

Stylometry and Automatic Attribution of Medieval Liturgical Monodies

Francesco Unguendoli*
Università di Modena e Reggio Emilia,
Modena

Giampaolo Cristadoro**
Università di Milano - Bicocca, Milano

Marco Beghelli†
Università di Bologna, Bologna

While automatic attribution of literary text as well as stylometry evaluation are nowadays well-established research areas, much less has been done in the field of musicology. Here we present the results of the implementation of an automatic stylometric attribution technique to a corpus of liturgical monodies of medieval origin (the so-called Gregorian Chant, Old Roman Chant and Ambrosian Chant). The 'unidimensional' nature of the musical repertoires investigated (rhythm-free melody without accompaniment) allows the adoption of a known method based on a pseudo-distance between frequency-vectors of n -gram of consecutive symbols. Finally, we show that some specific features of musicological interest inside the three liturgical families can be naturally extracted using a statistical analysis of n -gram distributions. The results presented show that a quantitative approach is well suited to support and accompany the investigation of refined problems in musicology.

1. Attributionism as stylometry

The problem of attribution in art, that is to say, the identification of the author of an anonymous work, is common for the visual and literary arts (attribution of unsigned paintings and texts). As a philological issue it is no less important for the historians of music, who often deal with more or less ancient compositions of uncertain paternity. However, the musicologists have no useful analytical tool to go beyond the simple *subjective* impression of listening, while stylistic investigation methodologies usually applied to music mostly work on macro-genre. That is because much of the research has aimed at the management, organisation, and access to large music databases (mainly those of the network), rather than a fine comparison of texts in the sphere of the so-called 'art music' (in which the musicologists are more interested): the Music Information Retrieval (Orion 2006) needs to automate and speed up the procedures handling large amounts of data, so renouncing to discriminate between pieces of the same genre or of stylistically close authors, which is exactly what instead is required for the solution of attributional issues.

* Dept. of Physics, Informatics, and Mathematics - Via Giuseppe Campi 213/a, 41125, Modena, Italy.
E-mail: francesco.unguendoli@unimore.it

** Dept. of Mathematics and Applications - Via Roberto Cozzi 55, 20125 Milano, Italy.
E-mail: giampaolo.cristadoro@unimib.it

† Dept. of Arts - Via Barberia 4, 40123 Bologna, Italy. E-mail: marco.beghelli@unibo.it

Before facing real problems of attribution in the musical field, it is therefore necessary to identify adequate analytical methods as *objective* as possible, that is to say supported by mathematical-statistical procedures implemented through appropriate algorithms. A model for the measurement of the style (*stylometry*) is not intended to substitute the ear and the discernment of the musicologists with a machine, but instead to offer them a further tool in the philological investigation, able to find further musical distinctive features, thus letting them evaluate stylistic aspects that they would not be able to perceive or might not give the right weight to, having not considered their statistical significance.

2. Strings of letters and notes

The emphasis on music considered as a language dates back to at least the eighteenth century. Methods based on the recognition of specific musical patterns have already been tested by various authors, with only partially satisfactory results. The computational comparison between the String Quartets of Haydn (1732-1809) and his emulous Mozart (1756-1791) obtained recognition values that reached a maximum of around 70 – 75%, see (Backer and van Kranenburg 2005), (Hillewaere, Manderick, and Conklin 2010), (Wolkowicz and Kešelj 2013), (Velarde et al. 2016). Even more difficult was the comparison between the champion of Renaissance polyphony Josquin Desprez (c. 1450-1521) and his contemporaries (Brinkman, Shanahan, and Sapp 2016): the results were decidedly better (over 90% of successful identifications) with a new method (McKay et al. 2017), but at the expense of the interpretation of data by musicologists.

Here we adapt to music the stylistic inspection used to textual area (Kešelj et al. 2003) and later refined for the well-known attribution of the anonymous articles by the politician Antonio Gramsci (1891-1937) (Basile et al. 2008). Their method is based on features at a lower level than the lexical or syntactic-semantic one, extracting from the texts the so-called '*n*-grams', i.e. sequences of *n* consecutive symbols (letters, spaces, and punctuation marks), whose frequencies are used to measure the stylistic distance between an anonymous text and the reference corpus (similarity measure):

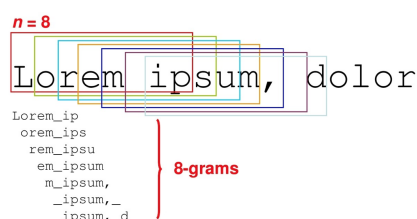


Figure 1
n-grams as stylistic verbal feature: an example of segmentation of the text in sequences of 8 consecutive symbols (8-grams).

This kind of sequential feature seems to adapt perfectly to a string of consecutive musical notes too, particularly when considering one-dimensional compositions, i.e. based on a simple succession of sounds at different heights: melodies with no rhythm and any concomitant harmonic aggregate or contrapuntal accompaniment, also lacking all those additional parameters that make it difficult to investigate Western music (note values, dynamic, agogic, phrasing, etc.). The great Christian monodic liturgical repertoires, developed during the Middle Ages, have such a characteristic. Here is an

example, transcribed in modern notation (a palaeographic image of the same passage can be seen in Fig. 6):



Figure 2

An example of Medieval monody in modern notation (Gregorian Chant, from the Easter Offertory *Terra tremuit et quievit*).

