COMPARING OFFERTORY MELODIES OF FIVE MEDIEVAL CHRISTIAN CHANT TRADITIONS

Peter van Kranenburg
Utrecht University, Meertens Institute
peter.van.kranenburg@meertens.knaw.nl

Geert Maessen
independent scholar
gmaessen@xs4all.nl

ABSTRACT

In this study, we compare the melodies of five medieval chant traditions: Gregorian, Old Roman, Milanese, Beneventan, and Mozarabic. We present a newly created dataset containing several hundreds of offertory melodies, which are the longest and most complex within the total body of chant melodies. For each tradition, we train n-gram language models on a representation of the chants as sequence of chromatic intervals. By computing perplexities of the melodies, we get an indication of the relations between the traditions, revealing the melodies of the Gregorian tradition as most diverse. Next, we perform a classification experiment using global features of the melodies. The choice of features is informed by expert knowledge. We use properties of the intervallic content of the melodies, and properties of the melismas, revealing that significant differences exist between the traditions. For example, the Gregorian melodies contain less step-wise intervals compared to the other repertoires. Finally, we train a classifier on the perplexities as computed with the n-gram models, resulting in a very reliable classifier.

1. INTRODUCTION

In 789 Charlemagne ordained the Roman rite normative for Christian worship throughout his Empire. The chant of this rite became widely known as Gregorian chant (GRE). The earliest manuscripts with pitch-readable notation date from the beginning of the eleventh-century, increasing in number until the Renaissance. Manuscripts with neumatic contour notation go back to the end of the ninth century, and manuscripts with only the texts of the chants to almost 800. Basically all these manuscripts exhibit the same chants for specific liturgical occasions [13].

Since the invention of book printing and the Reformation, this uninterrupted and almost omnipresent European chant tradition came to an end. The Council of Trent (1545–1563) seems the beginning of many emended and sometimes drastically refashioned traditions of Gregorian chant. Since the restoration of Gregorian chant in the late nineteenth century, remnants of non-Gregorian chant traditions have continued to intrigue scholars. By the thirteenth century most of these traditions had already been abolished and replaced by Gregorian chant.

To this day the only surviving non-Gregorian tradition is the Milanese chant (MIL) of the Ambrosian rite in Northern Italy. The earliest notated manuscripts date from the twelfth century. Several hundreds of MIL chants are melodically related to GRE chants [2]. The Old Roman chant (ROM) that once existed in Rome itself is preserved in three graduals, several antiphoners and fragments from the eleventh till thirteenth centuries. Nearly all ROM chants are melodically related to GRE chants, with similar liturgical assignments. ROM was abolished in the thirteenth century [14]. Nearly 200 chants of the Beneventan rite of Southern Italy survive in eleventh and twelfth-century manuscripts among the regular GRE chants. Old Beneventan chant (BEN) was abolished in 1058 [15].

On the Iberian Peninsula and Southern France the Mozarabic rite was dominant from the sixth till the eleventh century. Its chant is called Old Hispanic chant. It was abolished in 1085 and replaced by the Roman rite with its GRE. Six parishes in Toledo were allowed to continue the tradition. The oral Mozarabic tradition was notated in early sixteenth century musical notation (MOZ). However, we also have over 5,000 Old Hispanic chants preserved in neumatic contour notation from the tenth till thirteenth centuries. Unfortunately, the vast majority of these chants do not correspond with MOZ and remain pitch unreadable [19, 27].

Since the 1950s the central question in chant scholarship concerned the relationship between GRE and ROM. Which of these traditions was the earliest? Was there perhaps another tradition preceding both? Many hypotheses have been put forward, but hardly any conclusive positions have been reached. Most scholars, however, believe that both GRE and ROM are later developments of the Roman tradition that was known to the Carolingians in the second half of the eight century. So the question became: Which was closer to eight century Rome, GRE or ROM? Some scholars believe the formulaic character of ROM to hold the earliest evidence, although the surviving manuscripts are of later date than the earliest GRE sources [7]. Some believe GRE reflects the earlier tradition, having adjusted the Roman chant only slightly to the specific needs of the Carolingian world [22]. Some still believe a third, Gallican or Hispanic, tradition played a major role in the creation of

