International Winterschool Digital Musicology – Digitalisierung in der Musikwissenschaft

Musikwissenschaftliches Institut, Ebernhard Karls Universität Tübingen

18.–19. November 2016

Kate Helsen and Inga Behrendt

The Optical Neume Recognition Project

The original presentation of this project began with singing: the ideal way to at once situate our subject in its natural habitat and demonstrate the difficulty of transcribing *sound* with *signs* and then, for our purposes, extracting *meaning* from those signs, one thousand years later. By singing the chants as they are presented in the earliest notational form we have in Western Medieval culture, we turned our attention to the details of how these neumes are rendered, graphically, knowing this is the key to understanding the musical culture from which they came. The problem, historically, with achieving a reliable idea of this is the fact that there are too many neumes, all rendered according to various scribal traditions, for one person to get an accurate overview. Because of this, some notation traditions have been very well-studied, but others hardly at all.

The Optical Neume Recognition project uses software in a three step process: 1. to identify discrete neume shapes in scanned images (using a classifier), 2. interpret them as musical directives set out in a Neume Table and, 3., express these musical directives in MEI (the xml developed specifically for music notation, one schema of which has been adapted for staffless neumes.) The information gathered from this project can be used to speed up the time it takes researchers to compare old and new notations, isolate differences in chant melodies, and compare adiastematic neumes to early staff notation. The manuscript we have chosen to focus on in the development phase is St. Gallen manuscript 390 / 391 (Hartker's Antiphoner) because it contains a consistent, relatively clear primary hand, and a similarly consistent secondary hands added later. It is also well-known as one of the earliest and most complete examples of the important scribal tradition of St. Gallen and its notation has been relatively well-studied in the past, giving us a bench-mark with which to compare our results.

This project has grown over the past decade to include not only Inga and myself (the first, musicologist-dreamers,) but also Jennifer Bain (Dalhousie), Andrew Hankinson (Oxford), Ichiro Fujinaga (McGill), and other members of the development team at McGill University in Montreal. The project was initially funded by a Canadian government granting agency for two years, and is now part of a larger project called SIMSSA which also secured substantial funding from the same body.

The project's first challenge was to process the images of each folio in the most advantageous way for the neumes to be identified. This was accomplished using SIMSSA's 'Rodan' system.[1] Here, binarization of originally colour images is completed using algorithms specifically optimized for each manuscript it works on. As a chain of functions, this system cleans up the binarized image, crops the margins, and prepares each page for further analysis.

The more musicological considerations undertaken initially involve the interpretation of neumes as musical directives. For musicologists, the process of encoding Hartker's neumes into MEI has meant crystallizing an understanding of what these signs represent, and how best to distinguish

[1] Hankinson, A. 2015. "Optical music recognition infrastructure for large-scale music document analysis." Ph.D. diss., Schulich School of Music, McGill University

between musical information and musical interpretation. The MEI schema developed for twelfth-century neumes by Stefan Morent and his team[2] is an important step forward, but Hartker's neumes differ from this type in three fundamental ways: 1) since they are staffless, the notation contains no information about absolute pitch; 2) each neume is written with adjacent neumes more or less on a horizontal plane, rather than in vertical relationship to each other as is found in staff notation; and 3) minute alterations to the shape of each basic neume provides information about the number of pitches the neume represents, the melodic direction from one pitch to another, as well as non-pitch performance indications, such as speed of delivery and duration. A musicologist understands neumes by thinking about their relationship to the actual chant melody they depict, and how these neumes enrich the chant text by heightening its meaning. This assumes a kind of connectedness between neumes, from start to finish of both chant melody and text that is natural for our minds, but not for a computer. The challenge of describing not only what the neume itself means, but also any information it contains about its own context, must be very carefully met.

Before the meanings of neumes can be satisfactorily dealt with, however, the neumes must be successfully identified first. This is done using a neume classifier, developed at the Music Tech lab at McGill based on Gamera. The classifier works with the user to 'learn' how to 'recognize' what it 'sees' when it scans an image. First, it identifies all the graphic shapes it can recognize as such, using bounding boxes to delineate each one. Then it arranges these shapes so that the ones most similar are placed next to each other, and the more different ones are given farther down the list. By clicking on any particular shape, the user is brought to that particular shape in the larger image, in order to preserve context. The user is then charged with the task of naming, or

[2] http://www.dimused.uni-tuebingen.de/tuebingen.php

categorizing, each shape shown in the bounding boxes by clicking on the shape. After most of the shapes have been categorized, the user can resubmit the image to the classifier, helping it 'learn' how to categorize shapes on the next page. After several pages have been completed like this, the classifier is exponentially more adept at categorizing subsequent pages, although a human musicologist should still be present to check for errors and correctly categorize rare neumes that may not yet have been encountered by the system.

