Helsen (kate@helsen.org); September 26, 2011

'*Venite et videte*': First Results in the Optical Neume Recognition Project[1]

The media theorist and philosopher Marshall McLuhan once suggested, "It is the *framework* which changes with each new technology and not just the picture *within* the frame."[2] This idea of a changing framework describes the spirit with which my project partners and I began the Optical Neume Recognition Project a year ago. New technology helps recast some longstanding questions, and this change in "framework" results in a change in *how* we look at "the picture within the frame". The Optical Neume Recognition Project offers the chance to look at musical manuscripts differently and consequently, to ask different kinds of questions about them. This introduction to the project includes a discussion of previous research in the area; the project's methods, goals and challenges; and a description of its current successes.

To begin, I would like to give a clear definition of what the Optical Neume Recognition Project is, and what it is *not*. Our goal is to provide the musicologist with a new *tool* for research, one which offers an objective and efficient look at the graphic aspects of early notation. It does *not* seek to replace semiological studies with a piece of software. It does *not* propose to automate the analysis of a medieval musical document. It does *not* substitute for any inquiries into notation or scribal practices by the human, musical mind. Instead, it is meant to *expand* them. Simply put, our project uses optical document analysis software to identify discrete neume shapes in early medieval notation, in this case, St. Gallen scribal tradition. All aspects of categorization, analysis and interpretation of that notation remain the work of the

---

[1] A slightly edited transcript of the presentation at *Cantus Planus* 2011, Vienna, Austria

[2] Eric McLuhan & Frank Zingrone, eds. *Essential McLuhan*, (New York: Routledge, 1997), 273.

musicologist. In fact, a musicologist working with this kind of technology has no choice but to re-examine many aspects of notation. Software cannot ignore certain aspects of a problem "just for now" the way people can, nor does it jump from "what is there" to "what I *see* there" the way people are sometimes tempted to. It is up to the musicologist to define and determine *every* aspect of the image, even those which may not seem important initially. In a recent paper on the requirements for digital encoding of medieval liturgical manuscripts, Michael Norton wrote, "I needed not only to represent what I *knew*; I needed also to represent what I did *not* know (or did not care about). In a very real sense, I needed to represent my ignorance."[3] The Optical Neume Recognition Project offers a new opportunity to consider what we know, what we don't, and what *we don't know we don't know*, about early medieval notation.

Over the past twenty years, there have been several attempts to use computer technology to further our understanding of medieval notation. One of the first experiments involving the application of computer optics and images of medieval notation was undertaken at the University of Ottawa. In a 1991 article entitled, "The Optical Scanning of Medieval Music,"[4] McGee and Merkley described how eliminating the staff lines in images of late medieval square notation increased the computer's ability to recognize graphic patterns. They found that the result of stripping away the staff lines left the square notation looking "something very much like its cheironomic counterpart"[5] in which it was easier for the software to isolate discrete shapes. It seems that, despite subsequent interest in using optical software

[3] Michael Norton, "Representation, Interpretation, and Integration: A Layered Architecture for the Encoding of Medieval Liturgical Manuscripts" read at *Congress on Medieval Studies*, Western Michigan University (7 May 2005), 2.
[4] William McGee and Paul Merkley, "The Optical Scanning of Medieval Music" in *Computers and the Humanities* 25: 47-53, 1991.
[5] Ibid, 51.

on later musical notation, it was understood from the beginning that the earliest, staffless forms of notation had the advantage over later notation because of its graphic lay-out on the page.

Just over a decade ago, several projects were begun which used specific XML schemas, or mark-up languages, to represent musical notation (including the medieval kind) in standardized machine-readable formats. In 1999, the University of Virginia developed an XML schema called the Musical Encoding Initiative (or MEI) for encoding everything from medieval notation on staff lines to modern, printed scores.[6] That MEI was recently put to use by Stefan Morent and Gregor Schräder[7], working on a project called *TüBingen,* which depicts images of Hildegard of Bingen's chants digitally and connect them to other types of media, like video and transcription software.[8] At the same time, another project entitled NEUMES[9], housed at Oxford, began designing a different XML schema for the digital transcription of early chant. The researchers behind NEUMES, Barton, Caldwell and Jeavons, were interested in standardizing neume graphics digitally, because they maintained that optical character recognition software is not useful for *non*-standardized graphics. For NEUMES, the kind of hand-written neumes one encounters in a medieval manuscript had to be rendered in a standard way in order for the computer to process them. However, our Optical Neume Recognition Project focuses precisely