Attribution: Peter van Kranenburg, Geert Maessen. “Comparing Offertory Melodies of Five Medieval Christian Chant Traditions”, 18th International Society for Music Information Retrieval Conference, Suzhou, China, 2017.

tradition	century	chants	offertories	parts	avg. notes/part	std. notes/part
GRE: Gregorian Chant	XI–XII	1,000	115	344	162.54	62.65
ROM: Old Roman Chant	XI–XIII	700	94	285	170.14	69.04
MIL: Milanese Chant	XII–XIII	800	104	147	177.63	98.50
BEN: Beneventan chant	XI–XII	100	39	41	152.98	64.00
MOZ: Mozarabic chant	XI–XVI	400	71	139	127.94	52.74

Table 1. Estimation of date and number of mass proper chants in the main sources of five traditions, number of offertory-chants in our data set, number of offertory-parts, and average and standard deviation of the lengths of the parts.

the differences between GRE and ROM [17, 18]. A matter of debate also is the date when the chants were created. McKinnon [22] argues, primarily based on the liturgical assignment of the chant texts, that the Roman repertoire was composed according to a plan in the last decades of seventh-century Rome. Pfisterer [26] on the other hand argues, primarily based on the comparison of Latin Bible translations, that the repertoire has grown in accordance with the solemnity of the feasts between the fifth and early seventh centuries.

An important contribution to the discussion has been made by Rebecca Maloy in her 2010 monograph on the most complex of all chant genres in both traditions: the offertory [20]. Her book (including a digital edition of 94 cognate pairs of GRE and ROM offertories) provides a fascinating insight in modern scholarship and a highly sophisticated analysis of the offertory genre in both traditions. Basically, however, she does not reach conclusive arguments for a best hypothesis. In this paper, also, we do not pretend to present a conclusive position. Instead, we present the first results of a computational analysis of melodic similarities and differences between chant traditions, illustrating directions of research that may give new input to the longstanding questions. To this end we improved the musicological approach of traditional styles in terms of melismas, intervallic steps and leaps [13] to perplexities. We also verified and transformed Maloy's edition into a data set, and expanded the set with all offertories from the main sources of GRE, ROM, MIL, BEN and MOZ. Table 1 provides an overview with estimated numbers of mass proper chants in each tradition and the number of offertory chants included in our data set.

Important chant studies using computational techniques were published by several authors. However, most of the data are no longer available [10], represent only part of a single tradition [9], or a genre not easily available in five traditions [6, 12], or were not meant as exact data sets [30]. Some of the procedures used, however, need further investigation. Hansen's [10] distinction of different tonalities for different layers in GRE is one of these, as is the segmentation procedure used by Halperin [9] and Haas [8]. In fact this last approach can be seen as a precursor of the n-gram method we use in the current paper.

Maloy [20] does not use computational techniques, but she does with the offertory present a genre that is clearly available in five different traditions.

In this paper we demonstrate the importance of a computational approach for two longstanding and complementary questions in chant research. Based on local melodic structure, our n-gram method presents relations between different traditions (Section 3). Given a set of traditions, it shows which tradition has most characteristics in common with all (or most) traditions. This clearly relates to the musicological question of "origin". Based on global features of the chants, our decision-tree based classification method shows differences between the traditions, and is able to identify with high reliability the traditional "home" of single chants (Section 4). This can be helpful in identifying chants not corresponding to the catalogues in use.

2. DATA SET

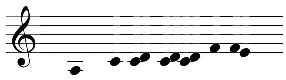
The contents of our data set is summarized in Table 1, showing the number of offertory chants included in our set. The first column lists codes and names of the separate traditions. The second and third columns give an estimate of period and number of the total preserved mass proper chants to which our offertories belong. In most cases, one offertory is divided in parts, the first part being the antiphon, and the subsequent parts the verses. Throughout this paper, we take the parts as basic units for analysis and classification. We include the number of parts per tradition in the table. We also include basic statistics on the length of chant-parts in number of notes.

