GREGORIAN MELODY, MODALITY, AND MEMORY: SEGMENTING CHANT WITH BAYESIAN NONPARAMETRICS

Vojtěch Lanz **Jan Hajič jr.**
Charles University, Faculty of Mathematics and Physics
Institute of Formal and Applied Linguistics
{lanz,hajicj}@ufal.mff.cuni.cz

ABSTRACT

The idea that Gregorian melodies are constructed from some vocabulary of segments has long been a part of chant scholarship. This so-called "centonisation" theory has received much musicological criticism, but frequent re-use of certain melodic segments has been observed in chant melodies, and the intractable number of possible segmentations allowed the option that some undiscovered segmentation exists that will yet prove the value of centonisation, and recent empirical results have shown that segmentations can outperform music-theoretical features in mode classification. Inspired by the fact that Gregorian chant was memorised, we search for an optimal unsupervised segmentation of chant melody using nested hierarchical Pitman-Yor language models. The segmentation we find achieves state-of-the-art performance in mode classification. Modeling a monk memorising the melodies from one liturgical manuscript, we then find empirical evidence for the link between mode classification and memory efficiency, and observe more formulaic areas at the beginnings and ends of melodies corresponding to the practical role of modality in performance. However, the resulting segmentations themselves indicate that even such a memory-optimal segmentation is not what is understood as centonisation.

1. INTRODUCTION

Gregorian chant, the liturgical monody of the Latin church, has been a pillar of European musical identity since the Carolingian period in the 8th-9th century A.D. And yet we have no constructive theory of chant melody like we have e.g., tonality for classical composition. No good explanation exists for why Gregorian melodies *should* be the way they are, or how to write a plausible one from scratch. "They are not meant to be composed at all" would be an answer, had no Gregorian melodies been newly composed or heavily edited. However, new chants were composed [1, p. 463], for example, for new feasts [2]. Thus, the question remains: how is Gregorian melody structured?

1.1 Modality

The main theoretical framework for Gregorian melody is modality, which classifies melodies into eight basic modes 1–8, [1] also known as the church modes and today most often referenced by their Greek names: dorian, hypodorian, phrygian, etc. Modes are primarily identified by their *finalis* (the final note, typically d, e, f, or g) and their range (*authentic* or *plagal*). This "theoretical" definition is passed down from early medieval sources throughout the Middle Ages [3, 4, Ch. 1].

The medieval definition of modes says little about how Gregorian melody should be constructed. Modality involved more than just the beginnings, ends, and ranges of melodies: an evolving understanding of modality [4, ch. 6] led to revisions, most notably, the Cistercian order made extensive revisions based on their notion of modal "purity" [1, p. 610] [5, p. 72]. Tonaries [2] [3, 6] show that the repertoire often diverged from the "theoretical" definition. At least 1 in 10 antiphons and responsories would be misclassified using the "theoretical" features [7]. Some tonaries even disagree with each other [8]. Mode, therefore, goes beyond its medieval definition.

An attractive approach to modality is the idea that modes are hidden "vocabularies" of characteristic melodic segments. This follows from observations that many melodic gestures are re-used across different melodies [9–11]. Melodies within a mode tend to be similar [3, 12], suggesting these segments may be specific to modes. Text-based and other naive segmentations of chant melodies have been used for mode classification [7, 13], showing promising results on the CantusCorpus v0.2 dataset [14, 15] and outperforming "theoretical" approaches, as well as pitch profiles. Taken to the extreme, the theory of "centonisation" postulates entire melodies are constructed by concatenating mode-specific melodic segments [16, 17].

Centonisation has faced strong criticism in Gregorian chant scholarship [1, 18, p. 74–75]; however, this largely concerns its framing as a deliberate compositional strategy, not the recurrence of melodic material across chants [18].

1.2 Memory

Gregorian chant was originally an oral tradition [4, 18, 19], requiring monks to memorise thousands of melodies with-

[1] Leaving aside transposed modes etc.
[2] Manuscripts that explicitly organise melodies by mode.

out exact pitch notation for over 300 years. The challenge of memorisation led to the rapid adoption of staff notation, introduced by Guido of Arezzo in the 11th century as a teaching aid [1, 4, p. 388].

Memory constraints, along with cultural-evolutionary and information-theoretic principles, strongly support the emergence of centonisation in the melodies. Oral transmission is imperfect [20, 21], and melodies tend to evolve toward lower entropy [22]. Since melodies were rarely written the same way twice [23, 24], despite efforts to preserve them [1, p. 611], they changed during centuries of oral transmission before getting written down with exact pitch notation. Given the memorisation demands on singers, centonisation would be expected under the Minimum Description Length (MDL) principle. Supporting this, a recent study [21] found that oral transmission led to *"using fewer building blocks (intervals, contours) that are increasingly reused and combined"* over time, which is exactly the type of structure centonisation refers to.

Many monodic oral traditions show centonisation – for example, Arab-Andalusian [25] and Byzantine chant [26], and high formulaicity has been observed in Old Roman [27, 28] and Beneventan chant [29]. Melodic formulas appear in Hindustani Ragas [30], and we are certainly omitting tens, if not hundreds, of other such traditions here.