[6] Roland, Perry and J. Stephen Downie, 'Recent Developments in the Music Encoding Initiative Project: Enhancing Digital Musicology and Scholarship' *Digital Humanities* 2007 (poster). See website: http://www2.lib.virginia.edu/innovation/mei/Intro/

[7] Gregor Schräder, 'Ein XML-Datenformat zur Repräsentation kritischer Musikedition unter besonderer Berücksichtigung von Neumennotation' Studienarbeit, Uni-Tübingen (2007): http://www.dimused.uni-tuebingen.de/downloads/studienarbeit.pdf

[8] http://www.adwmainz.de/fileadmin/adwmainz/MuKo_Veranstaltungen/S2-Digitale_Medien/TueBingen.pdf

[9] Louis W.G. Barton, John A. Caldwell, Peter G. Jeavons, 'E-Library of Medieval Chant Manuscript Transcriptions' in JCDL (Denver, Colorado, 2005).

on that non-standardization which NEUMES considers to be a flaw: our software attempts to recognize each neume shape in *whatever way* it has been drawn.

In the last several years, the attention of those working with optical document analysis software has been drawn to *printed* sources of early music, because of this advantage of standardization. The project Aruspix, at the University of Geneva, has developed music scanning software for early music prints which "learns" dynamically from corrections made by the user.[10] At McGill University in Montreal, a project called Gamut[11] is currently using Aruspix, as well as another type of open-source software called Gamera[12], to investigate whether optical music recognition software can be applied to microfilm images instead of the original score.

Surprisingly, the project closest in aim and scope to our Optical Neume Recognition Project focuses on an entirely different type of musical notation: Byzantine neumes. Results of this study on 'Optical Recognition of Psaltic Byzantine Chant Notation' were published in 2008 by Dalitz, Michalakis and Pranzas.[13] Again, standardization is a factor here, as Byzantine neumes were converted to print in the early 19th century, so it is significantly easier to identify each discrete neume sign out of its set of about one hundred. The researchers also based their software on the Gamera framework[14], which proved quite successful: the recognition rates on individual neumes were between 95% and 98.5%.

[10] L. Pugin, 'Aruspix: an Automatic Source-Comparison System' in Hewlett, W. B. and Selfridge-Field, E. (Eds.), *Music Analysis East and West* (Computing in Musicology vol. 14). Cambridge, MA: MIT Press, 2006, pp. 49-60.
[11] http://coltrane.music.mcgill.ca/GEMM/index.php
[12] MacMillan, K, M. Droettboom, and I. Fujinaga, 'Gamera: Optical music recognition in a new shell.' Proceedings of the International Computer Music Conference (2002): 482–5.
[13] Christoph Dalitz, Georgios K. Michalakis and Christine Pranzas, 'Optical Recognition of Psaltic Byzantine Chant Notation', International Journal of Document Analysis and Recognition vol. 11, no. 3 (Dec. 2008): 143 - 158.
[14] http://gamera.informatik.hsnr.de/index.html