For the GRE and ROM offertories we could have used the data set of Haas [8]. However, we preferred the critical edition of Maloy, because the Offertoriale Triplex [23] used by Haas is notably unreliable. One of the problems with the offertory concerns the many transpositions to avoid non-diatonical pitches. In selecting the best single manuscript for each separate chant Maloy chose, in our view, the best option. We converted Maloy's Finale scores to Volpiano strings and again carefully checked all details. We manually encoded the remaining offertories from the facsimile of Maloy's most important manuscript, Ben 34 [1] and one, GRE-115, *Audi Israhel*, from her book.

The Volpiano truetype font was developed by David Hiley and Fabian Weber at Regensburg University. [1] It is a typeface for note heads on the five line staff for monophonic music. It is perfectly suitable for our data set. It affords an encoding of each score as a string of characters. Characters `a` to `p` represent the notes A till a", while the

[1] Downloadable from: `http://www.uni-regensburg.de/Fakultaeten/phil_Fak_I/Musikwissenschaft/cantus/`

`i` represents the flat sign for b'. Small and capital `w`, `x`, `y` and `z` representing other alterations; some as defined in the font, some by new convention. Three dashes `---` separate the notes for the different words, two dashes `--`, for different syllables, and one dash marks a new neume within a syllable. Numerals indicate clefs and breaks. For clarification, Figure 1 shows an example of a Volpiano string and the rendering of it with the Volpiano font.



`1---a---c-cd-cdcd-f-fe---`

Figure 1. Example of a string in volpiano encoding and its rendering in the font.

We manually encoded the MIL, BEN and MOZ offertories from the best available sources; the Milanese mass book [29], the recent critical edition of Beneventan chant [16], and the facsimile of Mozarabic chant books [5].

3. COMPARING TRADITIONS USING N-GRAM MODELS

To examine the interconnections between the chant traditions concerning small-scale melodic fragments, we take an n-gram approach. n-gram modeling has been developed in computational linguistics [21]. It employs repetitive structures of a language to construct a probabilistic model allowing to compute the probability of occurrence of a word in its local context within a sentence. Concretely, let w be a word in vocabulary V belonging to a language L, and let $s = w_1, w_2, \ldots, w_l$ be a sentence of l words, also belonging to language L. Then, for a word w_i in s, an n-gram model allows to compute the conditional probability of w_i given the preceding context of $n-1$ words: $p(w_i|w_{i-(n-1)}, \ldots, w_{i-1})$. Previous application of n-gram modeling of music, notably include the IDyOM model [24], which combines long and short term models. For our purpose the basic n-gram approach suffices.

Typically, an n-gram model is derived from a large collection of training sentences belonging to the language of interest. In the most basic approach, for each unique context $w_{i-(n-1)}, \ldots, w_{i-1}$ in these training sentences, an inventory is made of all possible continuations w_i. This results in a distribution over the vocabulary, indicating the probability of each possible continuation of the context. The full model consists of the collection of distributions for the continuations of all unique contexts.

One of the uses of an n-gram model is to evaluate to what extent a given sentence fits in a given language. This is the way in which we employ n-gram models of the chant traditions. Since sentences are of variable length, it is not possible to simply compute the probability of a sentence as product of the probabilities of each word. Therefore, we will use the measure of *perplexity*, which indicates the degree to which the sentence 'fits' in the language:

$$PP = p(w_1, w_2, \ldots, w_l)^{-\frac{1}{l}}. \quad (1)$$

One particular problem in computing $p(w_i|w_{i-(n-1)}, \ldots, w_{i-1})$ occurs if the n-gram $w_{i-(n-1)}, \ldots, w_i$ has no occurrences in the training data. In that case, the probability of w_i is evaluated as zero, rendering the probability of the entire sentence zero. To circumvent this problem, several approaches exist. We use modified Kneser-Ney smoothing [4] as implemented in the KenLM Language Model Toolkit [11]. This method is widely accepted as the preferred method to deal with zero-counts.

3.1 Application on Chant Data

To derive an n-gram model from our chant data, we need to redefine some linguistic terms. We consider the traditions as languages. We consider each part of a chant as sentence, and we consider the intervals between the notes as words, where an interval is represented by a signed integer number indicating the direction (pos/neg) and the size of the interval in semi-tones. Because each of the chant traditions uses the same melodic intervals, the traditions have the same vocabulary, which allows us to compute the perplexity of a given chant for all five traditions. Also, since the vocabulary is very small compared to the vocabulary of any natural language, we need much less training data than typically is needed for natural language modeling.