The n -gram style-recognition method was indeed already been tested on Renaissance polyphony (Doraisamy and Ruger 2003) and on classical string quartets (Hillewaere, Manderick, and Conklin 2010), not to mention the studies of Kešelj et al. (2008 and 2013), more interested in research and categorization of large music databases. But both the polyphony and the string quartets are based on a multidimensional compositional process that moves away significantly from the linearity of the verbal language on which the n -gram method has been conceived, making it difficult to define the same ‘1-gram’.

For this reason we took a step back, leaving out the most complex musical forms to test the potential of the method on the medieval monody, consisting exactly of a simple sequence of symbols, like a string of letters. Really the analytic method based on ‘ n -gram’ would even seem theoretically more suitable to a linear sequence of notes than letters:

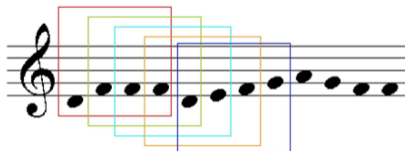


Figure 3

n -grams as stylistic musical feature: an example of segmentation of the melody in sequences of 4 consecutive symbols (from the Easter Offertory of Fig. 2).

In the case of a sequential melody, in fact, the n -grams isolate discrete fragments of it which (unlike the verbal fragments of text) maintain a certain syntactic and semantic identity, even for melodic fragments of just two notes: notes proceeding by step or by leap or by repercussion of the same sound represent three very distinct and identifying melodic motions; a third note that continues the ascending or descending direction of the two previous notes produces a ‘musical sense’ different from a note that reverses the motion of the previous two.

3. A quantitative model at the service of the musicologists

The research examined the three largest and most famous medieval liturgical corpora: the so-called Gregorian Chant (developed in the French area to spread then throughout Christian Europe), the so-called Old Roman Chant (limited to the non-pontifical Roman churches), and the so-called Ambrosian Chant (developed simultaneously in the Milanese diocese). Other historical repertoires could be also added (for example the

Beneventan, Mozarabic, and Aquileian traditions), spread elsewhere but characterized by similar monodies.

In general, the different geographic families of monodies are made up of hundreds of written musical pieces that resulted from a collective oral creation based on common roots. They originated well before the end of the first millennium AD and stratified over time, until they were established in later manuscripts intended to widely spread the chants through-out the territory. All families belong to the same great Western Christian tradition and are based on largely identical Latin verbal texts with similar liturgical functions (Offertory, Communion, Alleluia, etc.), but covered with different music.

It must be stressed that, in spite of multiple authorship and long stratification, each geographical tradition has assembled a musical repertoire in itself homogeneous and stylistically different from the others. But the stylistic differences between the various families of chants are hardly perceivable at first listening or reading, even for a common musicologist, not specialized in those repertoires.

These circumstances have been considered ideal for testing the quantitative attributional method of n -gram on linear sequences of sound symbols: we defined an automatic measurement of the stylistic distance of each single piece of music from each specific geographic repertoire. The aim is therefore not to trace the authorship of hopelessly anonymous music, but to test the validity of the analytical method based on n -gram in identifying distinctive *features* of the various geographical families, and to provide the musicologists with *objective* and innovative data to evaluate some heuristic assumptions behind them. For example:

- the real stylistic difference between the various families of liturgical monodies;
- the real spurious nature of certain chants within the family to which they belong, because they are the result of a contamination, or are very ancient (created before a stylistic fixation occurred for the geographic family they belong to), or perhaps very late, or the result of an anomalous origin (i.e. far from the stylistic course of their geographic tradition);
- the actual stylistic sub-identity of chants belonging to defined liturgical genres that are a breed apart within the family they belong to (e.g. Graduals and Offertories).

The isolation of n -grams of melodic nature allows also to make a further hermeneutic step within the genetic profile of the medieval monodies, which is perhaps more interesting than the aforementioned aesthetic macro outlines. It is in fact widely believed that in many cases such anonymous melodies resulted from a centonization, i.e. the juxtaposition of recurring melodic cells also responsible for the resulting stylistic unity that is clearly perceptible to listeners, despite its lack of a unique or circumscribed paterity. These cells inevitably coincide with the more or less long n -gram isolated from the analysis method, on which the computer can easily produce a series of statistical calculations for the musicologist who wants to extrapolate the distinctive traits of each family of Medieval liturgical monodies. For example, it is possible to identify:

- which are the most recurrent melodic cells within a defined geographic family, whose presence provides a sort of distinctive sound;

- which are the melodic cells (if any) common to all the geographical families, able to bring out some traces of a hypothetical common repertoire that changed over time;
- what are the melodic cells absolutely absent in some geographical families but present in other families, information that could provide useful elements for the philological restoration of an uncertain melodic passage, to correct a textual corruption, to reveal a stylistic contamination process;
- what is the average length of the most recurrent melodic cells, an element that can illuminate the compositional centonization process (a juxtaposition of short ‘musical lemmas’ or long melodic spans?).