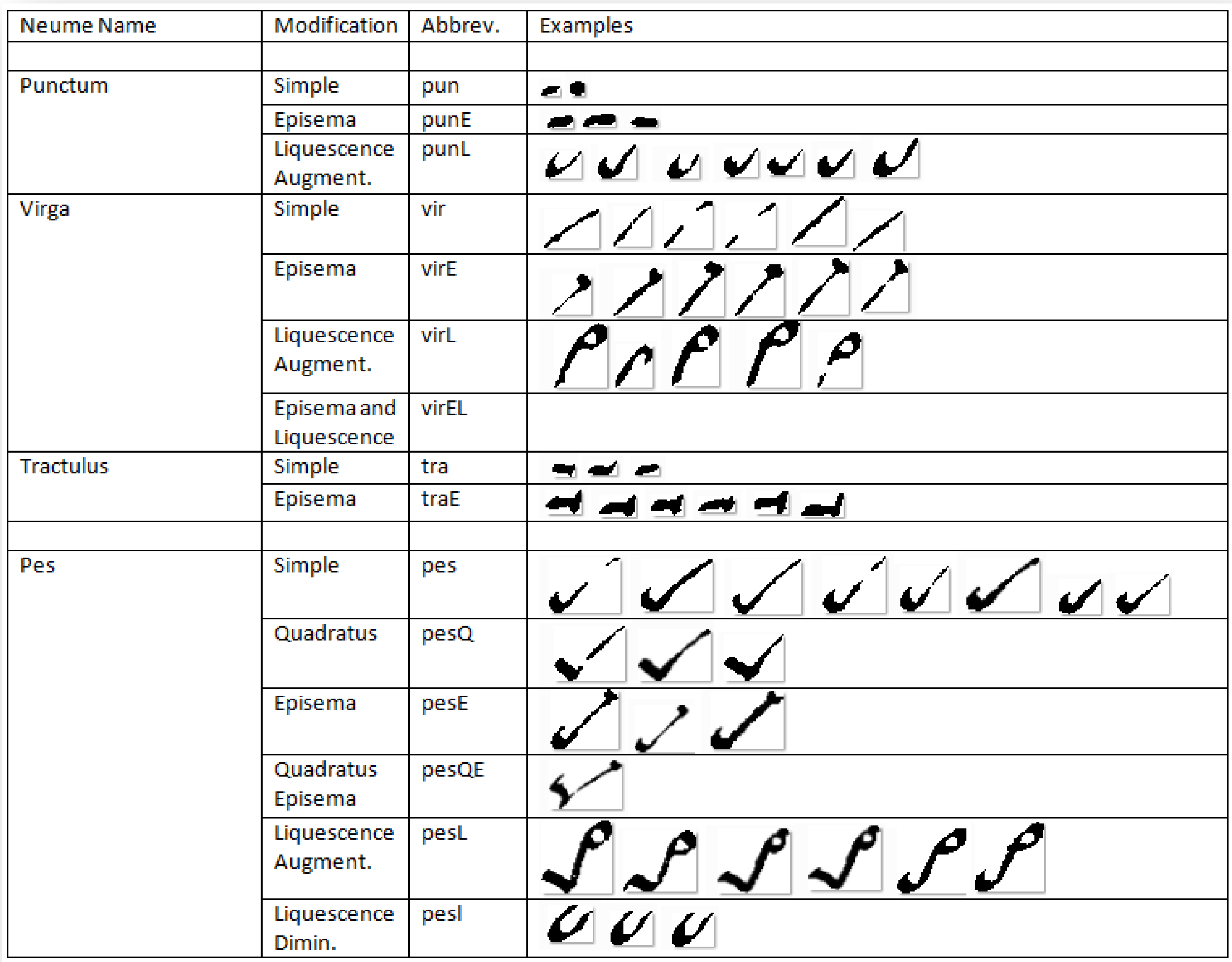
The subject of this paper, the Optical Neume Recognition Project, is a trans-Atlantic undertaking by two musicologists, Inga Behrendt and Kate Helsen, and a computer scientist with a specialisation in document analysis, Alan Sexton. Alan was brought on board in early 2010 while at the University of Western Ontario, where he was on sabbatical leave from the University of Birmingham. After exploring previous projects and the various possible avenues through which to begin our work, it was decided to focus our efforts on the notation in Codex Sang. 390 / 391. This seemed like a good choice for several reasons. First, it is a familiar source of a well-studied notation type with a facsimile published 40 years ago.[15] Second, the abbey library at St. Gall had joined the Swiss project entitled 'e-codices', making JPG images of many valuable books in Swiss libraries available for free on the internet. Every single folio of St. Gall 390 / 391 is therefore accessible online. Third, the abbey library was interested enough in our project to provide a sample selection of higher quality (TIF) images for initial testing upon our request, and recently invited our project to be featured in their year-long exhibition 'Musik im Kloster St. Gallen'. Finally, by focusing on this source, we could arrange a sort of initial test for the validity of our project, using the work of Ike de Loos and Kees Pouderoijen[16] who found that there are actually at least five scribal hands represented in the musical notation of St. Gall 390 / 391. We plan to use our comparatively objective method to confirm or question their conclusions by arranging the neumes in groups from most similar to least similar, as generated by the neume recognition software.