3.2 Choosing *n*

An important question is which value to choose for n. For each $n \in \{2, 3, \ldots, 10\}$, and for each tradition we compute for each chant in that tradition the perplexity given its own tradition. To avoid overfitting, we follow a 10-fold evaluation, successively taking one subset of the data to compute the perplexities and taking the other 9 subsets for training, making sure that all parts of the same chant always are in the same subset. By visual inspection of the distributions of perplexities, we observe that for GRE, MIL, and MOZ, no further decrease of average perplexity is noticeable for $n > 5$, while for BEN and ROM slight improvement is achieved for respectively 7-gram and 8-gram models. Based on these findings, we choose $n = 5$ throughout this paper.

3.3 Comparing Chant Traditions using *n*-gram models

3.3.1 Method

As we are interested in the differences and commonalities of the five chant traditions, we perform an exhaustive evaluation in which we compute for each chant-part five perplexity values, one for each of the five traditions. In the case of the perplexity of a chant-part given its own tradition, we need to derive an n-gram model from all other chants of that tradition, excluding the chant that contains the chant-part. To include this chant in deriving the n-gram model would result in a too optimistic value for the perplexity. The chant must be 'unseen' by the model. The resulting perplexity reflects the extent to which the chant-part fits in its own tradition. For the perplexities given the other four traditions, we take n-gram models that have been derived from the entire sets of chants from those traditions. These resulting four perplexity values reflect to what extent the chant-part fits in the respective other four traditions.

After obtaining all perplexity values, we visualize the distributions for the various conditions as box-and-whisker plots [31]. The median is indicated with the red horizontal line. The box extends from the first to the third quartile, which is the interquartile range (IQR). The lower vertical whisker extends to the lowest data point still within 1.5 IQR from the first quartile and the upper whisker extends to the highest data point still within 1.5 IQR from the third quartile. The data points past the whiskers are considered outliers and are individually plotted as circles.

We evaluate whether two distributions differ significantly by performing a Kolmogorov-Smirnov test [28], and we evaluate the magnitude of the difference by computing the effect-size according to

$$e = \frac{\bar{x}_1 - \bar{x}_2}{\max(s_1, s_2)} \tag{2}$$

in which x_1 and x_2 are the averages of the perplexities, and s_1 and s_2 are the standard deviations. By taking the max of s_1 and s_2, the resulting value for the effect size is a pessimistic estimation.

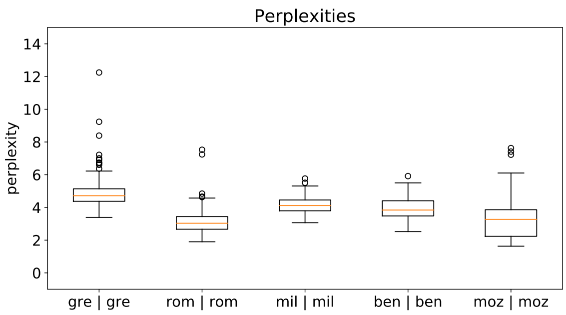


Figure 2. The distributions of perplexities of the chant-parts given their own respective tradition, represented as box-plots.

3.3.2 *Results and Interpretation*

The differences between the traditions are noticeable in Figure 2. The higher the perplexities, the higher the internal diversity of the repertoire. GRE is most divers. The outliers show specific chants of a single tradition most alien to this tradition. In GRE the two verses of GRE-63, *Oravi Deum meum,* are most extreme. This conforms to the fact that this is the most "chromatic" GRE chant. Its problematic pitches were already discussed by John of Afflighem (early twelfth century; [20]). The next GRE outlier is the antiphon of GRE-95, *Elegerunt apostoli*. *Oravi* and *Elegerunt* are considered two of only five offertories with possible Gallican origin, since they have cognate pairs in Old Hispanic chant. In ROM both antiphon and verse of ROM-92, *Domine Jesu Christe,* are most extreme. This conforms to the fact that this chant is almost identical to its GRE counterpart, GRE-92, *Domine Jesu Christe.* As Maloy demonstrates on textual evidence this chant in fact should be considered a GRE chant. As she puts it, "it is one of the few demonstrable instances of reverse, Frankish-to-Roman transmission in the offertory repertory" [20]. MIL and BEN hardly show outliers. However, in BEN the most extreme outlier, BEN-70, *Tunc imperator,* is the most syllabic chant of BEN. In MOZ, finally, the most extreme outlier is MOZ-13, *Offerte Domino,* the only MOZ chant in the fourth church mode.