4. The corpus of the investigation

There are a few dozen medieval manuscripts that hand down the individual repertoires of liturgical chants with the use of various and different forms of musical notation. The most long-lived repertoires, such as the Gregorian Chant (nominally still used today as the official liturgical singing of the Roman Catholic Church), have been progressively established in printed publications since the fifteenth century, for a widespread diffusion throughout the territory; other traditions are handed down by a small number of manuscript sources, due to their local diffusion (such as the Ambrosian Chant) or the limited use over time (such as the Old Roman Chant).

The modern editions are rare and partial; in some cases the individual repertoires are accessible only through the ancient manuscripts. Sources used in this survey: for the Gregorian Chant, the critical editions of the *Graduale Triplex* (Graduale 1979) and the *Offertoriale Triplex* (Offertoriale 1985) produced at Solesmes; for the Roman Chant, the diplomatic edition of the *Graduale Vat. lat. 5319* published in the *Monumenta Monodica Medii Aevi* (Stablein 1970); for the Ambrosian Chant, the *Antiphonale missarum* (Suñol 1935). Without discussing any philological problem, the choice of these editions was considered useful for their breadth, which allowed a wide availability of musical pieces to work on, offered in modern transcriptions recognized as reliable (apart from a handful of obvious typos that have been tacitly corrected).

The musical pieces used are 360, of various liturgical and formal nature. For the Gregorian Chant, 60 Offertories and 50 Graduals were considered, plus an additional 30 pieces with various and varied liturgical functions; the same was done with the Roman Chant, and with this database of 280 musical pieces we made the first measurements (Unguendoli et al., 2017). The comparison with the Ambrosian Chant clashed with a double difficulty: the easily accessible repertoire is small in quantity and length, and if you want to limit yourself to reasonably long chants (approximately more than 300 notes), you can only rely on 80 pieces, of which 30 Offertories, 38 Psalmelli (the equivalent of the Gregorian and Roman Graduals) and 12 other items of various liturgical use.

We ignored the verbal texts intoned from the individual pieces, the interest of the investigation being addressed exclusively to the musical dimension; in many cases the same liturgical texts are common among the three repertoires (those referable to psalms or other biblical passages), but for our purposes this circumstance is purely accidental, so that the incipit of every individual text is used here only to identify a chant.

In the transformation of the musical symbolic representation into searchable data, each note got the same standardized duration value. Reluctantly, we could not take into account any difference between single and grouped notes (the so-called *neumes*),

although fundamental for the phrasing of these melodies and important for some more detailed statistical surveys (see the discussion about the *tristropa* in Section 7). In the play of the refrains between the various antiphons a formal normalization was made to avoid excessive and unjustified differences in length of the chants.

5. Methods

Different possible strategies can be employed to represent a melody as a mathematical and computationally treatable object. For example, one can encode each note separately using its absolute height. Another, more refined choice, uses the distance of each note from the *finalis* (the last note of the piece which, in these repertoires, boasts an attractive, teleological strength, and at the same time constitutes the base on which the whole melody typically grows). We prefer not to follow these strategies in order to avoid the dependence on the octoechos: a system of ‘melodic modes’ in which the Gregorian chant was lately caged, but which has been scarcely adopted in the other families of monodies that we investigate. To stress the idea (see also Section 2) that not the single sounds but rather their *syntactic* relations are mostly relevant in our case, we used the difference of height between consecutive notes. With this choice the *unigram* (or 1-gram) corresponds to the melodic interval between two consecutive notes, computed in number of semitones. For example D–E=2, E–F=1, D–F=3, F–D=-3, etc¹.

5.1 (Pseudo-)distance between melodies

Fixed the length n , the (pseudo-)distance $d_n(x, y)$ between two pieces x and y is given by (Basile et al., 2008) :

$$d_n(x, y) = \frac{1}{|D_n(x)| + |D_n(y)|} \sum_{\omega \in D_n(x) \cup D_n(y)} \left(\frac{f_x(\omega) - f_y(\omega)}{f_x(\omega) + f_y(\omega)} \right)^2 \quad (1)$$

where $D_n(x)$ is the set (dictionary) of all different n -grams appearing in x and $f_x(\omega)$ is the relative frequency of the n -gram ω in x . In the following we scan different lengths n , in the range $1 \leq n \leq 15$, that correspond to melodic fragments of lengths varying from 2 to 16 notes. Note however (see also section Results) that the repetition of melodies of length $n > 10$ should be understood as a ‘citation’ – a melodic ‘reuse’ – rather than repetitions of common basic melodic units.

In the repertoires under investigation, the linear simplicity of the music is contrasted by the difficulties – for computational purposes – coming from the shortness of each sequence (only a few hundred musical notes for each piece) compared to those of a typical literary text. Moreover, the stylistic nature of this music enforces a pronounced melodic similarity among pieces of the same family as well as among different families. Such similarity is a consequence of their genesis, as noted before, with the prevalence of a collective authorship and repeated contaminations.

¹ We explicitly recall that in the natural scale, two neighbouring notes are separated by a note, but in the case E–F and B–C, that are separated by a semitone only.