In working out a description system for Hartker's neumes for a computer, we needed to figure out a way to describe the meaning of the neumes without having knowledge about the absolute pitches they represent, when sung. Since they do not arrange themselves vertically in such a way that would encourage this thinking – indeed, some types of neumes (and Hartker's are one such notation style) have particular axes to the slant of the scribal hand, which would obscure this further. But if we use characteristics that give information not about absolute, but relative pitch height, such as N (neutral, or unknown), H (higher), L (lower), S (same, or unison), A (same, or higher), and U (same, or lower), it will be then possible in the future to compare what has been rendered in adiastematic notations such as Hartker with later notation on staff lines. The *Punctum* (), for example, represents a note that is undefined (neutral) as to its relationship (in pitch height) to the following note and is therefore named "N". We begin each name with the number of pitches the neume represents overall, so the Punctum is "1-N". Lower-case letters represent descriptions concerning the rendering of parts of the neume in question: w (wavy), b (curved anticlockwise), c (curved clockwise), a (angled), e (episema), f (flat), j (jagged), l (liquescent), x (extended), y (diagonal right up), k (diagonal right-down), q2 (quilisma 2 curves), q3 (quilisma 3 curves). A *Tractulus* () is given as "1-Nf" because it represents one, neutral pitch and it is flat, graphically. A *Tractulus with an episema* () is „1-Nfe“ because in addition to being one neutral pitch that is flat, graphically, but it ends with an episema. The *Gravis* () is described as „1-Nfk“ because it is a single, neutral pitch that is flat and slanted down and to the right. The *Stropha* () as „1-Nc“ because it is a single, neutral pitch that is curved clockwise.

In the 12th century we already see neumes made out of several single components be understood as one Neume, with one name.[3] The graphic  is not called *Virga two single Puncta,* but got the name *Climacus.* It represents three single descending notes. In our codes we show the separation of single components within neumes with the letter g for “gapped”. This means that the climacus is expressed: “3-Nyg Lg L”.

Furthermore, neumes, after which the melody continues higher or lower, are given the final letter d (down, afterwards lower) or u (up, afterwards higher). Some scribes have the habit to add little indications, in the form of alphabetic letters, to their neumes; these letters – *litterae significative* – contain performance directives for the notes they belong to. For example, the letter “c” stands for *celeriter* (quick). These added little letters are recognised in our codes, in case of “c” it would be given as “p:c“.

Although we may think that Hartker has yet to reveal all his secrets, those who study other chant repertoires deal with even deeper mysteries. For example, the Old Hispanic chant from the

[3] Vgl. Constantin Floros, *Universale Neumenkunde. Ursprung und Deutung der Lateinischen Neumen* (Band 2), Kassel 1970, S. 184-207.

Iberian Peninsula, cannot be melodically recovered like other Roman chant repertoire, because there is no notation system in that tradition that was allowed to develop beyond the *campo aperto* stage; we cannot simply retrace its evolution from chants depicted in staff notations, with accurate and clear intervals, etc. Old Hispanic chant exists in five extant manuscripts, the earliest of which dates to the 10th century. There are only about a dozen chants in the entire repertoire for which there is a known melody; the rest lies seemingly irrecoverable. This is the challenge that Emma Hornby and Rebecca Maloy have undertaken in their research projects called *Compositional Planning, Musical Grammar and Theology in Old Hispanic Chant* (from 2009 to 2011,) culminating in a book on the subject in 2013.[4] This project is devoted to isolating and documenting the frequency of usage for every glyph and glyph combinations, looking for patterns and positions that might suggest meaning. Through this kind of 'detective work', they are now able to piece together some of the rules or 'grammar' for Old Hispanic notation, without the help of absolute pitches. Using melodic contour shaped from assigning parts of the Old Hispanic glyphs letters that represent relative pitch heights, Hornby and Maloy could describe each shape in a unique way and catalogue it using a computerized system similar to our classifier. Once all the glyphs were described like this, one could "zoom in" to study only the modifications of the certain shapes or the use of certain groups, which will lead to new discoveries. If the same melodic contour is represented by different shapes, what does this mean? By isolating all the various ways to depict a certain melodic motion (i.e., neutral – lower – higher) they could determine if there were patterns that came more frequently, or patterns that could theoretically appear but did not. They could also look at the melodic motion paired with the chant text to see if certain words were consistently set to particular melodic gestures. Overall,

[4] Emma Hornby and Rebecca Maloy, *Music and Meaning in Old Hispanic Lenten Chants: Psalmi, Threni and the Easter Vigil Canticles* (Suffolk, UK: Boydell & Brewer, 2013)

it is clear that understanding the concept of the chant notation is possible even when the melodies themselves may be lost, if one asks the right questions.

Hornby and Maloy’s project on Old Hispanic neumes has been running simultaneously with our project on Hartker’s neumes for several years and we have been actively sharing ideas and solutions for common problems. While our projects are currently separated by the repertoire we focus on, we hope that our description system is broad and comprehensive enough that one day our digital tools will be one-size-fits-all.