Figure 3 shows the interrelations between the traditions. Comparing the five traditions to the five models we have 25 comparisons, resulting in 25 distributions of perplexity values. The Kolmogorov-Smirnov test for independence shows that only six out of the 300 possible pairs of distributions do not differ significantly ($p > 0.028$). Only 15 pairs of distributions have an effect size less than medium ($e < 0.5$). This indicates that the vast majority of the differences we see in the diagrams, are of significance.

Most striking is the top figure, showing the perplexities given the GRE model. The five box-plots there show that all five traditions are pretty close to the GRE model. BEN being most alien. However, compared to the four other figures, we see BEN being even more alien to ROM and MOZ. As is apparent from the diagrams, GRE gives the best overall model for all traditions. Second best is MIL. The worst model for all is ROM, followed by MOZ.

These findings can be related to the longstanding question about origin. Assuming that the process of oral transmission generally results in decreasing complexity, it is well conceivable that all traditions stem from GRE, while it seems inconceivable that ROM was the root of all.

4. CLASSIFICATION WITH GLOBAL FEATURES

4.1 Feature Set

We also examine the differences between the traditions in terms of a set of global features. A global feature summarizes the entire melody in one value. The feature set we use relates to earlier musicological approaches to characterize the traditions. There are two groups of features: features that describe the intervallic contents of a melody, and features that are related to the length of melismas. We measure the following features: the frequencies of occurrence of each of the melodic intervals from -12 to 12 semitones, where the sign indicates the direction; `aleaps`, `asteps`, `dleaps`, and `dsteps`, which measure the fraction of intervals that respectively are ascending leaps, ascending steps, descending leaps, and descending steps; `leaps` and `steps` are the fractions of leaps and steps disregarding direction; `unison` is the fraction of note repetitions; `melis1-1`, `melis2-2`, `melis3-5`, `melis6-10`, `melis11-20`,

