$^{\pm 2.0}$	34.0$^{\pm 1.2}$
8-mer	59.8$^{\pm 1.6}$	51.8$^{\pm 1.8}$	53.3$^{\pm 2.2}$	39.5$^{\pm 0.8}$	36.4$^{\pm 1.0}$
10-mer	51.7$^{\pm 1.7}$	43.5$^{\pm 1.3}$	46.9$^{\pm 1.7}$	39.5$^{\pm 1.1}$	38.2$^{\pm 2.9}$
12-mer	45.2$^{\pm 1.9}$	40.1$^{\pm 1.4}$	41.2$^{\pm 2.6}$	37.6$^{\pm 1.9}$	36.3$^{\pm 2.9}$
14-mer	39.5$^{\pm 0.7}$	34.9$^{\pm 2.4}$	36.9$^{\pm 1.6}$	33.1$^{\pm 2.0}$	34.3$^{\pm 1.4}$
16-mer	36.8$^{\pm 2.0}$	29.9$^{\pm 1.2}$	31.5$^{\pm 1.7}$	30.3$^{\pm 1.3}$	32.4$^{\pm 1.1}$
poisson-3	73.7$^{\pm 2.3}$	49.2$^{\pm 1.1}$	42.6$^{\pm 1.8}$	23.1$^{\pm 3.0}$	16.4$^{\pm 2.8}$
poisson-5	62.5$^{\pm 1.1}$	43.3$^{\pm 1.6}$	41.8$^{\pm 0.8}$	23.6$^{\pm 2.5}$	20.6$^{\pm 4.7}$
poisson-7	53.0$^{\pm 2.2}$	38.3$^{\pm 3.0}$	37.1$^{\pm 1.6}$	25.2$^{\pm 2.1}$	18.8$^{\pm 3.3}$

D Distributional approach: antiphons

	pitch	dep. interval	indep. interval	dep. contour	indep. contour
neumes	88.0$^{\pm 0.9}$	71.6$^{\pm 0.8}$	42.0$^{\pm 0.9}$	34.4$^{\pm 1.6}$	25.4$^{\pm 0.8}$
syllables	88.0$^{\pm 0.7}$	72.4$^{\pm 0.6}$	44.4$^{\pm 1.1}$	38.0$^{\pm 0.5}$	29.0$^{\pm 1.4}$
words	88.7$^{\pm 0.4}$	83.2$^{\pm 1.2}$	**82.5**$^{\pm 0.8}$	**77.2**$^{\pm 1.1}$	**75.5**$^{\pm 1.1}$
1-mer	85.9$^{\pm 1.6}$	49.7$^{\pm 1.1}$	9.0$^{\pm 0.8}$	19.8$^{\pm 0.9}$	9.4$^{\pm 0.5}$
2-mer	88.9$^{\pm 1.2}$	73.2$^{\pm 0.7}$	34.4$^{\pm 1.3}$	23.9$^{\pm 1.7}$	17.4$^{\pm 0.6}$
3-mer	89.2$^{\pm 1.0}$	80.9$^{\pm 0.7}$	63.8$^{\pm 1.0}$	34.0$^{\pm 1.9}$	21.6$^{\pm 0.6}$
4-mer	**89.9**$^{\pm 0.6}$	**83.5**$^{\pm 1.1}$	76.7$^{\pm 1.3}$	43.7$^{\pm 1.5}$	29.4$^{\pm 1.1}$
5-mer	88.5$^{\pm 0.9}$	82.8$^{\pm 0.5}$	80.1$^{\pm 0.5}$	52.4$^{\pm 2.1}$	40.4$^{\pm 1.2}$
6-mer	86.4$^{\pm 0.5}$	80.9$^{\pm 1.1}$	79.9$^{\pm 0.7}$	57.0$^{\pm 1.6}$	48.1$^{\pm 1.1}$
8-mer	78.4$^{\pm 1.1}$	73.8$^{\pm 1.7}$	74.9$^{\pm 1.6}$	59.4$^{\pm 1.7}$	55.2$^{\pm 1.1}$
10-mer	69.0$^{\pm 2.3}$	62.3$^{\pm 1.3}$	63.1$^{\pm 0.8}$	54.9$^{\pm 2.3}$	54.1$^{\pm 1.6}$
12-mer	58.8$^{\pm 0.5}$	53.3$^{\pm 0.6}$	53.4$^{\pm 0.7}$	50.2$^{\pm 0.8}$	49.3$^{\pm 1.0}$
14-mer	50.4$^{\pm 0.7}$	44.8$^{\pm 1.1}$	45.2$^{\pm 1.4}$	43.7$^{\pm 2.2}$	43.7$^{\pm 2.0}$
16-mer	47.5$^{\pm 1.3}$	42.2$^{\pm 1.6}$	41.2$^{\pm 1.8}$	38.2$^{\pm 1.1}$	39.2$^{\pm 2.0}$
poisson-3	82.5$^{\pm 1.3}$	67.9$^{\pm 1.1}$	55.6$^{\pm 1.5}$	28.2$^{\pm 1.2}$	20.2$^{\pm 1.1}$
poisson-5	75.8$^{\pm 1.1}$	64.0$^{\pm 1.0}$	57.7$^{\pm 1.2}$	34.1$^{\pm 1.8}$	29.3$^{\pm 1.8}$
poisson-7	68.7$^{\pm 1.4}$	59.6$^{\pm 1.4}$	56.5$^{\pm 0.6}$	39.2$^{\pm 2.1}$	34.2$^{\pm 1.6}$

Figure S13. **Main results on subset.** Cantus often contains several variants of the same melody, as shown in Figure S11. As discussed in the main text, this is a difficult issue that for example also applies to the Essen folk-song collection. We decided to repeat our experiments on a subset of the data where we excluded melody variants. We heuristically identified melody variants by randomly picking one chant from all sets of chants that have the same Cantus ID and mode. This resulted in a set of 1763 responsories and 4159 antiphons. This figure shows the main classification results on this subset of the data. Clearly, the performance of all models drops. The drop is greatest for responsories across models. The main result that only natural units maintain high performance, even on contour representations, nevertheless stand. It should be noted that in antiphons, some *n*-grams now outperform the natural units (by less then 2%) when using pitch and dependent interval representation. This does not reflect a large change in performance: in the main results, these *n*-grams are also deviated from top performance by no more than 2%. These results are further discussed in the main text.