1.3 Outline

Since testing all segmentations is intractable, it remains possible that *some* segmentation could reveal that Gregorian chant is a "centonate" tradition.

In Section 3, we formalise chant memorisation as a computational segmentation problem, searching for an "optimally centonised" segmentation of chant by exploiting the connection between the MDL principle [31, 32] and Maximum a Posteriori (MAP) estimation [33, 34], and "borrowing" the nonparametric Bayesian Nested Hierarchical Pitman-Yor Language Model (NHPYLM) [35] to segment chant melody instead of natural language.

After briefly introducing the datasets we use (Section 4), using mode classification as a proxy for segmentation quality [7, 13], we show that the memorisation-based method outperforms existing segmentation baselines, both on pitch and interval representations (Section 5).

Imitating a monk learning from one manuscript (Section 6), we find that efficient memorisation relates to modality, though its influence varies across melody parts, depending on how they were performed in liturgy

We also contribute a Cython implementation of the NHPYLM model, and a class-conditioned version thereof. [3]

2. NHPYLM–RELATED WORK

In tasks such as vocabulary induction, phoneme-to-word mapping, and word segmentation of text without whitespace (e.g., Mandarin or Japanese), nonparametric Bayesian methods avoid manual model selection by learning model complexity from the data as part of inference. In this context, the Hierarchical Pitman-Yor Language Model (HPYLM) [35] was introduced. It has been widely applied to unsupervised word segmentation, from speech recognition [36, 37] to topic models [38]. Mochihashi et al. [39] extended this approach with the Nested HPYLM (NHPYLM), incorporating character-level priors from Character HPYLM.

NHPYLM has been adapted for melody segmentation, modelling motifs in tonal music [40]. The segmentation results were compared with the Generative Theory of Tonal Music, demonstrating the effectiveness of Bayesian nonparametric models for structured sequence learning, which we extend in the context of Gregorian chant.

3. MEMORY-EFFICIENT SEGMENTATION

Efficient memorisation of the Gregorian melodies can be formalised with the MDL principle [31, 32]: the best code H for data D is one that minimises $L(D|H) + L(H)$, where $L(D|H)$ is the length of the data encoded using the codebook H, and $L(H)$ is the length of the codebook. The MDL principle thus encourages codebooks such that the more frequently occurring and the longer a subsequence, the shorter code it gets to minimise $L(D|H)$, and it encourages small codebooks via the $L(H)$ term. Importantly for inferring memory-efficent segmentations, choosing a hypothesis using the MDL principle is equivalent to choosing the MAP hypothesis under a Bayesian model [33], where the prior probability of a hypothesis (corresponding to $L(H)$) decreases with its length [34].

To find an optimally memory-efficient segmentation, we need an unsupervised segmentation model that: (1) infers the vocabulary, including its size, since we want the vocabulary of chant segments to be a dependent variable following from efficient memorisation; (2) can model the sequential nature of chant melody (as one sings one segment after the other); (3) has a prior that prefers smaller segment vocabularies; and (4) has tractable inference and MAP estimation. These conditions are fulfilled by the Nested Hierarchical Pitman-Yor Language Model (NHPYLM) [35].

3.1 An intuition on the NHPYLM

The Pitman-Yor process (PY) is a nonparametric Bayesian model that assigns probabilities to categorical distributions without setting the number of categories in advance [41]. The intuition behind PY starts with the Chinese Restaurant Process (CRP): [4] the categorical distributions are normalised customer counts at tables serving a different dish each. The key property of the CRP is that the n-th customer chooses to sit at a given table proportionally to how many customers are already eating there, with a parameter α that leaves a (decreasing) chance $\alpha/(\alpha + n - 1)$ of sitting at a new table. This gives rise to a rich-get-richer behaviour, and a prior that strongly prefers fewer tables – in the case of language models, fewer vocabulary elements and thus shorter hypotheses under the MDL principle. PY

[3] Available at `https://github.com/lanzv/nhpylm`

[4] Also: Dirichlet Process.

is a generalisation of CRP that provides more control over the rich-get-richer behaviour, to better model long-tail distributions such as those found in natural languages [41]. [5]

In sequence segmentation, PY finds the optimal vocabulary under this rich-get-richer paradigm over unigram probabilities of the resulting segments. In order to model the sequence ordering with bigrams $P(s_i|s_{i-1})$ of the inferred segments, we must add a "restaurant of restaurants", which models the probability of having the history s_{i-1}. The "inner" restaurant then models the conditional probability of s_i, via a yet another hierarchy of restaurants over the character n-grams in s_i. This is, finally, the Nested Hierarchical Pitman-Yor Process [35, 39].

In our case, the "inner" HPYLM operates on tones, while the "outer" operates on segments. The base distribution for the segment-level HPYLM is obtained from the tone-level HPYLM.

3.2 Segmentation Probability with NHPYLM

To compute the probability of a given segmentation of a melody, we only need to evaluate the probability of each segment in its context under the NHPYLM and multiply them together. For a segment s given a context h, the "outer" (segment-level) model computes the probability recursively as:

$$p(s|h) = \frac{c(s|h) - d \cdot t_{hs}}{\theta + c(h)} + \frac{\theta + d \cdot t_h}{\theta + c(h)} \cdot p(s|h'), \quad (1)$$

where $c(s|h)$ is the number of customers (segment occurrences) for segment s with context h, $c(h)$ is the total number of customers with that context, t_{hs} is the number of tables serving segment s in context h, and t_h is the total number of tables with context h. The discount d and concentration θ are hyperparameters of the Pitman-Yor process. The term $p(s|h')$ refers to the probability of segment s with a shorter context h'. We use a bigram model, so the initial context h always consists of a single preceding segment, and h' is the empty context.