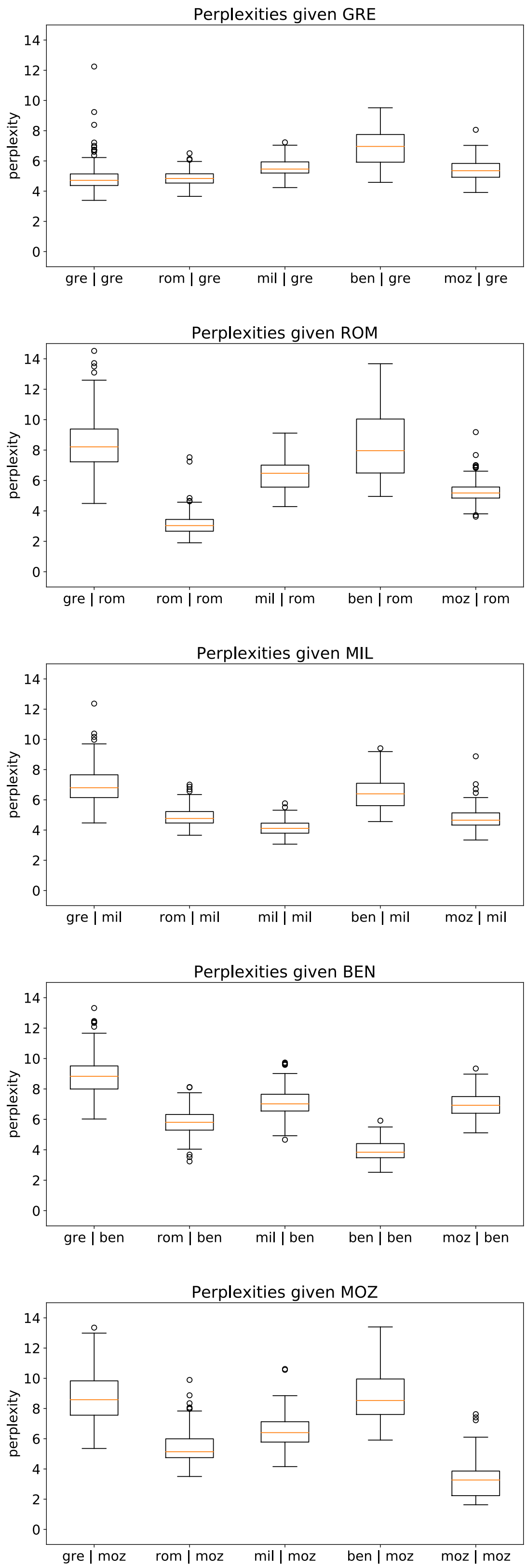


Figure 3. Distributions of the perplexities given the various traditions.

tradition	precision	recall	F1-score	support
BEN	0.38	0.22	0.28	41
GRE	0.90	0.89	0.89	344
MIL	0.64	0.65	0.65	147
MOZ	0.77	0.74	0.76	139
ROM	0.80	0.86	0.83	285
avg/total	0.79	0.79	0.79	956

Table 2. Classification results using the Decision Tree learner on the data set with global features.

`melis21-50`, `melis51-100`, `melis100-inf`, (`melisx-y` in general), represent the fraction of lyric syllables that have between x and y (y included) notes in the melody. `melis_mode` is the most common number of notes per syllable. `melis_longest`, `melis_secondlongest`, `melis_thirdlongest`, and `melis_fourthlongest` are the lengths of the four longest melismas. Finally, `melis_skewness`, and `melis_kurtosis` are the skewness and kurtosis of the distribution of melisma lengths.

We measure the values of these features in each of the 956 chant-parts. With the resulting dataset we perform a classification experiment to examine whether these features contain information for distinguishing between the five traditions.

4.2 Decision Tree Classification

Since we are not solely interested in classification accuracy, but we also want to understand the differences between the traditions, we prefer a learning algorithm that results in an interpretable model. Therefore, we learn a decision tree from our dataset with global features. We use the implementation of the tree learning algorithm as provided by the Python Scikit-learn library [25]. To prevent overfitting, and to obtain a relatively small tree, we set the minimum number of chant-parts per leaf to 10 and the maximum depth of the tree to 3.

To estimate the generalization of the learned tree, we perform 10-fold cross-validation, successively using one subset for testing and the other 9 subsets for learning a tree. For each chant in the current test-set, we record whether the classification was right. Again, we make sure to keep all parts from the same chant in either the test or the train set. After this procedure, we have a classification result for each of the chant-parts. Table 2 summarizes the resulting classification performance. While the overall-performance is not bad, discerning the chant-parts from BEN and MIL appears to be less successful.

There is no clear sign of overfitting. Therefore, we train a tree on the entire data set, which represents the differences between the traditions. The tree is depicted in Figure 4. It is apparent from the tree that the amount of step-wise motion in the melodies is one of the most important characteristics to isolate the GRE chants. These chants show the lowest amount of steps. Furthermore, the number of syllables with only one note, the amount of descending minor thirds, and the amount of descending minor seconds are of

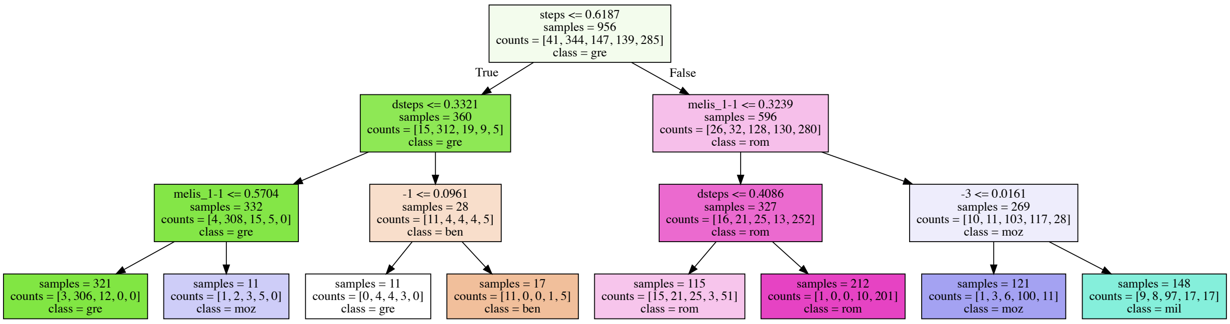


Figure 4. Decision tree as learned from the data set with global feature values. The order of the classes in the 'counts' field is: [BEN, GRE, MIL, MOZ, ROM]. The values indicate the number of chant-parts from the respective tradition that are 'in' the leave of tree.