5.2 Classification schemes

Our aim here is not to confront the performance of different classifications schemes in order to select the best one. On the contrary we here want to check if already the simplest schemes (both computationally and in their interpretability) are able to give meaningful and significant results from a musicological viewpoint. We thus tested two opposite schemes to attribute a piece x : one where only its nearest neighbour suffices, one where every piece contributes to the classification of x (see below for details). The rationale behind these two extreme cases are the following: from one side, the fact that the database at disposal is not very big suggested the use of a classification scheme (*voting scheme*) that uses almost all of it (and not, for example, only few k -neighbours); from the other side we wanted to check the musicological intuition that inside this kind of repertoire it is not unusual to find two pieces with almost plagiarised sections, so that they become very close (*nearest neighbour*).

To be more precise, given a piece x and two corpora A and B we tested the following classification schemes:

- 1) *nearest neighbour* – Assign x to the same class of its nearest neighbour;
- 2) *voting scheme* – List all pieces in both corpora² A and B by increasing distance from x . Assign to the j -th element (in such list) of A the value $k(j)/j - 1$, where $k(j)$ is its rank in the full list. Note that the closer a piece of the corpus A is to x , the smaller is its assigned value. Sum the values for all elements in A to obtain a score $\alpha(x)$ that quantifies the (dis)similarity of x to the corpus A. In a similar way, using pieces of B, we obtain the value $\beta(x)$. Finally the vote

$$v(x) = \frac{\beta(x) - \alpha(x)}{\alpha(x) + \beta(x)} \quad (2)$$

is used for the attribution task of x . The vote $v(x) \in [-1, 1]$ gives also an indication of the classification reliability: values close to 1(-1) indicate a strong attribution to the corpus A(B), while values close to 0 indicate a more uncertain classification.

Note that the ranking procedure is not well defined in the case in which two pieces have the same distance. In particular, given the shortness of the melodies investigated it is not unusual that two chants z and y do not have any n -grams in common with x and thus have distances $d_n(x, y) = d_n(x, z) = 1$. This happens in some cases even at moderate values of $n \sim 6$. For even larger n all distances become identical and the classification procedure is then ineffective for the given x . We addressed this problem in the voting procedure assigning the same rank to the pieces at the same distance to x , that thus contribute equally to the vote. With this modification if all distances are equal then $v(x)=0$ and the classification is not possible.

² The cardinality of the two corpora A and B should be the same to avoid spurious biases.

6. Results

We focused on three different tasks:

TASK 1 - Gregorian VS Roman, using only the Offertory liturgical function;

TASK 2 - Gregorian VS Roman VS Ambrosian;

TASK 3 - Offertories VS Graduals, inside both Roman and Gregorian families.

These different tasks have been selected for different purposes. In the first (simpler) one, we compare Gregorian and Roman families using only the Offertory type. This choice has been made for three distinctive reasons: Offertories are the most numerous pieces both in the Roman family and in Gregorian one; the length of each Offertory is typically higher than the other types; the corresponding intoned verbal texts is almost completely identical in all Offertories and thus all differences of the two corpora should lie in the melodic component only. This choice allowed us to test the method on a homogeneous group of pieces, without introducing potential variability dictated by the different liturgical functions. The next two tasks aim to classify the full three repertoires: in each of them different liturgical functions are represented (Introit, Alleluia, Gradual, Psalmellus, Tract, Sequence, Offertory, Communion, Antiphony, Hymn, Canticle) and thus the three groups are less homogeneous both melodically and in terms of lengths of the pieces. In addition, there are some cases (mostly in the Alleluia), where a long stretch of a melody is shared between different repertoires, making a meaningful automatic attribution more delicate in this case. The last task is divided into two parts: we compared Offertories and Graduals inside each of the two repertoires Gregorian and Roman. This task is thus intended to evaluate whether the quantitative analysis is able to confirm the stylistic differences between Graduals and Offertories (inside the Gregorian and the Roman family respectively) observed by musicologists: both liturgical genres are in fact well characterised at a stylistic level, to the point that they are considered to form homogeneous musical subgroups within the two repertoires.

Note that, as the voting scheme is a binary classifier, we split Task 2 (that compares the three families) into three different binary tests. Moreover, whenever the two classes have different cardinality, we randomly sample the larger one to obtain two corpora of the same size. We use the same procedures for the nearest neighbour scheme for consistency.

As expected, in all tasks, for large n -grams ($n \geq 10$) the number of unreliable classifications (i.e. $v = 0$) grows consistently and thus the accuracy rapidly drops to zero. For smaller n instead we obtained very high scores in all tasks examined (see Fig. 4). In particular in the first two tasks the percentage of correct attribution is above 90% for all n -gram lengths in the range $2 \leq n < 10$, with some peaks of 100% for $n = 2, 3$ and a slight drop (but still above 85%) for $n = 6$. Also in Task 3, aimed to distinguish between Offertories and Graduals, the results are all above 85% in the interval $2 \leq n \leq 8$ (with the only exception of $n = 2$ for the Roman family).