When no shorter context remains, the base distribution is the "inner" tone-level model, which computes the probability of a segment composed of tones $t_1, \ldots, t_k$ as:

$$p(t_1 \ldots t_k) = \frac{\prod_{i=1}^{k} p(t_i \mid t_1 \ldots t_{i-1})}{p(k)} \cdot Po(k \mid \lambda), \quad (2)$$

where $p(t_i \mid t_1 \ldots t_{i-1})$ is the tone-level probability computed by marginalizing over all possible context lengths n:

$$p(t \mid h) = \sum_{n=0}^{\infty} p(t \mid h, n) \cdot p(n \mid h), \quad (3)$$

Here, $p(t \mid h, n)$ is the probability of tone t given tone context h at depth n (i.e., the last n tones of the context are considered). This is computed using Equation 1, but applied at the tone level with a uniform base distribution over tones. The term $p(n \mid h)$ denotes the probability that context h has order n. $Po(k \mid \lambda)$ is the Poisson correction to prevent short segments from being over-favored [39]. Further mathematical details can be found in the original formulations of the model [35, 39].

The model is trained using blocked Gibbs sampling. Initially, each melody is assigned a random segmentation. In each sampling step, a randomly selected segmented melody is removed from the model, all its possible segmentations are evaluated with eqs. (1)–(3) using the current state of the model, and a new segmentation is sampled from them. The model is then updated with the newly sampled segmentation. During inference, optimal segmentation is computed using the Viterbi algorithm.

3.3 Mode-specific segmentation

Each mode may have its own characteristic melodic units and segmentation patterns [16, 17]. We extend the NHPYLM model to account for this by training separate segmentation models for each mode, as though one were learning the repertoire in 8 separate sub-corpora.

At inference time (i.e., when analyzing new chants), the mode of a chant c is unknown. However, we can use all eight mode-specific models (each estimating $p(\bar{c} \mid m)$, the likelihood of the optimal segmentation $\bar{c}$ under mode m) to determine the most probable mode $m^* = \arg\max_m p(m \mid c)$ given the melody c, using Bayes' Theorem. [6] We refer to the combined model as *NHPYLMClasses*.

3.4 From Melody Segments to Modes

Given the interaction between mode, melodic similarity, and memorisation [13], mode classification based on inferred segments can serve as a proxy for segmentation quality. However, aside from the NHPYLMClasses model, this "amount of information retained about mode" must be obtained in a separate downstream step.

To ensure comparability with previous work, we adopt the same mode classification pipeline used in earlier studies of chant segmentation and modality [7, 13]. Each melody is represented as a bag-of-segments vector. Although sequential information is not explicitly encoded, segments are inferred using a model that considers sequence context, so some information about order is implicitly preserved. TF-IDF weighting is applied to the vectors.

Following settings used in previous work on chant segmentation [7], only the 5000 most frequent segments are retained, with others discarded. The resulting vectorized segmentations from the training melodies are then used to train a Linear SVM classifier to predict chant modality.

4. DATASETS

We use the **CantusCorpus v0.2** dataset [15], derived from the Cantus Database [14], the most extensive digital collection of Gregorian chants. We focus on the two most abundant genres: antiphons and responsories. Antiphons

[5] PY differs from CRP by applying a constant discount to table counts, slowing the decay of the "density budget" for new tables.

[6] Mathematical and implementation details are available at: `https://github.com/lanzv/chant-modality-with-nhpylms`

CantusCorpus v0.2 data	Pitches		Intervals	
	Ant.	Res.	Ant.	Res.
Classical approach	*89.6*	*89.4*	–	–
4-gram	91.0	91.6	82.0	83.1
Syllables	89.3	***93.5***	72.9	89.3
Words	90.1	90.2	83.9	87.2
NHPYLM	91.7	93.3	86.7	89.7
NHPYLM-Cl, no SVM	***92.6***	***93.8***	**90.4**	**92.4**
NHPYLM-Cl	**92.7**	**93.9**	***90.2***	**92.4**
Overlapping n-grams, 1-7	*93.8*	*94.8*	*90.4*	*92.8*

Table 1: F1 scores for various mode classification methods applied to all antiphon and responsory melodies, encoded as both sequences of pitches and sequences of intervals. St.dev. across the 5 splits was between 0.001 and 0.005, so differences of an F1-score of 1.0 is "safe" at 1.96σ.

are simpler and shorter melodies, while responsories are more complex and longer. We apply the same preprocessing steps as used in previous work [7], keeping only completely transcribed melodies with simple mode annotations (1–8), and discarding non-pitch characters from Volpiano encoding. Crucially, we also remove *differentiae* from antiphons [13]. After the described filtration process [7, 13], a total of 13551 antiphons and 7031 responsories remain. Because responsories are longer (137.38 pitches on average vs. 53.97 for antiphons), the responsory dataset is longer, at 966k notes, with antiphons at 731k notes.