---

[15] Jacques Forger ed., 'Antiphonaire de Hartker.' *Paléographie musicale* II, 1 (Bern, 1970): 67 - 86.

[16] Kees Pouderoijen and Ike de Loos, 'Wer ist Hartker? Die Entstehung des Hartkerischen Antiphonars' in *Beiträge zur Gregorianik* 2009: 67 - 86.

The first step in creating software capable of recognizing neumes was to draw up a neume table in database format to which the software would match the glyphs it encountered initially. A sample of this table is shown in Figure 1.



Neume Name	Modification	Abbrev.	Examples
Punctum	Simple	pun	
	Episema	punE	
	Liquescence Augment.	punL	
Virga	Simple	vir	
	Episema	virE	
	Liquescence Augment.	virL	
	Episema and Liquescence	virEL	
Tractulus	Simple	tra	
	Episema	traE	
Pes	Simple	pes	
	Quadratus	pesQ	
	Episema	pesE	
	Quadratus Episema	pesQE	
	Liquescence Augment.	pesL	
	Liquescence Dimin.	pesl	

Figure 1: Sample selection from the Neume Table

The more instances of the same type of neume in the database, the more accurate the matching process would be. As anyone familiar with St. Gall neumes might imagine, the creation of this database wasn't simple. The basic set of neume signs was based on Cardine's table in *Semiologie Gregorienne*[17] and Corbin's *Die Neumen*[18] but it was also refined to include modifications such as episemas and liquescence. To this were added significative letters, as well as potential gestural and rhythmic indications. We were now ready to introduce optical software to the TIF images of selected folios donated by the St. Gall abbey library.

First, these colour images had to be turned into grey-scale images which were, in turn, made into black and white images through a process called binarisation. From a technological perspective, this is the most difficult step, even though the *human* eye has little difficulty distinguishing between a scribe's marks and a fleck on the page. For a computer, each pixel in a grey-scale image has a number between 0 (for black) and 255 (white). To turn a grey-scale image into black-and-white, each pixel must be determined as either 0 or 225. Problems with the image, such as bad lighting, warping, creasing and distortion, fading of or damage to the ink, and stains on the parchment makes binarisation more complicated than simply setting the threshold between black and white once for all the images, as that threshold will vary over different parts of the image. The algorithms we are using to carry out this process are among the most cutting-edge in their field.

Once the image has been binarised, all groups of black pixels must be analysed to determine if they make up what we would recognize as a neume. Since, at this stage, we are mostly interested in the neumes rather than the chant text, our software tool extracts whole lines of neumes at a time, shown in green in Figure 2, so that they can be analysed and classified.

[17] Eugène Cardine, 'Semiologie Gregorienne' in *Études Grégoriennes* XI (Solesmes, 1970).
[18] Solange Corbin, *Die Neumen,* (Cologne, 1977)