We further note that (see Fig. 5) in the Gregorian VS Roman task the best results are obtained for rather small n : this is in accordance with the fact that all repertoires are strongly characterised by short melodic units. The diminished accuracy for $3 \leq n \leq 6$ could be explained by the presence of pieces of different liturgical functions that are easily contaminated between different repertoires (Alleluias in particular). On the contrary in the more refined Task 3, as the short melodic units that were able to distinguish

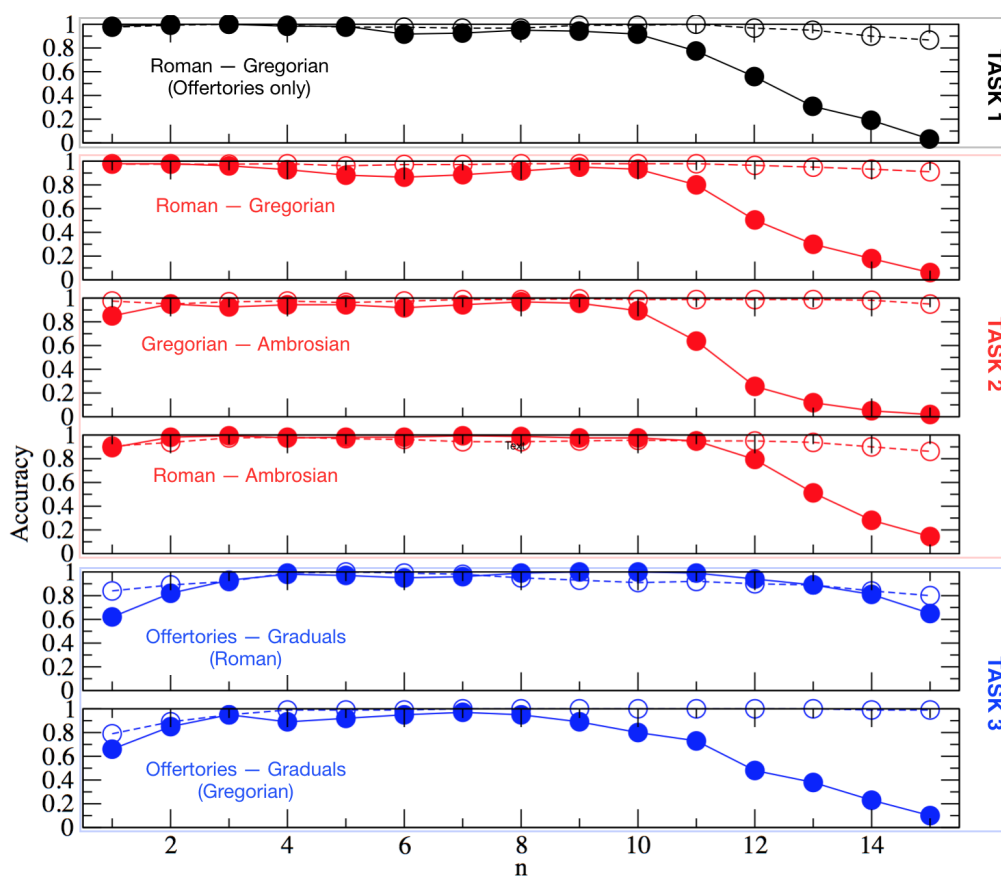


Figure 4 Accuracy (as percentage of correct attributions) for the three different tasks as a function of the length of the n -gram. Bold circles correspond to the voting scheme while open circles correspond to the nearest neighbour scheme.

between Gregorian and Roman chants are shared in the Offertories and Graduals, the distinction between these two liturgical functions should be made on longer melodic ‘phrases’. It is somehow surprising that the automatic quantitative method is capable of grasping differences that only in-depth philological studies are able to highlight.

It is also important to note that exceptionally good results (and in some cases the best ones) are obtained even with the simple nearest neighbour procedure. From one side this result is consistent with the observation that it is not unusual in these repertoires to adopt, re-use, and share the same melodic chunk, even of moderate length. On the other hand, we should not be tempted to conclude that the nearest neighbour procedure is always the best one in these kinds of repertoires: once compared with an attribution of a piece of uncertain monodic genre or of more distant origin, it is not at all clear that a very similar piece can be found. We thus consider more suitable to rely on the more robust voting scheme in the generic situation.

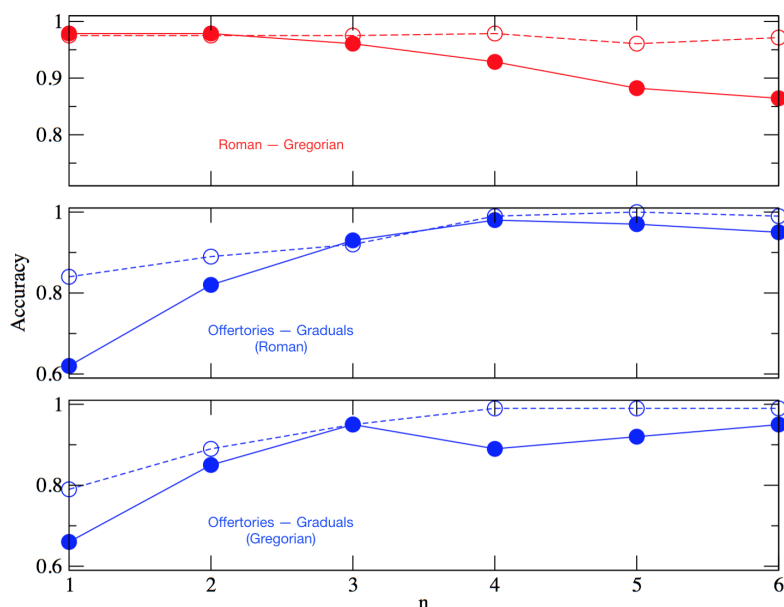


Figure 5

Accuracy for small n : accuracy diminishes as n grows in the general task Gregorian VS Roman, it increases with n in the more refined task Offertories VS Gregorians.