The CantusCorpus also contains multiple versions of the same melody if it is found in multiple transcribed sources. Although these are almost never identical [23,24], this still means closely related melodies could end up in both the test and training sets. While we could ensure this does not happen using the CantusID mechanism [14], the abundances of different melodies would still distort results if present in the test set, and distort optimisation if present in the training set. We propose using melodies from a single source. This resembles the situation of any chant practitioner: they would be expected to learn their local repertoire. We chose the largest available manuscript, **D-KA Aug. LX**,[7] , with 1965 antiphons and 907 responsories.

5. MODE CLASSIFICATION EXPERIMENTS

In this section, we describe experiments on the CantusCorpus dataset. Both NHPYLM-based methods are configured with segment lengths ranging from 1 to 7 tones using default hyperparameters: $d_0 = 0.5$; max. segment length of 7; initial $\theta = 2.0$ for inner hyperparameter updates [39]; Gamma priors on both d and θ with $\alpha = 1.0, \beta = 1.0$; and a Gamma prior for the Poison correction parameter with initial $\alpha_\lambda = 6.0, \beta_\lambda = 1.2$.[8] 10% of the training data is reserved as a validation set for checking for convergence for NHPYLM-based models.

Because relative pitch has greater saliency than absolute pitch [42, p.56,p.100], in addition to pitch sequences from cleaned CantusCorpus melodies, we also use sequences of intervals (which thus have a length of $n - 1$).

For all experiments, we use a 7:3 train-test split for both NHPYLM-based models and the SVM classifier. In the case of NHPYLM and NHPYLMClasses models, training data is used to learn the optimal probability distribution, and the resulting segmented training data is then used to train the SVM classifier. (This is why the train-test split is necessary also for the unsupervised methods.) During evaluation, the test set is first segmented by the NHPYLM models, after which the SVM classifier predicts the modes.

5.1 Evaluation

We evaluate how well the proposed segmentation methods retain modal information by using inferred segments as features in a mode classification task. This approach is musicologically justified, given the relationship between modality and memory [4, 13, ch. 3], and has been used before [7]. We report micro-averaged F1-scores in Tables 1 and 2. For NHPYLMClasses, we also report the model's "internal" classification performance, i.e., accuracy based on mode $m^* = \arg\max_m p(m \mid c)$ selected during inference, alongside SVM results using the inferred segments.

Baselines. Previous work [7] segmented chants using n-grams, neumes, syllables, or words. We adopt their best-performing segmentations as baselines: 4-grams, syllables, and words. We also include the *Classical approach* as a baseline, which classifies modes based on initial and final tones along with the melody range [7].

Upper bound. We do not know what the maximum achievable mode-classification performance over any segmentation is. The mode to which a melody belongs may not be solely determined by the melody itself: some melodies are assigned to different modes in different tonaries [8]. We at least make a rough estimate of the upper bound of mode classification performance that can be achieved using a "distributional approach" [7] and the SVM classifier as a reasonable "information extraction" black box. In this setting, we use all possible overlapping n-grams of lengths 1 to 7 as features. Thus, more information is available to the classifier than for any segmentation (with TF-IDF preprocessing to limit the effect of many uninformative features). We refer to this approach as *Overlapping n-grams, 1-7*. Already overlapping 4-grams outperformed all previous results [13].

Cross-validation. Each experiment was repeated five times with different random seeds for a 70-30 random split over the entire dataset. While the split is not explicitly stratified by mode, averaging results over multiple random splits helps mitigate potential imbalances and ensures more robust performance estimates.

5.2 Results

Mode classification for CantusCorpus antiphons and responsories is shown in Table 1. Two representations are used: sequences of pitches and sequences of intervals.

Across all four settings, NHPYLMClasses outperformed all baselines, though the syllable baseline on re-

[7] https://cantusdatabase.org/source/123612

[8] See also Supplementary materials – NHPYLM implementation.

D-KA Aug. LX data (Single manuscript)	Pitches		Intervals	
	Ant.	Res.	Ant.	Res.
Classical approach	***85.5***	*81.9*	–	–
4-gram	82.1	77.8	70.7	64.5
Syllables	83.3	82.8	61.0	71.7
Words	78.8	70.4	64.1	61.9
NHPYLM	***86.0***	***83.6***	73.3	72.6
NHPYLM-Cl, no SVM	85.3	**84.0**	**80.2**	**78.6**
NHPYLM-Cl	**86.1**	***83.6***	***79.9***	***78.5***
Overlapping n-grams, 1-7	*87.0*	*86.9*	*81.1*	*76.8*

Table 2: Mode classification scores for antiphon and responsory melodies from the manuscript D-KA Aug. LX, encoded both as sequences of pitches and intervals.

sponsories with absolute pitches comes close. Notably, the SVM showed no improvement over the model's internal $p(m|c)$, in line with the theory that modes serve to organise repertoire also in memory [4, ch. 3]. The base NHPYLM model was slightly worse, but did outperform all baselines except for syllables on responsories with absolute pitches (though again barely), further suggesting that despite having more data to estimate a good representation, conditioning on mode would have been a more effective principle for organising repertoire in memory. NHPYLMClasses also comes very close to the "upper bound" overlapping n-gram performance, suggesting that there may not be much room for improvement on mode classification.