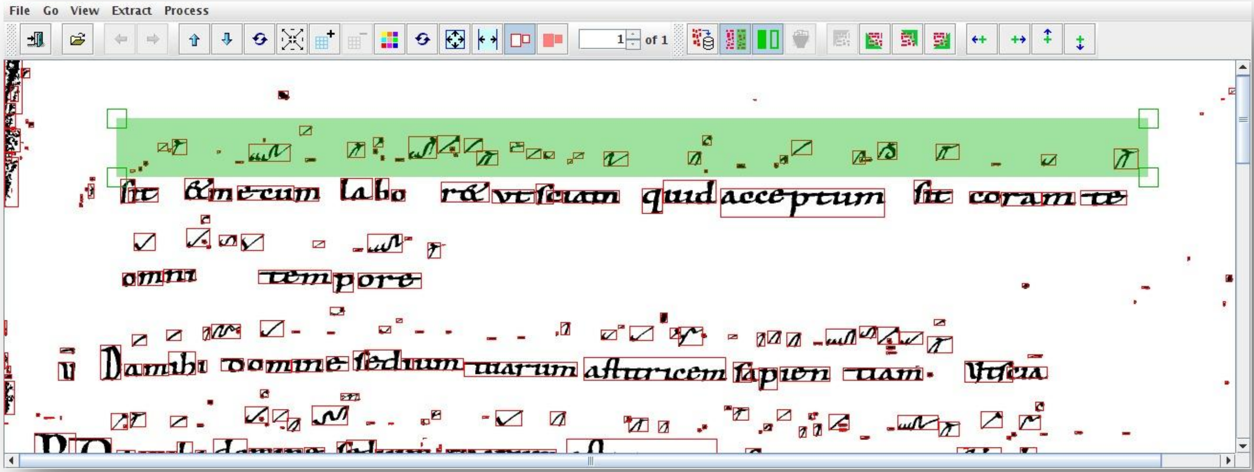


Figure 2: Neumes are identified in whole lines marked in green

It is also possible to leave the neumes and Latin text as one, and automatically identify which are lines of text and which are neumes by comparing the ratio of black to white pixels along any horizontal line on the folio. Those lines where the black / white ratio are higher will typically be text, while those lines where more white space is included are lines of neumes.

The next step is to identify the neumes. This is done by looking at the distribution of black pixels to white pixels in the smallest bounding box that can be drawn around the black pixels. The distribution is then compared to the neume database and the sign is allocated to a group with graphic similarities. In Figure 3 , we see two boxes with similarly distributed black and white pixels. The software will therefore rank these two shapes as more similar to each other than other shapes.

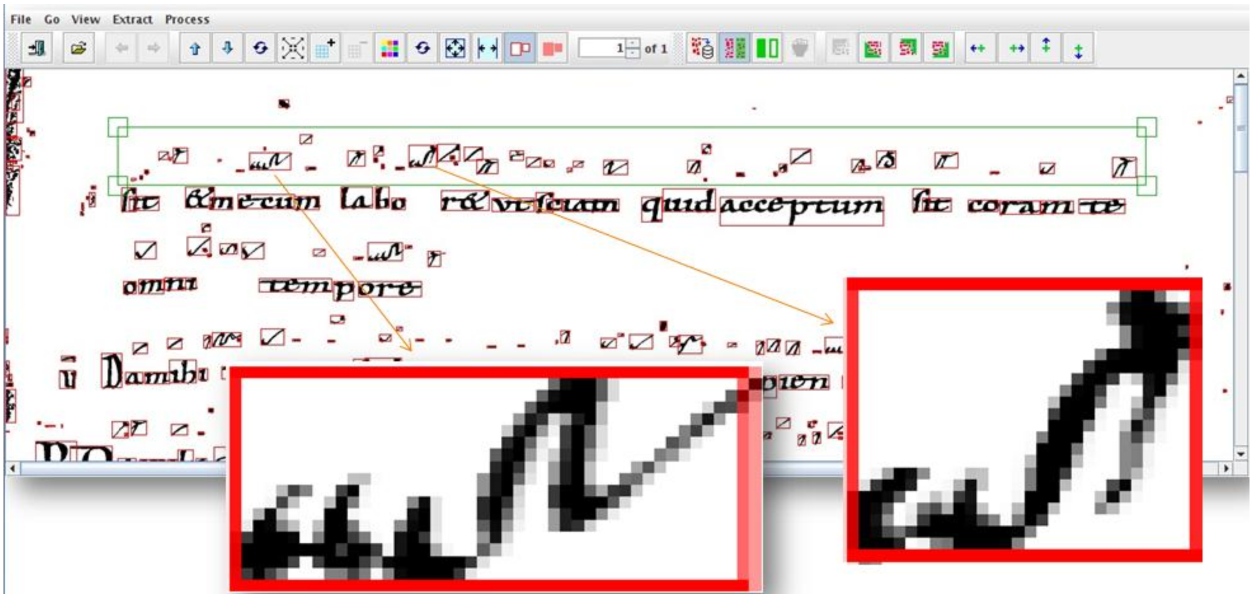


Figure 3: Pixel distribution in each bounding box around each neume

Once one neume has been identified, say, as a clivis, when the next sign with the same black/white pixel distribution is encountered it will be categorized as another clivis. Every allocation must be checked and approved by the musicologist working with the software. Although it is slow, the advantage to this way of working is that we can make the appropriate allowances for *variance* in each type of neume, rather than force them into one standard format and lose the nuances which neumes represent so well. Currently we are in the stage of developing and refining algorithms for this identification process. To give an idea of a possible interface that our tool might support, we show in Figure 4 how the Gamera toolkit displays the results of similar algorithms in its framework when applied to the neumes of the Hartker antiphoner.