7. n -gram statistics at the musicologists' disposal

While the results presented, which extend the already encouraging results of (Unguentoli, Cristadoro, and Beghelli 2017), showed the accuracy of the method based on the n -gram statistics for the automatic classification of simple unidimensional (monodic) musical pieces, no real authorship (or stylometry) problem was at stake. On the other hand, we expect that a closer inspection of some cases could offer to the musicologists stylistic insights, in particular for the few pieces that remain of uncertain classification in each of the inspected families³.

We expect that real uncertain attribution problems could benefit from a similar analysis, in particular when issues arise from composite manuscript codes that are born out of contaminations of different repertoires.

Beside that, even a simple statistical analysis could be useful to the expert of medieval monody, in particular on the aesthetic side concerning stylistic aspects of an entire liturgical family. They could indeed shed some light on the somehow surprising effectiveness of the automatic attribution previously discussed and highlight some aspects that could otherwise pass unnoticed. The potentiality of a simple statistical analysis can be appreciated in the following preliminary results:

³ As an example, the Gregorian Alleluia *Iuravit Dominus* [(Graduale Triplex, 1979), pag 489] is erroneously attributed to Roman tradition for all n -gram lengths from 2 to 4 in both classification schemes: this could be of interest for the expert that could eventually evaluate if this error is simply caused by its exceptional shortness (and thus simply due to lack of statistic) or eventually consider it as a signal of strong contamination of genres, and possibly use such data to reconsider the philological analysis of its genesis, authenticity, diffusion.

- The larger melodic interval present in all pieces is of 9 semitones both ascending and descending, in all three families of liturgic monodies. This is the amplitude that, in tonal register, is called ‘major sixth’ (C–A ascending, B–D descending, etc) with no differentiation inside Ambrosian, Gregorian, Roman. This observation confirms the accurate analysis done by the palaeographer Will Apel who noticed that in the Gregorian repertoire there is an unique interval of seventh minor precisely inside the Offertory *Domine Deus meus* [(Apel 1958), p. 364], a piece which is not inside our corpus; our data extends the validity of the major sixth as maximum interval to the other two monodic repertoires.
- For each of the three families, we merged all pieces to construct a single long sequence. We then compute their distance as an indication of the overall similarity of the three corpora. For $n = 2$ (but similar results are obtained for $n = 3, 4$) we obtain that the distance between Gregorian and Roman (0.412) is significantly greater than the distance between Ambrosian and Gregorian (0.325) or Ambrosian and Roman (0.333). Note that these last two distances are very similar. Ambrosian repertoire thus lies in between Gregorian and Roman ones, that are instead separated by a larger distance. Indeed we know that, over the centuries, contrasting pressures were imposed on the Milanese diocese to adapt its melodic style to the Roman or Gregorian tradition alternatively.
- A first quantitative estimate of the melodic richness of a corpus is given by the cardinality of the set of different n -grams it contains (its total dictionary). We find that the cardinality of the dictionary of the Gregorian is much bigger than the Roman one (starting from 20% more at $n = 2$, to 65% more at $n = 4$), while their total length is more or less the same. This is in perfect agreement with classical analyses made by experts of medieval monody. The Ambrosian appears richer than Roman even if its total length is smaller.
We also computed the peculiarity of a given Gregorian family by counting the number of n -grams that are unique, that is that they are not present in both the other two repertoires: we obtained that 25 over 1554 of the total different 2-grams of the Gregorian are unique, 259 over 819 of the 3-grams and 1071 over 3128 of the 4-grams; for comparison similar counts for Ambrosian and Roman gave between 3 and 4 times smaller results.
- It is interesting to compare also the most frequent n -grams among the three families. Note that, by construction, these frequent n -grams do not give a significant contribution to the n -gram distance we used in the attributions tasks. In the Ambrosian and Roman corpora, for the case $n = 3, 4$ (corresponding to melodic segments of 4 and 5 notes respectively) the first 10 most frequent n -grams are all given by sequence of movements by joint degree – tone or semitone – as $[2, -2, -2, 2]$ which can correspond for example to D–E–D–C–D. This last one is by far the most frequent n -gram in the Roman corpus and the second most frequent in the Ambrosian one. In the Gregorian corpus we have that for both 3-grams and 4-grams the most frequent segment is given by the repetition of the same note (4 or 5 times respectively). There are indeed neumes like the *bistropha* and the *tristropa* that are characterised by the repetition of the

same note on a single vowel: as an example in Fig. 6 we can see the three F that repeat the 'u' in 'dum' written in the medieval notation as a single neume of three notes (tristropha), that underline the unity of that tripartite melodic movement; on the other hand the three A that, right after, accompany the words 'in iudicio' are three distinct and independent notes that belong to the vowels of the three distinct syllables 'in', 'iu' and 'di'. It is interesting to observe that in the other two repertoires, such sequence of repeated notes are not among the 10 most frequent n -grams.