Setting up an encoding system that documents the variety and complexity in neumes is a challenge. However, the capacity of a computer to deal with large amounts of data (now called by industry, “Big Data”) has always excited medieval musicologists. At first, this capacity was generally restricted to database technology, but combining document analysis capabilities with metadata about the musical notation offers new avenues and an exciting future. In the lifetime of this project, literally hundreds of medieval musical manuscripts, fully scanned, have come online through both national initiatives like Switzerland's e-codices, and library- or archive-based projects like Gallica (BnF). These span the entire medieval European gamut, in terms of regions and scribal traditions represented. Musicologists will not be able to effectively use what is now given to them, free, unless we develop ways to automate the steps we take when dealing with this amount of information. If the goal is to achieve an accurate, useful overview of scribal methods and neume variation across geographical areas and time periods with the images now plentiful on the internet, we are going to rely on digital instruments to help us organize what we see. From there, of course, it is up to us, as musicologists, to make conclusions, test our hypotheses, and search for further clarification.

The encoding process starts with activating a software program that classifies neumes according to a list provided by the musicologist. In our case, this list of possible neumes, simply for Hartker, at one time exceeded 300 components or groups of components. Usually, there are broader headings for standard neume types and then subsets of that type, modified in some way (such as by the addition of an episema or a liquescence.) Once the neumes in the image are classified, the computer needs a system by which to encode information about what each component means. We have chosen the Music Encoding Initative (or MEI) to be this system. A short precis concerning MEI follows, for those who are less familiar with it.

MEI is an encoding system according to a universally agreed upon, but ideally flexible, 'grammar'. There is a wide-spread community of MEI 'speakers' – librarians and musicologists and technological developers, etc. - and even a small regulatory board who oversees the creation of new methods of expression, adjusted to the particulars of certain kinds of notation, such as tablature, mensural notation or, as Stefan Morent's TuBingen Project illustrated, staffed neumes. In fact, Stefan Morent was the first to use MEI for chant notation on staff lines.[5] The rules that govern MEI are hierarchical because it is part of the XML family of computer languages: 'Extensible Mark-up Language'. "Marking up" a piece of text for the purposes of good formatting is a hierarchical exercise; a page of text could include a title, under which there might be a subtitle, under which there will be a paragraph, within which are sentences, within which are words, within which are letters, which, on a computer, are made out of a particular arrangement of black and white pixels... and so on. The term "extensible" in XML means that this hierarchical thinking can be expanded to include other kinds of graphics, alphabets, pictures, etc.

[5] http://www.dimused.uni-tuebingen.de/hildegard_e.php

Anything that can be described as a set of hierarchical relationships can be rendered as XML. The figure below shows the MEI necessary to render the pair of eigth notes on the right.

```
<beam>
  <note xml:id="d1e129"
      stem.dir="up"
      pname="f"
      dur="8"
      oct="4" />
  <note xml:id="d1e130"
      stem.dir="up"
      pname="c"
      dur="8"
      oct="4" />
</beam>
```

In the MEI code, the line farthest to the left shows that the beam includes two pitches, and each note in that beam are described using characteristics of stem direction, pitch name, duration, and octave. When describing neumes in MEI, we cannot provide information about pitch name or duration, but we can describe the neume in terms of direction, angle, and graphic rendering (‘angled’, ‘rounded’, etc.) similarly to the way Hornby and Maloy describe Old Hispanic notation.

There are two main reasons why our project has chosen to adapt MEI specifics to the demands of neume notation. First, This language is 'open source'. This is not simply about cost; open source is a powerful, almost political, statement about our intentions for the future. Anyone is able to access the source material for this language, and use it, adapt it, build with it, and integrate it into whatever is coming next. No one owns the rights to use this language, nor can they sell them, or limit what others wish to do with it. Secondly, MEI is quickly becoming the standard for expressing musical scores in a machine-readable format and is the focus of many international conferences and publications on the subject. The more digital images of manuscripts freely available online, the more important it is to agree on how we will encode their contents. The advantage to digital encoding is that, one day, we will be able to compare vast amounts of musical data - chant with chant, whole manuscript with whole manuscript - and isolate variants from thousands - if not millions - of samples. However, this will only be possible if all the encoding meets certain standards before that day arrives. Preparing for this kind of world, where high res images from every page of the majority of medieval musical manuscript are available instantly means that more musicologists will feel capable of using these rich resources.

Beyond the philosophy behind MEI, our project's methods bear out the Open Source idea. We use GitHub and other open platforms to document our workflow; problems are identified, assigned to a specific person for resolving, and when a fix has been found, that is also 'committed' in a documented, open way. This method of working has obscured the idea of 'Intellectual Property' as belonging to a certain person and not to others, which is an accurate reflection of how we see our work; it belongs to all of us, and to those who want to use it, and to those who come after us. Because our project team members are spread over several Canadian provinces and two European countries, most of our interaction takes place over Skype and email, but we do see each other in person once or twice a year. We find that this open attitude is the most helpful in this online environment, where many projects are expanding their reach and capabilities, finding themselves cross-linked with other projects. The interplay between the specialization of some projects with the universally-applicable elements of others, all searchable on 'clearing-house' websites, or hyperlinked between them, means that the old model of one musicologist working on his or her single project is essentially dead. We must cooperate, share, help each other, and learn from each other's challenges in order to build a strong foundation for the next generation of digital-native musicologists.