6. REMEMBERING A MANUSCRIPT

We now consider a computational scenario inspired by a monk needing to memorise melodies for the liturgical environment he belongs to. In our case, this would be in late medieval Zwiefalten, documented by the liturgical book D-KA Aug. LX. Suppose the monk has already learned 70% of the melodies (training set): how difficult is it then to memorise the remaining ones?

6.1 Mode classification in a single manuscript

Table 2 shows that segmentation based on NHPYLMClasses is again the most accurate model for mode classification and, on responsories with interval representations, outperforms the "cheating" overlapping n-grams features, with its internal mode classifier performing slightly better than an SVM on top of the resulting segmentations.

The drop in performance compared to experiments on the full CantusCorpus may have two main reasons: first, it is a much smaller dataset, and second, only very few Cantus IDs have multiple instances in one liturgical book, so it eliminates issues with multiple versions of a melody randomly assigned to both the training and test sets. To isolate these influences, we ran two separate experiments: (1) assigning all instances of a Cantus ID only to one of the train/dev/test sets; (2) uniformly subsampling CantusCorpus to the size of D-KA Aug. LX. In experiment (1), 3% of performance was lost, and in (2), 4%. This corresponds surprisingly well to the approximately 7% drop overall (with pitch representation). [9]

Perplexity on D-KA Aug. LX	Antiphons	Resps.
NHPYLM—*intervals*	20.0	17.7
NHPYLMClasses—*intervals*	16.1	14.3
NHPYLM—*pitches*	15.4	13.5
NHPYLMClasses—*pitches*	**11.8**	**9.9**

Table 3: Perplexities for PY-based segmentations on pitch and interval melody encodings on D-KA Aug. LX.

6.2 Perplexity

We have so far been measuring segmentation quality indirectly, via mode classification. However, we can also measure its memorisation efficiency directly through *perplexity*, a common metric in language modelling [43] that follows directly from Shannon's coding theorem, and so is a measure of compression (lower is better). Intuitively, it reflects the model's average uncertainty - the number of equally likely choices it has at each step. It is defined as:

$$\text{perplexity}(s_1 \dots s_N) = 2^{-\frac{1}{N}\sum_{i=1}^{N} \log_2 p(s_i|h_i)},$$

where s_i is the i-th segment of the given segmentation and h_i is its context.

Table 3 compares the perplexity of the models. NHPYLMClasses leads to more efficient memorisation of melody than the standard NHPYLM method. [10] Interestingly, antiphons appear to be more challenging for effective segmentation than responsories. This is quite in line with the difficulty in identifying melodic families in antiphons [3, 12], and conversely the fact that melodic formulas were first described in responsories [1, 44, 45].

Additionally, pitch representation appears easier to remember compared to interval representation, even though their state space is larger. This may be because in the interval representation, the same segment can occur at different positions in the pitch system (f-g-f-f equivalent to d-e-d-d), so the conditional entropy of its continuation cannot decrease as much. It is also in line with performance practice: singers should have been always aware of which pitch within the system they were singing, as evidenced by prevalent techniques such as the Guidonian hand [4, ch. 6] [1, p.469] and solmisation [1, pp.467–468].

We find a correlation between perplexity values in Table 3 and the corresponding mode classification scores in Table 2 of -0.77 (and over -0.88 within the genres individually), offering initial empirical support for the hypothesised relationship between modality and efficient memorisation, consistent with historical perspectives [13].

6.3 Modal Identity and Memorization in Segments

Is modality, formulaicity, and thus "centonisability" more prominent in some parts of the melody than others, and does it relate to mode? This question relates to mode

[9] See `github.com/lanzv/chant-modality-with-nhpylms`.

[10] Though under MDL one should penalize it for having 8 vocabularies instead of a single shared one.

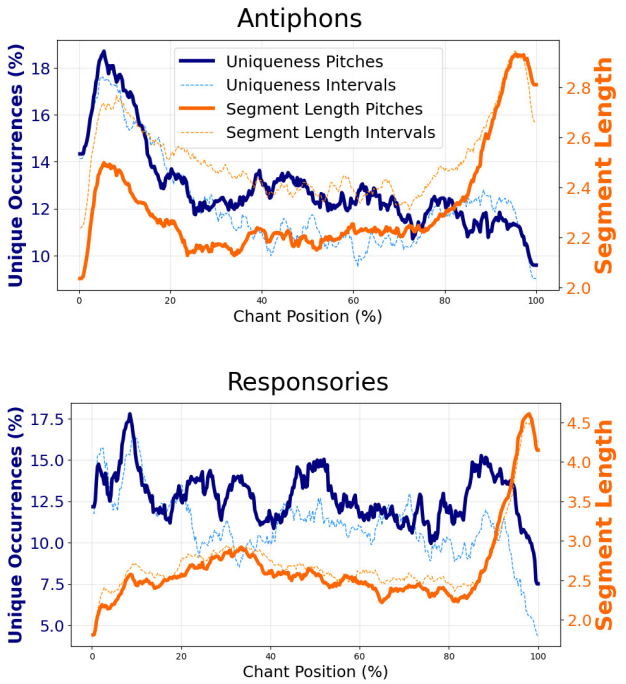

Figure 1: Distribution of average segment length (orange; values right) and average segment modal uniqueness (blue; values left) across relative positions in melodies. Note the difference in segment length scales – responsories are more formulaic. The legend in the antiphon plot applies to both.