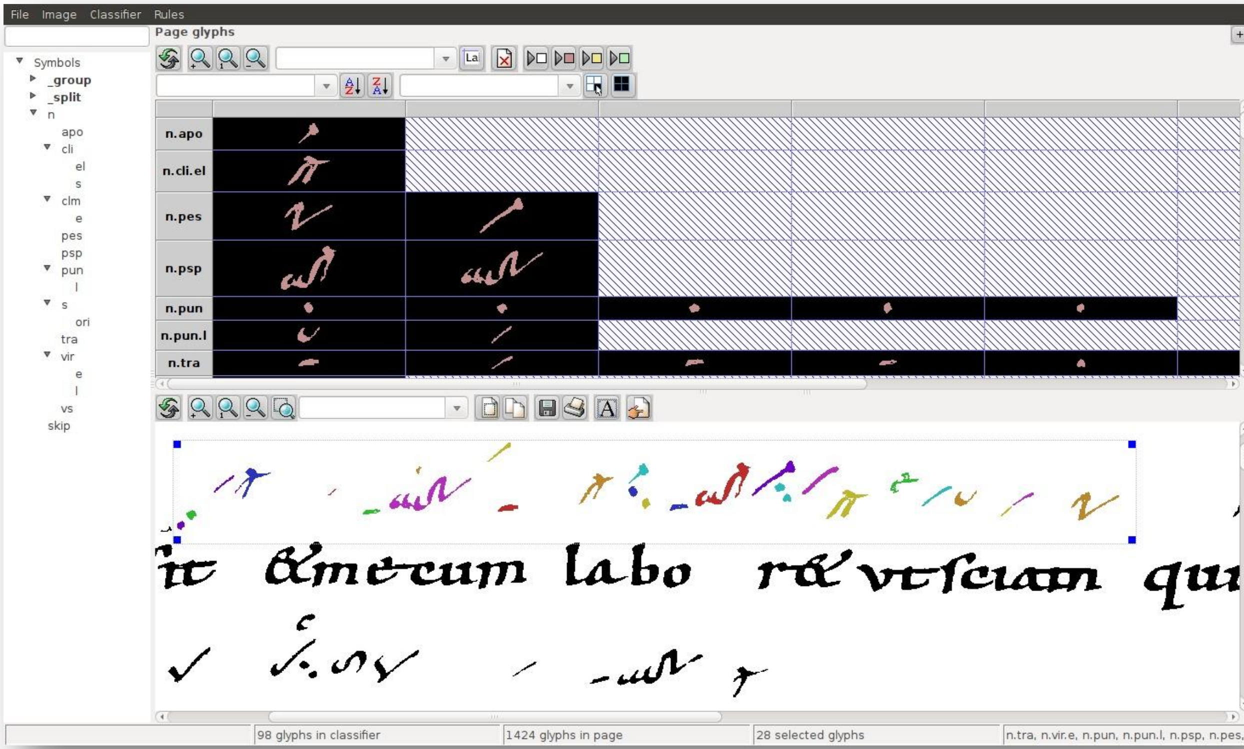


Figure 4: The Gamera Toolkit applied to neume classification

The Optical Neume Recognition Project, still in its nascent stage, has already obliged us to grapple with both technical and musicological challenges. We have had to look again at a familiar manuscript and consider anew the basic relationship between a neume and its meaning. We invite you to follow our progress through our website which details several recent papers given on the project by Kate and Inga in Canada, the United States, Germany, and St. Gallen, Switzerland itself. With the results we have achieved so far, we are hoping to be supported in the coming year by programs which encourage the integration of technology and the Humanities in Europe and North America.

Of course, there are still many questions to be resolved as we progress: What will happen when the software is applied to a different sort of early notation, such as Paleofrankish or Aquitanian? How might neume height be incorporated into the software's analysis? How can the syllables of Latin text be paired with each neume or neume grouping? What might we find when we are able to compare two adiastematic melodies quickly and accurately, without using a third, pitched source as the basis for the comparison? Indeed, the Optical Neume Recognition Project is altering the framework. This influences how we see the picture - that is, the notation - within it, and sparks further curiosity.