tradition	precision	recall	F1-score	support
BEN	0.71	0.59	0.64	41
GRE	0.90	0.92	0.91	344
MIL	0.73	0.73	0.73	147
MOZ	0.91	0.88	0.90	139
ROM	0.93	0.94	0.94	285
avg/total	0.88	0.88	0.88	956

Table 3. Classification results using the Random Forest classifier on the data set with global features.

tradition	precision	recall	F1-score	support
BEN	0.93	0.98	0.95	41
GRE	0.97	0.98	0.98	344
MIL	0.96	0.95	0.96	147
MOZ	0.95	0.94	0.95	139
ROM	1.00	0.98	0.99	285
avg/total	0.97	0.97	0.97	956

Table 4. Classification results using the Random Forest classifier on the perplexity data.

importance. With just these features, it appears possible to separate the traditions to a moderately high degree.

4.3 Random Forest Classification

To examine whether it is possible to get higher classification accuracy, we also train a Random Forest Classifier, which trains a number of trees on random subsets of the data [3]. This does not lead to an easily interpretable model, but this procedure is known to typically show higher performance than a single decision tree. We set the number of trees to 10 and we follow the same procedure using 10-fold cross validation. The results are presented in Table 3. The results show significant improvement, but still with weaknesses for BEN and MIL.

4.4 Classification with Perplexity values

Since the perplexity of a chant-part given the n-gram model of a tradition also can be considered a global feature, we assemble another data set with for each chant-part the five perplexities for the five traditions, as computed in Section 3, as features. The classification results for a Random Forest Classifier are shown in Table 4.

Based on the perplexity values, we obtain a very accurate classifier with a F1-score as high as 0.97. Even for the minority class BEN we obtain very good results. Such a classifier can be of particular interest in tracing chants whose origins are unclear.

We performed an analysis of the chant-parts that are mis-classified by our best-performing classifier, the random forest trained on the perplexity data. Due to space constraints, it is not possible to give a full account of the analysis here, but in general we can state that many of the mis-classified parts are remarkable cases, including the outliers that have been discussed in Section 3.3.2, but also some other chants with debatable origin.

5. CONCLUSION AND FUTURE WORK

We presented an n-gram method to examine relations between medieval chant repertories, touching on central questions in chant scholarship. Our method shows in a quantitatively precise way that the body of Gregorian offertory melodies is characterized by a higher internal diversity than the offertories from the other four traditions. We also presented a highly accurate classification method. Outliers and misclassifications in both cases pointed at known problems in chant scholarship. Future work will concentrate on the refinement of our approaches for separate chant genres within traditions.

6. REFERENCES

[1] *Le Codex VI 34 de la Bibliothèque capitulaire de Bénévent, Graduel avec Prosaire & Tropaire*. Number XV in Paléographie Musicale. Abbaye Saint-Pierre de Solesmes, Solesmes, 1992.

[2] Terence Bailey. Ambrosian chant. http://www.oxfordmusiconline.com/subscriber/article/grove/music/00754.

[3] L. Breiman. Random forests. *Machine Learning*, 45(1):5–32, 2001.

[4] Stanley F. Chen and Joshua Goodman. An empirical study of smoothing techniques for language modeling. *Computer Speech and Language*, 13:359–394, 1999.

[5] Ángel Fernández Collado, Alfredo Rodríguez González, and Isidoro Castañeda Tordera, editors. *Los Cantorales Mozárabes de Cisneros, Catedral de Toledo*. Cabildo de la Catedral Primada, 2011.

[6] Ike de Loos. *Duitse en Nederlandse Muzieknotaties in de 12e en 13e eeuw*. PhD thesis, Universiteit Utrecht, Utrecht, 1996.