and performance practice. An antiphon was sung before and after a psalm, and psalms were sung on one of just a few simple psalm tones highly specific to individual modes. So, it would have been appropriate for the end (1st antiphon performance) and beginning (2nd antiphon performance) to have a "compatible" melody at its interfaces with the psalm tone to which it was assigned (as evidenced, again, by tonaries). For the responsories, instead of a psalm there is a responsory verse, which still has relatively few melodies for each mode but more than psalm tones, and after the verse often only the second half of the responsory (the "respond") would be sung.

We plot the average segment length in which a note at a given relative position in the chant melody participates in Figure 1 (orange). High average segment length indicate more formulaic melodies. For antiphons, we see a small spike in segment length at both ends; for responsories, we see a more prominent spike at the end, corresponding well to the respective performance contexts.

The average uniqueness of the discovered segments to individual modes (Figure 1, in blue) is not as structured. For the antiphons, we see a spike in uniquess at the beginning, which corresponds to singing the antiphon after the psalm, but at the end of the antiphon (corresponding to the start of the psalm, and highest formulaicity) modal uniqueness drops yet lower. However, psalm tones for some mode pairs have an identical or near-identical start [1, p. 59–60], so the performance context makes unique formulas less likely.

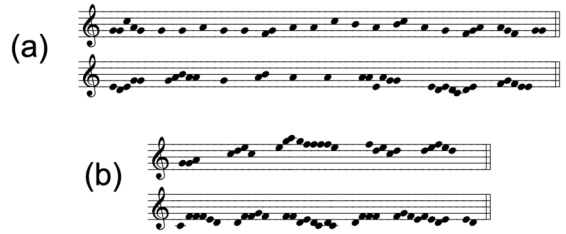

Figure 2: Comparing segmentation patterns: (a) two outputs of our model—a typical case and a rarer, more "centonised" one; (b) segmentation proposed by Levy [17].

7. CONCLUSIONS

Building on the memorisation of Gregorian chant, we take an information-theoretic approach and implement NHPYLMs for unsupervised melodic segmentation. The resulting segmentations achieve state-of-the-art mode classification on CantusCorpus v0.2, and in a "monk simulation" using D-KA Aug. LX, also reveal how performance contexts shape formulaicity and modal identity.

The correlation tentatively found between perplexity, as "ease of memorisation", and mode classification F1-score is initial empirical evidence for the relationship between melody segmentation and modality, in accordance with the music-historical argument for this link [4, 7, 13]. Pilot experiments with other melody encodings [11] suggest this holds across conditions, warranting further study.

Memory efficiency of baseline segmentations could be approximated with a bigram language model with Interpolated Kneser-Ney smoothing [46] trained on the optimal segmentations. KN smoothing is closely related to the PY process [35]. Pilot results showed its perplexity differed by no more than 0.5 from NHPYLM's on D-KA Aug. LX.

NHPYLMs can provide memory-based segmentations of other repertoires – both to detect centonisation where it is musicologically known to exist [25,26,29], and to enable information-theoretic comparisons with Gregorian chant. However, suitable datasets are needed first.

However, despite all the reasons to believe Gregorian chant could be centonised, and despite the empirical indications of a close relationship between segmentation, memory, and modality, the discovered optimal segments are far from a convincing centonisation of chant. Comparing a typical model output, and an output with maximum formulaicity, with centonisation proposed in chant scholarship in Figure 2, we see that the discovered segmentation patterns are far more fragmentary than what has been proposed for Gregorian chant [17] and what is known from Byzantine formulae [26]. [12] This opens the intriguing possibility that institutional rules and practices were able to significantly counteract processes inherent in oral transmission, making the evolutionary processes of Gregorian chant different from other musical traditions. The mystery of constructing Gregorian melodies remains unresolved.

[11] E.g., collapsing repeated notes – see results in the Supplementaries.
[12] Full segmentation results on D-KA Aug. LX are available in Supplementary materials.

8. ACKNOWLEDGMENTS

The authors are supported by the Social Sciences and Humanities Research Council of Canada by the grant no. 895-2023-1002, Digital Analysis of Chant Transmission, and by the SVV project number 260 698. The second author additionally acknowledges the support by the project "Human-centred AI for a Sustainable and Adaptive Society" (reg. no.: CZ.02.01.01/00/23_025/0008691), cofunded by the European Union. The computing infrastructure is provided by the LINDAT/CLARIAH-CZ Research Infrastructure (https://lindat.cz), supported by the Ministry of Education, Youth and Sports of the Czech Republic (Project No. LM2023062).

9. ETHICS STATEMENT

This study did not involve human participants, personal data, or sensitive information. All data consists of transcriptions of publicly available historical sources. No concerns related to privacy, consent, or bias apply. We, however, note that we aim to contribute to the understanding of chant without reinforcing any cultural or religious biases; our work in no way implies that Gregorian chant having different transmission patterns would be a value judgment.