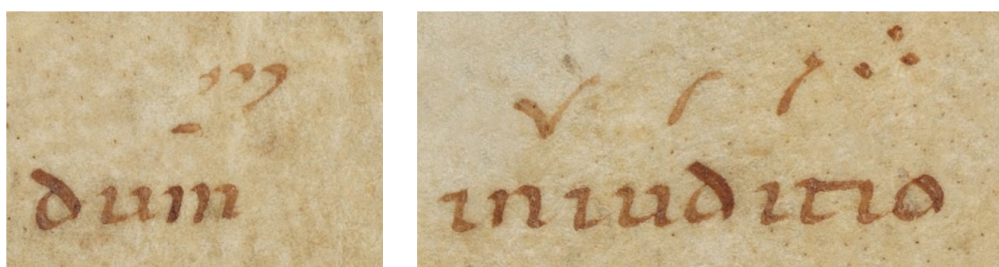


Figure 6

Two fragments of the Easter Offertory *Terra tremuit et quievit*, from the Gregorian manuscript Einsiedeln, Stiftsbibliothek, Codex 121 (f. 207), dating back to 960-970: in the paleographic adiaستمatic notation (written without staff lines) the three unisons of 'dum' are indicated with one single *tristropha*, a repercussive neume, while the three unisons of 'in iudicio' are indicated with three different *virgae* (/ / /), the first one joined at the previous note. Cf. Fig. 2, where the two different musical symbols are both transformed in the same one in the modern notation.

- Among many more, we selected two examples of melodic fragments that we believe are of particular interest. The first one is the concatenation of two descendant thirds, one minor, the other major: in our notation they correspond to the 2-grams $[-3, -4]$ (i.e.: G-E-C) and $[-4, -3]$ (i.e.: A-F-D). As already noted by Apel, these are fairly common sequences in the Gregorian tradition together with their corresponding symmetric ascending ones. In the other two repertoires instead, while the ascending sequences are quite frequent – but still less than in the Gregorian – the two descendants are very rare or even absent. The 2-gram $[-3, -4]$ appears 159 times in the Gregorian, 8 in the Ambrosian and 7 in the Roman, while $[-4, -3]$ appears 152 times in the Gregorian, 37 in the Ambrosian and is absent in the examined compositions of the Roman. The second example concerns two sequences, $[-2, 6]$ and $[6, -2]$, that are present respectively 31 and 29 times in the Roman, but they appear both only once in the Ambrosian and are absent in the Gregorian. The interval of six semitones is the so called tritone and the corresponding sequences are typically G-F-B and F-B-A. Note however that the edition of the Roman corpus inside the *Monumenta Monodica Medii Aevi* can be considered as a diplomatic version, and consequently no question about B flat is addressed: the presence of the mentioned 2-gram could thus indicate an effective melodic feature of the Roman, as well as a textual error due to the lack of a complete critical revision of the corpus.

8. Conclusions and perspectives

The investigation presented has to be considered as a first simplified test for the effectiveness of a quantitative approach to musicological issues, even if – as already mentioned – no real attribution problem was at stake. On the other hand, the results presented here, with accuracies around 90%, are encouraging and we hope that a similar approach can be adopted also for more complex musical compositions, characterised by different superimposed voices with the consequent formation of units of semantically meaningful vertical aggregates. To this aim, it is apparent that a crucial step will be the adaptation – in a non-trivial and useless way – of the n -gram coding to a dimension that is no longer simply linear. We plan to develop this program starting from cases of increasing complexity.

Note however that in order to test novel computational approaches we should firstly face the problem of the lack of well-formed databases of digitalised music. While databases of audio files for the main musical compositions of different epochs are accessible, the writing of such music is mostly available only in analogue format. Moreover some interesting corpora are in the musical symbolic representation used on paper less than four centuries ago (as we have seen, the manuscript of the liturgical monodies that we used use different writing systems from the modern one). The digitalisation of hundreds of musical compositions for a total of a few million notes is a large scale investment, in particular if the only available sources are ancient manuscripts. Note that, once the difficulties concerning databases availability are solved, several attribution problems are still open among musicologists where a contribution from a quantitative approach can be relevant. Among them we can just mention the possibility to distinguish between different authors in compositions that are the result of collaborations or to evaluate the degree of reliability of the attribution of a composition of uncertain paternity.

We list below a few concrete problems. They are identifiable in compositional genres much later than the medieval liturgical monody but with a slight increase in complexity concerning their musical structure (we plan to investigate them in forthcoming works):

- The *secco* recitative commonly adopted in Italian opera between the eighteenth and nineteenth centuries. In this case a very stereotyped vocal line is accompanied by long notes of the instrumental bass, that simply underline the concomitant harmony already implicit in the vocal part. Gioachino Rossini (1792-1868) was the most acclaimed opera composer among his contemporaries. He had to face so many requests that he was often forced to ask help from collaborators, mainly for the secondary parts (the recitatives and the arias of minor characters). Whenever the autograph manuscript is lost, and thus the graphic evidence of different alternating hands is missing, a quantitative approach could support the musicologist in identifying the recitatives not composed by Rossini.
- Opera at its origin (seventeenth century) has a slightly richer structure than that of the recitative of the next centuries: while the vocal line is more varied and personalised and the instrumental accompaniment is much more lively and complex, everything is almost always resolved in only two melodic lines (singing line and the bass). Claudio Monteverdi (1567-1643) is the most brilliant composer of the Baroque period. To him is traditionally attributed the opera *L'incoronazione di Poppea* (1643), staged

while he was ill, a disease that led to his death. No documentation attributes that work to him with certainty; on the other hand, we know for sure that some scenes were not written by him but by colleagues and disciples. Could a quantitative approach compare this opera with other works of Monteverdi and suggest attributions of the different parts of *Poppea*?