[7] Joseph Dyer. The roman offertory, an introduction and some hypotheses. In *The Offertory and its Verses: Research, Past, Present and Future*. Tapir Academic Press, Trondheim, 2007.

[8] Max Haas. *Mündliche Überlieferung und altrömischer Choral, Historische und analytische computergestützte Untersuchungen*. Peter Lang, Bern, 1997.

[9] David Halperin. *Contribution to a Morphology of Ambrosian Chant, A computer-aided analysis of the pars hiemalis according to the British Museum manuscript add. 34209 together with a package of computer programs for the analysis of monophonic music*. PhD thesis, Tel Aviv University, Tel Aviv, 1986.

[10] Finn Egeland Hansen. *The Grammar of Gregorian Tonality, An Investigation Based on the Repertory in Codex H 159, Montpellier*. Dan Fog Musikforlag, Copenhagen, 1979.

[11] Kenneth Heafield. Kenlm language model toolkit. http://kheafield.com/code/kenlm/.

[12] Katherine Eve Helsen. *The Great Responsories of the Divine Office, Aspects of Structure and Transmission*. PhD thesis, Universität Regensburg, Regensburg, 2008.

[13] David Hiley. *Gregorian Chant*. Cambridge University Press, Cambridge, 2009.

[14] Helmuth Hucke and Joseph Dyer. Old roman chant. http://www.oxfordmusiconline.com/subscriber/article/grove/music/11725.

[15] Thomas Forrest Kelly. Beneventan chant. http://www.oxfordmusiconline.com/subscriber/article/grove/music/02678.

[16] Thomas Forrest Kelly and Matthew Peattie, editors. *Monumenta Monodica Medii Aevi IX, The Music of the Beneventan Rite*. Bärenreiter, Kassel, 2016.

[17] Kenneth Levy. A new look at old roman chant. *Early Music History*, 19:81–104, 2000.

[18] Kenneth Levy. A new look at old roman chant - ii. *Early Music History*, 20:173–198, 2001.

[19] Geert Maessen and Peter Van Kranenburg. A semi-automatic method to produce singable melodies for the lost chant of the mozarabic rite. In *Proceedings of the 7th International Workshop on Folk Music Analysis*, Málaga, 2017.

[20] Rebecca Maloy. *Inside the Offertory, Aspects of Chronology and Transmission*. Oxford University Press, Oxford, 2010.

[21] Christopher D. Manning and Hinrich Schütze. *Foundations of Statistical Natural Language Processing*. MIT Press, Cambridge, Massachusetts, 1999.

[22] James McKinnon. *The Advent Project, The Later-Seventh-Century Creation of the Roman Mass Proper*. University of California Press, Berkeley, 2001.

[23] Karl Ott and Rupert Fischer, editors. *Offertoriale Triplex cum Versiculis*. Abbaye Saint-Pierre de Solesmes, Solesmes, 1985.

[24] Marcus Thomas Pearce. *The Construction and Evaluation of Statistical Models of Melodic Structure in Music Perception and Composition*. PhD thesis, City University, London, London, 2005.

[25] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[26] Andreas Pfisterer. *Cantilena Romana, Untersuchungen zur Überlieferung des gregorianischen Chorals*. Ferdinand Schöningh, Paderborn, 2002.

[27] Don Michael Randel. Mozarabic chant. http://www.oxfordmusiconline.com/subscriber/article/grove/music/19269.

[28] Nikolaĭ Vasilevich Smirnov. Tables for estimating the goodness of fit of empirical distributions. *Annals of Mathematical Statistics*, 19(2):279–281, 1948.

[29] Gregorio Maria Suñol, editor. *Antiphonale Missarum juxta ritum Sanctae Ecclesiae Mediolanensis*. Desclée et Socii, Rome, 1935.

[30] Jessica Thompson, Andrew Hankinson, and Ichiro Fujinaga. Searching the liber usualis: using couchdb and elasticsearch to query graphical musical documents. In *Proceedings of the 12th International Society for Music Information Retrieval Conference*, Miami, 2011.

[31] John Wilder Tukey. *Exploratory Data Analysis*. Addison-Wesley, Readin, PA, 1977.