10. DATA AND CODE ACCESSIBILITY

The CantusCorpus v0.2 dataset [15] that we used is available at: `https://github.com/bacor/cantuscorpus/releases/tag/v0.2` The NHPYLM model code is available at: `https://github.com/lanzv/nhpylm`. The experiment code, including more detailed results, is available at: `https://github.com/lanzv/chant-modality-with-nhpylms`.

11. REFERENCES

[1] D. Hiley, *Western plainchant: a handbook*. Oxford, United Kingdom: Clarendon Press, 1993.

[2] R. Hallas, *Two rhymed offices composed for the feast of the Visitation of the Blessed Virgin Mary: comparative study and critical edition*. Bangor University (United Kingdom), 2021.

[3] P. Merkley, "Tonaries and melodic families of antiphons," *Journal of the Plainsong and Mediaeval Music Society*, vol. 11, p. 13–24, Jan. 1988. [Online]. Available: http://dx.doi.org/10.1017/S0143491800001136

[4] C. M. Atkinson, *The critical nexus: tone-system, mode, and notation in early medieval music*. Oxford University Press, 2008.

[5] J. Glasenapp, *To Pray without Ceasing: A Diachronic History of Cistercian Chant in the Beaupré Antiphoner (Baltimore, Walters Art Museum, W. 759–762)*. Columbia University, 2020.

[6] M. Huglo, "Les tonaires," *Inventaire, Analyse, Comparaison. Paris: Société française de musicology*, pp. 132–40, 1971.

[7] B. Cornelissen, W. H. Zuidema, and J. A. Burgoyne, "Mode classification and natural units in plainchant." in *Proceedings of the 21st Int. Society for Music Information Retrieval Conf.*, Montreal, Canada, 2020, pp. 869–875.

[8] H. Mori, *Conflicting modal assignments of office antiphons: A comparative study of seven Germanic sources.* National Library of Canada= Bibliotheque nationale du Canada, Ottawa, 2001.

[9] K. Helsen, "The use of melodic formulas in responsories: constancy and variability in the manuscript tradition," *Plainsong & Medieval Music*, vol. 18, no. 1, pp. 61–76, 2009.

[10] T. Karp, *Aspects of orality and formularity in Gregorian chant*. Northwestern University Press, 1998.

[11] K. Helsen, M. Daley, and J. Schindler, "The sticky riff: Quantifying the melodic identities of medieval modes," *Empirical Musicology Review*, vol. 16, no. 2, p. 312–325, Mar. 2023. [Online]. Available: http://dx.doi.org/10.18061/emr.v16i2.7357

[12] F.-A. Gevaert, *La Mélopée Antique dans le Chant de l'Église Latine*. Ad. Hoste, 1895.

[13] V. Lanz and J. Hajič, "Text boundaries do not provide a better segmentation of gregorian antiphons," in *Proceedings of the 10th International Conference on Digital Libraries for Musicology*, 2023, pp. 72–76.

[14] D. Lacoste, "The Cantus Database and Cantus Index Network," in *The Oxford Handbook of Music and Corpus Studies*. Oxford University Press, 2022. [Online]. Available: https://doi.org/10.1093/oxfordhb/9780190945442.013.18

[15] B. Cornelissen, W. Zuidema, and J. A. Burgoyne, "Studying large plainchant corpora using chant21," in *7th International Conference on Digital Libraries for Musicology*, 2020, pp. 40–44.

[16] P. Ferretti, *Estetica gregoriana: ossia, Trattato delle forme musicali del canto gregoriano*, ser. Estetica gregoriana: ossia, Trattato delle forme musicali del canto gregoriano. Pontificio istituto di musica sacra, 1934, no. sv. 1. [Online]. Available: https://books.google.cz/books?id=vOWCnQEACAAJ

[17] K. Levy, "The italian neophytes' chants," *Journal of the American Musicological Society*, vol. 23, no. 2, pp. 181–227, 1970.

[18] L. Treitler, "Centonate" chant:" übles flickwerk" or" e pluribus unus?" *Journal of the American Musicological Society*, vol. 28, no. 1, pp. 1–23, 1975.

[19] H. Hucke, "Toward a new historical view of gregorian chant," *Journal of the American Musicological Society*, vol. 33, no. 3, pp. 437–467, 1980.

[20] D. Shanahan and J. Albrecht, "Examining the effect of oral transmission on folksongs," *Music Perception: An Interdisciplinary Journal*, vol. 36, no. 3, pp. 273–288, 2019.

[21] M. Anglada-Tort, P. M. Harrison, and N. Jacoby, "Studying the effect of oral transmission on melodic structure using online iterated singing experiments," *bioRxiv*, pp. 2022–05, 2022.

[22] T. Popescu and M. Rohrmeier, "Core principles of melodic organisation emerge from transmission chains with random melodies," *Evolution and Human Behavior*, vol. 45, no. 6, p. 106619, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1090513824000953

[23] D. J. Froger, "The critical edition of the roman gradual by the monks of solesmes," *Journal of the Plainsong & Mediaeval Music Society*, vol. 1, pp. 81–97, 1978.