- Structurally similar to opera at its origin are chamber cantatas, flowering during the seventeenth century; many are found in handwritten anthological collections without indication of author or with dubious, multiple attributions. Could a quantitative comparison with the cantatas of sure paternity tell us something more about the adespote ones?

As should be now clear, we expect that a quantitative approach will provide the musicologists with nothing more than an evaluation tool, and not definitive truths. Results should always be compared with the analyses offered by historical-aesthetic notions, documentary proofs, and traditional analytical methods based on reading the written text and listening to the music. For that reason we stress the importance of the use of interpretable methodologies as opposed to very performative ‘black boxes’: the automatic approach does not seem to match with the need of the investigations of the venerable but narrow *Musikwissenschaft* of positivist implantation.

References

- Apel, Willi. 1958. *Gregorian chant*. Indiana University Press, Bloomington.
- Backer, Eric and Peter van Kranenburg. 2005. On musical stylometry - a pattern recognition approach. *Pattern Recognition Letters*, 26:299–309.
- Basile, Chiara, Dario Benedetto, Emanuele Caglioti, and Mirko Degli Esposti. 2008. An example of mathematical authorship attribution. *Journal of Mathematical Physics*, 49(12):125211, 1–20.
- Brinkman, Andrew, Daniel Shanahan, and Craig Sapp. 2016. Musical stylometry, machine learning, and attribution studies: A semi-supervised approach to the works of Josquin. In *Proceedings of the 14th Biennial International Conference on Music Perception and Cognition*, pages 91–97, San Francisco, July 5–9.
- Doraisamy, Shyamala and Stefan Ruger. 2003. Robust polyphonic music retrieval with n-grams. *Journal of Intelligent Information Systems*, 21(1):53–70.
- Graduale. 1979. *Graduale Triplex seu Graduale Romanum Pauli PP. VI cura recognitum & rhythmicis signis a Soles-mensibus monachis ornatum neumis laudunensibus (cod. 239) et Sangallensibus (codicum San-gallensis 359 et Einsidlensis 121) nunc auctum*. Abbaye Saint-Pierre de Solesmes, Solesmes.
- Hillewaere, Ruben, Bernard Manderick, and Darrell Conklin. 2010. String quartet classification with monophonic models. In *Proceedings of the 11th International Society for Music Information Retrieval Conference*, pages 537–542, Utrecht, Netherlands, August 9–13.
- Kešelj, Vlado, Fuchun Peng, Nick Cercone, and Calvin Thomas. 2003. N gram-based author profiles for authorship attribution. In *Proceedings of the Pacific Association for Computational Linguistics Conference (PACLING '03)*, pages 255–264, Halifax, Canada, August, 22–25.
- McKay, Cory, Tristano Tenaglia, Julie Cumming, and Ichiro Fujinaga. 2017. Using statistical feature extraction to distinguish the styles of different composer. In *Medieval and Renaissance Music Conference 2017*, Prague, Czech Republic, July 4–8.
- Offertoriale. 1985. *Offertoriale triplex cum versiculis*. Abbaye Saint-Pierre de Solesmes, Solesmes.
- Orio, Nicola. 2006. Music retrieval: A tutorial and review. *Foundations and Trends in Information Retrieval*, 1(1):1–90.
- Stablein, Bruno. 1970. *Die Gesänge des altromischen Graduale*, Vat. lat. 5319. Bärenreiter, Kassel.
- Suñol, Gregorio Maria. 1935. *Antiphonale missarum juxta ritum Sanctae Ecclesiae Mediolanensis*. Desclée, Rome.
- Unguendoli, Francesco, Giampaolo Cristadoro, and Marco Beghelli. 2017. Applicazione di un metodo attribuzionistico quantitativo alla monodia liturgica medievale. In *Proceedings of the*

- Fourth Italian Conference on Computational Linguistics (CLiC-it 2017)*, pages 327–332, Rome, Italy, December 11-13.
- Velarde, Gissel, Tillman Weyde, Carlos C. Chacon, David Meredith, and Maarten Grachten. 2016. Composer recognition based on 2d-filtered piano-rolls. In *Proceedings of the 17th International Society for Music Information Retrieval Conference*, pages 115–121, New York City, USA, August 7-11.
- Wolkowicz, Jacek and Vlado Kešelj. 2013. Evaluation of n-gram-based classification approaches on classical music corpora. In *Mathematics and Computation in Music. Proceedings of the 4th International Conference, MCM 2013*, pages 213–225, Montreal, Canada, June 12-14.