[24] J. Hajič jr., G. A. Ballen, K. H. Mühlová, and H. Vlhová-Wörner, "Towards Building a Phylogeny of Gregorian Chant Melodies," in *Proceedings of the 24th International Society for Music Information Retrieval Conference*. ISMIR, Dec. 2023, pp. 571–578. [Online]. Available: https://doi.org/10.5281/zenodo.10340442

[25] T. Nuttall, M. García Casado, V. Núñez Tarifa, R. Caro Repetto, and X. Serra, "Contributing to new musicological theories with computational methods: the case of centonization in Arab-Andalusian music," in *20th Conference of the International Society for Music Information Retrieval (ISMIR 2019): 2019 Nov 4-8; Delft, The Netherlands.[Canada]: ISMIR; 2019. p. 223-8.* International Society for Music Information Retrieval (ISMIR), 2019.

[26] E. Wellesz, "Words and music in byzantine liturgy," *The Musical Quarterly*, vol. 33, no. 3, pp. 297–310, 1947.

[27] T. H. Connolly, "Introits and archetypes: Some archaisms of the old roman chant," *Journal of the American Musicological Society*, vol. 25, no. 2, pp. 157–174, 1972.

[28] J. Dyer, "Tropis semper variantibus: Compositional strategies in the offertories of old roman chant," *Early Music History*, vol. 17, pp. 1–60, 1998.

[29] M. Huglo, "The old beneventan chant," *Studia Musicologica Academiae Scientiarum Hungaricae*, vol. 27, no. Fasc. 1/4, pp. 83–95, 1985.

[30] S. Chakraborty, S. Tewari, A. Rahman, M. Jamal, A. Lipi, A. Chakraborty, A. Nanda, and P. Shukla, *Hindustani classical music: a historical and computational study*. Sanctum Books, 2021.

[31] J. Rissanen, *Stochastic complexity in statistical inquiry*. World scientific, 1998, vol. 15.

[32] P. D. Grünwald, *The minimum description length principle*. MIT press, 2007.

[33] A. R. Barron and T. M. Cover, "Minimum complexity density estimation," *IEEE transactions on information theory*, vol. 37, no. 4, pp. 1034–1054, 1991.

[34] S. Goldwater, T. L. Griffiths, and M. Johnson, "A bayesian framework for word segmentation: Exploring the effects of context," *Cognition*, vol. 112, no. 1, pp. 21–54, 2009. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0010027709000675

[35] Y. W. Teh, "A bayesian interpretation of interpolated kneser-ney," School of Computing, National University of Singapore, Technical Report TRA2/06, 2006.

[36] O. Walter, R. Haeb-Umbach, S. Chaudhuri, and B. Raj, "Unsupervised word discovery from phonetic input using nested pitman-yor language modeling," in *ICRA Workshop on Autonomous Learning*, 2013.

[37] R. Takeda, K. Komatani, and A. I. Rudnicky, "Word segmentation from phoneme sequences based on pitman-yor semi-markov model exploiting subword information," in *2018 IEEE Spoken Language Technology Workshop (SLT)*. IEEE, 2018, pp. 763–770.

[38] T. Araki, T. Nakamura, T. Nagai, S. Nagasaka, T. Taniguchi, and N. Iwahashi, "Online learning of concepts and words using multimodal lda and hierarchical pitman-yor language model," in *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2012, pp. 1623–1630.

[39] D. Mochihashi, T. Yamada, and N. Ueda, "Bayesian unsupervised word segmentation with nested Pitman-Yor language modeling," in *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP*. Suntec, Singapore: Association for Computational Linguistics, Aug. 2009, pp. 100–108. [Online]. Available: https://aclanthology.org/P09-1012

[40] S. Sawada, K. Yoshii, and K. Hirata, "Unsupervised melody segmentation based on a nested Pitman-Yor language model," in *Proceedings of the 1st Workshop on NLP for Music and Audio (NLP4MusA)*, S. Oramas, L. Espinosa-Anke, E. Epure, R. Jones, M. Sordo, M. Quadrana, and K. Watanabe, Eds. Online: Association for Computational Linguistics, 16 Oct. 2020, pp. 59–63. [Online]. Available: https://aclanthology.org/2020.nlp4musa-1.12/

[41] J. Pitman and M. Yor, "The two-parameter poisson-dirichlet distribution derived from a stable subordinator," *The Annals of Probability*,

vol. 25, no. 2, Apr. 1997. [Online]. Available: http://dx.doi.org/10.1214/aop/1024404422

[42] S. Hallam, I. Cross, and M. Thaut, Eds., *The oxford handbook of music psychology*, 2nd ed., ser. Oxford Library of Psychology. London, England: Oxford University Press, Dec. 2017.

[43] F. Jelinek, "Self-organized language modeling for speech recognition," *Readings in speech recognition*, pp. 450–506, 1990.

[44] W. H. Frere, *Antiphonale Sarisburiense: a reproduction in facsimile of a manuscript of the 13th century, with a dissertation and analytical index*. Gregg Press Limited, 1901.

[45] K. E. Helsen, "The great responsories of the divine office: aspects of structure and transmission," Ph.D. dissertation, 2008.

[46] H. Ney, U. Essen, and R. Kneser, "On structuring probabilistic dependences in stochastic language modelling," *Computer Speech & Language*, vol. 8, no. 1, p. 1–38, Jan. 1994. [Online]. Available: http://dx.doi.org/10.1006/csla.1994.1001